Some Contributions to Decision Making in Complex Information Settings with Imprecise Probabilities and Incomplete Preferences: Theoretical and Algorithmic Results

Christoph Manuel Jansen

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

Eingereicht am 27.04.2018

Some Contributions to Decision Making in Complex Information Settings with Imprecise Probabilities and Incomplete Preferences: Theoretical and Algorithmic Results

Christoph Manuel Jansen

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

vorgelegt von
Christoph Manuel Jansen
aus München

München, den 27.04.2018

Erstgutachter: Prof. Dr. Thomas Augustin
Zweitgutachter: Prof. Dr. Matthias Troffaes
Drittgutachter: Prof. Dr. Lev Utkin
Tag der mündlichen Prüfung: 20.07.2018

Acknowledgement

I would like to express my sincere gratitude to everyone who contributed to this dissertation! Special thanks go to ...

... *Thomas Augustin for introducing me to the exciting world of decision theory, for giving me the opportunity to write this dissertation, for always (even in the most busy periods) taking time for me, for big enthusiasm for all kinds of new research ideas, for many insights into the world of imprecise probabilities from a "man of the first hour", and for "borrowing" your lecture on decision theory to me. Thank you very much, Thomas, for the excellent supervision and collaboration!*

... *Matthias Troffaes and Lev Utkin for their willingness to act as external reviewers and to Christian Heumann and Volker Schmid for their willingness to be part of the examination committee.*

... *Thomas Augustin and Georg Schollmeyer for the collaboration on several different research projects that has been productive and relaxed at the same time.*

... *all former and current collegues of the working group "Foundations of Statistics and their Applications" (Aziz, Cornelia, Eva, Georg, Johanna, Julia, Patrick, Paul and Thomas) for stimulating joint research seminars, for equally stimulating joint lunches and dinners, and for making my time here at the institute a very pleasant and special one!*

... *all remaining former and current colleagues from the Department of Statistics for the pleasant and relaxed general atmosphere.*

... *the colleagues of the Munich Center for Mathematical Philosophy, especially to Jean Baccelli, Jürgen Landes, Arthur Merin, Barbara Osimani, Roland Pöllinger, and Gregory Wheeler for inspiring vivid exchange that helped to broaden my horizon.*

... *Julia and Georg for proof-reading parts of my thesis, and for all the years you shared an office with me. I could not have imagined any better office mates! Thanks.*

... *Susi and my family for constant backing and support and for sharing me with my work for such a long time. Thanks for all the patience!*

Zusammenfassung

Ein erheblicher Teil der Herausforderungen der modernen Wissenschaften kann als Problem der *Entscheidungstheorie unter Unsicherheit* aufgefasst werden. Dies umfasst insbesondere Parameterschätzungen und Hypothesentests in der Statistik, die Modellierung der Präferenzen und des Wahlverhaltens eines Akteurs in Philosophie und Ökonomie oder die Formalisierung spieltheoretischer Fragestellungen. Das Grundmodell der Theorie ist gleichermaßen einfach und ausdrucksstark: Ein Entscheidungsträger (oder Akteur) kann eine Alternative X aus einer Menge $\mathcal{G}$ verschiedener Optionen wählen. Jedoch hängt die Konsequenz der Wahl von $X \in \mathcal{G}$ vom wahren Umweltzustand aus einer Menge S ab: Die Wahlalternativen entsprechen also Zufallselementen $X : S \to A$, wobei A eine Menge von Konsequenzen bezeichnet, für welche üblicherweise zusätzliche Ordnungsstruktur verfügbar ist. Für dieses, sehr allgemeine, Modell wurden viele verschiedene Konzepte zur Bestimmung optimaler Alternativen X_* aus $\mathcal{G}$ vorgeschlagen. Jedoch stützen sich nahezu alle diese Konzepte auf die folgenden klassischen Annahmen: (I) Der Akteur kann die Unsicherheit über die Zustände aus S durch ein klassisches Wahrscheinlichkeitsmaß π charakterisieren und (II) die Präferenzen des Akteurs lassen sich angemessen durch eine kardinale Nutzenfunktion beschreiben. Durch (III) die Beschränkung des Modells auf einzelne Akteure wird zudem der Fall ausgeschlossen, dass Entscheidungen durch Gruppen mit inhomogenen Präferenzen getroffen werden müssen.

Die vorliegende kumulative Dissertation operationalisiert Entscheidungsprobleme unter (potentiell) sehr schwach strukturierter oder *unvollständiger* Information, in welchen eine oder mehrere der klassischen Annahmen (I), (II) und (III) verletzt sind. Überdies werden Kriterien zur optimalen Entscheidungsfindung vorgeschlagen, welche in solchen Situationen immer noch anwendbar sind. Der Hauptfokus der Arbeit liegt dabei auf der Gewinnung theoretischer Erkenntnisse über Eigenschaften optimaler Entscheidungen sowie der Entwicklung von Algorithmen zur optimalen Entscheidungsfindung.

Beitrag 1 behandelt Verletzungen von (I). Wir entwickeln Algorithmen zur Berechnung optimaler Entscheidungen bezüglich Kriterien basierend auf unsicheren Wahrscheinlichkeiten sowie auf Imprecise Probabilities. Zudem werden Bedingungen für die Sinnhaftigkeit randomisierten Entscheidens gegeben, welche durch ungünstigste a priori Verteilungen charakterisiert sind.

Beitrag 2 widmet sich erneut Verletzungen von (I). Wir schlagen ein Entscheidungskriterium und einen Algorithmus zu dessen Berechnung vor. Überdies werden Maße zur Quantifizierung des Ausmaßes an E-Zulässigkeit eingeführt, für deren Berechnung ebenfalls Algorithmen bereitgestellt werden. Schließlich werden Konzepte ordinaler Entscheidungstheorie diskutiert.

Beitrag 3 untersucht simultane Verletzungen von (I) und (II). Wir betrachten Alternativen X mit Werten in Präferenzsystemen und schlagen drei Ansätze zur Konstruktion von Entscheidungskriterien vor: i) Verallgemeinerte Erwartungsintervalle, ii) globale Zulässigkeit und iii) lokale Zulässigkeit. Es werden zudem Algorithmen zur Auswertung der Kriterien entwickelt.

Beitrag 4 behandelt Verletzungen von (III). Wir führen ein Kriterium zur Bewertung von Präferenzaggregationsverfahren ein, das auch die Homogenität der Gruppe einbezieht. Es wird gezeigt wie dieses Kriterium approximiert werden kann und wie diese Näherungen aus Daten geschätzt werden. Gängige Aggregationsregeln werden in einer Simulationsstudie verglichen.

Beitrag 5 entwickelt ein lineares Programm zur Überprüfung stochastischer Dominanz für Zufallsvariablen mit Werten in einer partiell geordneten Menge. Wir studieren das duale Programm und seine Eigenschaften und behandeln zudem die Frage der Inferenz durch Verwendung von Resampling-Methoden sowie die Anwendung der Vapnik-Chervonenkis-Theorie.

Summary

A significant amount of the challenges arising in the modern sciences can be reformulated as some suitable problem belonging to the theory of *Decision Under Uncertainty*. This includes parameter estimation and hypothesis testing in Statistics, modeling an agent's preferences and choice behavior in Philosophy and Economics or the formalization of game theoretic problems. The basic model of this theory is equally simple and expressive: A decision maker (or agent) can choose an alternative (or act) X from a set $\mathcal{G}$ of different available options. However, the consequence that choosing $X \in \mathcal{G}$ yields depends on which potential state of the world from a set S turns out to be the true one. Formally, the objects of choice are represented by random elements $X : S \to A$, where A is a set of consequences for which, usually, some additional order structure is available. For this very general setup, several concepts for obtaining an optimal alternative X_* from $\mathcal{G}$ have been proposed. However, almost all of these concepts (more or less directly) rely on the following assumptions: (I) the ability of the agent to characterize her beliefs about the states of the world from S by some classical probability measure π and (II) the availability of a cardinal utility function that adequately characterizes the agent's preferences. Additionally, by (III) restricting the model to single agents, also the case where decisions have to be formed by inhomogeneous groups is excluded in the classical framework.

The present cumulative PhD project models decision problems in settings with (potentially) very weakly structured, or *incomplete*, information available, i.e. in settings where one or more of the classical assumptions (I), (II) and (III) are violated. Moreover, we propose criteria for optimal decision making that are applicable in such weakly structured situations. Here, we lay a special focus on obtaining new theoretical insights into the properties of these optimal decisions and on providing algorithms capable of finding optimal decisions in these complex situations.

Contribution 1 deals with decision problems where only assumption (I) is violated. We propose algorithms for computing optimal acts with respect to criteria based on uncertain classical probabilities as well as on imprecise probabilities. Further, we give conditions under which randomization in decision making pays out and characterize these in terms of least favorable priors.

Contribution 2 again deals with violations of assumption (I). We introduce a new decision criterion and provide a simple algorithm for evaluating it. Further, we propose two measures for quantifying the extent of E-Admissibility and give linear programming algorithms for computing these. Finally, we discuss some ideas in the context of ordinal decision theory.

Contribution 3 investigates simultaneous violations of (I) and (II). We consider acts taking values in so-called preference systems and propose three approaches for constructing decision criteria: i) generalized expectation intervals, ii) global admissibility, and iii) local admissibility. Whenever suitable, we provide linear programming based algorithms for checking optimality.

Contribution 4 deals with violations of (III). We propose a criterion for evaluating the adequateness of preference aggregation procedures reflecting the group's homogeneity. We show how to approximate our criterion if information is only imperfectly given and how to estimate these approximations from data. Finally, we compare common aggregation rules in a simulation study.

Contribution 5 develops a linear programming method for detecting stochastic dominance for random variables with values in a partially ordered set. We study the dual program and discuss its properties. Subsequently, we address the question of inference by utilizing resampling methods as well as conservative bounds that are given by the application of Vapnik-Chervonenkis theory.

Contents

Contributions of the thesis

The present PhD project is composed of the following five contributions that are referred to as *Contribution 1* to *Contribution 5* throughout the rest of this work:

1. *Jansen, C.*; Augustin, T.; Schollmeyer, G. (2017a): Decision theory meets linear optimization beyond computation. In: Antonucci, A.; Cholvy, L.; Papini, O. (eds): *Symbolic and Quantitative Approaches to Reasoning with Uncertainty. ECSQARU 2017.* Lecture Notes in Computer Science, vol 10369. Springer.

2. *Jansen, C.*; Schollmeyer, G.; Augustin, T. (2017b): Quantifying degrees of E-admissibility in decision making with imprecise probabilities. To appear in: Cozman, F. and Wheeler, G. (eds): *Festschrift in Honor of Teddy Seidenfeld.* Theory and Decision Library A, Springer.

   As the book is not published yet, the following identical technical report is included in Chapter 3 of this thesis instead:

   Jansen, C.; Schollmeyer, G.; Augustin, T. (2018a): Quantifying degrees of E-admissibility in decision making with imprecise probabilities. Department of Statistics (LMU München): *Technical Reports*, Nr. 215, March 2018.
   Available under: https://epub.ub.uni-muenchen.de/42674/

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

4. *Jansen, C.*; Schollmeyer, G.; Augustin, T. (2018c): A probabilistic evaluation framework for preference aggregation reflecting group homogeneity. Revision under review for *Mathematical Social Sciences.*

   As the paper is still under review, the following technical report that is identical to the submitted revision is included in Chapter 3 of this thesis instead:

   Jansen, C.; Schollmeyer, G.; Augustin, T.. (2016/2018): A Probabilistic Evaluation Framework for Preference Aggregation Reflecting Group Homogeneity. Department of Statistics (LMU München): *Technical Reports*, Nr. 193, last updated March 2018. Available under: https://epub.ub.uni-muenchen.de/42751/

5. Schollmeyer, G.; *Jansen, C.*; Augustin, T. (2017): Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems. Department of Statistics (LMU München): *Technical Reports*, Nr. 209, August 2017. Available under: https://epub.ub.uni-muenchen.de/40416/

Declaration of the author's specific contributions

All contributing papers are the result of a fruitful collaboration with several co-authors. By separately referring to each of the papers, in the following the own contribution of the author of this dissertation is clarified:

- *Contribution 1:* The paper builds up on results that the author developed for a poster for ISIPTA '15 with the same title. The author drafted and wrote the whole paper with the exception of the paragraph about the connection to pignistic probabilities directly following Proposition 3, which was written by Thomas Augustin, and the paragraph introducing to Section 3 of the paper, which was jointly written by Georg Schollmeyer and the author. The propositions and proofs in the paper are due to the author. The connection between the decision criterion by Hodges and Lehmann and the Gamma-Maximin criterion was worked out in joint discussions of all contributing authors. Additionally, several rounds of discussing preliminary versions the paper with all authors lead to improvements of the presentation. All authors contributed to revising the paper according to the reviewers' comments.

- *Contribution 2:* The whole paper was drafted and, in most parts, written by the author. Exceptions are given by Footnote 2, parts of Footnote 3, Footnote 4, Footnote 7, Footnote 8, Footnote 10, and Footnote 14. These were written by Thomas Augustin and mainly contain hints to further relevant literature. The idea for Proposition 1 and the corresponding proof are due to the author. The idea for the concept of E_ε-admissibility (see Definition 1) and the linear programming problem for evaluating it (see Proposition 2) were jointly developed by Georg Schollmeyer and the author in discussions about extending Proposition 2 of *Contribution 1.* The idea of investigating concepts for quantifying the extent of E-admissibility is due to joint discussions by all authors. The concept of maximal extent (see Definition 2) as well as the series of linear programming problems for computing it (see Proposition 3) was, in a first variant, developed jointly by all authors and worked out in detail by the author. The concept of uniform extent (see Definition 3) and the algorithm for computing it (see Proposition 4) are, in a first variant, due to Georg Schollmeyer and the author. The author then worked out the details for Proposition 4 and its proof. In Section 4, the criterion of joint statistical preference is due to Thomas Augustin. The linear programming problem for checking joint stochastic dominance in the imprecise version is due to the author, relying on results from Contribution 3. The example in Section 5 was constructed and implemented by the author. In joint discussions the presentation of the paper was improved.

- *Contribution 3:* This contribution is based on a conference paper for the ISIPTA '17 conference (see Jansen et al. (2017)), drafted by the author and revised according to remarks of the other authors as well as review comments of the three anonymous ISIPTA '17 referees. Following the invitation to the conference's special issue in the

International Journal of Approximate Reasoning, the conference version was significantly extended to the present version. The whole extended version was drafted and, in most parts, written by the author. Exceptions are the paragraph directly following Definition 6, the paragraph directly following the proof of Proposition 3, the second paragraph following Definition 8, and the first two paragraphs directly following the proof of Proposition 6. These were drafted by Georg Schollmeyer. The idea of the proof of Proposition 3 was jointly developed by Georg Schollmeyer and the author. The concept of granularity (see Definition 3) was developed in discussions of all three authors. The name granularity is due to Thomas Augustin, who also contributed to embedding the paper into the relevant literature. Additionally, several rounds of discussing preliminary versions of the paper with all authors lead to improvements of the presentation. All authors contributed to revising the paper according to the reviewers' comments.

- *Contribution 4:* The whole paper was mainly drafted and, in most parts, written by the author. Exceptions are the section about preference homogeneity in related work (see Section 2.2 of the paper), the insight that the measure W of Kendall and Smith (see Kendall and Smith (1939)) does not satisfy our minimal set of conditions for preference homogeneity measures (see Definition 1 of the paper) and the corresponding counter-example, the part about the non-locality of the maximum consensus homogeneity δ_n (see Definition 2 of the paper) that is discussed in Appendix A 3, and the section about the commonality sharing rule (see Section 4.2 of the paper). These were drafted by Georg Schollmeyer. All authors contributed to proof-reading and revising the paper in several discussion rounds. In particular, the current form of Section 3.1 is influenced by comments of Thomas Augustin.

- *Contribution 5:* The paper was mainly drafted and, in most parts written, by Georg Schollmeyer. The author's main contribution to the manuscript is the section about the dual programming problem to the linear program for detecting stochastic dominance and its characterization via a mass transportation problem. Further, the author investigated the interrelations of this dual programming problem and the algorithms discussed in Range and Osterdal (2013). All authors contributed to proof-reading and revising the paper in several discussion rounds.

1 Motivation: Why to bother about imprecision in decision making?

The present dissertation is concerned with *decision making under uncertainty*, sometimes also called *choice under uncertainty*, for situations in which the available information on both the decision maker's preferences and the mechanism generating the states of nature is (potentially) very weakly structured. Specifically, this has the consequence that many of the classical approaches to decision making under uncertainty can no longer be applied or only be applied in a modified form appropriately taking into account the weak structure of the available information. In this first chapter, we start by introducing some basic concepts that are required for understanding the broader context of the works contributing to this thesis. Importantly, note that neither the selection of the presented concepts nor the presentation of concepts themselves is intended to be exhaustive by any means. Particularly, if not directly needed, many of the mathematical details are left out in order to preserve the introductory character of the chapter.[1] Accordingly, the chapter can be rather viewed as a guideline that helps placing the works contributing to this thesis into the field they are contributing to. Moreover, note that in terms of the assumptions (I), (II), and (III) mentioned in the summary at the beginning of the present dissertation, Chapter 1.3 recalls theories that are in accordance with all three assumptions, Chapter 1.4 recalls theories capable of dealing with violations of (I) and (II), and Chapter 1.5 recalls some few details on group decision making, i.e. deals with violations of assumption (III).[2] Previously, in the Sections 1.1 and 1.2, we briefly sketch the basic model of decision making under uncertainty and the formalism for determining optimal acts.

1.1 The basic setting

A significant amount of the challenges arising in the modern sciences can be reformulated as some suitable problem belonging to the theory of *Decision Under Uncertainty*. Particularly, this includes parameter estimation and hypothesis testing in statistics (see, e.g., Augustin (1998, 2001); Hable (2009); Cattaneo and Wiencierz (2012); Cattaneo (2013); van Ommen

[1] This includes in particular the concepts recalled in Section 1.3, where we do not bother too much about whether the probability measures in the different representation results are finitely or countably additive and about how the integrals with respect to these measures are defined. We refer to references providing details on these aspects at the corresponding places in the text.

[2] Note that the assumptions (I), (II), (III) will be replaced by a more refined classification into information sources (see Section 1.1) throughout the rest of this thesis.

(2017)), modeling an agent's preferences and choice behavior in Philosophy and Economics (see, e.g., Kaplan (1983); Braun and Gautschi (2011)) or the formalization of game theoretic problems (see, e.g., Milnor (1951); Luce and Raiffa (1957)). The basic model of this well-established theory, referred to as *basic decision model (BDM)* in the following, is equally simple and expressive: A decision maker (or agent) can choose an alternative (or act) X from a set $\mathcal{G}$ of different available options. However, the consequence that choosing $X \in \mathcal{G}$ has for the agent is fraught with uncertainty; it depends on which potential state of the world from a set S turns out to be the true one. Formally spoken, the objects of choice are represented by random elements $X : S \to A$, where A is some set of consequences (or *consequence space*) and, for $s \in S$, we interpret $X(s)$ as the consequence that choosing act X yields given the state s turns out to correspond to the true description of the world. Here, we implicitly make the common assumptions that one, and only one, of the states collected in S is the true state (see, for instance, Fishburn (1965, p. 217)).[3]

More often than not, some information about the agent's preferences among the consequences collected in A as well as on the underlying uncertainty about the states collected in S will be available. Of course, such information can take various forms, for instance depending on the field of interest and the concrete application in mind. Specifically, we mainly distinguish between two different sources of information coming along with a BDM, namely *information on the agent's preferences on* A and *information about the uncertainty on* S. The following description of these information sources is consciously kept rather general at this point of the work and will be made more concrete at several points throughout this chapter:

$\mathcal{I}_P$: Information about the agent's preferences on A

The information source $\mathcal{I}_P$ contains all information about the agent's preferences among the consequences in A or objects that are somehow related to those consequences. Such information can, for instance, be given in form of some (not necessarily complete) binary relation $R \subset A \times A$ on the consequence space A that models the agent's ordinal preferences among the elements of A. If in addition knowledge on *preference strength* is available, this can be formalized in different ways. For instance, the information in $\mathcal{I}_P$ can consist of some pair (R_1, R_2) of (not necessarily complete) relations, where R_1 models the ordinal and R_2 models the *cardinal* part of the agent's preferences. This is the approach, for instance, followed in Krantz et al. (1971, Chapter 4) or French (1986, Section 3.4). Mathematically closely related, ordinal and cardinal information are sometimes simultaneously modeled by a so-called quaternary relation on the consequence set A (see, e.g., Pivato (2013)).

Another way of encoding preference strength is by extending the space: The information in $\mathcal{I}_P$ is then assumed to consist of a binary relation not only on the consequence space A itself, but on some extended space somehow related to it (for instance $\mathcal{L}_A$, see Section 1.3).

[3] It should, however, be mentioned that appropriately defining the state space is a highly non-trivial task which has provoked a number of paradoxes that most commonly relate to the problem of act-state dependence. The probably most famous one among these paradoxes is the one of *Newcomb* (see for instance Nozick (1969) or Bar-Hillel and Margalit (1972)).

Such approaches were pioneered in von Neumann et al. (1944) and generalized for instance in Aumann (1962), Fishburn (1970, Chapter 10), and Rubin (1987). More applied approaches, where the information on the ordinal and the cardinal part of the preferences is directly formulated in terms of different types of statements about utilities of the consequences, have for instance been followed in Danielson and Ekenberg (1998), Danielson et al. (2003), and Danielson (2005). Of course, also other types of information are imaginable to be part of $\mathcal{I}_P$ (for instance information on the agents attitude towards ambiguity, see Gilboa and Schmeidler (1989)). Some of these approaches will be recalled in some more detail in Sections 1.3 and 1.4 of this chapter.[4]

$\mathcal{I}_U$: Information about the uncertainty on S

The information source $\mathcal{I}_U$ contains all information (subjective, objective, or both) about the mechanism generating the states of nature or, in other words, about the agent's uncertainty about the occurrence of the different states $s \in S$. Most commonly, it is assumed that this information is structured enough to allow for the specification of a classical probability measure π on the state space. This assumption corresponds, at least when subjective probabilities are considered, to the classical framework pioneered by de Finetti (see, e.g., de Finetti (1974)). However, often the information will turn out to be too weakly structured to allow for specifying a classical probability measure. For instance, $\mathcal{I}_U$ could consist only of statements about probabilities of certain events $D, D_1, D_2 \subseteq S$, such as "D is no more probable than 0.6" or "D_1 is at least as probable as D_2". In this case, several related theories for modeling the information have been proposed, all somehow relating to sets of probability measures as central ingredients. Most prominent examples for such models include *linear partial information*[5], *credal sets*[6], *lower previsions*[7], and *interval probability*[8]. Some of these approaches will be recalled in some more detail in Sections 1.3 and 1.4 of this chapter.

Of course, also settings not allowing for a separate treatment of the information sources $\mathcal{I}_P$ and $\mathcal{I}_u$, but rather lending themselves for a "simultaneous model" are imaginable. Among the most prominent examples for such simultaneous models are the frameworks of *Savage*

[4]Here, it is important to note that our notion of a basic decision model and in particular the structure of the information source $\mathcal{I}_P$ makes an implicit assumption: Since the information contained in $\mathcal{I}_P$ is about the agent's preferences on A, it is assumed that the agent possesses *fixed* (potentially partial) preferences on A. Specifically, these preferences are therefore assumed to be independent of what the true state of nature $s \in S$ turn out to be: The knowledge of s does not influence the agent's preferences on A. We will refer to this as *state-independent preferences* or, if preferences are expressed by utility functions, as *state-independent utility*. This is an assumption that is made throughout the whole thesis. For more information on state-dependent utility see, e.g., Karni et al. (1983), Seidenfeld et al. (1995, Section 4), and Baccelli (2017).

[5]See Kofler and Menges (1976); Kofler (1989) for textbooks and Kofler et al. (1984) for a more compact presentation of the theory.

[6]See Levi (1974) and Levi (1980).

[7]See Walley (1991) or Troffaes and de Cooman (2014).

[8]See Weichselberger and Pöhlmann (1990) and Weichselberger (2001) for textbooks or Weichselberger (2000) for a compact presentation of the basic concepts.

(see Savage (1954)) and *Anscombe and Aumann* (see Anscombe and Aumann (1963)) and their many relaxations and modifications. Also some of these models for the information sources will be discussed in some more detail in Section 1.3 and 1.4. Before, we briefly recall an important concept in order to obtain a formal framework for determining optimal decision, namely the concept of *choice functions*.

1.2 Formalism for determining optimal decisions

Given some basic decision model $\mathcal{G}$, the goal is to obtain some act $X^* \in \mathcal{G}$ or, more generally, some subset of acts $\mathbb{O}^*_{\mathcal{G}} \subseteq \mathcal{G}$, that is optimal in the sense that it is derived by the application of some selection rule that best possibly utilizes the available information sources $\mathcal{I}_P$ and $\mathcal{I}_U$. Formally, every such selection rule can be viewed as a *choice function*[9]

$$\mathrm{ch} : \Sigma \to \Sigma \quad , \quad \mathcal{D} \mapsto \mathrm{ch}(\mathcal{D})$$

with the property that $\mathrm{ch}(\mathcal{D}) \subseteq \mathcal{D}$ for every $\mathcal{D} \in \Sigma$, where $\Sigma := 2^{\mathcal{G}}$. For $\mathcal{D} \in \Sigma$, we call $\mathrm{ch}(\mathcal{D})$ the *choice set* of $\mathcal{D}$ (with respect to the function ch) and $\mathrm{re}(\mathcal{D}) := \mathcal{D} \backslash \mathrm{ch}(\mathcal{D})$ the *reject set* of $\mathcal{D}$ (with respect to the function ch).[10] Here, at least two different interpretations of choice sets should be distinguished: If the information sources $\mathcal{I}_P$ and $\mathcal{I}_U$ provide enough structure, for instance if they allow for a complete ordering of the available acts in $\mathcal{G}$, one can directly interpret the choice set $\mathrm{ch}(\mathcal{D})$ as the set of acts from $\mathcal{D}$ that are a worthy choice for the agent. Between these acts the agent should be indifferent; they are all equally best acts, as Bradley (2015) puts it. This view on choice functions is sometimes called the *strong* interpretation (see, e.g., Bradley (2015, p. 58)).

If contrarily the information sources do not provide as much structure, for instance if they allow only for a partial ordering of the available acts in $\mathcal{G}$, one often goes for the *weak* interpretation of the choice function: The acts in the choice set $\mathrm{ch}(\mathcal{D})$ are then not interpreted as all equally good, but rather as not comparable given the information sources. Specifically, the agent is then not necessarily indifferent between the acts in $\mathrm{ch}(\mathcal{D})$, however, is still able to tell that each of them is preferable to each of the acts in the corresponding reject set $\mathrm{re}(\mathcal{D})$. Accordingly, the weak interpretation corresponds to interpreting rather the reject set $\mathrm{re}(\mathcal{D})$ instead of the choice set itself, namely as the set of all acts that are definitely not a worthy choice.

Due to the generality of the concept, choice functions can be constructed in various ways. Most easily, if the available information sources $\mathcal{I}_P$ and $\mathcal{I}_U$ display enough structure, it may be possible to define a meaningful real-valued criterion function[11] of the form $cr : \mathcal{G} \to \mathbb{R}$ that assigns a real number to every available act that somehow quantifies its value for the

[9] See Sen (1971) for an original source and Chambers and Echenique (2016, Chapter 2) for a textbook.

[10] Note that we also allow for empty choice sets in order to be prepared for the general discussions to come. Of course, it seems generally desirable to construct choice functions with non-empty choice sets.

[11] Note that the term criterion function is sometimes used to denote functions that map from $\mathcal{G}$ to $\mathbb{R}^q$, i.e. to vectors of real numbers. We differ from this convention here.

decision maker on the basis of what she knows from $\mathcal{I}_P$ and $\mathcal{I}_U$. We then obtain a choice function ch_{cr} with respect to the criterion function cr by simply defining $\mathrm{ch}_{cr} : \Sigma \to \Sigma$ with

$$\mathrm{ch}_{cr}(\mathcal{D}) := \Big\{ Y \in \mathcal{D} : \mathrm{cr}(Y) \geq \mathrm{cr}(X) \text{ for all } X \in \mathcal{D} \Big\} \tag{1.1}$$

for all $\mathcal{D} \in \Sigma$. For choice functions ch_{cr} induced by some criterion function cr it will often be the case that the choice set $\mathrm{ch}_{cr}(\mathcal{D})$ of optimal acts from $\mathcal{D}$ is actually a singleton and, accordingly, that there exists a uniquely defined optimal act X^* in $\mathcal{D}$. However, even if the choice set $\mathrm{ch}_{cr}(\mathcal{D})$ of optimal acts from $\mathcal{D}$ contains more than just one element, the agent will not have any disadvantage of choosing arbitrarily from $\mathrm{ch}_{cr}(\mathcal{D})$, since for any two acts $X, Y \in \mathrm{ch}_{cr}(\mathcal{D})$ the criterion function assigns identical values, i.e. it holds that $\mathrm{cr}(X) = \mathrm{cr}(Y)$. The acts contained in $\mathrm{ch}_{cr}(\mathcal{D})$ are not incomparable, they are completely equivalent with respect to the decision maker's criterion function. Thus, choice sets that are assigned by choice functions induced by some meaningful criterion function can be interpreted in the strong sense of the above classification, namely as being sets of equally best acts from the corresponding set $\mathcal{D}$ of available acts.

However, in many real-world decision problems it will be the case that the available information sources $\mathcal{I}_P$ and $\mathcal{I}_U$ do not provide enough structure in order to define a criterion function cr: $\mathcal{G} \to \mathbb{R}$ as above in a meaningful and non-arbitrary manner. In these situations the construction of an appropriate choice function is clearly not as obvious. A common way to proceed in such situations is to utilize the available information sources in order to specify a, generally not complete, binary relation $\succeq \subseteq \mathcal{G} \times \mathcal{G}$ on the set of available acts $\mathcal{G}$ that models the rank order between such pairs (X, Y) of acts that can be compared on the basis of the available information. Specifically, the expression $(X, Y) \in \succeq$ then is interpreted as X being at least as preferable as Y. Every such binary relation $\succeq$ then naturally induces a choice function $ch_{\succeq}$ by assigning to each subset $\mathcal{D}$ of acts from $\mathcal{G}$ the set $\mathrm{ch}_{\succeq}(\mathcal{D})$ containing exactly the maximal, e.g. non-dominated, elements of $\mathcal{D}$ with respect to $\succeq$. Formally we arrive at a choice function $\mathrm{ch}_{\succeq} : \Sigma \to \Sigma$ defined by setting

$$\mathrm{ch}_{\succeq}(\mathcal{D}) := \Big\{ Y \in \mathcal{D} : \nexists X \in \mathcal{D} \text{ such that } (X, Y) \in P_{\succeq} \Big\} \tag{1.2}$$

for all $\mathcal{D} \in \Sigma$.[12] Note that, by construction, choice functions induced by some possibly incomplete binary relation only allow for the weak interpretation according to the classification discussed above: The elements collected in $\mathrm{ch}_{\succeq}(\mathcal{D})$ are, in general, incomparable with respect to the relation $\succeq$ and, therefore, incomparable on the basis of the available information sources. Contrarily, the corresponding reject set $\mathrm{re}_{\succeq}(\mathcal{D}) = \mathcal{D} \setminus \mathrm{ch}_{\succeq}(\mathcal{D})$ possesses a strong interpretation. It contains all the acts from $\mathcal{D}$ that definitely should not be chosen by the decision maker, since for each of them there exists an act in the corresponding choice set $\mathrm{ch}_{\succeq}(\mathcal{D})$ that is strictly better with respect to the relation $P_{\succeq}$.[13]

[12] Here, $P_{\succeq}$ denotes the *strict part* of $\succeq$, that is $(X, Y) \in P_{\succeq} \Leftrightarrow (X, Y) \in \succeq \ \wedge (Y, X) \notin \succeq$.

[13] Formally, the set $\mathrm{ch}_{cr}(\mathcal{D})$ forms a *minimal complete class* w.r.t. $\succeq$ in $\mathcal{D}$ (see, e.g., French and Insua (2000, Section 5.11) or, in a non-statistical context, e.g., Ferschl (1975, Definition 9.3)).

Finally, it could be the case that the agent can best possibly utilize the available information sources neither by specifying a criterion function *cr* nor by specifying some binary relation $\succeq$, but by directly defining a choice function ch: $\Sigma \to \Sigma$. This for instance appears naturally, whenever the agent is able to tell his choice sets for certain subsets of $\mathcal{G}$ that, however, may contain more than just two elements. Note that, in the opposite way as discussed above, such a choice function then naturally induces a binary relation $\succeq_{\text{ch}} \subset \mathcal{G} \times \mathcal{G}$ on the set of available acts $\mathcal{G}$ by ranking an act X at least as preferable as another act Y, whenever X is an element of the choice set of $\{X, Y\}$ with respect to the function ch. Formally, we arrive at $(X, Y) \in \succeq_{\text{ch}}$ if and only if $X \in \text{ch}(\{X, Y\})$ (see, e.g., Chambers and Echenique (2016, Section 2.1) for more details on this construction).[14]

The connection between a choice function and its associated binary relation plays an important role in some parts of the works contributing to this thesis, especially in *Contribution 3*, since it allows us to classify choice functions into two different classes: *global* and *local* ones. A choice function ch: $\Sigma \to \Sigma$ will be called local, whenever it is completely characterized by its associated binary relation $\succeq_{\text{ch}}$ according to the following rule

$$\text{ch}(\mathcal{D}) = \text{ch}_{\succeq_{\text{ch}}}(\mathcal{D}) = \bigcup\nolimits_{X \in \mathcal{D}} \bigcap\nolimits_{Y \in \mathcal{D}} \text{ch}(\{X, Y\}) \tag{1.3}$$

for all $\mathcal{D} \in \Sigma$. If this is the case, one also says that the choice function ch is generated by pairwise choices (see, e.g., Kadane et al. (2004)): Knowing the images of the function for sets containing exactly two elements is equivalent to knowing the complete function. Contrarily, a choice function will be called global, whenever property (1.3) is not satisfied. In such cases, as Seidenfeld et al. (2010) puts it, the choice function does not reduce to pairwise comparisons. See, e.g., van Camp (2018, Section 2.8) for an easy example.

In summary, we have recalled that choice functions are an appropriate concept for formalizing the search for optimal acts in basic decision models. Afterwards, we have briefly looked at three different ways of constructing choice functions: (1) choice functions induced by some criterion function $cr : \mathcal{G} \to \mathbb{R}$ assigning real numbers to every available act, (2) choice functions induced by some, generally not complete, binary relation $\succeq \subset \mathcal{G} \times \mathcal{G}$, and (3) choice functions that are directly determined by information on the agent's choice behavior on certain subsets of $\mathcal{G}$, generally containing more than two elements. However, note that we kept our considerations about when to use which construction method purposely vague: We solely argued to use that method that best possibly uses the available information sources $\mathcal{I}_P$ and $\mathcal{I}_U$. In the next section, we get a bit more concrete about how exactly this can be done and, in the course of this process, recall some classic assumptions on the structure of the information sources $\mathcal{I}_P$ and $\mathcal{I}_U$ and the choice function that result from these assumptions.

[14] Two aspects should be briefly mentioned: First, there are further possibilities of generating a binary relation from a given choice function (see, e.g., Sen (1971, Definitions 5 and 9) for formal definitions). Second, which properties are satisfied by the relation $\succeq_{\text{ch}}$, of course, heavily depends on the properties of the underlying choice function ch. Specifically, if no further constraints on the construction of ch are imposed, in general, the strict part $P_{\succeq_{\text{ch}}}$ of the resulting relation $\succeq_{\text{ch}}$ doesn't even have to be transitive (take any choice function ch satisfying $\text{ch}(\{x, y\}) = \{x\}$ and $\text{ch}(\{y, z\}) = \{y\}$ and $\text{ch}(\{x, z\}) = \{z\}$ for a simple example).

1.3 Classical assumptions about the information sources and the associated choice functions

In the following, we present a number of classical assumptions and theories that have been made concerning the information sources in the wide literature on decision making under uncertainty. There, we mainly distinguish between three types of classical assumptions, namely assumptions on the information source $\mathcal{I}_P$ alone, assumptions on the information source $\mathcal{I}_U$ alone, and simultaneous assumptions on both information sources.

1.3.1 Classical assumptions on the agent's preferences

Maybe the most classical assumption on the structure of the information source $\mathcal{I}_P$ is that it allows the agent to specify a *preference relation* $R \subseteq A \times A$ on the consequence space A.[15] This means that the agent can compare arbitrary pairs of alternatives by preference, and that these pairwise comparisons yield a transitive order of the consequences. Under relatively weak conditions[16], a theorem originally proven by *Cantor* then guarantees the existence of an *ordinal utility function* $o : A \to \mathbb{R}$ with the property $(a, b) \in R$ if and only if $o(a) \geq o(b)$, which moreover is unique up to strictly monotone transformations.[17] Note that the latter implies that utility differences with respect to the function o are meaningless apart from their sign and, therefore, that there is no information on the agent's *strength of preferences* available in this model. In the following, we denote the set of all ordinal utility functions with respect to R by $\mathcal{O}_R$.

The classical assumption on the information source $\mathcal{I}_P$ that captures also strength of preferences is that it allows the agent to specify a *cardinal utility function* $u : A \to \mathbb{R}$ on A. Such function u is characterized by exactly two properties. (1) For all consequences $a, b \in A$ we know that the agent does strictly prefer a to b whenever $u(a) > u(b)$ and that the agent is indifferent between a and b whenever $u(a) = u(b)$. And (2): This function u is unique up to positive linear transformations. Of course, condition (2) implies that there is some information on the strength of the agent's preferences. Such information cannot be captured by assumptions on a binary relation on A alone. There are at least two different classical theories about the information source $\mathcal{I}_P$ that guarantee the existence of a cardinal utility function in the above sense:

von Neumann and Morgenstern's theory (vNM theory)

The agent is not only required to specify a binary relation on the consequence space A, but even on the space $\mathcal{L}_A$ of all *simple lotteries* on A. Here, a simple lottery is a mapping $\ell : A \to [0, 1]$ with the properties $|supp(\ell)| < \infty$ and $\sum_{a \in supp(\ell)} \ell(a) = 1$, where $supp(\ell) := \{a \in A : \ell(a) > 0\}$. A simple lottery ℓ is then interpreted as receiving

[15] A preference relation $R \subseteq X \times X$ on a space X is a *complete* (i.e. $(x, y) \in R \vee (y, x) \in R$ for all $x, y \in X$) and *transitive* (i.e. $(x, y) \in R \wedge (y, z) \in R$ implies $(x, z) \in R$ for all $x, y, z \in X$) binary relation.

[16] The preference relations needs to be *separable* (see, e.g., Gilboa (2009, p. 51)).

[17] See Cantor (1915) or Chambers and Echenique (2016, Theorem 1.1) for a short proof.

consequence $a \in A$ with probability $\ell(a)$. Denote the agent's binary relation on $\mathcal{L}_A$ by $\succsim_{\mathcal{L}_A}$. This relation is then required to satisfy three consistency conditions, namely (i) being a preference relation, (ii) *independence*, and (iii) *continuity*. If all three conditions are satisfied, the famous theorem of von Neumann and Morgenstern states the following: There exists a utility function $u : A \to \mathbb{R}$ that is unique up to positive linear transformation such that for all simple lotteries $\ell_1, \ell_2 \in \mathcal{L}_A$ it holds that

$$(\ell_1, \ell_2) \in \succsim_{\mathcal{L}_A} \iff \sum_{a \in supp(\ell_1)} u(a) \cdot \ell_1(a) \geq \sum_{a \in supp(\ell_2)} u(a) \cdot \ell_2(a) \tag{1.4}$$

Thus, if $\mathcal{I}_P$ allows for a complete, transitive, independent and continuous ordering of the simple lotteries on the consequence space, the agent can be assigned a cardinal utility function. See von Neumann et al. (1944) for the original result, Fishburn (1970) for a very compact proof, and Gilboa (2009) for a less technical discussion of the theorem and its implications for the theory of decision making under uncertainty. Moreover, see Fishburn (1970, Chapter 10) or Fishburn (1982, Chapter 3) for an extension of von Neumann and Morgenstern's result to probability measures instead of simple lotteries (of course, under an adaptation of the vNM axioms to this more general setting).

Krantz, Luce, Suppes, and Tversky's theory

A rather different approach for obtaining a cardinal utility function was followed by the authors in Krantz et al. (1971, Chapter 4).[18] Instead of coding the cardinal part of the information in the source $\mathcal{I}_P$ by means of a relation on the space $\mathcal{L}_A$, here the authors go a more direct way: They assume that the agent can specify a pair (R_1, R_2) of relations, where $R_1 \subseteq A \times A$ is a binary relation on the consequence space A and $R_2 \subseteq (A \times A) \times (A \times A)$ is a binary relation on the set $A \times A$. For $a, b \in A$ the expression $(a, b) \in R_1$ is then interpreted as a being at least as preferable as b. Moreover, the expression $((a, b), (c, d)) \in R_2$ is interpreted as exchanging b by a being at least as preferable as exchanging d by c. Thus, while R_1 models the ordinal part of the agent's preferences in A, the relation R_2 explicitly models a notion of strength of preferences, that is the cardinal part. The authors then show that if both relations R_1 and R_2 are complete and transitive and they further satisfy a list of compatibility axioms, then there exists a function $u : A \to \mathbb{R}$ that is unique up to positive linear transformations that satisfies the following two properties: For all $a, b, c, d \in A$ it holds

$$(a, b) \in R_1 \quad \Leftrightarrow \quad u(a) \geq u(b) \tag{1.5}$$

$$((a, b), (c, d)) \in R_2 \quad \Leftrightarrow \quad u(a) - u(b) \geq u(c) - u(d) \tag{1.6}$$

The result and the concrete necessary assumptions can be found in Krantz et al. (1971, Sections 4.2 to 4.4) or French (1986, Section 3.7). In summary, if the information source $\mathcal{I}_P$

[18] Note that in Krantz et al. (1971), the authors actually deal with measurement theory rather than utility theory. However, as these theories are closely connected, their results can be interpreted in the light of utility theory. This is, for instance, done in Suppes and Winet (1955), Fishburn (1970, Chapter 6), and French (1986, Section 3.7).

allows to specify of a pair of relations (R_1, R_2) as just discussed, the agent can be assigned a cardinal utility.

1.3.2 Classical assumptions on the agent's beliefs

Let us turn to information source $\mathcal{I}_U$ now, that is the information about the mechanism generating the states of nature $s \in S$. Here, the most common assumption is that the information in $\mathcal{I}_A$ is sufficiently structured in order to be able to specify a probability measure π on the measurable space S (see for instance Luce and Raiffa (1957, Section 2.4), Ferschl (1975, Part 2), and Rüger (1999, Section 2.4)). Clearly, there are several theories about the information source $\mathcal{I}_U$ that give rise to such a probability measure:

Ideal stochasticity

Under ideal stochasticity, we know, for instance by substance matter considerations, that the states of nature are generated by a perfect random mechanism with known objective probabilities given by the probability measure π on the state space S. For example, the states could be generated by some phenomenon that has been repeatedly observed for a very long time and whose relative frequencies of ending up in a certain state can be regarded as sufficiently good estimates for the true probabilities. Of course, this assumption about the information source $\mathcal{I}_U$ is of particular interest for the field of *statistical decision theory* (see Section 2.2 for some more information and hints to relevant literature on this discipline). Probably one of the first researchers who investigated decision problems under ideal stochasticity (namely in gambling) was Daniel Bernoulli (see Bernoulli (1954) for an English translation of the original (Latin) manuscript).

Ideal probabilistic beliefs

The agent's *beliefs* about the occurrence of the different states of nature can be characterized by a probability measure π on the state space. This sometimes is referred to as (first) *paradigm of Bayesian statistics* (see, e.g. Augustin (2003)). Note that this assumption for instance corresponds to the classical framework of subjective probability as pioneered by de Finetti (see de Finetti (1974)): The information in $\mathcal{I}_U$ needs to be well-behaving enough that the agent can specify a binary relation $\succsim_F$ on the set $\mathbb{R}^S$ of all real-valued functions $f : S \to \mathbb{R}$, where every $f \in \mathbb{R}^S$ is interpreted as a bet winning $f(s)$ units of money if $s \in S$ occurs. If this relation satisfies a number of consistency properties, then the agent can be shown to make comparisons between the different bets by comparing their expectations with respect to a uniquely defined probability π on the states (see (Gilboa, 2009, Chapter 9) for details). Alternatively to the betting approach of de Finetti, in Kraft et al. (1959) the authors prove that if $\mathcal{I}_U$ allows the decision maker to specify a *qualitative probability*[19] on the subsets of a finite state space S that satisfies some additional conditions, then the agent can be ascribed a uniquely defined probability measure. For infinite (more

[19] A qualitative probability is a binary relation on the subsets on S that satisfies certain properties capturing the intuition of a plausibility ranking, see Kreps (1988, p. 118) for a definition.

precisely uncountable) state spaces S, conditions on a qualitative probability that allow for specifying the agent's unique probability are, for instance, discussed in Kreps (1988, Chapter 8).

Note that if S is assumed to be finite and if a probability measure π on the state space together with a von Neumann-Morgenstern utility function $u : A \to \mathbb{R}$ is available, then the theorem of von Neumann and Morgenstern teaches us that there is only one way of defining a choice function on $\mathcal{G}$. To see that, first note that, given π, every act $X : S \to A$ from $\mathcal{G}$ is uniquely associated with a simple lottery ℓ_X from $\mathcal{L}_A$ by setting $\ell_X(a) = \pi(X^{-1}(\{a\}))$ for all $a \in A$. The vNM-theorem then tells us that, for two acts $X, Y \in \mathcal{G}$, the agent should prefer the lottery ℓ_X associated with X to the lottery ℓ_Y associated with Y if and only if

$$\mathbb{E}_\pi(u \circ X) = \sum_{a \in supp(\ell_X)} u(a) \cdot \ell_X(a) \geq \sum_{a \in supp(\ell_Y)} u(a) \cdot \ell_Y(a) = \mathbb{E}_\pi(u \circ Y) \tag{1.7}$$

Thus, if one accepts the (hardly deniable) assumption that a decision maker should prefer act X to act Y whenever she prefers ℓ_X to ℓ_Y (and vice versa), this implies that the decision maker necessarily ranks the acts according to the criterion function $cr_{u,\pi} : \mathcal{G} \to \mathbb{R}$ defined by $cr_{u,\pi}(X) = \mathbb{E}_\pi(u \circ X)$ for all $X \in \mathcal{G}$, which is then well-defined since u is unique up to positive linear transformations and the expectation operator is a linear functional. Taking into account our considerations about the construction of choice functions from Section 1.2, this leads us to the choice function $\mathrm{ch}_{cr_{u,\pi}} : \Sigma \to \Sigma$ defined by

$$\mathrm{ch}_{cr_{u,\pi}}(\mathcal{D}) := \Big\{ Y \in \mathcal{D} : \mathbb{E}_\pi(u \circ Y) \geq \mathbb{E}_\pi(u \circ X) \text{ for all } X \in \mathcal{D} \Big\} \tag{1.8}$$

for all $\mathcal{D} \in \Sigma$. Obviously, the choice function $\mathrm{ch}_{cr_{u,\pi}}$ selects from every subset $\mathcal{D}$ of $\mathcal{G}$ exactly these acts that maximize the expected utility with respect to the decision maker's uniquely defined cardinal utility function and the externally given unique probability measure π.[20] Note that, if the cardinal utility is obtained from a pair of relations (R_1, R_2) as discussed in the second approach above, then applying the choice function $\mathrm{ch}_{cr_{u,\pi}}$ is no longer a direct implication from the axioms. However, it still is a very plausible (and well-defined) selection rule.

Contrarily, if solely an ordinal utility representation $o : A \to \mathbb{R}$ induced by some preference relation $R \subseteq A \times A$ is available, applying Equation (1.8) will generally not lead to a well-defined choice function: It is easy to construct situations in which there exist $X, Y \in \mathcal{G}$ and strictly monotone transformations $o_1, o_2 \in \mathcal{O}_R$ of o such that $\mathbb{E}_\pi(o_1 \circ X) > \mathbb{E}_\pi(o_1 \circ Y)$ but $\mathbb{E}_\pi(o_2 \circ X) < \mathbb{E}_\pi(o_2 \circ Y)$. A straightforward fix of this problem is the following: Label X superior to Y, whenever $\mathbb{E}_\pi(o \circ X) \geq \mathbb{E}_\pi(o \circ Y)$ for all $o \in \mathcal{O}_R$. This induces

[20] Note that there exist generalizations of von Neumann and Morgenstern's framework to non-simple lotteries, i.e. probability measures (see , e.g., Fishburn (1970, Chapter 10) and Kreps (1988, Chapter 5) or Delbaen et al. (2011)). Using these, the construction of the choice function in (1.8) can be done in a completely analogous way as just discussed, however, without the assumption of a finite state space S.

a binary relation on $\mathcal{G}$ which (essentially) coincides with classical first order stochastic dominance (see, e.g., Lehmann (1955) or Mosler and Scarsini (1991)). The choice function $\mathrm{ch}_{SD(\pi)} : \Sigma \to \Sigma$ induced by this relation then has the form

$$\mathrm{ch}_{SD(\pi)}(\mathcal{D}) := \left\{ Y \in \mathcal{D} : \nexists X \in \mathcal{G} \text{ such that } \mathbb{E}_\pi(o \circ X) > \mathbb{E}_\pi(o \circ Y) \text{ for all } o \in \mathcal{O}_R \right\} \quad (1.9)$$

for all $\mathcal{D} \in \Sigma$. Clearly, the choice sets induced by this choice function do not allow for the strong interpretation. Since the relation between the acts is generally not complete, the acts contained in the choice set $\mathrm{ch}_{SD(\pi)}(\mathcal{D})$ are simply not comparable given the information. Again, one should rather go for an interpretation of the rejection sets: An act that no expectation maximizing agent compatible with the relation R would pick is definitely not a worthy choice. Moreover, it should be mentioned that, given a precise probability π along with an ordinal utility o, using the choice function (1.9) is not without alternative: A competing way of proceeding would be to order the acts from $\mathcal{G}$ by *statistical preference* (see, e.g., Montes et al. (2014a)) and then define the corresponding version of (1.9) accordingly. However, note that, in general, ordering the available acts by statistical preference need not produce a transitive relation (see Montes (2014, Remark 2.17)).

1.3.3 Classical simultaneous assumptions on the information sources

Until now, we have only recalled a number of classical approaches that make assumptions on both information sources separately, that is assumptions on the decision maker's preferences separate from assumptions about the uncertainty about the mechanism generating the states of nature. However, in the classical literature on decision theory there also exist a number of approaches making simultaneous assumptions on both information sources. Most prominent, or at least most relevant for our purposes, are here the approaches of *Savage* (see Savage (1954)) and *Anscombe and Aumann* (see Anscombe and Aumann (1963)). We now briefly recall the basic idea of both these approaches and briefly comment on the choice function they naturally induce.

Savage's theory

In Savage's theory, the joint information in the sources $\mathcal{I}_P$ and $\mathcal{I}_U$ is assumed to provide enough structure that the decision maker is able to specify a preference relation $\succsim_S$ on the space $A^S = \{X : S \to A\}$ of all potential acts (not only the available ones). If this relation satisfies a list of seven consistency conditions, namely Savage's famous, yet heavily discussed, axioms P1 to P7, then Savage shows that there exists a unique (finitely additive) probability measure π on the state space and a function $u : A \to \mathbb{R}$ that is unique up to positive linear transformations such that for all acts $X, Y \in A^S$ it holds that

$$(X, Y) \in \succsim_S \iff \int_S u \circ X d\pi \geq \int_S u \circ Y d\pi \quad (1.10)$$

where the integral in the equation is the integral with respect to finitely additive measures

as defined in Fishburn (1970).[21] Similarly as in the case of a vNM-utility together with a known precise probability, also here exists only one meaningful way of defining a choice function on the set of available acts $\mathcal{G}$: According to above, the decision maker necessarily ranks the act by the criterion function $cr_S : \mathcal{G} \to \mathbb{R}$ defined by $cr_S(X) = \int_S u \circ X d\pi$, where u is a utility function and π is the unique probability measure obtained by the decision maker's preferences on A^S. Note that the criterion is well-defined since u is unique up to positive linear transformations and the integral is linear. According to our considerations about choice functions in Section 1.2 this criterion function induces a choice function $\mathrm{ch}_{cr_S} : \Sigma \to \Sigma$ by applying the construction principle from Equation (1.1). Note, however, that Savage's theory has a severe drawback: His axioms imply the state space to be uncountable (see, e.g., the discussion in Fishburn (1970, p. 193)).[22] Thus, it does not teach us a lot about decision problems with finite or countable state spaces.

Anscombe and Aumann's theory

In Anscombe and Aumann's theory, the main objects are so-called *horse lotteries*. A horse lottery brings together the ideas of von Neumann and Morgenstern's and Savage's theories: It is a mapping h from the state space S into the set of the simple lotteries, formally $h : S \to \mathcal{L}_A$. We denote the set of all horse lotteries by $\mathcal{H}_{(A,S)}$. The information in the sources $\mathcal{I}_P$ and $\mathcal{I}_U$ is then assumed to provide enough structure for the decision maker to be able to specify a preference relation $\succsim_{AA}$ on the set $\mathcal{H}_{(A,S)}$ that satifies four consistency conditions, namely the axioms of Anscombe and Aumann. If these axioms are satisfied, Anscombe and Aumann show that there exists a utility function $u : A \to \mathbb{R}$ that is unique up to a positive linear transformation as well as a unique (finitely additive) probability measure on the state space S such that for all $h_1, h_2 \in \mathcal{H}_{(A,S)}$ it holds that

$$(h_1, h_2) \in \succsim_{AA} \iff \int_S \mathbb{E}_{h_1(s)}(u) d\pi(s) \geq \int_S \mathbb{E}_{h_2(s)}(u) d\pi(s) \tag{1.11}$$

where $\mathbb{E}_{h(s)}(u) = \sum_{a \in supp(h(s))} u(a) \cdot h(s)(a)$ for arbitrary horse lotteries $h \in \mathcal{H}_{(A,S)}$. Note that this representation result already induces a uniquely defined choice function on the set $\mathcal{G}$ of available acts: Since every act $X \in \mathcal{G}$ is uniquely associated with a (degenerated) horse lottery $h_X \in \mathcal{H}_{(A,S)}$ by setting $h_X(s)(a) = \mathbb{1}_{\{X(s)\}}(a)$ for all $a \in A$ and $s \in S$, the theorem of Anscombe and Aumann implies that the agents ranks the acts according the criterion function $cr_{AA} : \mathcal{G} \to \mathbb{R}$ that is defined by setting $cr_{AA}(X) = \int_S \mathbb{E}_{h_X(s)}(u) d\pi(s)$, which further simplifies to $cr_{AA}(X) = \int_S u \circ X(s) d\pi(s)$. The resulting choice function $\mathrm{ch}_{cr_{AA}} : \Sigma \to \Sigma$ is then again induced by applying the construction principle from Equation (1.1). Importantly, note that in the theory of Ancsombe and Aumann, contrarily to Savage's theory, the state space may be countable or even finite. However, this has the

[21] This integral coincides with the *natural extension* of a linear prevision as discussed in Walley (1991) and is also discussed in Denneberg (1994). See moreover de Cooman et al. (2008) for a work on a unifying approach to integration theory with respect to finitely additive probability measures.

[22] See, however, Hens (1992) and Gul (1992) for different modifications of Savage's original axioms that allow for finite state spaces.

price that a decision maker having well-behaving preferences on the set of all horse lotteries may seem like an even stronger assumption than having the same on the set of all potential acts (compare also the hypothetical example discussed in Chapter 2.1 of this thesis).

1.4 Selected relaxations of the classical assumptions

Up to now, we have seen a number of classical assumptions that can be made about the information sources in a decision making problem in order to obtain appropriate choice functions on the set of available acts. In all cases except for the function $\mathrm{ch}_{SD(\pi)}$ from (1.9), the corresponding choice functions even allowed for the stronger of the two common interpretations: The acts selected by them form a set of equally best acts between which the decision maker is truly indifferent. However, all of these approaches required pretty strong consistency conditions about the structure of the information sources, most commonly in the form of a well-behaving (i.e. compatibility with certain axioms) binary relation on some set (for instance on $\mathcal{L}_A$, on $\mathbb{R}^S$, or on $\mathcal{H}_{(A,S)}$) that is somehow connected to the consequence space A or the state space S or both. Of course, this gives rise to the question: What if the information sources are not as well-behaving? What if the information contained in them is simply not enough to be sure that any of the axiom systems mentioned earlier can clearly assumed to be satisfied? Indeed, there exits much work questioning the classical assumptions[23] and, accordingly, much theory about relaxing the approaches discussed in Section 1.3.

Let's start by the relaxation of the assumptions on the source $\mathcal{I}_P$ alone: In Aumann (1962), the author generalizes the framework of von Neumann and Morgenstern to the case where the agent's preference relation on the set $\mathcal{L}_A$ of simple lotteries no longer needs to be complete, i.e. where not all simple lotteries have to be comparable. The remaining axioms stay the same. Aumann then goes on showing that if this modified set of axioms is satisfied, then there exists a (generally non-unique) utility function $u : A \to \mathbb{R}$ such that the equivalence in (1.4) is still satisfied in the $\Rightarrow$ direction. Specifically, if there is a precise probability measure π on the states available, we still can define a choice function in the style of the one from (1.8). However, the resulting choice sets do not allow for the strong, but only for the weak interpretation of choice sets: The selected acts are not all equally best, but only incomparable given the information source $\mathcal{I}_P$. We rather interpret the rejection sets as sets of acts that are not a worthy choice.

Also the classical theories about the information source $\mathcal{I}_U$ have been heavily questioned. Here, the first main direction are direct generalizations of de Finetti's behavioral approach of preferences over bets by allowing that the agent's supremum buying prices for bets may

[23] The criticism ranges from works (see, e.g., Luce (1956), Beja and Gilboa (1992) and Nishimura and Ok (2016)) questioning the transitivity of the agent's indifference relation to works (see, e.g., Ellsberg (1961)) that doubt the empirical validity of Savage's axioms. Also certain axioms of von Neumann and Morgenstern theory have been criticized for not being realistic (see the discussions in Luce and Raiffa (1957, p. 23-31) or the *Allais paradox* in Allais (1953)).

differ from her infimum selling prices (in de Finetti's theory these have to coincide).[24] If this is the case, the agent can no longer be ascribed a precise probability on the state space but, very roughly, assigns each event $E \subseteq S$ the interval $[\underline{p}_E, \overline{p}_E]$ ranging from her supremum buying price $\underline{p}_E$ to her infimum selling price $\overline{p}_E$ for the bet $\mathbb{1}_E$. The second main direction is to assume that $\mathcal{I}_U$ allows for the specification of a generalized probability assignment, a so-called *interval probability*.[25] Specifically, such interval probability is a mapping from the subsets of S to the set of closed intervals contained in $[0, 1]$ that satisfies a generalized version of the classical Kolmogorov axioms for probabilities. Importantly, note that interval probabilities are free from any (behavioral) interpretation: For their definition it does not matter whether the interval assignments result from subjective preferences over bets or from some partially known frequentist phenomenon.[26] This framework therefore also allows for modeling information sources $\mathcal{I}_U$ arising from partially known random phenomena. Finally, note that both ways of generalizing probability formally (essentially) correspond to models using suitable *sets of probabilities* on S.[27]

Also the horse lottery framework of Anscombe and Aumann (see Anscombe and Aumann (1963)) and, therefore, simultaneous modeling both information sources, has been generalized in several directions. In Schmeidler (1989) (building on results from Schmeidler (1986)), the author weakens Anscombe and Aumann's independence axiom by demanding independence only for such horse lotteries that are *comonotonic*[28]. He then shows a representation result looking very similar as the one given Equation (1.11), with the difference that, instead of a probability measure π, there now exists a uniquely defined *capacity*[29] v and the integral in the representation result is replaced by the Choquet integral with respect to this capacity v. A choice function for our basic decision model $\mathcal{G}$ can then be constructed completely similar as in the classical Anscombe and Aumann case, with the difference that the corresponding criterion function now uses the representation via the Choquet integral (see Choquet (1954) for the original source).

An even weaker version of Anscombe and Aumann independence axiom, has been proposed in Gilboa and Schmeidler (1989), where independence was only assumed for mixtures with constant horse lotteries. Here, in addition to the remaining axioms of Anscombe and Aumann, a further axiom is required: ambiguity aversion. The authors then go on showing that if a binary relation $\succsim_{GS}$ on the horse lotteries $\mathcal{H}_{(A,S)}$ satisfies this modified (and extended) set of axioms, then there exists a unique (closed and convex) set $\mathcal{M}$ of probability measures on the states S and a utility function $u : A \to \mathbb{R}$ that is unique up

[24] See Williams (1975, 1976, 2007) for pioneering approaches in this direction which have been extended to a complete mathematical and philosophical theory in Walley (1991). See moreover Walley (2000) for a survey.

[25] See Weichselberger and Pöhlmann (1990) and Weichselberger (2001) for textbooks or Weichselberger (2000) for a compact presentation of the basic concepts. See moreover Augustin and Seising (2017) for an overview on Weichselberger's contribution to imprecise probabilities.

[26] For a frequentist approach to interval valued probability assignments see Walley and Fine (1982).

[27] For other approaches that use such sets as their model primitive see, e.g., Levi (1974) or Kofler and Menges (1976). This also is the approach followed in most parts of our *Contributions 1* to *4*.

[28] See, e.g., Gilboa (2009, Section 16.4) for a definition.

[29] See Destercke and Dubois (2014b, Section 4.2) for a definition.

to positive linear transformations such that

$$(h_1, h_2) \in \succsim_{GS} \iff \min_{\pi \in \mathcal{M}} \int_S \mathbb{E}_{h_1(s)}(u) d\pi(s) \geq \min_{\pi \in \mathcal{M}} \int_S \mathbb{E}_{h_2(s)}(u) d\pi(s) \tag{1.12}$$

A choice function for our basic decision model $\mathcal{G}$ can then be constructed completely similar as in the classical Anscombe and Aumann case, with the difference that the corresponding criterion function now uses the representation via the minimum over all integrals with respect to measures from $\mathcal{M}$. This choice function will play an important role from a more applied point of view in our *Contribution 1.*

Besides the independence axiom, mainly one other of the primitives of Anscombe and Aumann's model has been relaxed, namely the completeness of the relation. Here, three works seem particularly relevant: In Seidenfeld et al. (1995) the authors propose a modified set of Anscombe and Aumann's axioms allowing for a strict partial order on the set $\mathcal{H}_{(A,S)}$ of horse lotteries. They then establish a notion of indifference on this space that allows them to distinguish between incomparable and equivalent horse lotteries. Relying on this indifference relation, they then show that the original strict partial relation can be extended to a preference relation (while preserving the remaining axiom) for which there exists an unique utility function $u : \mathcal{H}_{(A,S)} \to [0, 1]$ on $\mathcal{H}_{(A,S)}$ that respects the original partial order for constant horse lotteries (i.e. simple lotteries) and "almost" respects it on all horse lotteries (see (Seidenfeld et al., 1995, Theorem 3)). Since the extension of the partial order is non-unique, there exists a different uniquely defined utility function for every extension. The authors show that the set of all such utility functions is convex and "almost" characterizes the original partial order by applying a Pareto type principle (see their Theorem 5). Moreover, they show that the set of utilities can be decomposed as a set of probability and state-dependent utility pairs (see their Corollary 4.1) and give two additional axioms under which the state-dependent utility can be replaced by a so-called *almost state-independent utility* (see their Theorem 6).

In Nau (2006), the author goes a slightly different way. Instead of defining a notion of indifference, he treats incomparability of horse lotteries completely serious and judges that "the point of dropping the completeness assumption is precisely to permit the decision maker to judge some alternatives to be noncomparable" (see Nau (2006, p. 2443)). In order to obtain his representation result, he therefore proposes an alternative modification of the Anscombe and Aumann axioms, namely a strengthened version of their independence requirement. Utilizing this, the author manages to show that a partial order on the horse lotteries satisfying his axioms is represented by a set of utility functions which is the convex hull of a set of utility functions that are decomposable by a pair of a probability and a state-*independent* utility (see Nau (2006, Theorem 4)). Finally, in Galaabaatar and Karni (2013), the authors choose another modification of the classical Anscombe and Aumann axioms. The main conceptual difference to the approach in Nau (2006) is that their axioms allow for representation results very explicitly separating the sets of probabilities and utility functions that are involved in the representation of the partial order. For the corresponding representation results see, in particular, Galaabaatar and Karni (2013, Theorems 3 and 4).

1.5 Some short comments about group decision making

Up to now, we throughout implicitly assumed that the decision has to be made by *one single* agent. Clearly, all the theory discussed trivially extends to groups of decision makers in which all the members share identical preferences on A and identical beliefs about the mechanism generating the states $s \in S$. However, if this is not the case, a generalization of the theory is far from trivial, but presents a whole field of its own: *group decision making*. Both preferences and beliefs have to be appropriately *aggregated* before any decision can be made by the group.

In this thesis, we deal with the aggregation of preferences rather than the aggregation of beliefs.[30] Specifically, suppose that there is a group G_n of decision makers consisting of n members. Each of the group's members has stated a (possibly non-complete) relation $R_i \subset A \times A$ on the consequence space A, expressing her personal opinion about how the consequences should be ranked. The task is then to aggregate the tuple of relations $\underline{R} = (R_1, \ldots, R_n)$ to some *consensus order* $S(\underline{R})$ that each of the group members considers fair. Once having obtained such a consensus order, the group can make decisions by simply acting like a single agent with preferences given by $S(\underline{R})$. However, finding good formal criteria for defining the term "fair" in the first place has proven a cumbersome task, since already seemingly very weak requirements lead to very general impossibility results (like most prominently the one of Arrow (1950)). In *Contribution 4* of this thesis, we investigate whether information on the considered group's homogeneity can ease the choice of an appropriate aggregation rule (see Section 3.4).

Finally, it should be mentioned that (at least formally) there is a close connection between group decision making and individual decsion making with partial preferences: If we, given the group member's relations $R_1, \ldots, R_n$, define a new binary relation $\tilde{R} \subset A \times A$ on A by setting $(a, b) \in \tilde{R}$ if and only if $(a, b) \in R_i$ for all $i = 1, \ldots, n$, that is $\tilde{R} = \cap_i R_i$, then instead analyzing the corresponding individual decision problem with $\mathcal{I}_P$ consisting of $\tilde{R}$ will be in accordance with the preferences of all group members simultaneously. Thus, analyzing decision problems with partial preferences available also teaches us something about group decision problems. However, note that in most situations, especially for very heterogeneous groups (compare also our *Contribution 4*), the relation $\tilde{R}$ will be rather sparse (in the sense of containing only few pairs) or even empty, such that choice functions solely based on $\tilde{R}$ will often not yield a satifactory decision.

[30] For works dealing also with aggregation of beliefs see, e.g., Bacharach (1975) and Seidenfeld et al. (1989).

2 Aim of this work

2.1 The view on decision making in this thesis

In the previous chapter we recalled a variety of (classical and rather sophisticated) theories about the structure of the informations sources $\mathcal{I}_P$ and $\mathcal{I}_U$ coming along with a basic decision model. However, almost all of these theories had a very strong *axiomatic* and, therefore, normative character in the sense that they all involved statements like "if the agent has well-behaving preferences over some very large set of fancy mathematical objects, then there is a straightforward way of defining a choice function suitable for the agent". Most commonly, the resulting choice function is then somehow related to an utility function that adequately represents the agent's preferences and to a probability measure that adequately characterizes the agent's beliefs about the states of nature.

From this point of view, axiomatic theories about the information sources could be argued to be rather interesting for investigating the behavioral implications of applying certain choice functions (like, for instance, expected utility maximization), than for constructing choice functions in the first place. In other words, if an agent decides to make her decisions by maximizing the expected utility with respect to a pair (u, π) of utility and probability, Savage's theorem will tell her what this means for her preferences amongst arbitrary acts. If she isn't contented with the behavioral implications that the choice of (u, π) yields, she should try another pair better suiting her behavior. Hence, axiomatic theories about information sources play a crucial role when it comes to the *falsification* of certain choice functions as adequate descriptions of the agents decision behavior.

Using the representation results in the opposite direction, that is for constructing choice functions rather than falsifying them, usually turns out to be very hard: In reality it will almost never be the case that an agent has fully specified preferences over any of these complicated spaces. This may have several different reasons. First, the spaces the agent is required to articulate her preferences on (most commonly) contain uncountably many different objects, like for instance simple lotteries, horse lotteries, bets, or acts. So, even if in principle the agent of interest was able to order arbitrary pairs of the objects of interest by preference, such task can never be finished in a reasonable time.[1] Often, the best one can do in such situations is to design experiments in which the agent successively ranks such pairs of objects that encode "much" information about the agents preferences on the considered

[1] To make it even more dramatic, note that also if one ignores this "practical" problem, one still has to face the fact that most of the representation results recalled in the previous chapter are pure existence results rather than constructive ones. That is, even in a highly idealized case such results don't teach us too much about, for instance, constructing appropriate utility functions.

space. Stopping this procedure at a certain point will, evidently, not give us a unique probability or utility representation, however, will nevertheless invoke certain constraints on the sets of possible such representations. These considerations directly give rise to the question of how to *elicit* the probability and utility functions of an agent of interest, that is how to design procedures efficiently restricting the set of possible representations by asking the agent of interest meaningful questions. A lot of work has been done in this direction (see, e.g., Savage (1971), Walley (1991, Chapter 4), Gilboa (2010), Smithson (2014), Troffaes and Sahlin (2017) to only name a few).

Second, it may also be the case that certain pairs of objects are simply incomparable for the agent, might it be for substance matter or for cognitive reasons (see also the discussion at the beginning of Chapter 1.4 of the present work). Such concerns are surely not too far-fetched: Suppose the consequence space consists of all the 106 different dishes on the menu of some Vietnamese restaurant, all the 6523 different books in a German book store, and all the 27 études by Frédéric Chopin. How do you like the horse lottery giving you, in case of sunshine, *Pho Bo* with probability 0.1, Fontane's *Effi Briest* with probability 0.5, and Chopin's *The Horseman* with probability 0.4 and, in case of no sunshine, Ende's *Momo* with probability 0.6, Nabokov's *Pnin* with probability 0.05, and Chopin's *Toccata* with probability 0.35? Do you prefer it to the horse lottery yielding, in case of sunshine, *Pho Ga* with probability 0.1, *Nem Cuon Tom* with probability 0.1, Kracht's *Imperium* with probability 0.5, and Chopin's *Chromatique* with probability 0.3 and, in case of no sunshine, *Banh Xeo* for sure? (If you feel that you can indeed compare these horse lotteries, just extend the consequence space or the state space or both until you can't.) See moreover Trautmann and Wakker (2018) for critical discussion of the Anscombe and Aumann axioms.

Independent of the concrete reason for the agent not having fully specified preferences and probabilities on the spaces of interest (partial elicitation or incomparability or both), both of the shortcomings mentioned above do, in general, not leave us with unique utility or probability representations of an agent's preferences and beliefs. Instead, in such situations we will rather obtain sets $\mathcal{U}$ of compatible utility representations[2] as well as sets $\mathcal{M}$ of compatible probability measures.[3] Consequently, also the construction of adequate choice functions in such situation will have to be based on the the sets $\mathcal{U}$ and $\mathcal{M}$ alone. This is where the representation results recalled in this thesis' Chapter 1 play an important role: They are viewed as ideal types to be used if information was perfectly structured. Specifically, as most of these theories more or less directly result in maximizing the expected utility of acts for some pair (u, π) of utility and probability, the choice functions based on the pair of sets $(\mathcal{U}, \mathcal{M})$ should generalize this idea in the sense that they reduce to it

[2]Such sets are commonly termed *multi-utility representations*, see, e.g., Dubra et al. (2004) or Evren and Ok (2011). An agent's partial preferences are said to be represented by the set $\mathcal{U}$, if he prefers a consequence to another one, whenever the first one receives a greater or equal real value by every member of $\mathcal{U}$.

[3]Such sets are often (see, e.g., Levi (1974, 1980)), and also in the contributions of this thesis, termed *credal sets*. These, in particular, arise as central components of all theories summarized under the umbrella term *imprecise probabilities*. See also the discussion in Chapter 1.4 of the present thesis.

whenever $\mathcal{U}$ and $\mathcal{M}$ turn out to be singletons. Apart from this "minimal requirement", there is quite much freedom of definition left, depending among other things on the attitude of the decision maker towards the complete ignorance between the elements of $\mathcal{U}$ and $\mathcal{M}$, sometimes called *ambiguity* (see, e.g., Ellsberg (1961) in particular) in this context. Consequently, many different proposals for defining choice functions in such situations have been discussed. These range from Levi's criterion of *E-admissibility*[4] over Walley's *maximality*[5] and Bewley's *structural dominance* (see Bewley (2002) or Etner et al. (2012, p. 243)), to the *Gamma-Maximin* and the *Imprecise Hurwicz* criterion (to only name some).[6]

Almost as important as defining meaningful criteria for the weakly structured settings just discussed, is the development of efficient algorithms that are capable to evaluate them. In many situations, *linear programming theory*[7] turns out to offer an adequate toolbox for doing so. This is also one of the major interests of the present dissertation: We intend to obtain decision criteria that are still applicable in situations with very weakly structured information and to provide evaluation algorithms guaranteeing the practical usefulness of these criteria. Of course, much work in this direction has already been done. Of particular interest here seem to be the following works: In Utkin and Augustin (2005) and Kikuti et al. (2011) the authors propose algorithms for decision making under different decision criteria with sets of probabilities. In Utkin and Kozine (2001), linear programming approaches for characterizing natural extensions are discussed, while in Nakharutai et al. (2017) the authors present efficient linear programs for checking avoiding sure loss. In Hable and Troffaes (2014), the authors collect algorithms for a number of concepts relevant to the theory of imprecise probabilities.

2.2 Statistics and Decision Theory

Finally, we now briefly comment on the interrelations of the theory of decision under uncertainty as presented here and classical problems from the field of statistics. There are at least two different points of view. The first, more direct, interrelation is to view statistical methodology as a toolbox that might help to gain insights on the mechanism generating the states of nature in some decision problem by collecting suitable data. Of course, this view is limited to decision problems in which this state generating process

[4] In the presence of a cardinal utility function u, the E-admissibility criterion (see Levi (1980)) labels such acts optimal that maximize expected u-utility for at least one probability measure $\pi \in \mathcal{M}$. Note that, in general, the choice function induced by it does not satisfy the locality property (1.3).

[5] If a cardinal utility function u is available, maximality (see Walley (1991)) labels such acts optimal, for which there does not exist another available act that has strictly greater expected u-utility for every probability $\pi \in \mathcal{M}$. The maximality criterion does indeed satisfy the locality property (1.3).

[6] The Gamma-Maximin criterion, relying on the idea of maximizing the expected utility under the worst possible probability from $\mathcal{M}$, can, under the name *Max E Min*, already be found in Kofler and Menges (1976). The first ones to give an axiomatic justification of the criterion were Gilboa and Schmeidler (1989), compare also the discussion in Chapter 1.4 of this thesis. The imprecise Hurwicz criterion is for instance discussed in Huntley et al. (2014).

[7] See, e.g., Vanderbei (2015) for a textbook.

can be related to some observable experiment that can be repeated for a large number of runs under constant conditions. If this is the case, statistical methods can help to obtain reliable estimates for the probabilities of the different states and, therefore, facilitate the problem of choosing optimal acts.

The second point of view originates from a much deeper relation: Most of the classical problems in statistics can formally be embedded into the theory of decision under uncertainty. These embeddings are commonly summarized under the term *statistical decision theory* (see, e.g., French and Insua (2000)) and, in particular, involve *likelihood theory* (see, e.g., Cattaneo (2013)), *regression analysis* (see. e.g., Utkin and Coolen (2011); Cattaneo and Wiencierz (2012)), *Bayesian statistics* (see, e.g., Berger (1980)), *hypothesis testing* (see, e.g., Wald (1949); Augustin (1998)), and *classification* (see, e.g., Utkin et al. (2015)). The common idea underlying these embeddings is quickly explained: Consider a basic decision model $\mathcal{G}$ as defined in Chapter 1.1 of this dissertation together with a cardinal utility function $u : A \to \mathbb{R}$ on the underlying consequence space. Moreover, suppose we observe some random variable $\mathbf{Z}$, our *data*, for which we know that $\mathbf{Z} \sim \xi_s$ if $s \in S$ is the true state of nature. That is, the state space S parametrizes our statistical model for the data $\mathbf{Z}$. Then, instead of the original available acts $X \in \mathcal{G}$, we now consider *decision functions* $d : \mathcal{Z} \to \mathcal{G}$, where $\mathcal{Z}$ denotes the space the random variable $\mathbf{Z}$ maps to. Collect all sensible such function in a set $\mathbb{D}$.[8] We then evaluate the choice of $d \in \mathbb{D}$ under state $s \in S$ by the expression $U(d, s) := \mathbb{E}_{\xi_s}(g_{(d,s)})$, where $g_{(d,s)} : \mathcal{Z} \to \mathbb{R}$ is defined by $g_{(d,s)}(z) = u(d(z)(s))$ for all $z \in \mathcal{Z}$. Thus, a decision function d under state s is evaluated by the expectation of utility that choosing acts with respect to d yields if the data is distributed by ξ_s.[9]

The construction just described demonstrates that, from a purely formal point of view, basic decision models with additional statistical information, that is data, can again be represented by (more complex) basic decision models without additional statistical information. While on the one hand such construction may seem like an unnecessary complication of the original problem, on the other hand it has some advantage that is hard to overestimate: Whatever statement holds true for arbitrary basic decision models, does also hold true for basic decision models with additional statistical information. This should always be kept in mind when reading the contributions of this thesis.

[8] Of course, which functions d are sensible depends heavily on the concrete statistical context. A minimal requirement is to demand measurability of the induced functions $g_{(d,s)}$ for all $s \in S$ (see next line in the main text) for the expectation to be well-defined. Beyond that, in statistical estimations problems one often restricts analysis to *unbiased* decision functions d.

[9] Note that a similar embedding for a slightly modified decision model is also briefly discussed in the last paragraph of Section 2.1 of our *Contribution 2.*

3 About the contributing material: Relations, summaries and outlooks

In this chapter we take a closer look at the five contributions that form the core of this cumulative PhD thesis. For each of the contributions, we start by giving a summary of its contents and main results and, afterwards, add some comments and discuss possible perspectives for future research that is related to them. Additionally, we discuss interrelations between the different contributions whenever this seems suitable. Importantly, note that from now on parts of the present work are referred to as "Chapters", whereas parts of the contributions are referred to as "Sections".

3.1 Decision theory meets linear optimization

3.1.1 Summary of Contribution 1: "Decision theory meets linear optimization beyond computation"

In *Contribution 1*, we discuss linear programming based algorithms for decision making with uncertain precise probabilities (Section 3 of the contribution) and imprecise probabilities (Section 4 of the contribution). Moreover, we discuss a linear programming problem for determining least favorable prior distributions from a given credal set and investigate what can be learned by studying the dual of this linear programming problem.

Throughout *Contribution 1*, we consider decision making problems with a finite state space Θ[1], a finite set of available acts, and a cardinal utility function adequately representing the agent's preferences on the consequence space. In the spirit of the classification from Chapter 1, we thus assume the information in the source $\mathcal{I}_P$ to be structured enough to either satisfy the axioms of von Neumann and Morgenstern or the axioms of Krantz, Suppes, Luce, and Tversky (see Chapter 1.3.1). With the information in the source $\mathcal{I}_P$ structured like that, the decision problem reduces to a much simpler form than the general decision models considered in Chapter 1 of this dissertation, namely to a triplet

$$\mathfrak{A} := (\mathbb{A}, \Theta, u) \tag{3.1}$$

consisting of a finite set of available acts $\mathbb{A} = \{a_1, \ldots, a_n\}$, a finite set of states of nature $\Theta = \{\theta_1, \ldots, \theta_m\}$, and a cardinal utility function $u : \mathbb{A} \times \Theta \to \mathbb{R}$, where $u(a, \theta)$ is interpreted as the (cardinal) utility of choosing act $a \in \mathbb{A}$ given $\theta \in \Theta$ is the true state. Then, every

[1] In this contribution, the state space is denoted by Θ instead of S.

act $a \in \mathbb{A}$ is naturally associated with a random variable $u_a : \Theta \to \mathbb{R}$ on the state space that is defined by $u_a(\theta) := u(a, \theta)$ for all $\theta \in \Theta$.[2]

In some parts of the paper, the concept of randomization plays an important role (see, e.g., Fishburn (1965) or Augustin (2003)). The original decision problem is then extended for randomized acts, which are classical probability measures λ on $(\mathbb{A}, 2^{\mathbb{A}})$. Here, choosing λ is interpreted as implementing act a_i with probability $\lambda(\{a_i\})$. We denote the set of randomized acts by $G(\mathbb{A})$ and extend the original utility function u on $\mathbb{A} \times \Theta$ to a utility function $G(u)$ on $G(\mathbb{A}) \times \Theta$ by mapping each pair (λ, θ) onto $G(u)(\lambda, \theta) := \mathbb{E}_\lambda\left[u^\theta\right]$, where $u^\theta : (\mathbb{A}, 2^{\mathbb{A}}) \to \mathbb{R}$ with $u^\theta(a) := u(a, \theta)$. Every pure act $a \in \mathbb{A}$, is uniquely identified with $\delta_a \in G(\mathbb{A})$, the Dirac-measure in $\{a\}$, and we have $u(a, \theta) = G(u)(\delta_a, \theta)$ for all pairs $(a, \theta) \in \mathbb{A} \times \Theta$. Again, for every $\lambda \in G(\mathbb{A})$ fixed, the extended utility function $G(u)$ is associated with a random variable $G(u)_\lambda$ on Θ by setting $G(u)_\lambda(\theta) := G(u)(\lambda, \theta)$ for all $\theta \in \Theta$. The *randomized extension* of the triplet $(\mathbb{A}, \Theta, u)$ is then given by

$$G(\mathfrak{A}) := \Big(G(\mathbb{A}), \Theta, G(u)\Big) \tag{3.2}$$

In the whole paper, the imprecision underlying the decision situation is assumed to solely arise from the weak structure of information source $\mathcal{I}_U$. Specifically, we mainly distinguish two different scenarios: In Section 3 of the contribution, we consider the case that a precise probability measure π on the state space is available, however, the decision maker (for some reason) doubts its full appropriateness, i.e. the case of *uncertainty about precise probabilities*.[3] For such situations, a decision criterion proposed by Hodges and Lehmann [4] seems to be well-suited: Instead of directly maximizing expected utility with respect to the probability measure π, one maximizes a convex combination of the worst case utility[5] and the π-expected utility. The weight $\alpha \in [0, 1]$ that falls on the expectation part of the criterion is then called the agent's degree of trust, or trust parameter.

This leads us to the first result of the paper (see Proposition 1 of *Contribution 1*): We present an algorithm for determining an optimal randomized act with respect to the criterion of Hodges and Lehmann that requires to solve only one single linear programming problem. Utilizing this, the paper's second result (see Corollary 1) is obtained by considering the dual linear programming problem of the one introduced in Proposition 1. We

[2] A short note on how this framework relates to the more general framework discussed in Chapter 1: Given a finite basic decision model $\mathcal{G} = \{X_1, \ldots, X_n\}$ with acts $X_i : \Theta \to A$ together with a cardinal utility function $\tilde{u} : A \to \mathbb{R}$, one can instead directly consider the finite basic decision model $\tilde{\mathcal{G}} = \{\tilde{X}_1, \ldots, \tilde{X}_n\}$, where $\tilde{X}_i := \tilde{u} \circ X_i : \Theta \to \mathbb{R}$ are now acts that directly map states to utilities. If we now define a function $u : \mathcal{G} \times \Theta \to \mathbb{R}$ by setting $u(X_i, \theta) = \tilde{X}_i(\theta)$, the triplet $(\mathcal{G}, \Theta, u)$ gives as a decision problem just of the structure as the one in Equation (3.1) of the present chapter.

[3] For instance, suppose you have a pretty good idea about the probabilities that a certain experiment under fixed conditions ends up in a certain result. Now you slightly change the experimental setup. The probabilities from the original experiment will still be available, but their appropriateness for the new experiment will depend on how strong the change in conditions actually influenced the experiment.

[4] See Hodges and Lehmann (1952) for the original source and Equation (1) in *Contribution 1* for the adaptation of the criterion into our setup.

[5] Directly maximizing the worst case utility is the idea of *Wald's maximin criterion*, see Wald (1949).

show that an agent's decisions in terms of the decision criterion by Hodges and Lehmann can be reinterpreted as expectation maximal decisions with respect to a modified prior distribution, which is directly obtained as an optimal solution of the dual program.

Subsequently, in Section 4, we assume that $\mathcal{I}_U$ is compatible with a whole credal set[6] of probability measures that are in accordance with our information about the states of nature, i.e. we consider the case of *imprecise probabilities*. Specifically, we assume $\mathcal{I}_U$ to be characterized by a polyhedrical set $\mathcal{M}$ of probability measures on $(\Theta, 2^\Theta)$ of the form

$$\mathcal{M} := \left\{\pi |\ \ \underline{b}_s \leqslant \mathbb{E}_\pi(f_s) \leqslant \overline{b}_s \ \forall s = 1, ..., r\right\} \tag{3.3}$$

where, for all $s = 1, ..., r$, we have $(\underline{b}_s, \overline{b}_s) \in \mathbb{R}^2$ such that $\underline{b}_s \leqslant \overline{b}_s$ and $f_s : \Theta \to \mathbb{R}$, i.e. to be describable by lower and upper bounds for the expected values of a finite number of random variables on the space of states.[7] For this case, many different ways of defining decision criteria exist of which two are of particular interest for the results in the paper: Walley's $\mathcal{M}$-maximality and the Gamma-Maximin criterion with respect to $\mathcal{M}$ (compare also the discussions in Chapter 2.1 of this dissertation).

The next result of the paper (see Proposition 2 of *Contribution 1*) is an algorithm for checking whether a pure act $a_z \in \mathbb{A}$ is $\mathcal{M}$-maximal among the available pure acts by solving one single linear programming problem. Note that Kikuti et al. (2011) also introduce such an algorithm that, however, requires to solve a series of linear programming problems. Afterwards, in Proposition 3, we propose a linear programming approach for determining a *least favorable prior distribution* from $\mathcal{M}$, i.e. an element $\pi^- \in \mathcal{M}$ that yields the minimal maximal expected utility among all elements of the credal set. Utilizing this linear program, we then show a very close connection between least favorable priors and the Gamma-Maximin criterion (see Proposition 4 of the contribution). Specifically, we demonstrate that optimal randomized acts with respect to the Gamma-Maximin criterion can only assign strictly positive probability mass to such acts from $\mathbb{A}$ that are optimal with respect to every least favorable prior distribution. This immediately has the following, very interesting, corollary (see Corollary 2 of the contribution): If there exists a least favorable prior for which there exists a unique optimal act, then randomization is unnecessary if optimality is defined in terms of the Gamma-Maximin criterion.

3.1.2 Comments and perspectives

As already mentioned, contrarily to the approaches in Kikuti et al. (2011), Proposition 2 of the contribution allows to check $\mathcal{M}$-maximality of an act a_z by solving one single linear programming problem. This provides a very nice opportunity for generalizing the algorithm given in the proposition: We can easily add linear constraints controlling that the members of the family of probability measures with respect to which the act a_z dominates the

[6] The name credal set is attributed to Isaac Levi (see Levi (1974) and Levi (1980)).

[7] This essentially is the common framework of decision making under an imprecise probabilistic model as it is used, for instance, in Fishburn (1965), Kofler et al. (1984), Walley (1991) and surveyed in Troffaes (2007) or Huntley et al. (2014).

remaining acts in expectation do not differ from each other too much. Denoting the maximum acceptable deviation of the involves measures by ε, this directly leads to the idea of a new decision criterion, E_ε-*admissibility*, allowing for an adjustable trade-off between $\mathcal{M}$-maximality and E-admissibility with respect to $\mathcal{M}$. This idea is worked out in some more detail in our *Contribution 2* (see Definition 1, Remark 1 and Proposition 2 of the contribution in particular).

A further interesting aspect is the following: In the contribution, we very briefly discuss a close mathematical connection between the decision criterion of Hodges and Lehmann and the Gamma-Maximin criterion (see Footnote 3 of *Contribution 1*). Specifically, if the credal set of interest is given as an ε-contamination model $\mathcal{M}_{(\pi_0,\varepsilon)} := \{(1-\varepsilon)\pi_0 + \varepsilon\pi : \pi \in \mathcal{P}(\Theta)\}$, where $\mathcal{P}(\Theta)$ is the set of all probability measures on $(\Theta, 2^\Theta)$, $\varepsilon > 0$ is a fixed contamination parameter and $\pi_0 \in \mathcal{P}(\Theta)$ is the central distribution[8], then the Hodges and Lehmann criterion with trust parameter $(1-\varepsilon)$ and prior distribution π_0 coincides with the Gamma-Maximin criterion with respect to $\mathcal{M}_{(\pi_0,\varepsilon)}$. This is remarkable from several different perspectives: Firstly, it shows that the algorithm proposed in Proposition 1 of the contribution can indeed also be used (as a simple method) for determining optimal randomized Gamma-Maximin acts with respect to ε-contamination models. This seems of particular interest also in the light of the reformulation of Walley's *Imprecise Dirichlet Model (IDM)* as some suitable ε-contamination model (see Herron et al. (1997)).

Secondly, one can utilize this connection in the opposite direction. If the Hodges and Lehmann criterion is just a special case of the Gamma-Maximin criterion, then of course all results obtained for the latter do also hold for the first. Particularly, this implies Corollary 2 of the contribution: If there exists a least favorable prior from $\mathcal{M}_{(\pi_0,\varepsilon)}$ which yields an unique expectation-optimal act, then randomization does not pay out if optimality is defined in terms of the Hodges and Lehmann criterion with respect to the prior distribution π_0 and the trust parameter $(1-\varepsilon)$.

3.2 Quantifying degrees of E-admissibility

3.2.1 Summary of Contribution 2: "Quantifying degrees of E-admissibility in decision making with imprecise probabilities"

Contribution 2 uses an identical formal setup as *Contribution 1*. Specifically, we again consider finite decision problems of the form (3.1) with a cardinal utility function available as well as their randomized extensions as defined in (3.2). Moreover, throughout the whole paper, the information about the uncertainty about the states of nature is assumed to be characterized by a credal set $\mathcal{M}$ of the form (3.3). After having introduced into the required concepts, the paper is composed of three main parts: In the first part, we introduce a new decision criterion, E_ε-*Admissibility*, that can be viewed as a strengthened version of Walley's maximality or as a weakened version of Levi's E-admissibility criterion,

[8] See Huber (1981, p. 12) for details.

respectively. For this criterion, we also provide an efficient and simple algorithm that is based on linear programming theory. In the second part of the paper, we propose two new measures for quantifying the extent of E-admissibility of a given E-admissible act, i.e. for quantifying the size of the set of probabilities for which the corresponding act maximizes expected utility. Also here, for both measures, we give linear programming algorithms capable to deal with them. In the third part of the paper, we discuss some ideas in the context of ordinal decision theory.[9]

Let us start by summarizing the first main part (Section 3.1 of the contribution): Our first result (see Proposition 1 of *Contribution 2*) is about randomization in the context of Levi's E-admissibility criterion. It states that the set $G(\mathbb{A})_{\mathcal{M}}$ of randomized acts that are E-admissible with respect to $\mathcal{M}$ can be constructed from the sets $\mathbb{A}_\pi$, where $\pi \in \mathcal{M}$, of pure acts that maximize expected utility with respect to π by taking the union over all of the convex hulls of the sets $\{\delta_a : a \in \mathbb{A}_\pi\}$. In particular, this demonstrates that (generally) the sets $G(\mathbb{A})_{\mathcal{M}}$ and $\text{conv}(\mathbb{A}_{\mathcal{M}})$ do not coincide, but that the first is a proper subset of the second. Afterwards, we define a new decision criterion, called E_ϵ-admissibility (see Definition 1 of *Contribution 2*). The idea is to weaken the concept E-admissibility in the sense that there no longer needs to exist one single measure with respect to which the act under consideration simultaneously maximizes expected utility among all available acts, but, similar as for maximality, the measure may differ for every competing act. However, contrarily to maximality, the involved measures may not differ too much from each other, but are restricted to have a maximal distance of $\varepsilon \geq 0$ with respect to some distance measure induced by a suitable norm. Accordingly, for $\varepsilon = 0$ the concept coincides with E-admissibility, whereas for ε sufficiently large we arrive at maximality.[10] In Proposition 2 of *Contribution 2* we provide an algorithm for checking E_ε-admissibility (with respect to the distance induced by the supremum norm) by solving one single linear programming problem. In Example 1 we demonstrate how E_ε-admissibility could be used as a second order decision criterion if none of the optimal acts selected by the original criterion is E-admissible.

In the second part of the paper (Section 3.2 of the contribution), we propose two different approaches for quantifying the extent of E-admissibility of an E-admissible act of interest. For this purpose, we introduce two different measures: The *maximal extent* (see Definition 2 of the contribution) and the *uniform extent* (see Definition 3 of the contribution). For $a \in \mathbb{A}$, denote by $\mathcal{M}_a = \{\pi \in \mathcal{M} : a \in \mathbb{A}_\pi\}$ the subset of measures of the credal set for which act a maximizes expected utility. Both measures relate to characteristics of this set. For an E-admissible acts a, the maximal extent $ext_{\mathcal{M}}(a)$ is defined as the maximal

[9]Note that the focus on discussing concepts related to E-admissibility in this contribution is by no means coincidence. The contribution is part of a Festschrift for Teddy Seidenfeld, who advocates the concept of E-admissibility against other decision criteria (see, e.g., Seidenfeld (2004)). One main reason is that this criterion not solely relies on *pairwise comparisons* of acts, but additionally depends on which other acts are available.

[10]Importantly, note that the concept of E_ε-admissibility makes only sense when considering pure acts, since for randomized acts the concepts of maximality and E-admissibility (and therefore also the concept of E_ε-admissibility) coincide (see Walley (1991, p. 163)).

distance that two measures within the set $\mathcal{M}_a$ can have (with respect to a distance measure induced by some suitable norm). Contrarily, the uniform extent $uxt_{\mathcal{M}}(a)$ measures half the diameter of largest barycentric cube that can be inscribed into $\mathcal{M}_a$. For an illustration of the two concepts and how they differ see Figure 1 and Figure 2 in *Contribution 2*, respectively. Moreover, we provide algorithms for evaluating both measures: Proposition 3 of the contribution gives an algorithm that allows to compute the maximal extent with respect to the maximum norm by a series of linear programming problems. In Remark 3 of the contribution it is briefly discussed how this computation (and also the one with respect to the 1-norm) could be alternatively done by solving one *bi*linear problem. Proposition 4 describes an algorithm for evaluating the uniform extent of an act by solving one single linear programming problem.

In the last main part of the paper (Section 4 of the contribution), we turn to a slightly different setting: the *ordinal* case. We suppose that the utility function u is solely an ordinal representation of the agent's preferences and, therefore, that utility differences are meaningless apart from their sign (compare also the discussions in Chapter 1.3.1). For this setting, we first list a number of decision criteria, classified in global and local as well as precise and imprecise ones. We then mainly focus on one of these criteria, namely the imprecise version of joint stochastic dominance.[11] Specifically, we describe an algorithm that allows for evaluating the criterion by solving one single linear programming problem, given the extreme points of the underlying credal set $\mathcal{M}$ are known. In Section 5 of the paper, we demonstrate the results and algorithms developed in the paper by means of a stylized application example, which underlines their applicability in real world problems.

3.2.2 Comments and perspectives

In the paper, among other things, we proposed different measures for quantifying the extent of E-admissibility of E-admissible acts. Clearly, such measures would be of interest also for other criteria from the theory of decision making with imprecise probabilities. One such criterion is $\mathcal{M}$-maximality. Suppose $a^* \in \mathbb{A}$ is an $\mathcal{M}$-maximal act. Now, what would a good measure for the extent of a^*'s maximality look like? Therefore, first recall that if a^* is $\mathcal{M}$-maximal, then, by definition, for all $a \in \mathbb{A}$ there exists a measure $\pi_a \in \mathcal{M}$ such that $\mathbb{E}_{\pi_a}(u_{a^*}) \geq \mathbb{E}_{\pi_a}(u_a)$. Now, denote the (non-empty) set of all measures from $\mathcal{M}$ with respect to which a^* expectation dominates a by $\mathcal{M}_{(a^*,a)}$, that is

$$\mathcal{M}_{(a^*,a)} = \left\{ \pi \in \mathcal{M} : \mathbb{E}_\pi(u_{a^*}) \geq \mathbb{E}_\pi(u_a) \right\} \tag{3.4}$$

In order to measure the extent of a^*'s maximality, a very natural idea would be using the following procedure: In a first step, we somehow "quantify" the size of the sets $\mathcal{M}_{(a^*,a)}$ for every $a \in \mathbb{A} \setminus \{a^*\}$. Subsequently, in a second step, we use the resulting values for

[11] Note that a generalized version of this criterion is also discussed in our *Contribution 3* under the name $\mathcal{A}$-admissibility, see Definition 7ii) in *Contribution 3*. If the relation R_1 is chosen to be complete and the relation R_2 is chosen to be empty, the two criteria coincide.

constructing a sensible measure. For the first step, one can proceed completely analogous as for measuring the extents of the sets $\mathcal{M}_a$ in *Contribution 2* (see Definitions 2 and 3 of the contribution in particular), namely computing the maximal or the uniform extents of every set $\mathcal{M}_{(a^*,a)}$ separately. For instance, suppose we want to go for the uniform extent. Equation (4) of Defintion (3) of the contribution would then translate to

$$\widehat{uxt}_{\mathcal{M}}(a^*, a) = \max\Big\{\varepsilon : \exists \pi \in \mathcal{M}_{(a^*,a)} \; s.t. \; B_\varepsilon(\pi) \subset \mathcal{M}_{(a^*,a)}\Big\} \tag{3.5}$$

for every $a \in \mathbb{A} \setminus \{a^*\}$, where $B_\varepsilon(\pi)$ denotes the barycentric ε-cube around π exactly as defined in Definition 3 of the contribution. This leaves us with a set of $n-1$ values.

Concerning the second step of the procedure, taking the minimum over all these values seems to be the most plausible choice: The performance of the act a^* should be measured with respect to its toughest competitor (otherwise, one very bad performing act would have too strong influence on the whole measure). Applying this, we would arrive at the following measure for quantifying the extent of a^*'s maximality:

$$\widehat{uxt}_{\mathcal{M}}(a^*) = \min_{a \in \mathbb{A}\setminus\{a^*\}} \widehat{uxt}_{\mathcal{M}}(a^*, a) \tag{3.6}$$

Obviously, for computing the measure $\widehat{uxt}_{\mathcal{M}}(a^*)$, a modified version of the algorithm provided in Proposition 4 of the contribution can be applied. Instead of solving one linear programming problem, one now has to solve $n-1$ such problems, however, less complex ones: The program (5) of the contribution now has to be solved for every act $a \in \mathbb{A}$ separately, but the constraints given at bullet point five and six reduce to a number of m instead of nm both times. In summary, we thus have to solve $n-1$ linear programs with $1 + 2mr + 2m$ constraints instead of one linear program with $1 + 2mr + 2mn$ constraints.

Beyond extending the proposed measures to other criteria, there are also other promising directions for further research: In the discussion directly following Definition 1 of the contribution, we argued in favor of the concept of E_ε-Admissibility, since it allows to take into account more than only one expert opinion while simultaneously allowing to control how far the involved experts may differ in opinion. This idea could easily be extended. Firstly, one could control the differences of opinion in both directions. Consider a political decision maker with an advisory body of experts (represented by the credal set). Moreover, suppose the politician intends to make very well-balanced decisions, in the sense that she wants to take into account different points of view, however, also wants to make sure to avoid considering too extreme expert opinions. This would directly lead to a criterion of $E_{(\underline{\varepsilon},\overline{\varepsilon})}$-*admissibility*, where $\underline{\varepsilon}$ is the minimal and $\overline{\varepsilon}$ is the maximal deviation that the involved expert opinions are allowed to have.

Moreover, instead of (solely) controlling how far the involved experts may differ from each other in terms of opinion, one could also control how far their opinions differ from some externally given criterion. If we take again our example of some politician with an advisory body of experts, the external criterion could for instance be the opinion of the politician herself (formulated in terms of some suitable probability measure), so that she only takes

those expert opinions into account that do not differ too much from her own one. Of course other examples for external criteria are imaginable.

Finally, it would be also promising to further investigate some of the criteria that were only sketched in Section 4 of the contribution. While some of these have already been investigated in both their precise and their imprecise version[12], the criterion of *joint statistical preference* (see pp. 15-16 of the contribution) has, to the best of our knowledge, not been considered yet. Currently, a more efficient implementation of the algorithms discussed in the contribution (among other tasks) is worked out by Florian Baier (LMU Munich) as part of a Master's thesis under supervision of our working group.

3.3 Partial ordinal and partial cardinal preferences

3.3.1 Summary of Contribution 3: "Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences"

Contribution 3 uses a significantly more general notion of decision problems under uncertainty as the one used in *Contribution 1* and *Contribution 2*. Exactly as discussed in the introductory Chapter 1.1 of this dissertation, our basic choice objects are now acts $X : S \to A$ that map states collected in some non-empty state space S to consequences collected in some non-empty consequence space A. After a short introduction (Section 1 of the contribution) and a section recalling some fundamental concepts underlying our contribution (Section 2 of the contribution), the paper essentially consists of three main parts: In its Section 3, we introduce our basic method for modeling the information source $\mathcal{I}_A$, so called *preference systems*. Its Section 4 introduces three different approaches for decision making with acts taking values in a preference system by proposing decision criteria based on generalized expectation intervals (Section 4.2), on global comparisons of acts (Section 4.3), and on pairwise (or local) comparisons of acts (Section 4.4). Section 5 of the contribution is devoted to an application of the theory that can be viewed as a prototypical example for a whole class of similar applications.

We start our summary by briefly surveying the main ideas of Section 3 of the contribution. There, we first introduce a very natural and convenient way of formalizing information sources $\mathcal{I}_P$ that may contain both ordinal and cardinal information on the decision maker's preferences on A, however, both possibly partial in nature: *preference systems* (see Definition 1 of the contribution). Formally, a preference system is a triplet $\mathcal{A} = [A, R_1, R_2]$ containing, besides the consequence space A, a pre-order[13] $R_1 \subseteq A \times A$ on the consequence

[12] See in particular our *Contribution 3*, Montes (2014), Montes et al. (2014a), and Montes et al. (2014b).

[13] A pre-order $R \subseteq A \times A$ on a non-empty set A is a binary relation that is both reflexive (i.e. $(a, a) \in R$ for all $a \in A$) and transitive (i.e. $(a, b) \in R$ and $(b, c) \in R$ imply $(a, c) \in R$ for all $a, b, c \in A$). In the following, we associate to every pre-order two further binary relations, the *indifference* part $I_R \subseteq A \times A$ defined by $(a, b) \in I_R$ if $(a, b) \in R$ and $(b, a) \in R$ for all $a, b \in A$ as well as the *strict* part $P_R \subset A \times A$ defined by $(a, b) \in P_R$ if $(a, b) \in R$ and $(b, a) \notin R$ for all $a, b \in A$. While I_R is again reflexive and

space itself and a pre-order $R_2 \subseteq R_1 \times R_1$ on the set of pairs that are in relation with respect to R_1. The relation R_1 models the ordinal part of the information in $\mathcal{I}_P$ and we interpret $(a, b) \in R_1$ as a being at least as desirable as b, $(a, b) \in I_{R_1}$ as a being equally desirable as b, and $(a, b) \in P_{R_1}$ as a being strictly more preferable than b. Contrarily, the relation R_2 models the cardinal part of the information in $\mathcal{I}_P$, i.e. the information on the strength of the agent's preferences. We interpret $((a, b), (c, d)) \in R_2$ as the exchange of b by a being at last as desirable as the exchange of d by c. The interpretation of I_{R_2} and P_{R_2} is alike. One natural interpretation of a preference system is to view it as data on observed choices of the agent. If $(a, b) \in R_1$ this then would mean that the agent once chose a in a situation where she was presented exactly the options a and b. If this behavior has been observed many times, one may even conclude $(a, b) \in P_{R_1}$, whereas if another time the agent chose b in the presence of exactly a and b, one may conclude $(a, b) \in I_{R_1}$. In contrast, R_2 is not directly observable, but needs to be gained by hypothetical comparisons in interviews and polls by asking questions like: "Imagine you have objects b and d. Would you rather be willing to accept the exchange of b by a or the exchange of d by c?"

Of course, a preference system elicited in the manner just described is not guaranteed to satisfy any condition of rationality (the observed agent might behave completely irrational). Therefore, we go on defining a notion of *consistency* of preference systems (see Definition 2 of the contribution) by demanding the existence of at least one utility function $u : A \to [0, 1]$ that represents the preference system in the sense of the conditions i) and ii) of Definition 2. We denote the set of all such representations by $\mathcal{U}_\mathcal{A}$. This leads us to the first result of the paper (see Proposition 1 of the contribution): We propose an algorithm for checking consistency of a preference system, i.e. for checking that $\mathcal{U}_\mathcal{A}$ is non-empty, by solving one single linear programming problem.

In Section 4 of the contribution, we turn to decision problems involving uncertainty. Specifically, we assume that the source $\mathcal{I}_U$ containing the information on the mechanism generating the states of nature is characterized by a credal set $\mathcal{M}$ on the measurable space $(S, \sigma(S))$, where $\sigma(S)$ denotes some suitable σ-algebra. We then define the notion of a *decision system* $\mathcal{G}$ (see Definition 4 of the contribution) as a subset of the set $\mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ (see Equation (2) of the contribution) of all acts $X : S \to A$ such that $u \circ X$ is $\sigma(S) - \mathbb{B}_\mathbb{R}$-measurable for all $u \in \mathcal{U}_\mathcal{A}$. For such decision systems we then present three different approaches for constructing choice rules, that basically follow the classification of the construction of choice functions discussed in the introductory Chapter 1.2 of this dissertation. In Section 4.2, Definition 5 of the contribution we define a notion of interval expectation that is suitable for our setup, the *generalized interval expectation*. Based on this, we present three different types of criterion functions $cr : \mathcal{G} \to \mathbb{R}$ that can be derived from this generalized interval expectation, given the agent's attitude towards the ambiguity is

transitive and, additionally, symmetric (i.e. $(a, b) \in I_R$ implies $(b, a) \in R$ for all $a, b \in A$), the relation P_R is transitive and asymmetric (i.e. $(a, b) \in P_R$ implies $(b, a) \notin P_R$ for all $a, b \in A$). I_R is thus an *equivalence relation*, while P_R is a *strict partial order*. Importantly, note that since R is not assumed to be complete, there might exist elements $a, b \in A$ for which it holds that $(a, b) \notin I_R$ and $(a, b) \notin P_R$ and $(b, a) \notin P_R$, i.e. elements that are neither ranked equivalent nor can be put in a strict rank order.

known (see Definition 6 of the contribution). Moreover, in Proposition 3 of the contribution, we provide an algorithm for computing the generalized interval expectation of an act by solving a series of linear programs.

In Section 4.3, we introduce a number of concepts of *global* decision criteria (see Definition 7 of the contribution).[14] All of these concepts rely on the idea that, given perfect information, maximizing expected utility should be the criterion of choice. However, they differ in the way they handle the ambiguity underlying the involved sets $\mathcal{M}$ and $\mathcal{U}_{\mathcal{A}}$, by demanding expected utility maximization for *at least one*, for *certain*, or for *all* pairs (u, π) of compatible utility-probability representations. Proposition 4 provides an algorithm that allows checking an act of interest for optimality with respect to one of these criteria, $\mathcal{A}$-admissibility, by solving one single linear programming problem. A possible algorithm for checking $\mathcal{A}|\mathcal{M}$-dominance is described at the end of the contribution's Section 4.4.

In Section 4.4, we set focus on criteria of *local* admissibility.[15] Therefore, we first define a number of binary relations on the set $\mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ (see Equations (4) to (9) of *Contribution 3*) which are all based on the idea of pairwise expected utility comparisons of acts, however, differ in whether expected utility has to dominate for *at least one*, for *certain*, or for *all* pairs (u, π) of compatible utility-probability representations. The locally admissible acts from $\mathcal{G}$ are the defined as the maximal elements from $\mathcal{G}$ with respect to the corresponding relation restricted to $\mathcal{G} \times \mathcal{G}$ (see Definition 8 of *Contribution 3*). Proposition 5 provides linear programming based algorithms for checking if two acts are in relation with respect to the relations $R_{\exists\exists}$ and $R_{\forall\forall}$. Proposition 6 gives an additional condition for improving the algorithm for $R_{\exists\exists}$. Finally, Section 5 discusses a prototypical example applying our theory where the orders R_1 and R_2 naturally arise from acts that map into some bivariate product space with one cardinal and one (potentially partial) ordinal dimension.

3.3.2 Comments and perspectives

Maybe despite the first impression, *Contribution 3* has a very strong focus on applications in mind. Contrarily to the axiomatic and normative approaches to decision theory that were recalled in the introductory Chapter 1 of this dissertation, we here do not primarily intend to give an axiomatic explanation of how an agent should behave in oder to guarantee some form of expected utility characterization of his choices. Instead, we propose a framework that, on the basis of very weakly structured information, helps the agent to exclude certain choices if he agrees that expectation maximization is in principle the right criterion of choice.

[14] The term "global" is meant in the sense that the choice functions that are induced by the decision criteria introduced in Section 4.3, generally, do not satisfy the locality property (1.3) from Chapter 1.2 of the present dissertation (for the special case of a cardinal utility this is shown in Schervish et al. (2003)). An exception is given by the choice function induced by $\mathcal{A}|\mathcal{M}$-dominance, which satisfies property (1.3). This is also discussed in the very last paragraph of Section 4.4 of *Contribution 3*.

[15] The term "locality" is meant in the sense that the choice functions induced by the criteria of local admissibility from Definition 8 in *Contribution 3* do indeed satisfy the locality condition (1.3) from Chapter 1.2 of this dissertation.

From my point of view, one major issue to be addressed in future research in the context of *Contribution 3* is to make the theory computationally tractable and to apply it to real world data. Clearly, finding datasets that are suitable for our purposes, i.e. data sets with partial ordinal and partial cardinal scaled observations, is not an easy task: Data is always collected in order to be analyzed and so there will be few data sets suitable for trying out new methods, since at the time such data would have been collected, there wouldn't have been any method for analyzing it. It's a pity... Nevertheless, there do exist some datasets that could be modified in order to be suitable for our purposes. For instance in Horwitz et al. (2017), the authors analyze data including (among other para-data variables) also the time a respondent needed to decide between different options. Generally, such kind of data could be utilized as a very nice way of eliciting both relations R_1 and R_2 of a preference system simultaneously. Specifically, suppose we have a finite set $A = \{a_1, \dots, a_n\}$ of consequences. Start with two empty relations $R_1 = \emptyset$ and $R_2 = \emptyset$. Next, you ask the agent successively about her preferences between certain (not necessarily all) pairs of (a_i, a_j) of consequences. There are three possibilities:

i) The agent judges a_i and a_j incomparable. In this case R_1 and R_2 remain unchanged.

ii) The agent considers a_i at least as desirable as a_j. In this case we add (a_i, a_j) to R_1 and note down the time t_{ij} the decision between the two consequences took him.

iii) The agent considers a_j at least as desirable as a_i. In this case we add (a_j, a_i) to R_1 and note down the time t_{ji} the decision between the two consequences took him.

This procedure leaves us with a (potentially partial) relation R_1. Subsequently, we can utilize the time data that we collected during the above procedure. For that, we first fix a threshold $\xi > 0$ below which we do not want to account for time differences. Then, we pick successively pairs of pairs (a_i, a_j), $(a_k, a_l) \in R_1$ and add $((a_i, a_j), (a_k, a_l))$ to our relation R_2 if and only if $t_{kl} - t_{ij} \geq \xi$, i.e. if the decision between a_k and a_l took notably longer than the decision between a_i and a_j. Finally, this procedure produces an preference system whose consistency can be checked by means of Proposition 1 of the contribution.

Concerning the computational aspects, some work has already been done: (Essentially) all linear programming based algorithms introduced in the paper have been implemented in the statistical software R (and applied to the application example in Section 5 of the contribution). However, the code at the moment is far from being efficient and could be improved at several points. For that reason, currently an R package containing (more efficient versions of) these algorithms and several refinements is designed as part of a Master's thesis written by Florian Fendt (LMU Munich) under the supervision of our working group. As another part of this Master's thesis, it is planned to implement a decision aid tool that, for a given (fixed and finite) decision system $\mathcal{G}$, elicits the preference system step by step by asking pairwise preferences (according to the procedure sketched above) however, after each step, checks whether one act is clearly preferable among the other acts. If this is the case, the elicitation procedure can be stopped. Note that decision aid tools relying on a somewhat similar procedure have already been discussed in Danielson and Ekenberg (1998), Danielson et al. (2003), and Danielson (2005).

3.4 Preference aggregation reflecting group homogeneity

3.4.1 Summary of Contribution 4: "A probabilistic evaluation framework for preference aggregation reflecting group homogeneity"

In this contribution we propose a new criterion for group specific evaluation of preference aggregation functions in the presence of partial probabilistic information on the considered group's homogeneity. As a preparation, we discuss a minimal axiomatization for quantifying homogeneity and give a concrete proposal for a homogeneity measure. Finally, we compare a number of preference aggregation functions with respect to our criterion in a simulation study.

Contribution 4 deals with a rather different topic than the *Contributions 1* to *3*, namely the evaluation of methods for preference aggregation. The paper's main object of study are so-called *preference aggregation function*, which are formally mappings

$$S : \mathcal{R}^n \to \mathcal{Q} \tag{3.7}$$

where $\mathcal{R}$ denotes the set of all asymmetric and negatively transitive relations $R \subset C \times C$ on some non-empty and finite set of consequences C and $\mathcal{Q}$ denotes the set of all asymmetric relations $Q \subset C \times C$ on C.[16] Every $\underline{R} = (R_1, \ldots, R_n) \in \mathcal{R}^n$ is called a *preference profile* (or short: *profile*) on C and formalizes one possible way of how the n members of a fixed group G_n under consideration could form their opinions about ranking the elements of C by preference. The order $S(\underline{R})$ is then called *consensus order* with respect to the aggregation function S given the group G_n has constituted the profile $\underline{R}$. Of course, the construction of a meaningful preference aggregation function should not be arbitrary, but needs to be done in a way which all group members of G_n consider fair or appropriate.

Due to the generality of this task, it hardly surprises that many different rules have been proposed since the pioneering works by de Borda (1781), de Condorcet (1785), and Hare (1857) (see Brams and Fishburn (2002) for a survey). Moreover, several rather axiomatic characterizations of "fair" or "adequate" preference aggregations functions have been proposed, like most prominently the approaches of Black (1948), Arrow (1950), Inada (1964) and Sen (1966). However, almost all these are non-group-specific and intended to be valid independently of the concretely considered group. But shouldn't the adequateness of an aggregation procedure, beyond compatibility with non-group-specific criteria, also depend on certain characteristics of the specific group?

We propose a quality criterion for preference aggregation functions that also takes into account information on the group's homogeneity. This involves several steps: First, in Sections 2 of the contribution, we deal with measuring the homogeneity inherent in a preference profile by so-called homogeneity measures $A_n : \mathcal{R}^n \to [0, 1]$. Therefore, in Section 2.3, we begin by stating and discussing a set of three relatively weak conditions

[16] A very similar formal setup is used, e.g., in Ha and Haddawy (2003).

that every reasonable such measure A_n should satisfy (see Definition 1 of the contribution). Subsequently, in Section 2.4, we propose a concrete measure of profile homogeneity, the *maximum consensus homogeneity* δ_n (see Definition 2 of the contribution), show that this measure satisfies the conditions from Section 2.3 (see Proposition 1), and discuss what makes this measure a reasonable choice even beyond its compatibility with these conditions.

Next, suppose that the members of G_n constitute the profiles from $\mathcal{R}^n$ with respect to some precise, yet unknown, probability measure $\mathbb{P}_{G_n}$. This measure represents perfect information of the group's homogeneity structure.[17] The idea for our quality measure for preference aggregation functions is then very simple: We first define a *similarity measure* $Y_S^u : \mathcal{R}^n \to \mathbb{R}$ that, for each profile $\underline{R} \in \mathcal{R}^n$ fixed, measures the similarity of the orders collected in $\underline{R}$ and the corresponding consensus order $S(\underline{R})$ (see Definition 3). To evaluate the quality of procedure S for group G_n we then simply compute the expected similarity of Y_S^u with respect to $\mathbb{P}_{G_n}$, i.e. we define the criterion as $m_{G_n}^u(S) := \mathbb{E}_{\mathbb{P}_{G_n}}(Y_S^u)$ (see Definition 4).

However, since the measure $\mathbb{P}_{G_n}$ is unknown, the criterion cannot be directly evaluated and needs to be approximated best possibly utilizing the available information on G_n's homogeneity structure. To do so, we use the ideas of Section 2 of the contribution and assume that our information can be described as the distribution of some homogeneity measure $A_n : \mathcal{R}^n \to [0,1]$ taking values $k_1 < k_2 < \cdots < k_\xi \in [0,1]$, i.e. by a vector $\alpha := (\alpha_1, \ldots, \alpha_\xi)$ such that $\mathbb{P}_{G_n}(A_n = k_j) = \alpha_j$ for all $j = 1 \ldots, \xi$.[18] This assumption naturally gives rise to a set $\mathcal{M}_\alpha$ of probability measures on $\mathcal{R}^n$ which are all equally plausible candidates for approximating the true measure $\mathbb{P}_{G_n}$ (see Equation (6) of the contribution). Using this set, we then propose two different approaches for approximating our criterion $m_{G_n}^u(S)$: The first one initially obtains the measure $\mathbb{P}_\alpha^*$ from $\mathcal{M}_\alpha$ (see Equation (7) of the contribution) that yields maximum Shannon-entropy (see Jaynes (1957) for the *principle of maximum entropy*) and then approximates the original expectation $\mathbb{E}_{\mathbb{P}_{G_n}}(Y_S^u)$ by $\mathbb{E}_{\mathbb{P}_\alpha^*}(Y_S^u)$ (see Equation (8) of the contribution). The second approach uses the set $\mathcal{M}_\alpha$ as a whole and approximates the expectations $\mathbb{E}_{\mathbb{P}_{G_n}}(Y_S^u)$ by the covering interval ranging from the lowest to the highest expectation of Y_S^u that is compatible with a measure from $\mathcal{M}_\alpha$ (see Equation (9) o the contribution). Finally, in Section 3.4, we discuss different approaches for statistically estimating the probabilities $\alpha_1, \ldots, \alpha_\xi$ from data or by expert knowledge or both. For instance, see Equation (11) of the contribution, this can be done the Dirichlet-Categorical Model (see Berger (1980) or Gelman et al. (2004)).

In Section 4.1, we briefly recall some common preference aggregation rules from literature, namely *Borda count* (see de Borda (1781)), *Condorcet's rule* (see de Condorcet (1785)), *Hare's method* (see Hare (1857)), *Coombs' rule* (see Coombs and Cohen (1984)), and *Kemeny's rule* (see Kemeny (1959) or Kemeny and Snell (1962)). In Section 4.2, we in more detail describe the less common *commonality sharing* rule that was recently proposed

[17] This is meant in the following sense: Suppose we have fixed some homogeneity measure. Then, if we know the probabilities for each profile, we also know whether the group tends to constitute homogeneous or heterogeneous profiles.

[18] Substantially, this relates to the assumption that, even if the full group-specific measure $\mathbb{P}_{G_n}$ is unknown, we still know the probabilities α that the group G_n constitutes a certain degree of homogeneity.

in Schollmeyer (2017a). Subsequently, in Section 5, we compare these aggregation rules with respect to our criterion in a hypothetical study for groups along varying degree of homogeneity. The results of this study show that the quality of an aggregation function does indeed depend on the considered groups homogeneity structure (see, in particular, Figures 1 and 2 and the discussions of the results in Section 5.2).

3.4.2 Comments and perspectives

From my view, there are mainly two generalizations of the paper that should be further investigated in future research. The first aspect relates to the estimation of the probability values $\alpha_1, \dots, \alpha_\xi$ that is discussed in Section 3.4 of the paper. There, the proposed application of the Dirichlet-categorical model requires the specification of a precise Dirichlet distribution with hyper parameters $\gamma_1, \dots, \gamma_\xi$ according to which the vector α is a priori distributed. If no expert knowledge for specifying the hyper-parameters is available, a so-called near-vacuous prior model, such as the Imprecise Dirichlet Model (IDM), can be chosen (see Walley (1996) for the original work or Bernard (2005); Utkin and Augustin (2007); Augustin et al. (2014) for further interesting properties). In this model, instead of specifying an arbitrary Dirichlet prior for α, one goes for an imprecise probabilistic model and assumes that α is distributed by the credal set of all Dirichlet priors (more precisely, one assumes that all Dirichlet priors are equally plausible candidates to be the prior distribution of α). Clearly, assuming a whole set of prior distribution also yields a set of posterior distributions, namely the set of all posterior distributions that are obtained by updating all the priors with respect to the observed data. Consequently, this also gives interval-valued posterior expectations for the values of α_j: The estimates $\hat{\alpha}_j$ from Equation (11) of the contribution have to replaced by intervals $\hat{I}_j$ ranging from the lowest to the highest posterior expectation that α_j can have under some Dirichlet prior.

This leaves us with a collection of intervals $\hat{I}_1, \dots, \hat{I}_\xi$. To proceed, we need a little detour. First, it is important to note that, for every probability vector $\alpha = (\alpha_1, \dots, \alpha_\xi)$, the set $\mathcal{M}_\alpha$ as defined in Equation (6) of *Contribution 4* very naturally induces a *basic probability assignment* b_α[19] on the space $(\mathcal{R}^n, 2^{\mathcal{R}^n})$ by setting

$$b_\alpha : 2^{\mathcal{R}^n} \to [0,1] \quad , \quad D \mapsto \begin{cases} \alpha_j & \text{if } D = A_n^{-1}(k_j) \text{ for some } j = 1, \dots, \xi \\ 0 & \text{else} \end{cases} \tag{3.8}$$

which is simply a probability mass function on the power set $2^{\mathcal{R}^n}$ of the set of preference profiles $\mathcal{R}^n$. Accordingly, the collection of intervals $\hat{I}_1, \dots, \hat{I}_\xi$ then very naturally induces a set $\mathcal{B}$ of such basic probability assignments by setting

$$\mathcal{B} = \left\{ b_\alpha : \alpha_j \in \hat{I}_j, \forall j = 1, \dots, \xi \ \wedge \ \sum\nolimits_{j=1}^{\xi} \alpha_j = 1 \right\} \tag{3.9}$$

[19] This is meant in the sense of the concept of basic probability assignments known from the Dempster-Shafer theory of belief functions (see Shafer (1976)). See also Destercke and Dubois (2014a, Section 5.2.1) for a brief introduction.

Observe that this set $\mathcal{B}$ is non-empty and closed and, therefore, defines a so-called *generalized basic probability assignment* in the sense of Augustin (2005, Definition 3.2). It then directly follows from Augustin (2005, Theorem 3.3) that if we set

$$L : 2^{\mathcal{R}^n} \to [0,1] \quad , \quad D \mapsto \min_{b_\alpha \in \mathcal{B}} \sum\nolimits_{V \subseteq D} b_\alpha(V) \tag{3.10}$$

then the triplet $(\mathcal{R}^n, 2^{\mathcal{R}^n}, L)$ defines an *F-probability field* in the sense of Weichselberger's theory.[20] Since such an F-probability field can equivalently be represented by a suitable credal set (its *structure*), or by a *coherent lower probability* in the sense of Walley's theory (see Walley (1991)), considering the IDM for estimating the probabilities α_j of the homogeneity classes leads to an usual imprecise probabilistic model. This allows for applying all the machinery from these very rich theories for the proposed generalization.[21]

The second main aspect that should be addressed in future research is that of partially ordered individual preferences. Throughout *Contribution 4*, we modeled the individual group members' opinions about the consequences collected in C by asymmetric and negatively transitive binary relations. This is a very common way of modeling fully specified preferences, since incomparability with respect to such relations can be interpreted as indifference (the incomparability relation is transitive due to negative transitivity, see Section 2.1 and Footnote 2 of the contribution for details). However, as has been argued in situations of individual decision making before, there are good reasons to assume incomparability different from indifference in certain situations. Thus, a natural way to generalize the paper would be to allow for group members with opinions that correspond to *strict partial orders*, i.e. binary relations that are irreflexive and transitive. Essentially, the whole framework proposed in the paper would straightforwardly extend to such situations and could still be applied. However, two modifications would be needed. First, the maximum consensus homogeneity δ_n (see Definition 2 of the contribution) needs to be modified: The number $e_{\underline{R}}(a,b)$ of group members being indifferent between a and b used in its definition, now corresponds to the number of group members that are either indifferent between a and b or who cannot compare the two at all. Thus, a high value of $e_{\underline{R}}(a,b)$ does no longer (necessarily) indicate a homogeneous opinion about the pair (a,b) in the considered profile $\underline{R}$. A straightforward and simple modification of δ_n for partial individual orders is to base it solely on the numbers $c_{\underline{R}}(a,b)$ of members strictly preferring a to b.

The second modification that would be needed concerns the definition of the preference aggregation rules in Section 4 of the contribution. Generally, there seems to be no unique and straightforward way of extending them to partial orders, but such extensions can be done in many plausible ways. For instance, how to compute and compare the ranks of consequences in a partially ordered set when generalizing Borda count? A discussion of generalizing preference aggregation functions to partially ordered individual preferences can be found, e.g., in Pini et al. (2011). An even more general discussion of preference aggregation under incomplete information is discussed in Chambers and Hayashi (2014).

[20] See Weichselberger and Pöhlmann (1990) and Weichselberger (2001) for textbooks or Weichselberger (2000) for a more compact description of some basic concepts.

[21] An (essentially) equivalent result has been obtained in Miranda et al. (2005).

3.5 Detecting stochastic dominance

3.5.1 Summary of Contribution 5: "Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems"

Contribution 5 develops a linear programming method for detecting stochastic dominance for random variables with values in a partially ordered set. We study the corresponding dual programming problem and discuss its properties. Subsequently, we address the question of inference by utilizing resampling methods as well as conservative bounds that are obtained by the application of Vapnik-Chervonenkis theory.

In this paper we take a more statistical point of view. We deal with (first order) stochastic dominance for random variables taking values in some partially ordered space $(V, \leq)$.[22] In contrast to the characterization of stochastic dominance used in *Contribution 3*, we here use a different characterization: For random variables $X, Y : \Omega \to V$, we say that Y stochastically dominates X, denoted by $X \leq_{SD} Y$, if and only if it holds that $P(X \in A) \leq P(Y \in A)$ for every *upset* A, where a subset $A \subseteq V$ is called upset if and only if $x \in A$ and $x \leq y$ implies $y \in A$ for arbitrary elements $x, y \in V$.

Intuitively, this characterization of stochastic dominance can be interpreted as follows: Every upset A can be viewed as a possible (meaningful) concept for defining the subset of large values from V with respect to $\leq$. Given this view, Y stochastically dominates X if and only if Y has higher probability of attaining large values than X for no matter what meaningful concept for defining large values is chosen. However, rather than dealing with the distributions of the random variables X and Y directly, we deal with their empirical analogues associated to the concrete samples. Specifically, suppose we have i.i.d. samples $x = (x_1, \ldots, x_{n_x})$ and $y = (y_1, \ldots, y_{n_y})$ of X and Y, respectively. Moreover, let $\hat{P}_{n_x,x}$ and $\hat{P}_{n_y,y}$ denote the empirical distributions associated with the samples x and y.[23] We are then interested in whether the sample y stochastically dominates the sample x, denoted by $X \leq_{\hat{SD}} Y$, or, in other words, whether it holds that $\hat{P}_{n_x,x}(X \in A) \leq \hat{P}_{n_y,y}(Y \in A)$ for arbitrary upsets $A \subseteq V$.

This leads us to the first main part of the paper: We propose an algorithm for detecting whether a sample y stochastically dominates a sample x by solving one single linear programming problem. This is discussed in Section 3.2 of the contribution. The construction involves several steps. Since we deal with finite samples, we without loss of generality can assume that V is finite, say $V = \{v_1, \ldots, v_k\}$. Then, one has to notice that the set of all upsets of V is in one-to-one correspondence with the set of all monotone increasing indicator functions on V (w.r.t. $\leq$) and, therefore, to the set $M_{\leq}$ of all vectors

[22] Specifically, in this contribution we assume the relation $\leq \subseteq V \times V$ to be reflexive (i.e. $(v, v) \in \leq$ for all $v \in V$), transitive (i.e. $(u, v) \in \leq$ and $(v, w) \in \leq$ implies $(u, w) \in \leq$ for all $u, v, w \in V$), and antisymmetric (i.e. $(u, v) \in \leq$ and $(v, u) \in \leq$ implies $u = v$ for all $u, v \in V$).

[23] Specifically, we have $\hat{P}_{n_x,x}(X \in A) = \frac{1}{n_x} \sum_{i=1}^{n_x} \mathbb{1}_A(x_i)$ and $\hat{P}_{n_y,y}(Y \in A) = \frac{1}{n_y} \sum_{i=1}^{n_y} \mathbb{1}_A(y_i)$.

$m = (m_1, \ldots, m_k) \in \{0,1\}^k$ corresponding to such indicator functions.[24] The elements of $M_{\leq}$ can easily be described via linear constraints (see Equation (3) of the contribution). This allows to check whether $X \leq_{\hat{SD}} Y$ by solving the optimization problem

$$\langle w^x - w^y, m \rangle \longrightarrow \max_{m \in M_{\leq}} \tag{3.11}$$

where, for $z \in \{x, y\}$, we have $w^z = (w_1^z, \ldots, w_k^z)$ and w_i^z denotes the share of observations of v_i in the sample z. It then holds that $X \leq_{\hat{SD}} Y$ if and only if the optimal outcome of the above problem (3.11) is at most 0. Moreover, it even turns out that one can equivalently replace the constraint $m = (m_1, \ldots, m_k) \in \{0,1\}^k$ by $m = (m_1, \ldots, m_k) \in [0,1]^k$ (see Section 4.1 of the contribution for more details), which makes problem (3.11) a classical linear programming problem.

In Section 3.2.2 of the contribution, we then investigate the dual program of the linear programming problem obtained in Section 3.1 (see Equation (10) of the contribution in particular). Here, it turns out that this dual program can be interpreted as a certain type of (probability) mass transportation problem that has some close connections to the algorithms discussed in Tarp and Osterdal (2007) and Range and Osterdal (2013). We illustrate this dual programming problem by means of a concrete toy example and give visualization of it in the Figures 1 and 2 of the contribution.

In Section 3.4 and 4 of the contribution, we utilize the fact that the set of all upsets of a partially ordered set V is a particular example from a more general class of set systems, so-called *closure systems*.[25] We therefore investigate in how far the optimization techniques developed in Section 3 extend to this more general setting or, more precisely, whether optimizing linear functionals on closure systems still reduces to linear programming problems. It turns out that if the closure system $\mathcal{S}$ is sufficiently easy to characterize[26], then this is still true. However, for more complex closure systems, the integrality constraints for the vectors representing the indicator function can no longer be equivalently replaced by interval constraints and, thus, we arrive at linear programming problems with *binary* linear constraints. That such problems nevertheless can be feasibly solved for practically relevant situations is shown in Sections 4.1 and 4.2 of the contribution.

In Section 5, we turn to the question of inference or, more precisely, the question of how to statistically test stochastic dominance for random variables taking values in a partially ordered set. After discussing related existing approaches[27] and emphasizing why a reasonable consistent classical statistical test is not reachable at all, we essentially propose

[24] If $i : V \to \{0,1\}$ is a monotone increasing indicator function, the corresponding vector is given by $m = (i(v_1), \ldots, i(v_k))$.

[25] A closure system on V is a set system $\mathcal{S} \subseteq 2^V$ that contains V and is closed under arbitrary intersections.

[26] Precisely, if $\mathcal{S}$ is generated by an *implication base* consisting solely of *formal implications* with a singleton *premise* (see Sections 2.2 and 3.4 of the contribution). Equivalently, such closure systems can be characterized as being additionally closed under arbitrary unions.

[27] See in particular Barrett and Donald (2003) for an univariate test with switched hypothesis or Babbar (1955), Sengupta et al. (1963), and Prékopa (1966) for results on the asymptotic behavior of random linear programming problems.

two slightly modified approaches. Therefore, let D^+ denote the maximal outcome of the above problem (3.11) and let D^- denote the minimal outcome of (3.11) in minimum form. In Section 5.1, we propose a permutation-based test with the test statistic D^+ for a more conservative test hypothesis than the one we are actually interested in, namely $H_0 : P_X = P_Y$ instead of $\tilde{H}_0 : \sup_A P(X \in A) - P(Y \in A) = 0$ (where the supremum is taken over all upsets A). In Section 5.2, we rely on Vapnik-Chervonenkis theory (see, e.g., Vapnik (1982)) for obtaining bounds for the test statistic D^+ which can be used as conservative critical values for tests relying on D^+. The obtained bounds for D^+ then depend on the complexity of set of all upsets and are usually very conservative. In Section 5.3, we propose a procedure for reducing the complexity of the set of all upsets by systematically excluding upsets that are "too complex", namely upsets that are generated by maximal anti-chains with respect to the relation $\geq$. This procedure can be viewed as a kind of "regularization" procedure for the test statistic in the sense that the supremum in D^+ is now taken over a smaller class of monotone indicator functions. Section 6 of the paper is devoted to several applications of the theory, namely multivariate inequality analysis (Section 6.1), item response theory (Section 6.2), cognitive diagnosis models (Section 6.3), and a geometrical characterization of the Kolmogorov-Smirnov test (Section 6.4).

3.5.2 Comments and perspectives

Looking closely, there are some tight interrelations between *Contribution 3* and *Contribution 5* of this PhD dissertation. Specifically, in *Contribution 3* we are concerned with acts $X : S \to A$, where A is some non-empty set of consequences equipped with a pre-order R_1 on A and a pre-order R_2 on R_1.[28] Accordingly, if we choose $R_2 = \emptyset$ and R_1 to be additionally antisymmetric, we are essentially back in the framework of *Contribution 5*, namely in the context of random variables taking values in a partially ordered space (A, R_1).[29] Moreover, assume that the uncertainty about the states from S is described by the classical probability measure π, that is $\mathcal{M} = \{\pi\}$ is a singleton. Under mild regularity conditions for the underlying space A[30], it then holds for arbitrary acts X and Y that

$$\pi(X \in B) \leq \pi(Y \in B) \quad \text{for all (measurable) upsets } B \text{ of } (A, R_1) \tag{3.12}$$

is (essentially) equivalent to

$$\mathbb{E}_\pi(u \circ X) \leq \mathbb{E}_\pi(u \circ Y) \quad \text{for all utility functions } u \in \mathcal{U}_\mathcal{A} \tag{3.13}$$

where $\mathcal{U}_\mathcal{A}$ is the set of utility function $u : A \to [0, 1]$ that has been defined in Chapter 3.3 of the present work (or in Definition 2 of *Contribution 3*, respectively). That is, for the special case of a preference system $\mathcal{A} = [A, R_1, R_2]$ as defined above and a credal set $\mathcal{M} = \{\pi\}$ that is a singleton, first order stochastic dominance with respect to π and the relation $R_{\forall\forall}$

[28] See the summary in Chapter 3.3 of this dissertation for further details.

[29] Of course, this connection exists rather on a formal level, since *Contribution 5* is not primarily about decision under uncertainty.

[30] See, e.g., Mosler and Scarsini (2012) for details.

from Definition 8 of *Contribution 3* coincide (obviously, in that case also the relations $R^2_{\exists\forall}$ and $R^1_{\forall\exists}$ coincide, since these fall together with the relation $R_{\forall\forall}$ for $\mathcal{M}$ being a singleton).

The above equivalence offers a very nice way for generalizing the concept of first order stochastic dominance in two (not mutually excluding) different directions. Firstly, one could easily extend stochastic dominance to sets of probability measures. To do so, one only needs drop the assumption that $\mathcal{M}$ has to be a singleton by demanding that the above property (3.13) holds, for instance, for all compatible or at least one compatible probability measure. This is, among other things, essentially done in Section 4.4 of our *Contribution 3* (see the relations $R^2_{\exists\forall}$, $R^1_{\forall\exists}$, and $R_{\forall\forall}$ of the contribution in particular). Compare also the summary of our *Contribution 3* discussed in Chapter 3.3 of the present dissertation.[31].

Secondly, one could drop the assumption that the relation R_2 belonging to the preference system has to be empty and, accordingly, demand the above property (3.13) only for such utility functions u that weakly represent[32] the preference system $\mathcal{A}$ as a whole. This would give the opportunity for extending the concept of stochastic dominance from random variables taking values in (partial) ordinal scaled spaces to random variables that take values in partial ordinal and partial cardinal scaled spaces, i.e. preference systems. Specifically, such an extension seems very naturally for the following situation: Suppose we are interested in comparing random variables taking values in some q-dimensional space $A^* \subset \mathbb{R}^q$, where $q \geq 2$ is some fixed number. That is, $A^* = A_1 \times \cdots \times A_q$ for suitable real subsets $A_1, \ldots, A_q$. Moreover, suppose on each of the sets A_i, where $i = 1, \ldots, q$, we have a cardinal scale in the sense that comparisons of differences allow for a meaningful interpretation. However, elements belonging to different sets A_i and A_j do not allow for a comparison at all (neither ordinal nor cardinal). Then a partially ordinal relation $R_1^* \subset A^* \times A^*$ is very naturally given by the component-wise order on A^* defined by $(a, b) \in R_1^*$ if and only if $a_i \geq b_i$ for all $i = 1, \ldots, q$. However, simply comparing random variables with values in A^* by first order stochastic with respect to R_1^* would ignore the information about the cardinal structure in the different dimensions. To use this information, one can then proceed and define a relation R_2^* on $R_1^* \times R_1^*$ by setting $((a, b), (c, d)) \in R_2^*$ if and only if $a_i - b_i \geq c_i - d_i$ for all $i = 1, \ldots, q$, i.e. to weakly prefer the exchange of b by a to the exchange of d by c whenever the former has an at least as high difference as the latter in every dimension. If we then, for two random variables $X, Y : S \to A^*$, demand the above property (3.13) only for $u \in \mathcal{U}_{\mathcal{A}^*}$, where $\mathcal{A}^* = [A^*, R_1^*, R_2^*]$, we arrive at a generalized notion of stochastic dominance additionally allowing to incorporate partial cardinal information. The example just discussed can, in some more detail, be found in Schollmeyer (2017b, pp. 28-29) who also discusses the interrelations to the so-called *scaled convex order* that is introduced in Koshevoy and Mosler (2007).

[31] For more approaches generalizing stochastic dominance to credal sets see Denoeux (2009); Couso and Dubois (2012); Montes et al. (2014b); Couso and Destercke (2015).

[32] See Definition 2 of our *Contribution 3* for details.

4 General concluding remarks

As the more detailed and paper-specific research perspectives were already discussed under the headline "comments and perspectives" in the sections directly following the summary of the corresponding contribution, this seems to be the right place for some more general concluding remarks. To introduce these, let us also briefly summarize the five papers building the core of this thesis.

In this work, we proposed, presented, and discussed several theoretical and algorithmic approaches for extending the classical theory of decision making under uncertainty to situations in which the available information is very weakly structured. Specifically, we presented approaches that are capable of dealing with non-complete individual preferences as well as partially specified probabilities of the different states of nature. In *Contribution 1* and *Contribution 2*, where only the probabilities are assumed to be partially specified, we presented algorithms for determining optimal acts with respect to certain decision criteria from the theory of imprecise probabilities, introduced a new decision criterion, and studied measures for quantifying the "extent of optimality" of optimal acts with respect to the criterion of E-admissibility. In *Contribution 3* we assumed that both the utilities and the probabilities are only partially specified. Moreover, we introduced a model for partial ordinal and partial cardinal preferences. In this setup, we proposed three new approaches for decision making in such weakly structured decision problems, one based on generalized expectation intervals, the other two based on local and global expectation comparisons of acts, respectively. Any of these approaches is complemented by algorithms that allow the evaluation of the decision criteria.

In *Contribution 4* we investigated if, when considering the problem of aggregating the preferences of the members of a fixed group, information on the considered groups homogeneity structure in form of partial probabilistic knowledge can help to better choose between different (common as well as recently proposed) preference aggregation functions. Indeed, it showed that the appropriateness of such aggregation procedures does depend on the group's homogeneity. In *Contribution 5* we proposed an algorithm for checking stochastic dominance in samples of partially ordered data relying on linear programming theory. Further, we investigated in how far such type of algorithm can be extended to the general task of optimizing linear functions on closure systems. Moreover, we addressed the question of inference, i.e. what can be learned about stochastic dominance for the underlying true random variables from the samples.

So, what remains to say? As argued already in Chapter 1, decision theory under uncertainty is a very general theory with applications in many different disciplines. This, in particular, involves more applied fields where the rather mathematical assumptions justi-

fying or dismissing certain decision criteria to be suitable are very hard to communicate and often not questioned at all (for instance, one here could think of medical diagnosis systems in medicine or software packages automatically calculating the insurance premium for some customer in an insurance company). In such fields, people rely on the adequacy of such decision support systems even if a wrong decision may have dramatic consequences. Taking this into account, the wide use of decision theoretic methods gives rise to (at least) two different aspects worth thinking about, one *optimistic* the other rather *pessimistic*:

- *Optimistic view:* Further developing the theory of decision making under uncertainty very naturally provides a field where scientific progress is always much needed and important. It is hard to think of any other scientific discipline that has intersections with comparably many (scientific) fields.

- *Negative view:* Decision theory relying on hardly (or not at all) justifiable assumptions about utilities or probabilities or both, i.e. on what we called the information sources $\mathcal{I}_P$ and $\mathcal{I}_U$, may have dramatic consequences that aren't solely of academic interest, but might (negatively) affect the lives of real people.

At first sight, it seems hard to take both of these two extreme points of view into account simultaneously. But it does not seem impossible. However, every attempt to solve this problem necessarily has to begin with the researchers working on decision theory. Only if they develop decision models based on justifiable assumptions, practitioners can indeed rely on such models without risking (too) wrong decisions. Thus, bringing together the two different extreme points of view ideally could result in the following compromise:

- *Idealistic view:* Further developing the theory of decision making under uncertainty is relevant and important for a variety of disciplines, however, this needs to be done in a way avoiding hyper-idealized assumptions and instead reflecting the real nature of the available information sources.

This is exactly where many alternative, non-classical, works on decision theory have been contributing very valuable progress in the last decades. Some of these works have explicitly been mentioned in the present dissertation. I truly hope that with the works collected in the present PhD dissertation we could contribute some small steps (to the many big steps that already have been taken) to reaching such "idealistic" compromise and, if not, at least could raise the awareness for the necessity of such compromise. A first step in this direction would be to actually implement the preference elicitation schemes discussed in Chapter 3.3.2 of this dissertation.

Further references

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21:503–546.

Anscombe, F. and Aumann, R. (1963). A definition of subjective probability. *The Annals of Mathematical Statistics*, 34:199–205.

Arrow, K. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, 58:555–566.

Augustin, T. (1998). *Optimale Tests bei Intervallwahrscheinlichkeit*. Vandenhoeck und Ruprecht.

Augustin, T. (2001). On decision making under ambiguous prior and sampling information. In de Cooman, G., Fine, T., Moral, S., and Seidenfeld, T., editors, *Proceedings of the Second International Symposium on Imprecise Probability: Theories and Applications*, pages 9–16. Shaker.

Augustin, T. (2003). On the suboptimality of the generalized bayes rule and robust bayesian procedures from the decision theoretic point of view – a cautionary note on updating imprecise priors. In Bernard, J.-M., Seidenfeld, T., and Zaffalon, M., editors, *Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications*, pages 31–45. Carleton Scientific.

Augustin, T. (2005). Generalized basic probability assignments. *International Journal of General Systems*, 34:451–463.

Augustin, T. and Seising, R. (2017). Kurt Weichselberger's contribution to imprecise probabilities. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 13–24. PMLR.

Augustin, T., Walter, G., and Coolen, F. (2014). Statistical inference. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors, *Introduction to Imprecise Probabilities*, pages 135–189. Wiley, Chichester.

Aumann, R. (1962). Utility theory without the completeness axiom. *Econometrica*, 30:445–462.

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

Baccelli, J. (2017). Do bets reveal beliefs? A unified perspective on state-dependent utility issues. *Synthese*, 194:3393–3419.

Bacharach, M. (1975). Group decisions in the face of differences of opinion. *Management Science*, 22:182–191.

Bar-Hillel, M. and Margalit, A. (1972). Newcomb's paradox revisited. *The British Journal for the Philosophy of Science*, 23:295–304.

Barrett, G. and Donald, S. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71:71–104.

Beja, A. and Gilboa, I. (1992). Numerical representations of imperfectly ordered preferences (a unified geometric exposition). *Journal of Mathematical Psychology*, 36:426–449.

Berger, J. (1980). *Statistical Decision Theory and Bayesian Analysis*. Springer.

Bernard, J. (2005). An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39:123–150.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22:23–36. The original paper from 1738 was translated from Latin into English by Louise Summer.

Bewley, T. (2002). Knightian decision theory. Part I. *Decisions in Economics and Finance*, 25:79–110.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, 56:23–34.

Bradley, S. (2015). How to choose among choice functions. In Augustin, T., Doria, S., Miranda, E., and Quaeghebeur, E., editors, *Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 57–66. Aracne.

Brams, S. and Fishburn, P. (2002). Voting procedures. In Arrow, K., Sen, A., and Suzumura, K., editors, *Handbook of Social Choice and Welfare, Vol. 1*, pages 173–236. North-Holland.

Braun, N. and Gautschi, T. (2011). *Rational-Choice-Theorie*. Juventa.

Cantor, G. (1915). *Contributions to the Founding of the Theory of Transfinite Numbers*. The Open Court Publishing Company.

Cattaneo, M. (2013). Likelihood decision functions. *Electronic Journal of Statistics*, 7:2924–2946.

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

Chambers, C. and Echenique, F. (2016). *Revealed Preference Theory*. Econometric Society Monographs. Cambridge University Press.

Chambers, C. and Hayashi, T. (2014). Preference aggregation with incomplete information. *Econometrica*, 82:589–599.

Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, 5:131–295.

Coombs, C. and Cohen, J. (1984). Social choice observed: five presidential elections of the American Psychological Association. *The Journal of Politics*, 46:479–502.

Couso, I. and Destercke, S. (2015). Ranking of fuzzy intervals seen through the imprecise probabilistic lens. *Fuzzy Sets and Systems*, 278:20–39.

Couso, I. and Dubois, D. (2012). An imprecise probability approach to joint extensions of stochastic and interval orderings. *Communications in Computer and Information Science*, 299:388–399.

Danielson, M. (2005). Generalized evaluation in decision analysis. *European Journal of Operational Research*, 162:442–449.

Danielson, M. and Ekenberg, L. (1998). A framework for analyzing decision under risk. *European Journal of Operational Research*, 104:474–484.

Danielson, M., Ekenberg, L., Johansson, J., and Larsson, A. (2003). The DecideIT decision tool. In Bernard, J.-M., Seidenfeld, T., and Zaffalon, M., editors, *Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications*, pages 204–217. Carleton Scientific.

de Borda, J. (1781). *Memoire sur les elections au scrutin*. Historie de l'Academie Royale des Sciences.

de Condorcet, N. (1785). *Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix*. Paris.

de Cooman, G., Troffaes, M., and Miranda, E. (2008). A unifying approach to integration for bounded positive charges. *Journal of Mathematical Analysis and Applications*, 340:982 – 999.

de Finetti, B. (1974). *Theory of Probability*. Wiley.

Delbaen, F., Drapeau, S., and Kupper, M. (2011). A von Neumann-Morgenstern representation result without weak continuity assumption. *Journal of Mathematical Economics*, 47:401–408.

Denneberg, D. (1994). *Non-additive Measure and Integral*. Kluwer.

Denoeux, T. (2009). Extending stochastic ordering to belief functions on the real line. *Information Sciences*, 179:1362–1376.

Destercke, S. and Dubois, D. (2014a). Other uncertainty theories based on capacities. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors, *Introduction to Imprecise Probabilities*, pages 93–113. Wiley, Chichester.

Destercke, S. and Dubois, D. (2014b). Special cases. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors, *Introduction to Imprecise Probabilities*, pages 79–92. Wiley, Chichester.

Dubra, J., Maccheroni, F., and Ok, E. (2004). Expected utility theory without the completeness axiom. *Journal of Economic Theory*, 115:118 – 133.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75:643–669.

Etner, J., Jeleva, M., and Tallon, J.-M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, 26:234–270.

Evren, O. and Ok, E. (2011). On the multi-utility representation of preference relations. *Journal of Mathematical Economics*, 47:554 – 563.

Ferschl, F. (1975). *Nutzen- und Entscheidungstheorie: Einführung in die Logik der Entscheidungen*. Westdeutscher Verlag.

Fishburn, P. (1965). Analysis of decisions with incomplete knowledge of probabilities. *Operations Research*, 13:217–237.

Fishburn, P. (1970). *Utility Theory for Decision Making*. Wiley.

Fishburn, P. (1982). *The Foundations of Expected Utility Theory*. Springer.

French, S. (1986). *Decision Theory: An Introduction to the Mathematics of Rationality*. Halsted Press.

French, S. and Insua, D. (2000). *Statistical Decision Theory*. Wiley.

Galaabaatar, T. and Karni, E. (2013). Subjective expected utility with imcomplete preferences. *Econometrica*, 81:255–284.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall.

Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge University Press, New York.

Gilboa, I. (2010). *Rational Choice*. MIT press.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141—153.

Gul, F. (1992). Savage's theorem with a finite number of states. *Journal of Economic Theory*, 57:99 – 110.

Ha, V. and Haddawy, P. (2003). Similarity of personal preferences: Theoretical foundations and empirical analysis. *Artificial Intelligence*, 146:149 – 173.

Hable, R. (2009). Data-based decisions under imprecise probability and least favorable models. *International Journal of Approximate Reasoning*, 50:642 – 654.

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

Hare, T. (1857). *The Machinery of Representation*. Maxwell.

Hens, T. (1992). A note on Savage's theorem with a finite number of states. *Journal of Risk and Uncertainty*, 5:63–71.

Herron, T., Seidenfeld, T., and Wasserman, L. (1997). Divisive conditioning: Further results on dilation. *Philosophy of Science*, 64:411–444.

Hodges, J. and Lehmann, E. (1952). The use of previous experience in reaching statistical decisions. *The Annals of Mathematical Statistics*, 23:396–407.

Horwitz, R., Kreuter, F., and Conrad, F. (2017). Using mouse movements to predict web survey response difficulty. *Social Science Computer Review*, 35:388–405.

Huber, P. (1981). *Robust Statistics*. Wiley.

Huntley, N., Hable, R., and Troffaes, M. (2014). Decision making. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors, *Introduction to Imprecise Probabilities*, pages 190–206. Wiley, Chichester.

Inada, K. (1964). A note on the simple majority decision rule. *Econometrica*, 32:525–531.

Jansen, C., Schollmeyer, G., and Augustin, T. (2017). Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., 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.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106:620–630.

Kadane, J., Schervish, M., and Seidenfeld, T. (2004). A Rubinesque theory of decision. *IMS Lecture Notes-Monograph Series*, 45:45–55.

Kaplan, M. (1983). Decision theory as philosophy. *Philosophy of Science*, 50:549–577.

Karni, E., Schmeidler, D., and Vind, K. (1983). On state dependent preferences and subjective probabilities. *Econometrica*, 51:1021–1031.

Kemeny, J. (1959). Mathematics without numbers. *Quantity and Quality*, 88:577–591.

Kemeny, J. and Snell, J. (1962). *Mathematical Models in the Social Sciences*. Ginn and Company.

Kendall, M. and Smith, B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, 10:275–287.

Kikuti, D., Cozman, F., and Filho, R. (2011). Sequential decision making with partially ordered preferences. *Artificial Intelligence*, 175:1346–1365.

Kofler, E. (1989). *Prognosen und Stabilität bei unvollständiger Information*. Campus.

Kofler, E., Kmietowicz, Z., and Pearman, A. (1984). Decision making with linear partial information (l.p.i.). *The Journal of the Operational Research Society*, 35:1079–1090.

Kofler, E. and Menges, G. (1976). *Entscheiden bei unvollständiger Information*. Springer.

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

Kraft, C., Pratt, J., and Seidenberg, A. (1959). Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, 30:408–419.

Krantz, D., Luce, R., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement. Volume I: Additive and Polynomial Representations*. Academic Press.

Kreps, D. (1988). *Notes on the Theory of Choice (Underground Classics in Economics)*. Westview Press Incorporated.

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

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, 71:391–418.

Levi, I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT Press.

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

Luce, R. and Raiffa, H. (1957). *Games and Decision-Introduction and Critical Survey*. Wiley.

Milnor, J. (1951). Games against nature. *RAND Project Air Force Santa Monica, California*. Available under: `https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM679.pdf`.

Miranda, E., de Cooman, G., and Couso, I. (2005). Lower previsions induced by multivalued mappings. *Journal of Statistical Planning and Inference*, 133:173 – 197.

Montes, I. (2014). *Comparison of Alternatives under Uncertainty and Imprecision*. PhD thesis, Oviedo. `digibuo.uniovi.es/dspace/bitstream/10651/28953/6/TD_IgnacioMontesGutierrez.pdf`.

Montes, I., Miranda, E., and Montes, S. (2014a). Decision making with imprecise probabilities and utilities by means of statistical preference and stochastic dominance. *European Journal of Operational Research*, 234:209 – 220.

Montes, I., Miranda, E., and Montes, S. (2014b). Stochastic dominance with imprecise information. *Computational Statistics and Data Analysis*, 71:867–885.

Mosler, K. and Scarsini, M. (1991). Some theory of stochastic dominance. In Mosler, K. and Scarsini, M., editors, *Stochastic Orders and Decision under Risk*, pages 203–212. Institute of Mathematical Statistics, Hayward, CA.

Mosler, K. and Scarsini, M. (2012). *Stochastic Orders and Applications: A Classified Bibliography*. Springer.

Nakharutai, N., Troffaes, M., and Caiado, C. (2017). Efficient algorithms for checking avoiding sure loss. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 241–252. PMLR.

Nau, R. (2006). The shape of incomplete preferences. *The Annals of Statistics*, 34:2430–2448.

Nishimura, H. and Ok, E. (2016). Utility representation of an incomplete and nontransitive preference relation. *Journal of Economic Theory*, 166:164 – 185.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G. Hempel*, pages 114–146. Springer.

Pini, M., Rossi, F., Venable, K., and Walsh, T. (2011). Incompleteness and incomparability in preference aggregation: Complexity results. *Artificial Intelligence*, 175:1272 – 1289.

Pivato, M. (2013). Multiutility representations for incomplete difference preorders. *Mathematical Social Sciences*, 66:196–220.

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

Range, T. and Osterdal, L. (2013). Checking bivariate first order dominance. *Technical report, Discussion Papers on Business and Economics.* Available under: `http://static.sdu.dk/mediafiles/0/C/2/%7B0C290F51-8643-421F-A9EF-F38D17A7F570%7Ddpbe9_2013.pdf`.

Rubin, H. (1987). A weak system of axioms for "rational" behavior and the non-separability of utility from prior. *Statistics and Risk Modelling*, 5:47–58.

Rüger, B. (1999). *Test-und Schätztheorie: Band I: Grundlagen.* de Gruyter.

Savage, L. (1954). *The Foundations of Statistics.* Wiley.

Savage, L. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66:783–801.

Schervish, M., Seidenfeld, T., Kadane, J., and Levi, I. (2003). Extensions of expected utility theory and some limitations of pairwise comparisons. In Bernard, J.-M., Seidenfeld, T., and Zaffalon, M., editors, *Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications*, pages 496–510. Carleton Scientific.

Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, 97:255–255.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587.

Schollmeyer, G. (2017a). Application of lower quantiles for complete lattices to ranking data: analyzing outlyingness of preference orderings. *Technical Report 208, Department of Statistics, LMU Munich (2017).* Available under: `https://epub.ub.uni-muenchen.de/40452/`.

Schollmeyer, G. (2017b). *Reliable statistical modeling of weakly structured information: Contributions to partial identification, stochastic partial ordering and imprecise probabilities*. PhD thesis, Ludwig-Maximilians Universität München. Available under: `https://edoc.ub.uni-muenchen.de/21367/`.

Seidenfeld, T. (2004). A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus E-admissibilty. *Synthese*, 140:69–88.

Seidenfeld, T., Kadane, J., and Schervish, M. (1989). On the shared preferences of two Bayesian decision makers. *The Journal of Philosophy*, 86:225–244.

Seidenfeld, T., Kadane, J., and Schervish, M. (1995). A representation of partially ordered preferences. *The Annals of Statistics*, 23:2168–2217.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2010). Coherent choice functions under uncertainty. *Synthese*, 172:157–176.

Sen, A. (1966). A possibility theorem on majority decisions. *Econometrica*, 34:491–499.

Sen, A. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, 38:307–317.

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

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

Smithson, M. (2014). Elicitation. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors, *Introduction to Imprecise Probabilities*, pages 318–328. Wiley, Chichester.

Suppes, P. and Winet, M. (1955). An axiomatization of utility based on the notion of utility differences. *Management Science*, 1:259–270.

Tarp, F. and Osterdal, L. (2007). Multivariate discrete first order stochastic dominance. *Discussion Papers 07-23, University of Copenhagen. Department of Economics.* Available under: `http://www.economics.ku.dk/research/publications/wp/2007/0723.pdf`.

Trautmann, S. and Wakker, P. (2018). Making the Anscombe-Aumann approach to ambiguity suitable for descriptive applications. *Journal of Risk and Uncertainty*, 46:83–116.

Troffaes, M. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45:17–29.

Troffaes, M. and de Cooman, G. (2014). *Lower Previsions*. Wiley.

Troffaes, M. and Sahlin, U. (2017). Imprecise swing weighting for multi-attribute utility elicitation based on partial preferences. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 333–345. PMLR.

Utkin, L. and Augustin, T. (2005). Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In Cozman, F., Nau, R., and Seidenfeld, T., editors, *Proceedings of the Fourth International Symposium on Imprecise Probability: Theory and Applications*, pages 349–358. SIPTA.

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

Utkin, L., Chekh, A., and Zhuk, Y. (2015). Classification SVM algorithms with interval-valued training data using triangular and Epanechnikov kernels. In Augustin, T., Doria, S., Miranda, E., and Quaeghebeur, E., editors, *Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 295–303. Aracne.

Utkin, L. and Coolen, F. (2011). Interval-valued regression and classification models in the framework of machine learning. In Coolen, F., de Cooman, G., Fetz, T., and Oberguggenberger, M., editors, *Proceedings of the Seventh Symposium on Imprecise Probability: Theory and Applications*, pages 371–380. SIPTA.

Utkin, L. and Kozine, I. (2001). Different faces of the natural extension. In de Cooman, G., Fine, T., Moral, S., and Seidenfeld, T., editors, *Proceedings of the Second International Symposium on Imprecise Probability: Theories and Applications*, pages 316–323. Shaker.

van Camp, A. (2018). *Keuzefuncties als onzekerheidsmodellen*. PhD thesis, Ghent University. Available under: `https://biblio.ugent.be/publication/8546928/file/8546932.pdf`.

van Ommen, T. (2017). Computing minimax decisions with incomplete observations. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 358–369. PMLR.

Vanderbei, R. (2015). *Linear Programming*. Springer.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*. Springer.

von Neumann, J., Morgenstern, O., Kuhn, H., and Rubinstein, A. (1944). *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press.

Wald, A. (1949). Statistical decision functions. *The Annals of Mathematical Statistics*, 20:165–205.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:3–57.

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125 – 148.

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

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

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

Weichselberger, K. and Pöhlmann, S. (1990). *A Methodology for Uncertainty in Knowledge-based Systems*. Springer.

Williams, P. (1975). Coherence, strict coherence and zero probabilities. In Butts, R. and Hintikka, J., editors, *Proceedings of the Fifth International Congress on Logic, Methodology and Philosophy of Science*, pages 29–33. Kluwer.

Williams, P. (1976). Indeterminate probabilities. In Przelkecki, M., Szaniawski, K., Wójcicki, R., and Malinowski, G., editors, *Formal Methods in the Methodology of Empirical Sciences*, pages 229–246. Springer.

Williams, P. (2007). Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383.

Attached contributions

Contribution 1:

Jansen, C.; Augustin, T.; Schollmeyer, G. (2017a): Decision theory meets linear optimization beyond computation. In: Antonucci, A.; Cholvy, L.; Papini, O. (eds): *Symbolic and Quantitative Approaches to Reasoning with Uncertainty. ECSQARU 2017.* Lecture Notes in Computer Science, vol 10369. Springer.

The original publication is available at Springer via

`https://doi.org/10.1007/978-3-319-61581-3_30`

Decision Theory Meets Linear Optimization Beyond Computation

Christoph Jansen( ), Thomas Augustin, and Georg Schollmeyer

Department of Statistics, LMU, Munich, Germany
{christoph.jansen,augustin,georg.schollmeyer}@stat.uni-muenchen.de

Abstract. The paper is concerned with decision making under complex uncertainty. We consider the Hodges and Lehmann-criterion relying on uncertain classical probabilities and Walley's maximality relying on imprecise probabilities. We present linear programming based approaches for computing optimal acts as well as for determining least favorable prior distributions in finite decision settings. Further, we apply results from duality theory of linear programming in order to provide theoretical insights into certain characteristics of these optimal solutions. Particularly, we characterize conditions under which randomization pays out when defining optimality in terms of the Gamma-Maximin criterion and investigate how these conditions relate to least favorable priors.

Keywords: Linear programming · Decision making · Least favorable prior · Duality · Maximality · Imprecise probabilities · Gamma-maximin · Hodges & Lehmann

1 Introduction

Many problems arising in modern sciences, e.g. estimation and hypothesis testing in statistics or modeling an agent's preferences in economics, can be embedded in the formal framework of *decision theory under uncertainty*. However, as the specification of a *precise* (i.e. classical) probability measure on the space of uncertain states often turns out to be too restrictive from an applicational point of view, decision theory using *imprecise probabilities* (for a survey see, e.g., [12]) has become a more and more attractive modeling tool recently. For determining optimal decisions with respect to the complex decision criteria particularly (but not exclusively) arising in the context of the theory of imprecise probabilities, *linear programming theory* (see, e.g., [15]) often turns out to be well-suited: By embedding decision problems into this general optimization framework, one can draw on the whole theoretical toolbox of this well-investigated mathematical discipline. Particularly, this allows for a computational treatment of complex decision making problems in standard software (e.g. MATLAB or for statisticians R) and, therefore, helps in order to make the abstract theory applicable for practitioners. Accordingly, there exists plenty of literature on linear optimization driven algorithms for facing complex decision problems. Examples include [6,13]. A survey is given in [5].

A. Antonucci et al. (Eds.): ECSQARU 2017, LNAI 10369, pp. 329–339, 2017.
DOI: 10.1007/978-3-319-61581-3_30

However, quite similar to characterizations of imprecise probabilities and natural extensions in [17, Chap. 4] and [14], the opportunities of using linear programming in decision theory are by far not exhausted by producing powerful algorithms (see [18, p. 402]). Instead, applying basic results on duality from linear programming theory (such as, e.g., the *complementary slackness* property, see, e.g., [15, Sect. 5.5]) can often provide theoretical insights on both the connection between different decision criteria and the specific properties shared by all optimal solutions with respect to a certain criterion.

The paper is structured as follows: In Sect. 2, we recall the classical model of finite decision theory as well as the extended version of the model allowing for randomized acts. In Sect. 3, we give a linear program for determining optimal randomized acts with respect to a decision criterion of Hodges and Lehmann which tries to cope with uncertain prior probabilistic information and investigate the corresponding dual programming problem. In Sect. 4, we consider the case of decision making under imprecise probabilistic information. Particularly, we present an algorithm for checking maximality of pure acts in one single linear program in Sect. 4.1 and use duality theory for deriving connections between least favorable prior distributions and the Gamma-Maximin criterion in Sect. 4.2. Finally, Sect. 5 is preserved for concluding remarks.

2 The Basic Model

Throughout the paper, we consider the standard model of *finite* decision theory: An *agent* (or *decision maker*) has to decide which *act* a_i to pick from a finite set $\mathbb{A} = \{a_1, \ldots, a_n\}$. However, the *utility* of the chosen act depends on which *state of nature* from a finite set $\Theta = \{\theta_1, \ldots, \theta_m\}$ corresponds to the true description of reality. Specifically, we assume that the utility of every pair $(a, \theta) \in \mathbb{A} \times \Theta$ can be evaluated by a *known* real-valued *cardinal utility function* $u : \mathbb{A} \times \Theta \to \mathbb{R}$. For simplicity, we will often use the notation $u_{ij} := u(a_i, \theta_j)$, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The structure of the basic model and a running example repeatedly considered throughout the paper are visualized in Table 1. For every act $a \in \mathbb{A}$, the utility function u is naturally associated with a random variable $u_a : (\Theta, 2^\Theta) \to \mathbb{R}$ defined by $u_a(\theta) := u(a, \theta)$ for all $\theta \in \Theta$. Similarly, for every $\theta \in \Theta$, we can define a random variable $u^\theta : (\mathbb{A}, 2^\mathbb{A}) \to \mathbb{R}$ by setting $u^\theta(a) := u(a, \theta)$ for all $a \in \mathbb{A}$.

Depending on the context, we also allow for *randomized acts*, i.e. classical probability measures λ on $(\mathbb{A}, 2^\mathbb{A})$. Choosing λ is then interpreted as leaving your final decision to a random experiment which yields act a_i with probability $\lambda(\{a_i\})$. We denote the set of randomized acts on $(\mathbb{A}, 2^\mathbb{A})$ by $G(\mathbb{A})$.

The utility function u on $\mathbb{A} \times \Theta$ is then extended to a utility function $G(u)$ on $G(\mathbb{A}) \times \Theta$ by assigning each pair (λ, θ) the expectation of the random variable u^θ under the measure λ, i.e. $G(u)(\lambda, \theta) := \mathbb{E}_\lambda[u^\theta]$, which corresponds to the expectation of utility that choosing the randomized act λ will lead to, given θ is the true description of reality. Every *pure* act $a \in \mathbb{A}$ then can uniquely be identified with the *Dirac-measure* $\delta_a \in G(\mathbb{A})$, and we have $u(a, \theta) = G(u)(\delta_a, \theta)$

Table 1. Basic model (left) and running example with acts $\mathbb{A} = \{a_1, a_2, a_3\}$, states $\Theta = \{\theta_1, \ldots, \theta_4\}$ (right) and the credal set $\mathcal{M} := \{\pi : 0.3 \leqslant \pi(\{\theta_2\}) + \pi(\{\theta_3\}) \leqslant 0.7\}$ additionally considered in the Sects. 4.1 and 4.2.

$u(a_i, \theta_j)$	θ_1	$\cdots$	θ_m
a_1	$u(a_1, \theta_1)$	$\cdots$	$u(a_1, \theta_m)$
$\vdots$	$\vdots$	$\ldots$	$\vdots$
a_n	$u(a_n, \theta_1)$	$\cdots$	$u(a_n, \theta_m)$

$u(a_i, \theta_j)$	θ_1	θ_2	θ_3	θ_4
a_1	20	15	10	5
a_2	30	10	10	20
a_3	20	40	0	20

for all $(a, \theta) \in \mathbb{A} \times \Theta$. Again, for every $\lambda \in G(\mathbb{A})$ fixed, the extended utility function $G(u)$ is associated with a random variable $G(u)_\lambda$ on $(\Theta, 2^\Theta)$ by setting $G(u)_\lambda(\theta) := G(u)(\lambda, \theta)$ for all $\theta \in \Theta$. Finally, we refer to the triplet $(\mathbb{A}, \Theta, u)$ as the *(finite) decision problem* and to the triplet $(G(\mathbb{A}), \Theta, G(u))$ as the corresponding *randomized extension*.

Within this framework, our goal is to determine an *optimal* act (depending on the context, either randomized or pure). However, any appropriate definition of optimality depends on (what we assume about) the *mechanism generating the states of nature*. Here, traditional decision theory mainly covers two *extremes*: The mechanism follows a *known* probability measure π on $(\Theta, 2^\Theta)$ or it can be compared to a *game against an omniscient enemy*. In this cases optimality is almost unanimously defined by either *maximizing expected utility* with respect to π (also known as *Bayes-criterion*) or applying the *Maximin-criterion* (i.e. choosing an act that has maximal utility under the worst possible state of nature).

In contrast, defining optimality of acts becomes less obvious if the prior π is only *partially* known (case of *imprecise probabilities*) or there is uncertainty about the complete appropriateness of it (case of *uncertainty about precise probabilities*). The following sections are concerned with these two situations.

3 Handling Uncertain Precise Probabilistic Information: The Hodges and Lehmann-Criterion

Apart from the border cases of maximizing expected utility with respect to a precise prior π in the presence of perfect probabilistic information and the Maximin-criterion in complete absence of probabilistic information, classical decision theory tries to cope with decision making under uncertain probabilistic information, too: Anticipating ideas of *robust statistics*, Hodges and Lehmann proposed applying the Bayes-criterion only to such acts, whose worst possible utility does not fall below a certain amount of the Minimax utility (see [4]). Their idea is to utilize probabilistic information from previous experience while simultaneously distrusting the complete appropriateness of this information and restricting analysis to acts that are not too bad under the worst state. They also give the following alternative representation of their approach that has a

different, intuitively more accessible, interpretation[1]: The decision maker is allowed to model his *degree of trust* in the prior by a parameter $\alpha \in [0,1]$. Specifically, if π is a probability measure on $(\Theta, 2^\Theta)$, a randomized act $\lambda^* \in G(\mathbb{A})$ is said to be *Hodges and Lehmann*-optimal w.r.t. π and α (short: $\Phi_{\pi,\alpha}$-optimal), if $\Phi_{\pi,\alpha}(\lambda^*) \geqslant \Phi_{\pi,\alpha}(\lambda)$ for all $\lambda \in G(\mathbb{A})$, where

$$\Phi_{\pi,\alpha}(\lambda) := (1-\alpha) \cdot \min_\theta G(u)(\lambda,\theta) + \alpha \cdot \mathbb{E}_\pi\Big[G(u)_\lambda\Big] \tag{1}$$

Thus, the parameter α in (1) controls how the linear trade-off between expectation maximization w.r.t. π and applying the Maximin-criterion is actually made. The following Proposition 1 describes an algorithm for determining a randomized Hodges and Lehmann-optimal act for arbitrary pairs (π, α).[2]

Proposition 1. *Consider the linear programming problem*

$$(1-\alpha)\cdot(w_1 - w_2) + \alpha \cdot \sum_{i=1}^{n} \mathbb{E}_\pi(u_{a_i}) \cdot \lambda_i \longrightarrow \max_{(w_1,w_2,\lambda_1,\ldots,\lambda_n)} \tag{2}$$

with constraints $(w_1, w_2, \lambda_1, \ldots, \lambda_n) \geqslant 0$ *and*

- $\sum_{i=1}^n \lambda_i = 1$
- $w_1 - w_2 \leqslant \sum_{i=1}^n u_{ij} \cdot \lambda_i \quad$ *for all* $j = 1, \ldots, m$.

Then the following holds:

(i) Every optimal solution $(w_1^*, w_2^*, \lambda_1^*, \ldots, \lambda_n^*)$ *to (2) induces a* $\Phi_{\pi,\alpha}$*-optimal randomized act* $\lambda^* \in G(\mathbb{A})$ *by setting* $\lambda^*(\{a_i\}) := \lambda_i^*$.

(ii) There always exists an $\Phi_{\pi,\alpha}$*-optimal randomized act.* □

By computing the dual linear program of the linear program given in Proposition 1, we receive the following Corollary. It can be interpreted as a method to construct priors that take the agent's *scepticism* about the prior probability π (expressed by the parameter α) into account.

Corollary 1. *Let* $\lambda^* \in G(\mathbb{A})$ *denote a* $\Phi_{\pi,\alpha}$*-optimal randomized act. Then, there exists a probability measure* $\mu_{\pi,\alpha}$ *on* $(\Theta, 2^\Theta)$ *and a pure act* $a^* \in \mathbb{A}$ *such that*

$$\Phi_{\pi,\alpha}(\lambda^*) = \mathbb{E}_{\mu_{\pi,\alpha}}[u_{a^*}] \tag{3}$$

Proof. The dual of the optimization problem (2) is given by:

$$z_1 - z_2 \longrightarrow \min_{(z_1,z_2,\sigma_1,\ldots,\sigma_m)} \tag{4}$$

with constraints $(z_1, z_2, \sigma_1, \ldots, \sigma_m) \geqslant 0$ and

[1] A further mathematical characterization from the viewpoint of Gamma-Maximinity for certain imprecise probabilities is given in Footnote 3.

[2] The proofs of Propositions 1, 2 and 3 are straightforward and therefore left out.

- $\sum_{j=1}^{m} \sigma_j = 1 - \alpha$
- $z_1 - z_2 \geq \sum_{j=1}^{m} u_{ij} \cdot \sigma_j + \alpha \cdot \mathbb{E}_\pi(u_{a_i}) \quad$ for all $i = 1, \ldots, n$.

Let $(z_1^*, z_2^*, \sigma_1^*, \ldots, \sigma_m^*)$ denote an optimal solution to (4). Then the constraints guarantee that assigning $\mu_{\pi,\alpha}(\{\theta_j\}) := \alpha \cdot \pi(\{\theta_j\}) + \sigma_j^*$ for all $j = 1, \ldots, m$ induces a probability measure on $(\Theta, 2^\Theta)$ and that for all expectation maximal acts $a^* \in \mathbb{A}$ with respect to $\mu_{\pi,\alpha}$ it holds that $z_1^* - z_2^* = \mathbb{E}_{\mu_{\pi,\alpha}}[u_{a^*}]$. Further, by duality, we know that $z_1^* - z_2^*$ coincides with the optimal value of program (2) and, therefore, with $\Phi_{\pi,\alpha}(\lambda^*)$ where $\lambda^* \in G(\mathbb{A})$ denotes an Hodges and Lehmann-optimal randomized act. Thus, $\Phi_{\pi,\alpha}(\lambda^*) = \mathbb{E}_{\mu_{\pi,\alpha}}[u_{a^*}]$, as desired. □

Running Example (Table 1): Let π denote the prior on $(\Theta, 2^\Theta)$ induced by $(0.2, 0.7, 0.05, 0.05)$ and let our trust in π be expressed by $\alpha = 0.35$. Resolving the linear programming problem from Proposition 1 gives the optimal solution $(8, 0, 0.8, 0, 0.2)$. Thus, a $\Phi_{\pi,0.35}$-optimal randomized act $\lambda^* \in G(\mathbb{A})$ is induced by $(0.8, 0, 0.2)$. Next, we can use Corollary 1 to compute $\mu_{\pi,0.35}$. An optimal solution of problem (4) is given by the vector $(11.78, 0, 0, 0, 0.6385, 0.0115)$, and thus the measure $\mu_{\tau,0.35}$ is induced by the vector $(0.070, 0.245, 0.656, 0.029)$.

4 Handling Imprecise Probabilistic Information: The Gamma-Maximin View

We now turn to decision criteria taking into account the uncertainty in the prior information in a more direct way: For modeling prior knowledge, instead of one classical probability, we consider polyhedral sets of probability measures that are a common tool in different theories of imprecise probabilities, like e.g. *linear partial information* ([7]), *credal sets* ([8]), *lower previsions* ([16]) or *interval probability* ([17]) as well as in robust statistics, like e.g. *ε-contamination models* (see [3, p. 12]). Particularly, we assume probabilistic information is expressed by a polyhedrical set $\mathcal{M}$ of probability measures on $(\Theta, 2^\Theta)$ of the form

$$\mathcal{M} := \left\{\pi \middle|\ \underline{b}_s \leqslant \mathbb{E}_\pi(f_s) \leqslant \overline{b}_s \ \forall s = 1, ..., r\right\} \tag{5}$$

where, for all $s = 1, ..., r$, we have $(\underline{b}_s, \overline{b}_s) \in \mathbb{R}^2$ such that $\underline{b}_s \leqslant \overline{b}_s$ and $f_s : \Theta \to \mathbb{R}$. Specifically, the available information is assumed to be describable by lower and upper bounds for the expected values of a finite number of random variables on the space of states. Clearly, if uncertainty is described by a set of probability measures, defining meaningful criteria for decision making strongly depend on the agent's *attitude towards ambiguity*, i.e. towards the non-stochastic uncertainty between the measures contained in $\mathcal{M}$. Accordingly, many competing criteria exist (see [12] for a survey or [2,8,16] for original sources). In the following sections, we present linear programming based results for a selection of such criteria, namely *Walley's maximality* and the *Gamma-Maximin* criterion. For the latter, we also investigate some connections to least favorable priors.

4.1 Checking Maximality of Pure Acts

The idea behind maximality of an act $a^* \in \mathbb{A}$ is quite simple: One repeatedly compares an act a^* pairwise to all other acts and checks whether there exists an element of the set $\mathcal{M}$ with respect to which u_{a^*} dominates the corresponding other act in expectation. Formally, an act $a^* \in \mathbb{A}$ is said to be $\mathcal{M}$-*maximal*, if

$$\forall\, a \in \mathbb{A} \;\; \exists\, \pi_a \in \mathcal{M} : \quad \mathbb{E}_{\pi_a}(u_{a^*}) \geqslant \mathbb{E}_{\pi_a}(u_a) \tag{6}$$

Naturally, the above definition extends to randomized acts. However, when also considering randomized acts, the criterion of $\mathcal{M}$-Maximality coincides (see [16, p. 163]) with another well-investigated criterion known from IP decision theory contributed to *Levi*: E-admissibility. For a detailed discussion of connections between the two criteria see [11]. An algorithm for determining the set of all randomized E-admissible acts has been introduced in [13]. However, for finite $\mathbb{A}$, being $\mathcal{M}$-Maximal is a strictly weaker condition and, therefore, needs to be checked separately from E-admissibility. Other approaches for doing so have already been proposed in [6]. Proposition 2 describes an algorithm for checking $\mathcal{M}$-Maximality of a pure act $a_z \in \mathbb{A}$ by solving one single linear program.

Proposition 2. *Let $(\mathbb{A}, \Theta, u)$ denote a finite decision problem and let $\mathcal{M}$ be of the form (5). Further, let $a_z \in \mathbb{A}$ be any act. Consider the linear program*

$$\sum_{i=1}^{n}\Big(\sum_{j=1}^{m}\gamma_{ij}\Big) \longrightarrow \max_{(\gamma_{11},\ldots,\gamma_{nm})} \tag{7}$$

with constraints $(\gamma_{11}, \ldots, \gamma_{nm}) \geqslant 0$ and

- $\sum_{j=1}^{m} \gamma_{ij} \leqslant 1 \quad$ *for all* $i = 1, \ldots, n$
- $\underline{b}_s \leqslant \sum_{j=1}^{m} f_s(\theta_j) \cdot \gamma_{ij} \leqslant \overline{b}_s \quad$ *for all* $s = 1, ..., r$, $i = 1, \ldots, n$
- $\sum_{j=1}^{m} (u_{ij} - u_{zj}) \cdot \gamma_{ij} \leqslant 0 \quad$ *for all* $i = 1, \ldots, n$.

Then $a_z \in \mathbb{A}$ is $\mathcal{M}$-Maximal iff the optimal outcome of (7) equals n. □

If $(\gamma_{11}^*, \ldots, \gamma_{nm}^*)$ is an optimal solution to problem (7) yielding an value of n, we can construct $\pi_{a_i} \in \mathcal{M}$ for which act a_z dominates act a_i in expectation by setting $\pi_{a_i}(\{\theta_j\}) := \gamma_{ij}$. The problem possesses $n(3 + r)$ constraints and nm decision variables. Determining the set of all maximal acts requires to solve n such linear programs. Compared to this, the algorithm based on pairwise comparisons of acts proposed in [6] here translates to solving $n^2 - n$ linear programs with m decision variables, however, with only $r + 2$ constraints.

Running Example (Table 1): Resolving the linear programming problem from Proposition 2 for every act a_1, a_2 and a_3 separately gives optimal value 3 for each of them. Thus, all available acts are $\mathcal{M}$-Maximal.

4.2 Gamma-Maximin and Least Favorable Priors

In this section, we first present a linear program for identifying a *least favorable prior distribution* from the credal set $\mathcal{M}$ under consideration. Afterwards, we investigate the dual of this linear program and, in this way, provide a connection between pure acts $a \in \mathbb{A}$ that maximize expected utility with respect to a least favorable prior and randomized acts $\lambda \in G(\mathbb{A})$ that are optimal with respect to the Gamma-Maximin criterion.

Before we proceed, some additional notation is needed: For a credal element $\pi \in \mathcal{M}$, let $B(\pi)$ denote the maximal expectation with respect to π that an act from $\mathbb{A}$ can yield (that is $B(\pi) = \mathbb{E}_\pi(u_{a^*})$, where $a^* \in \mathbb{A}$ maximizes expected utility with respect to π). The set of all acts $a \in \mathbb{A}$ that maximize expected utility with respect to π is denoted by $\mathbb{A}_\pi$. Further, we call a credal element $\pi^- \in \mathcal{M}$ a *least favorable prior (lfp)* from $\mathcal{M}$ iff $B(\pi^-) \leqslant B(\pi)$ holds for all $\pi \in \mathcal{M}$. Specifically, π^- is a lfp, if it yields the minimal best possible expected utility under all concurring elements on the credal set. Proposition 3 describes a linear program for determining a lfp from $\mathcal{M}$.

Proposition 3. *Let $(\mathbb{A}, \Theta, u)$ denote a decision problem and let $\mathcal{M}$ be of the form (5). Consider the linear program*

$$w_1 - w_2 \longrightarrow \min_{(w_1, w_2, \pi_1, \ldots, \pi_m)} \tag{8}$$

with constraints $(w_1, w_2, \pi_1, \ldots, \pi_m) \geqslant 0$ and

- $\sum_{j=1}^m \pi_j = 1$
- $\underline{b}_s \leqslant \sum_{j=1}^m f_s(\theta_j) \cdot \pi_j \leqslant \bar{b}_s \quad$ *for all* $s = 1, ..., r$
- $w_1 - w_2 \geqslant \sum_{j=1}^m u_{ij} \cdot \pi_j$ *for all* $i = 1, \ldots n$.

Then the following holds:

(i) Every optimal solution $(w_1^*, \ldots, \pi_m^*)$ to (8) induces a least favorable prior $\pi^- \in \mathcal{M}$ by setting $\pi^-(\{\theta_j\}) := \pi_j^*$.
(ii) There always exists a least favorable prior. □

A lfp can be understood as a kind of "pignistic" probability, representing the decision problem under complex uncertainty in a way that is specific to the problem and the criterion under consideration, but in return gives the exact criterion value. This contrasts lfps from pignistic probabilities in Smets' spirit, who argued that a decision problem under complex uncertainty could be approached by distinguishing between a *credal level*, where the uncertain beliefs are to be expressed with all their ambiguity and scarceness by an imprecise probability (belief function in Smets' context), and a *decision level*, where eventually the imprecise probability is condensed into a traditional probability on which expected utility theory could be applied (see, e.g., [9,10], as well as, e.g., [1] for geometric techniques to represent belief functions by a single precise probability).

We now show some connections between least favorable priors and randomized Gamma-Maximin acts w.r.t. $\mathcal{M}$ ($\mathcal{M}$-Maximin). Recalling its definition, a randomized act $\lambda^* \in G(\mathbb{A})$ is said to be *$\mathcal{M}$-Maximin optimal* iff for all $\lambda \in G(\mathbb{A})$:

$$\underline{\mathbb{E}}_{\mathcal{M}}\left[G(u)_{\lambda^*}\right] \geqslant \underline{\mathbb{E}}_{\mathcal{M}}\left[G(u)_{\lambda}\right] \tag{9}$$

where $\underline{\mathbb{E}}_{\mathcal{M}}(X) := \min_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(X)$ for random variables $X : (\Theta, 2^{\Theta}) \to \mathbb{R}$.[3] It turns out that the linear program from Proposition 3 is *dual* to the one for determining a randomized $\mathcal{M}$-Maximin act described in [13, Sect. 3.2]. Together with *complementary slackness* (see, e.g., [15, Sect. 5.5]) from linear optimization theory, this allows to derive connections between lfps and the Gamma-Maximin.

Proposition 4. *Let $(\mathbb{A}, \Theta, u)$ denote a finite decision problem and let $\mathcal{M}$ be of the form (5). Then the following holds:*

(i) If π^- is a lfp from $\mathcal{M}$, then for all optimal randomized $\mathcal{M}$-Maximin acts $\lambda^* \in G(\mathbb{A})$ we have $\lambda^*(\{a\}) = 0$ for all $a \in \mathbb{A} \backslash \mathbb{A}_{\pi^-}$.

(ii) Let π^- denote a lfp from $\mathcal{M}$ and let $\lambda^* \in G(\mathbb{A})$ denote a randomized $\mathcal{M}$-Maximin act. Then for all $a \in \mathbb{A}_{\pi^-}$ we have

$$\mathbb{E}_{\pi^-}\left[u_a\right] = \underline{\mathbb{E}}_{\mathcal{M}}\left[G(u)_{\lambda^*}\right]$$

Proof. The dual programming problem of problem (8) is given by:

$$z_1 - z_2 + \sum_{s=1}^{r} (\underline{b}_s x_s - \overline{b}_s y_s) \longrightarrow \max_{(z_1, z_2, x_1, \ldots, x_r, y_1, \ldots, y_r, \lambda_1, \ldots, \lambda_n)} \tag{10}$$

with constraints $(z_1, z_2, x_1, \ldots, x_r, y_1, \ldots, y_r, \lambda_1, \ldots, \lambda_n) \geqslant 0$ and

- $\sum_{i=1}^{n} \lambda_i = 1$
- $z_1 - z_2 + \sum_{s=1}^{r} f_s(\theta_j)(x_s - y_s) \leq \sum_{i=1}^{n} u_{ij} \cdot \lambda_i$ for all $j = 1, \ldots, m$.

The resulting linear program (10) is exactly the one for determining a randomized act $\lambda^* \in G(\mathbb{A})$ which is optimal with respect to the $\mathcal{M}$-Maximin criterion as proposed and proven in [13, Sect. 3.2]. We now can use standard results on duality and complementary slackness (see, e.g., [15, Chap. 5]) to proof the proposition:

[3] For the special case of an ε-contamination model (a.k.a. *linear-vacuous model*) of the form $\mathcal{M}_{(\pi_0, \varepsilon)} := \{(1-\varepsilon)\pi_0 + \varepsilon\pi : \pi \in \mathcal{P}(\Theta)\}$, where $\mathcal{P}(\Theta)$ denotes the set of all probability measures on $(\Theta, 2^{\Theta})$, $\varepsilon > 0$ is a fixed contamination parameter and $\pi_0 \in \mathcal{P}(\Theta)$ is the central distribution, Gamma-Maximin is mathematically closely related to the Hodges and Lehmann-criterion: For fixed $X : (\Theta, 2^{\Theta}) \to \mathbb{R}$ we have $\underline{\mathbb{E}}_{\mathcal{M}_{(\pi_0, \varepsilon)}}(X) = \min_{\pi \in \mathcal{P}(\Theta)}((1-\varepsilon)\mathbb{E}_{\pi_0}(X) + \varepsilon\mathbb{E}_{\pi}(X)) = (1-\varepsilon)\mathbb{E}_{\pi_0}(X) + \varepsilon \min_{\pi \in \mathcal{P}(\Theta)} \mathbb{E}_{\pi}(X) = (1-\varepsilon)\mathbb{E}_{\pi_0}(X) + \varepsilon \min_{\theta \in \Theta} X(\theta)$. Thus, maximizing the lower expectation w.r.t. the ε-contamination model is equivalent to maximizing the Hodges and Lehmann-criterion with trust parameter $(1-\varepsilon)$ and prior π_0.

Part (i): Let $\pi^- \in \mathcal{M}$ denote a lfp and let $a_z \in \mathbb{A} \backslash \mathbb{A}_{\pi^-}$. Then

$$(\max\{B(\pi^-), 0\}, -\min\{B(\pi^-), 0\}, \pi^-(\{\theta_1\}), \ldots, \pi^-(\{\theta_m\})) \tag{11}$$

defines an optimal solution to (8) for which it holds that $B(\pi^-) > \mathbb{E}_{\pi^-}(u_{a_z})$. Thus, there exists an optimal solution to (8), for which the constraint $w_1 - w_2 \geq \sum_{j=1}^{m} u_{zj} \cdot \pi_j$ holds strictly and, therefore, the corresponding slack variable is strictly greater 0. Hence, by complementary slackness, the corresponding variable in the dual problem (10), that is λ_z, equals 0 for every optimal solution of problem (10). Finally, note that $\{\lambda_z^* : \lambda_z^* \text{ appears in optimal solution}\} = \{\lambda^*(\{a_z\}) : \lambda^* \in G(\mathbb{A})\ \mathcal{M}\text{-Maximin optimal}\}$, since, as (implicitly) shown in [13, Sect. 3.2], every $\mathcal{M}$-Maximin optimal $\lambda^* \in G(\mathbb{A})$ induces an optimal solution to (10), namely

$$(z_1^*, z_2^*, x_1, \ldots, x_r^*, y_1^*, \ldots, y_r^*, \lambda^*(\{a_1\}), \ldots, \lambda^*(\{a_n\})) \tag{12}$$

where $(z_1^*, z_2^*, x_1, \ldots, x_r^*, y_1^*, \ldots, y_r^*)$ denotes an optimal solution to a reduced version of problem (10) with $(\lambda_1, \ldots, \lambda_n) := (\lambda^*(\{a_1\}), \ldots, \lambda^*(\{a_n\}))$ fixed.

Part (ii): Let $\pi^- \in \mathcal{M}$ denote an lfp and $\lambda^* \in G(\mathbb{A})$ denote an $\mathcal{M}$-Maximin act. Use (11) and (12) to construct optimal solutions to (8) and (10). As the optimal value of (8) equals $B(\pi^-)$ and the optimal value of (10) equals $\underline{\mathbb{E}}_{\mathcal{M}}\left[G(u)_{\lambda^*}\right]$, the result follows by the duality theorem. □

As an immediate consequence of Proposition 4 (i), we can specify a condition under which randomization cannot improve utility, if optimality is defined in terms of the Gamma-Maximin criterion. Specifically, we have the following corollary.

Corollary 2. *If there exists a lfp π^- from $\mathcal{M}$ such that $\mathbb{A}_{\pi^-} = \{a_z\}$ for some $z \in \{1, \ldots, n\}$, then $\delta_{a_z} \in G(\mathbb{A})$ is the unique randomized $\mathcal{M}$-Maximin act. Specifically, considering randomized acts is unnecessary in such situations.* □

Running Example (Table 1): Algorithm 8 leads to the optimal solution vector $(13, 0, 0, 0, 0.7, 0.3)$. Thus, a lfp π^- from $\mathcal{M}$ is induced by $(0, 0.7, 0.3, 0)$. Simple computation gives $\mathbb{A}_{\pi^-} = \{a_2\}$. Hence, according to Corollary 2, a_2 is the unique $\mathcal{M}$-Maximin act (even compared to randomized acts) with utility 13.

5 Summary and Concluding Remarks

We presented linear programming based approaches for determining optimal randomized acts and investigated what can be learned by dualizing these. Future research includes the following issues: If $\mathcal{M}$ is non-degenerated, i.e. $\pi(\{\theta\}) > 0$ for all $(\pi, \theta) \in \mathcal{M} \times \Theta$, the same holds for every lfp π^-. Since every π^- induces an optimal solution to (8), complementary slackness implies that all constraints of problem (10) are binding for every optimal solution. This gives a system of linear *equations* that have to be satisfied by every randomized $\mathcal{M}$-Maximin act. A natural question is: Under which conditions is this system sufficient to

identify an optimal act without solving an optimization problem at all? A further interesting point is that algorithm (7) for checking maximality of an act a_z takes into account all other acts a_i in one linear program simultaneously. This could be used to modify the algorithm for finding maximal acts that are not too far from being E-admissible in the sense that the involved probabilities π_{a_i} that establish maximality of a_z differ not too much w.r.t. the L_1-norm which can be guaranteed by imposing further *linear* constraints.

Acknowledgement. The authors would like to thank the three anonymous referees for their helpful comments and their support.

References

1. Cuzzolin, F.: Two new Bayesian approximations of belief functions based on convex geometry. IEEE T. Syst. Man. Cy. B **37**, 993–1008 (2007)
2. Gilboa, I., Schmeidler, D.: Maxmin expected utility with non-unique prior. J. Math. Econ. **18**, 141–153 (1989)
3. Huber, P.: Robust Statistics. Wiley, New York (1981)
4. Hodges, J., Lehmann, E.: The use of previous experience in reaching statistical decisions. Ann. Math. Stat. **23**, 396–407 (1952)
5. Hable, R., Troffaes, M.: Computation. In: Augustin, T., Coolen, F., de Cooman, G., Troffaes, M. (eds.) Introduction to Imprecise Probabilities, pp. 329–337. Wiley, Chichester (2014)
6. Kikuti, D., Cozman, F., Filho, R.: Sequential decision making with partially ordered preferences. Artif. Intel. **175**, 1346–1365 (2011)
7. Kofler, E., Menges, G.: Entscheiden bei unvollständiger Information. Springer, Berlin (1976)
8. Levi, I.: The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance. MIT Press, Cambridge (1983)
9. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. Int. J. Approx. Reason. **38**, 133–147 (2005)
10. Smets, P.: Decision making in a context where uncertainty is represented by belief functions. In: Srivastava, R., Mock, T. (eds.) Belief Functions in Business Decisions, pp. 17–61. Physica, Heidelberg (2002)
11. Schervish, M., Seidenfeld, T., Kadane, J., Levi, I.: Extensions of expected utility theory and some limitations of pairwise comparisons. In: Bernard, J.-M., Seidenfeld, T., Zaffalon, M. (eds.) Proceedings of ISIPTA 2003, pp. 496–510. Carleton Scientific, Waterloo (2003)
12. Troffaes, M.: Decision making under uncertainty using imprecise probabilities. Int. J. Approx. Reason. **45**, 17–29 (2007)
13. Utkin, L., Augustin, T.: Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In: Cozman, F., Nau, R., Seidenfeld, T. (eds.) Proceedings of ISIPTA 2005, pp. 349–358 (2005)
14. Utkin, L., Kozine, I.: Different faces of the natural extension. In: de Cooman, G., Fine, T., Seidenfeld, T. (eds.) Proceedings of ISIPTA 2001, pp. 316–323 (2001)
15. Vanderbei, R.: Linear Programming: Foundations and Extensions. Springer, New York (2014)
16. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

17. Weichselberger, K.: Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept. Physica, Heidelberg (2001)
18. Weichselberger, K.: Interval probability on finite sample spaces. In: Rieder, H. (ed.) Robust Statistics, Data Analysis, and Computer Intensive Methods, pp. 391–409. Springer, New York (1996)

Contribution 2:

Jansen, C.; Schollmeyer, G.; Augustin, T. (2017b): Quantifying degrees of E-admissibility in decision making with imprecise probabilities. To appear in: Cozman, F. and Wheeler, G. (eds): *Festschrift in Honor of Teddy Seidenfeld*. Theory and Decision Library A, Springer.

As the book is not published yet, the following identical technical report is included in Chapter 3 of this thesis instead:

Jansen, C.; Schollmeyer, G.; Augustin, T. (2018a): Quantifying degrees of E-admissibility in decision making with imprecise probabilities. Department of Statistics (LMU München): *Technical Reports*, Nr. 215, March 2018.

Available under:

`https://epub.ub.uni-muenchen.de/42674/`

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Christoph Jansen, Georg Schollmeyer and Thomas Augustin

Quantifying Degrees of E-Admissibility in Decision Making with Imprecise Probabilities

Technical Report Number 215, 2018
Department of Statistics
University of Munich

http://www.statistik.uni-muenchen.de

Quantifying Degrees of E-admissibility in Decision Making with Imprecise Probabilities

C. Jansen[a,*], G. Schollmeyer[a], T. Augustin[a]

[a] *Department of Statistics, LMU Munich, Ludwigsstr. 33, D-80539 Munich, Germany*

Abstract

This paper is concerned with decision making using imprecise probabilities. In the first part, we introduce a new decision criterion that allows for explicitly modeling how far decisions that are optimal in terms of Walley's maximality are accepted to deviate from being optimal in the sense of Levi's E-admissibility. For this criterion, we also provide an efficient and simple algorithm based on linear programming theory. In the second part of the paper, we propose two new measures for quantifying the extent of E-admissibility of an E-admissible act, i.e. the size of the set of measures for which the corresponding act maximizes expected utility. The first measure is the maximal diameter of this set, while the second one relates to the maximal barycentric cube that can be inscribed into it. Also here, for both measures, we give linear programming algorithms capable to deal with them. Finally, we discuss some ideas in the context of ordinal decision theory. The paper concludes with a stylized application example illustrating all introduced concepts.

Keywords: Decision Making under Uncertainty; Imprecise Probabilities; E-admissibility; Maximality; Linear Programming; Ordinal Decision Theory; Stochastic Dominance

1. Introduction

A fair amount of the challenges arising in the modern sciences, e.g. parameter estimation and hypothesis testing in statistics, modeling an agent's preferences and choice behavior in philosophy and economics or the formalization of game theoretic problems, can be embedded into the formal framework of decision theory under uncertainty. If moreover the uncertainty underlying the decision situation is describable by some classical probability measure on the space of uncertain states of nature, we find ourselves within the framework of maximizing expected utility and we can draw on the whole toolbox of this well-investigated and elegant mathematical theory.

However, it is well known that in practice the necessity to specify a precise (i.e. classical) probability measure on the space of uncertain states might involve too strong consistency conditions regarding the beliefs of the decision maker of interest: It for instance might be the case that some decision maker finds it highly probable, say at least 0.8, that she will have had dinner in some restaurant by 9 p.m. tonight. However, since she doesn't know at all what the city she is traveling to has to offer, she cannot split this belief among different types of restaurants. That is even so if she made the (rather simplifying) assumption that the above probability exactly equals 0.8, there

*Corresponding author
Email address: christoph.jansen@stat.uni-muenchen.de (C. Jansen)

is too less information for inferring, for instance, the probability of the dinner to take place in a Chinese restaurant. For situations of this kind (and also for less artificial ones), working with imprecise probabilities (Walley (1991); Weichselberger (2001), see also, e.g., Augustin et al. (2014) for an introduction) has become more and more attractive recently, since these allow for utilizing also partial probabilistic knowledge without the necessity of making assumptions that aren't met.

Much work has been undertaken on decision making with imprecise probabilities, and several strategies for optimal decision making have been proposed. Surveys of the theory are given in Seidenfeld (2004b); Troffaes (2007); Etner et al. (2012); Huntley et al. (2014). For original sources see, e.g., Kofler and Menges (1976); Levi (1974); Walley (1991); Gilboa and Schmeidler (1989). In the present paper, we contribute some new insights especially in the context of Levi's E-admissibility.

The paper is organized as follows: In Section 2, we recall the basic model of finite decision theory (Section 2.1) and the most commonly applied decision principles from precise and imprecise decision making (Section 2.2) for reference. Section 3 is divided into two parts: In Section 3.1 we contrast the criteria $\mathcal{M}$-maximality and E-admissibility and introduce a new decision criterion that in some sense lies in between the two. In Section 3.2, we propose two measures, one optimistic and one pessimistic, for quantifying the extent of E-admissibility of some E-admissible acts under consideration. In Section 4 we discuss decision problems in which the utility function is only interpretable in terms of an ordinal utility representation, however, utility differences have no meaning. Again, we recall and discuss criteria for both the precise and the imprecise case. In Section 5, we analyze a stylized application example and apply the theory developed in the paper. Section 6 concludes.

2. The Basic Model

We start our discussion by recalling the classical setup of decision making under uncertainty in Section 2.1 and the most commonly applied decision criteria under different types of uncertainty in Section 2.2 for reference.

2.1. Framework

Throughout most parts of the paper[1], we will consider the common model of finite decision theory: Some agent (or decision maker) is asked to decide which act a_i to choose from a finite set $\mathbb{A} = \{a_1, \ldots, a_n\}$ of available acts. However, the utility of the chosen act is fraught with uncertainty: it depends on which state of nature from a finite set $\Theta = \{\theta_1, \ldots, \theta_m\}$ of possible states corresponds to the true description of reality. Specifically, we assume that the utility of every pair $(a, \theta) \in \mathbb{A} \times \Theta$ can be evaluated by some real-valued cardinal utility function $u : \mathbb{A} \times \Theta \to \mathbb{R}$ that is unique up to a positive linear transformation.[2] We denote by $u_{ij} := u(a_i, \theta_j)$ the utility of choosing a_i given θ_j is the true state. For every act $a \in \mathbb{A}$, the utility function u is naturally associated with a random variable $u_a : (\Theta, 2^\Theta) \to \mathbb{R}$ defined by $u_a(\theta) := u(a, \theta)$ for all $\theta \in \Theta$. Similarly, for every $\theta \in \Theta$, we can define a random variable $u^\theta : (\mathbb{A}, 2^\mathbb{A}) \to \mathbb{R}$ by setting $u^\theta(a) := u(a, \theta)$ for all $a \in \mathbb{A}$. The structure of the basic model is visualized in Table 1.

Depending on the situation, the standard model will sometimes be extended for randomized acts, which are classical probability measures λ on $(\mathbb{A}, 2^\mathbb{A})$. Choosing λ is then interpreted as

[1] The one exception is the discussion in Section 4, where we do not assume a cardinal utility representation.

[2] See Schervish et al. (2013) for the situation where multiple utilities through different currencies are available and exchange rates have to be taken into account.

Table 1: The basic model of finite decision theory.

$u(a_i, \theta_j)$	$\boldsymbol{\theta_1}$	$\cdots$	$\boldsymbol{\theta_m}$
$\mathbf{a_1}$	$u(a_1, \theta_1)$	$\cdots$	$u(a_1, \theta_m)$
$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\mathbf{a_n}$	$u(a_n, \theta_1)$	$\cdots$	$u(a_n, \theta_m)$

leaving the final decision to a random experiment which yields act a_i with probability $\lambda(\{a_i\})$. We denote the set of randomized acts on $(\mathbb{A}, 2^{\mathbb{A}})$ by $G(\mathbb{A})$.

If also randomized acts are considered, the original utility function u on $\mathbb{A} \times \Theta$ can straightforwardly be extended to a utility function $G(u)$ on $G(\mathbb{A}) \times \Theta$ by assigning each pair (λ, θ) the expectation of the random variable u^θ under the measure λ, i.e. $G(u)(\lambda, \theta) := \mathbb{E}_\lambda\big[u^\theta\big]$, i.e. the expectation of utility that choosing the randomized act λ will lead to, given θ is the true description of reality. Every act $a \in \mathbb{A}$, sometimes called pure act when the difference to randomized acts needs to be emphasized, then can uniquely be identified with the Dirac-measure $\delta_a \in G(\mathbb{A})$, and we have $u(a, \theta) = G(u)(\delta_a, \theta)$ for all $(a, \theta) \in \mathbb{A} \times \Theta$. Again, for every $\lambda \in G(\mathbb{A})$ fixed, the extended utility function $G(u)$ is associated with a random variable $G(u)_\lambda$ on $(\Theta, 2^\Theta)$ by setting $G(u)_\lambda(\theta) := G(u)(\lambda, \theta)$ for all $\theta \in \Theta$. We refer to the triplet $(\mathbb{A}, \Theta, u)$ as the (finite) decision problem and to the triplet $(G(\mathbb{A}), \Theta, G(u))$ as the corresponding randomized extension.

Finally, note that the standard model of decision theory also contains statistical estimation and hypothesis testing problems as special cases: If we, in addition to the basic problem $(\mathbb{A}, \Theta, u)$, observe some random variable $X : \Omega \to \mathcal{X}$ such that $X \sim P_\theta$ if $\theta \in \Theta$ is the true state of nature, that is we know the distribution of the random experiment if we know the true state, then statistical procedures can be viewed as decision functions $d : \mathcal{X} \to \mathbb{A}$ that map observed data on acts. The utility function u of the original problem then very naturally can be extended to a gain function $U : \mathcal{D} \times \Theta \to \mathbb{R}$ for evaluating decision functions by setting $U(d, \theta) := \mathbb{E}_{P_\theta}[u^\theta \circ d]$. Here, $\mathcal{D}$ denotes some appropriate set of possible statistical procedures. Formally, the resulting triplet $(\mathcal{D}, \Theta, U)$ then again can be viewed as a basic decision problem. Thus, even if we do not explicitly formulate our results for statistical procedures in the following, they always also can be interpreted in a statistical context.[3]

2.2. Criteria for Decision Making

Given a decision model $(\mathbb{A}, \Theta, u)$ of the form just recalled, the challenge is quickly explained: Determine an (in some sense) optimal act $a^* \in \mathbb{A}$ (or, depending on the context, optimal randomized act $\lambda^* \in G(\mathbb{A})$). The subtlety rather comes with the definition of the term optimality, since any

[3]It should, however, be emphasized that in the context of imprecise probabilistic models (like for instance credal sets or interval probabilities) the relationship between optimal decision functions in terms of prior risk and posteriori loss optimal acts may be more subtle than in the context of precise probability: the main theorem of Bayesian decision theory may fail (cf., e.g., Augustin (2003, Section 2.3)). This failure is in essence a variant of the general phenomenon of potential sequential incoherence in decision making and discrepancy between extensive and normal forms, as investigated in depth by Seidenfeld (e.g., Seidenfeld (1988, 1994)). Immediate counter-examples arise from the phenomenon of dilation, which has intensively been studied by Seidenfeld and co-authors (cf., e.g., Seidenfeld (1994), Seidenfeld and Wassermann (1993), Wassermann and Seidenfeld (1994)), see also, e.g., Liu (2015).

meaningful definition necessarily has to depend on (what the decision maker assumes about) the mechanism generating the states of nature. Here, traditional decision theory mainly covers two extreme poles: (I) The generation of the states follows a known classical probability measure π on $(\Theta, 2^\Theta)$ or (II) it can be compared to a game against an omniscient enemy. In these cases optimality is almost unanimously defined by the following two well-known principles:

(I) *Maximizing Expected Utility:* Label any act $a^* \in \mathbb{A}$ optimal that receives maximal expected utility with respect to π, i.e. for which $\mathbb{E}_\pi[u_{a^*}] \geq \mathbb{E}_\pi[u_a]$ for all other $a \in \mathbb{A}$. We denote by $\mathbb{A}_\pi$ the set of all acts from $\mathbb{A}$ that maximize expected utility with respect to π.

(II) *Wald's Maximin Principle:* Label any act $a^* \in \mathbb{A}$ optimal that receives highest possible utility value under that state that is worst possible for this particular act, i.e. for which $\min_{\theta\in\Theta} u(a^*, \theta) \geq \min_{\theta\in\Theta} u(a^*, \theta)$. We denote by $\mathbb{A}_W$ the set of all maximin acts.

Straightforwardly, principles (I) and (II) generalize to randomized acts, and we will denote the corresponding sets of optimal randomized acts by $G(\mathbb{A})_\pi$ and $G(\mathbb{A})_W$, respectively. In contrast, defining optimality of acts becomes less obvious if (A) the probability measure π is only partially known (case of imprecise probabilities) or (B) there is uncertainty about the complete appropriateness of it (case of uncertainty about precise probabilities). In situation (A), one commonly assumes that the available probabilistic information is describable by a polyhedral[4] set $\mathcal{M}$ of probability measures on $(\Theta, 2^\Theta)$ of the form

$$\mathcal{M} := \left\{\pi |\ \ \underline{b}_s \leqslant \mathbb{E}_\pi(f_s) \leqslant \overline{b}_s \ \forall s = 1, ..., r\right\} \tag{1}$$

where, for all $s = 1, ..., r$, we have $(\underline{b}_s, \overline{b}_s) \in \mathbb{R}^2$ such that $\underline{b}_s \leqslant \overline{b}_s$[5] and $f_s : \Theta \to \mathbb{R}$, which is an example for a imprecise probabilistic model. Specifically, the available information is assumed to be describable by lower and upper bounds for the expected values of a finite number of random variables on the space of states. Note that this also includes models in which the uncertainty arises from a variety of different (possibly precise) expert opinions: If, for instance, each from a bunch of experts gives precise expected payoff estimates for a number of stocks, we take $\mathcal{M}$ to be the set of probabilities that yield for every stock an expectation that ranges within the lowest and the highest expert guess. Most simply, this includes also the case where every expert specifies a precise probability measure on the state space, since a probability measure is always representable by a family of indicator functions. The picture of $\mathcal{M}$ being the opinions of a committee of experts will be used at different points in the paper (similarly as also done in, e.g., Bradley (2015)).

Under an imprecise probabilistic model of form (1), several optimality criteria for decision making had been proposed (cf., e.g., Troffaes (2007); Etner et al. (2012); Huntley et al. (2014) for general surveys and Utkin and Augustin (2005); Kikuti et al. (2011); Hable and Troffaes (2014); Jansen et al. (2017a) for computational aspects). We now briefly recall the ones among them which are most important for our purposes:

[4]See, however, e.g., Wheeler (2012), Majo-Wilson and Wheeler (2016, Section 2), and the references therein, for arguments to consider also non-convex sets of probabilities.

[5]For technical convenience we assume, wlog, that $0 \in [\underline{b}_s, \overline{b}_s]$ for all $s = 1, ..., r$ in the following. Note that if $\mathcal{M}$ is described by functions $(f_1, \ldots, f_r)$ and bounds $((\underline{b_1}, \overline{b}_1), \ldots, (\underline{b_r}, \overline{b}_r))$ not meeting this assumptions, we can always equivalently characterize it by functions $(f_1 - c_1, \ldots, f_r - c_r)$ and bounds $((\underline{b_1} - c_1, \overline{b}_1 - c_1), \ldots, (\underline{b_r} - c_r, \overline{b}_r - c_r))$, where, for all $s = 1, \ldots, r$, we set $c_s = \underline{b}_s$ if $\underline{b}_s > 0$ and $c_s = -\overline{b}_s$ if $\overline{b}_s < 0$ and $c_s = 0$ if $0 \in [\underline{b}_s, \overline{b}_s]$.

(A_1) *$\mathcal{M}$-Maximin* (*$\mathcal{M}$-Maximax*):[6] Label any act $a^* \in \mathbb{A}$ optimal that maximizes expected utility with respect to the worst (best) compatible probability measure, i.e. for which $\min_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u_{a^*}) \geq \min_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u_a)$ (resp. $\max_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u_{a^*}) \geq \max_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u_a)$) for all $a \in \mathbb{A}$. We denote by $\mathbb{A}_{\underline{\mathcal{M}}}$ (resp. $\mathbb{A}_{\overline{\mathcal{M}}}$) the set of $\mathcal{M}$-maximin (resp. $\mathcal{M}$-maximax) acts.

Clearly, $\mathcal{M}$-maximin is a rather pessimistic criterion that reflects the attitude of decision makers that react averse to the ambiguity between the different compatible probabilities from $\mathcal{M}$. Contrarily, $\mathcal{M}$-maximax reflects the attitude of ambiguity seeking agents. Note also that in the extreme cases where either the credal set $\mathcal{M}$ is the set of all precise probability measures (vacuousness) or it contains only one such measure (ideal stochasticity), the criterion $\mathcal{M}$-maximin reduces to Wald's maximin principle or precise expectation maximization, respectively.

(A_2) *$\mathcal{M}$-Maximality*:[7] Label any act $a^* \in \mathbb{A}$ optimal that dominates every other available act $a \in \mathbb{A}$ in expectation with respect to at least one probability measure $\pi_a \in \mathcal{M}$, i.e. if for every $a \in \mathbb{A}$ there exists $\pi_a \in \mathcal{M}$ such that $\mathbb{E}_{\pi_a}(u_{a^*}) \geq \mathbb{E}_{\pi_a}(u_a)$. We denote the set of all $\mathcal{M}$-maximal acts by $\mathbb{A}_{\max}$.

The idea of $\mathcal{M}$-maximality thus is to exclude every act a_0 from the decision problem for which there exists another act a_1 that dominates it with respect to every compatible probability measure. Note that $\mathcal{M}$-maximality can be viewed as a *local* decision criterion: The preference between the acts a_0 and a_1 is independent of the other available acts in $\mathbb{A} \setminus \{a_0, a_1\}$ or, as Schervish et al. (2003)[8] puts it, $\mathcal{M}$-maximality is induced by pairwise comparisons of acts in $\mathbb{A}$ only. Note further that, in the extreme case of $\mathcal{M}$ being a singleton, the criterion reduces to classical expectation maximization.

(A_3) *E-Admissibility*:[9] Label an act $a^* \in \mathbb{A}$ optimal if it maximizes expected utility among all other available acts with respect to a least one compatible probability measure, i.e. if there exists $\pi^* \in \mathcal{M}$ such that $a^* \in \mathbb{A}_{\pi^*}$. We denote by $\mathbb{A}_{\mathcal{M}}$ the set of all E-admissible acts from $\mathbb{A}$ with respect to the credal set $\mathcal{M}$.

In contrast to $\mathcal{M}$-maximality, the concept of E-admissibility can rather by viewed as a *global* decision criterion: In order to be able to build a preference between two acts a_0 and a_1, utility information for all the other available acts in $\mathbb{A} \setminus \{a_0, a_1\}$ is required. To put it in the words of Schervish et al. (2003) again: E-admissibility, in general, is not induced by pairwise comparisons of acts in $\mathbb{A}$ only. Again, in the case of ideal stochasticity the criterion reduces to classical expectation maximization. Contrarily, in the case of vacuousness every act that is not dominated by another act in every state is E-admissible.[10]

Again, if randomized acts are of interest, we denote the corresponding optimal sets by $G(\mathbb{A})_{\underline{\mathcal{M}}}$, $G(\mathbb{A})_{\overline{\mathcal{M}}}$, $G(\mathbb{A})_{\max}$ and $G(\mathbb{A})_{\mathcal{M}}$. As easy to see, it holds that $G(\mathbb{A})_\pi = \mathrm{conv}(\tilde{\mathbb{A}}_\pi)$, where we have $\tilde{\mathbb{A}}_\pi := \{\delta_a : a \in \mathbb{A}_\pi\}$ and $\mathrm{conv}(S)$ denotes the convex hull of S. Thus, we can easily construct

[6] See, for instance, Kofler and Menges (1976) and Gilboa and Schmeidler (1989). Many authors denote $\mathcal{M}$ by Γ, and thus the name Γ-maximin is common as well.

[7] This criterion is mainly advocated by Walley (1991) and work following him.

[8] Compare also Seidenfeld et al. (2010) and Vicig and Seidenfeld (2012, Section 3).

[9] This criterion is introduced by Levi (1974).

[10] Both criteria just discussed are also of high interest in forecasting with imprecise probabilities. While for imprecise probabilities there is no real-valued strictly proper scoring rule, it is possible to formulate an appropriate lexicographic strictly proper scoring rule with respect to $\mathcal{M}$-maximinity and E-admissibility, supplemented by $\mathcal{M}$-maximinity (Seidenfeld et al., 2012).

the set of randomized actions that maximize expected utility with respect to π by taking all convex combinations of pure acts with the same property. This fact is often used to argue that randomization does not pay out in the context of maximizing expected utility. Moreover, as shown by Walley (cf., Walley (1991, p. 163)) and emphasized in Schervish et al. (2003), we have that $G(\mathbb{A})_{\max} = G(\mathbb{A})_{\mathcal{M}}$, i.e. in the context of randomized acts the criteria $\mathcal{M}$-maximality and E-admissibility coincide in the sense of selecting the same optimal acts.

To complete the section, we now recall one criterion of optimality for situation (B), i.e. the case of an uncertain precise probability π: The criterion of Hodges and Lehmann (cf. Hodges and Lehmann (1952)). One motivation of this decision principle is to model the decision maker's skepticism in the available probability measure more directly. It is defined as follows:

(B_1) *Hodges and Lehmann Optimality:* Label any act $a^* \in \mathbb{A}$ optimal that maximizes the term $\alpha\mathbb{E}_\pi(u_{a^*}) + (1-\alpha)\min_\theta u(a^*,\theta)$ among all other acts $a \in \mathbb{A}$, that is which maximizes a weighted sum of the expected utility and the worst state utility. The value $\alpha \in [0,1]$ expresses the degree of trust that the agent assigns to the probability measure π.

Note that Hodges and Lehmann optimality can be viewed as a special case of $\mathcal{M}$-maximinity (cf., for instance, Jansen et al. (2017a)): If the underlying credal set is chosen to arise from an ε-contamination model (a.k.a. linear-vacuous mixture model) having the form

$$\mathcal{M}_{(\pi_0,\varepsilon)} := \big\{(1-\varepsilon)\pi_0 + \varepsilon\pi : \pi \in \mathcal{P}(\Theta)\big\}$$

where $\mathcal{P}(\Theta)$ is the set of all probabilities on $(\Theta, 2^\Theta)$, $\varepsilon > 0$ is a fixed contamination parameter and $\pi_0 \in \mathcal{P}(\Theta)$ is the central distribution, it holds

$$\begin{aligned}\underline{\mathbb{E}}_{\mathcal{M}_{(\pi_0,\varepsilon)}}(X) &= \min_{\pi\in\mathcal{P}(\Theta)}((1-\varepsilon)\mathbb{E}_{\pi_0}(X) + \varepsilon\mathbb{E}_\pi(X)) \\ &= (1-\varepsilon)\mathbb{E}_{\pi_0}(X) + \varepsilon\min_{\pi\in\mathcal{P}(\Theta)}\mathbb{E}_\pi(X) \\ &= (1-\varepsilon)\mathbb{E}_{\pi_0}(X) + \varepsilon\min_{\theta\in\Theta}X(\theta)\end{aligned}$$

for arbitrary random variables $X : (\Theta, 2^\Theta) \to \mathbb{R}$. Thus, maximizing the lower expectation w.r.t. the ε-contamination model is equivalent to maximizing the Hodges and Lehmann-criterion with trust parameter $(1-\varepsilon)$ and probability π_0. This connection is also of interest for Bayesian statistical inference with imprecise probabilities: As pointed out by Seidenfeld and Wassermann (1996) in the discussion of Walley (1996) and made explicit in Herron et al. (1997), the well-investigated Imprecise Dirichlet Model (IDM) for generalized Bayesian statistical learning is mathematically equivalent to an ε-contamination model with the relative frequencies as the central distribution π_0. Taking into account the above calculation, this also shows a very close relation between decision making in the IDM (e.g., Utkin and Augustin (2007)) and the criterion suggsted by Hodges and Lehmann.

3. *E*-admissibility, Maximality and a Criterion in between

We start our discussion by setting focus on the criteria $\mathcal{M}$-maximality and E-admissibility and develop some new ideas in this context. The discussion is divided in two main parts: In Section 3.1 we briefly compare the two criteria and then propose a new criterion providing an adjustable trade-off between them, for which we also derive a linear programming based algorithm. Afterwards, in Section 3.2, we discuss a new measure for quantifying the extent of E-admissibility of an E-admissible act of interest.

3.1. Comparing E-admissibility and Maximality

As already seen in the previous section, when considering also randomized acts, the concepts of $\mathcal{M}$-maximality and E-admissibility with respect to $\mathcal{M}$ induce the same optimal acts and therefore coincide. However, for a finite (or more general non-convex) set of acts $\mathbb{A}$ the criterion of $\mathcal{M}$-maximality is the strictly weaker condition in the sense that $\mathbb{A}_{\mathcal{M}} \subset \mathbb{A}_{\max}$. Our first result describes how to construct the set $G(\mathbb{A})_{\mathcal{M}}$ of all randomized E-admissible acts (and therefore also the set $G(\mathbb{A})_{\max}$ of randomized $\mathcal{M}$-maximal acts) from the set $\mathbb{A}_{\mathcal{M}}$ of pure E-admissible acts.

Proposition 1. *Let the decision problems $(\mathbb{A}, \Theta, u)$ and $(G(\mathbb{A}), \Theta, G(u))$ and the sets $\mathcal{M}$, $\mathbb{A}_\pi$, $\mathbb{A}_{\mathcal{M}}$, $G(\mathbb{A})_\pi$, $G(\mathbb{A})_{\mathcal{M}}$ be defined as before. The following holds:*

$$G(\mathbb{A})_{\mathcal{M}} = \bigcup\nolimits_{\pi \in \mathcal{M}} \operatorname{conv}(\tilde{\mathbb{A}}_\pi)$$

where $\tilde{\mathbb{A}}_\pi := \{\delta_a : a \in \mathbb{A}_\pi\}$ and $\operatorname{conv}(S)$ denotes the convex hull of a set S.

Proof. $\subset$: Let $\lambda^* \in G(\mathbb{A})_{\mathcal{M}}$. Choose $\pi^* \in \mathcal{M}$ such that $\lambda^* \in G(\mathbb{A})_{\pi^*}$. Assume, for contradiction, there exists $a_0 \in \mathbb{A}$ such that $\lambda^*(\{a_0\}) > 0$ and $a_0 \notin \mathbb{A}_{\pi^*}$. Pick then $a_1 \in \mathbb{A}_{\pi^*}$ and define a randomized act $\lambda_0 \in G(\mathbb{A})$ by setting $\lambda_0(\{a\}) := \lambda^*(\{a\})$ for $a \in \mathbb{A} \setminus \{a_0, a_1\}$, $\lambda_0(\{a\}) := 0$ for $a = a_0$ and $\lambda_0(\{a\}) := \lambda^*(\{a_0, a_1\})$ for $a = a_1$. Then, the following calculation are immediate:

$$\begin{aligned}
\mathbb{E}_{\pi^*}[G(u)_{\lambda_0}] &= \sum_{a \in \mathbb{A}} \lambda_0(\{a\}) \mathbb{E}_{\pi^*}(u_a) \\
&= \sum_{a \in \mathbb{A} \setminus \{a_0, a_1\}} \lambda^*(\{a\}) \mathbb{E}_{\pi^*}(u_a) \;+\; \lambda^*(\{a_0, a_1\}) \mathbb{E}_{\pi^*}(u_{a_1}) \\
&> \sum_{a \in \mathbb{A}} \lambda^*(\{a\}) \mathbb{E}_{\pi^*}(u_a) \;=\; \mathbb{E}_{\pi^*}[G(u)_{\lambda^*}]
\end{aligned}$$

This contradicts $\lambda^* \in G(\mathbb{A})_{\pi^*}$. Therefore, we have $\lambda^* \in \operatorname{conv}(\tilde{\mathbb{A}}_{\pi^*})$.

$\supset$: Let conversely $\lambda^* \in \bigcup_{\pi \in \mathcal{M}} \operatorname{conv}(\tilde{\mathbb{A}}_\pi)$. Then there exists $\pi^* \in \mathcal{M}$ such that $\lambda^* \in \operatorname{conv}(\tilde{\mathbb{A}}_{\pi^*})$ and we have $\mathbb{E}_{\pi^*}[G(u)_{\lambda^*}] = \mathbb{E}_{\pi^*}(u_a)$ for all $a \in \mathbb{A}_{\pi^*}$. Choose $a_0 \in \mathbb{A}_{\pi^*}$ and observe that for arbitrary $\lambda \in G(\mathbb{A})$ it holds that

$$\mathbb{E}_{\pi^*}[G(u)_\lambda] = \sum_{a \in \mathbb{A}} \lambda(\{a\}) \mathbb{E}_{\pi^*}(u_a) \;\leq\; \mathbb{E}_{\pi^*}(u_{a_0}) = \mathbb{E}_{\pi^*}[G(u)_{\lambda^*}]$$

Thus there exists $\pi^* \in \mathcal{M}$ with respect to which λ^* maximizes expected utility implying that $\lambda^* \in G(\mathbb{A})_{\mathcal{M}}$. □

Since Proposition 1 allows us to construct both sets $G(\mathbb{A})_{\mathcal{M}}$ and $G(\mathbb{A})_{\max}$ once having computed the set $\mathbb{A}_{\mathcal{M}}$ of pure E-admissible acts, we restrict analysis to non-randomized acts for the rest of the section. For this setting, we now propose a new decision criterion that allows for labeling only such $\mathcal{M}$-maximal acts optimal that are not too far from being E-admissible with respect to $\mathcal{M}$ in the sense that the probabilities for which the corresponding act expectation dominates the other acts differ not too much from each other. The deviation of an act from E-admissibility can be explicitly controlled by an additional parameter ε.

Definition 1. *Let $(\mathbb{A}, \Theta, u)$ and $\mathcal{M}$ be defined as before and let $\varepsilon \geq 0$. An act $a^* \in \mathbb{A}$ is called E_ε-admissible if there exists a family of probability measures $(\pi_a)_{a \in \mathbb{A}}$ from $\mathcal{M}$ such that the following two conditions are satisfied:*

i) $\mathbb{E}_{\pi_a}(u_{a^*}) \geq \mathbb{E}_{\pi_a}(u_a)$ *for all* $a \in \mathbb{A}$ *and*

ii) $\| \pi_a - \pi_{a'} \| \leq \varepsilon$ *for all* $a, a' \in \mathbb{A}$*, where* $\| \cdot \|$ *denotes a norm on* $\mathcal{M}$*.*

We denote by $\mathbb{A}_\varepsilon$ *the set of all* E_ε*-admissible acts from* $\mathbb{A}$*.*

Remark 1. *Obviously, the set of* E_0*-admissible acts coincides with the set of* E*-admissible acts with respect to* $\mathcal{M}$*, i.e.* $\mathbb{A}_0 = \mathbb{A}_\mathcal{M}$*. Moreover, for* ε^* *chosen sufficiently large, namely* $\varepsilon^* \geq b := \sup_{\pi,\pi' \in \mathcal{M}} \| \pi - \pi' \|$*, the set of* E_{ε^*}*-admissible acts coincides with the set of* $\mathcal{M}$*-maximal acts, i.e.* $\mathbb{A}_{\varepsilon^*} = \mathbb{A}_{max}$*. For* $\varepsilon \in (0, b)$*, it usually will hold that* $\mathbb{A}_\mathcal{M} \subsetneq \mathbb{A}_\varepsilon \subsetneq \mathbb{A}_{max}$ *and the set* $\mathbb{A}_\varepsilon$ *then exactly contains those* $\mathcal{M}$*-maximal acts that are not too far (controlled by* ε*) from being* E*-admissible. Moreover, the criterion is monotone in the sense that for* $\varepsilon_1 \geq \varepsilon_2$ *we have that* $\mathbb{A}_{\varepsilon_2} \subset \mathbb{A}_{\varepsilon_1}$*. Thus, it allows for ranking* $\mathcal{M}$*-maximal acts with respect to their deviation from being* E*-admissible.*

If we again take the point of view that $\mathcal{M}$ arises from different expert opinions, it turns out that the criterion of E_ε-admissibility is based on a quite convincing intuition: Consider for instance a political decision maker that consults an advisory body of experts when it comes to facing difficult decisions. In this situation, applying E-admissibility corresponds to only choosing acts which one fixed expert labels optimal among all other options. Contrarily, in terms of $\mathcal{M}$-maximality an act is already optimal if for each other act there is at least one expert preferring the former to the latter, no matter how different the involved experts are in opinion. Here, E_ε-admissibility builds a bridge between these two extremes: While the decision maker can still make use of opinions of different experts, she nevertheless can explicitly control by an additional parameter ε how strong the experts contributing to the decision process are allowed to differ in opinion.

We now provide an algorithm for checking whether an act in a given decision problem is E_ε-admissible for a fixed value ε. It turns out that this, provided the L_1-norm is used for measuring the distances between the elements of $\mathcal{M}$, can be done by solving one single, relatively simple, linear programming problem. We arrive at the following proposition.

Proposition 2. *Let* $(\mathbb{A}, \Theta, u)$ *and* $\mathcal{M}$ *be defined as before and let* $\varepsilon \geq 0$*. For some act* $a_z \in \mathbb{A}$*, consider the following linear programming problem:*

$$\sum_{i=1}^{n} \Big(\sum_{j=1}^{m} \gamma_{ij} \Big) \longrightarrow \max_{(\gamma_{11}, \ldots, \gamma_{nm})} \tag{2}$$

with constraints $(\gamma_{11}, \ldots, \gamma_{nm}) \geqslant 0$ *and*

- $\sum_{j=1}^{m} \gamma_{ij} \leqslant 1 \quad$ *for all* $i = 1, \ldots, n$
- $\underline{b}_s \leqslant \sum_{j=1}^{m} f_s(\theta_j) \cdot \gamma_{ij} \leqslant \overline{b}_s \quad$ *for all* $s = 1, ..., r$*,* $i = 1, \ldots, n$
- $\sum_{j=1}^{m} (u_{ij} - u_{zj}) \cdot \gamma_{ij} \leqslant 0 \quad$ *for all* $i = 1, \ldots, n$
- $\sum_{j=1}^{m} |\gamma_{i_1 j} - \gamma_{i_2 j}| \leq \varepsilon$ *for all* $i_1 > i_2 \in \{1, \ldots, n\}$

Then $a_z \in \mathbb{A}$ *is* E_ε*-admissible iff the optimal outcome of (2) equals* n*.*

Proof. Clearly, if (2) possesses an optimal solution $(\gamma_{11}^*, \ldots, \gamma_{nm}^*)$ yielding an objective value of n, then the constraints guarantee that setting $\pi_{a_i}(\{\theta_j\}) := \gamma_{ij}^*$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$ defines a family of probability measures $(\pi_{a_i})_{i=1,\ldots n,}$ from $\mathcal{M}$ satisfying the properties from Definition 1. Thus, $a_z \in \mathbb{A}_\varepsilon$.

If conversely $a_z \in \mathbb{A}_\varepsilon$, we can choose a family of probability measures $(\pi_{a_i})_{i=1,\ldots n,}$ from $\mathcal{M}$ satisfying the properties from Definition 1. One then easily verifies that setting $\gamma_{ij}^* := \pi_{a_i}(\{\theta_j\})$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$ defines an admissible solution $(\gamma_{11}^*, \ldots, \gamma_{nm}^*)$ to (2) that yields an objective value of n. □

Remark 2. *To see the linearity of the constraint* $\sum_{j=1}^m |\gamma_{i_1 j} - \gamma_{i_2 j}| \leq \varepsilon$ *for all* $i_1 > i_2 \in \{1, \ldots, n\}$ *in the above linear programming problem, one can proceed as follows: Add* $2m$ *decision variables* $l_1, \ldots, l_m$ *and* $o_1, \ldots, o_m$ *and replace the above constraints equivalently by the constraints* $l_j \leq \gamma_{ij} \leq o_j$ *for all* $i = 1, \ldots, n$ *and* $j = 1, \ldots, m$ *as well as* $\sum_{j=1}^m (o_j - l_j) \leq \varepsilon$*. In sum, the programming problem (2) thus possesses* $n+rn+n+nm+1 = n(1+r+m)+1$ *constraints and* $nm+m = m(n+1)$ *decision variables.*

We conclude the section by illustrating the results so far by a brief toy example, which in parts is also discussed in Seidenfeld (2004a, p. 2) in order to demonstrate that $\mathcal{M}$-maximinity does not imply E-admissibility with respect to $\mathcal{M}$ and vice versa. We additionally show how the proposed concept of E_ε-admissibility can help to clarify analysis in such situations. The example reads as follows:

Example 1. *Consider the basic decision problem* $(\mathbb{A}, \Theta, u)$ *that is defined by the following utility table*

u_{ij}	θ_1	θ_2
a_1	1	0
a_2	0	1
a_3	4/10	4/10
a_4	6/10	11/35

Moreover, suppose the uncertainty on the states is modeled by the credal set

$$\mathcal{M} = \Big\{\pi : 0.3 \leq \pi(\{\theta_1\}) \leq 0.8\Big\}$$

In this case, we have $\mathbb{A}_{\mathcal{M}} = \{a_1, a_2\}$, $\mathbb{A}_{\max} = \mathbb{A}$ *and* $\mathbb{A}_{\underline{\mathcal{M}}} = \{a_3, a_4\}$*. Thus, we have a situation with two different* $\mathcal{M}$*-maximin acts, which are both not* E*-admissible. In order to make a decision between the acts* a_3 *and* a_4*, we can apply the* E_ε*-criterion to see which of the two is closer to being* E*-admissible. We receive the following results:*

outcome of (2) for	$\varepsilon = 0.1$	$\varepsilon = 0.2$	$\varepsilon = 0.3$	$\varepsilon = 0.6$
a_3	≈ 0.76	≈ 1.51	≈ 2.27	4
a_4	2.3	4	4	4

The results show that act a_4 *is that act among the* $\mathcal{M}$*-maximin acts that is closer to* E*-admissibility, since it is* E_ε*-admissible already for an* ε*-level of* 0.2 *whereas act* a_3 *is not. Thus, it could be argued that it* a_4 *is preferable. Finally, if we additionally consider randomized acts, then Proposition 1 and the discussions in Section 2.2 imply that it holds*

$$G(\mathbb{A})_{\mathcal{M}} = G(\mathbb{A})_{\max} = G(\mathbb{A})_\varepsilon = \mathrm{conv}(\{\delta_{a_1}, \delta_{a_2}\})$$

for arbitrary values of $\varepsilon \geq 0$*.*

3.2. The Extents of E-admissible acts

In the previous section, we considered acts optimal that are not too far from being E-admissible. We accordingly weakened the concept of E-admissibility towards acts that are in some sense almost E-admissible. In this section we take rather the opposed direction and address the following question: Given an E-admissible act $a \in \mathbb{A}_{\mathcal{M}}$ with respect to some credal set $\mathcal{M}$, how large is the set of compatible probability measures from $\mathcal{M}$ for which act a maximizes expected utility? If we again use the picture of $\mathcal{M}$ modeling the opinions of some committee of experts, the question translates as follows: How diverse can these experts be in opinion while still all sharing the view that act a is optimal?

In order to answer this question, we propose two measures for the extent of E-admissibility of acts in the following: The maximal extent and the uniform extent. While the first concept measures the maximal diameter of the set of measures for which the considered act maximizes expected utility, the latter one searches for a maximal set that can be inscribed into this set. Together, the two measures will be shown to give a pretty good impression about the extent of E-admissibility. We start by defining the concept of maximal extent.

Definition 2. *Let $(\mathbb{A}, \Theta, u)$ and $\mathcal{M}$ be defined as before and let $\|\cdot\|$ denote some norm on $\mathcal{M}$. Moreover, let $a \in \mathbb{A}_{\mathcal{M}}$ be an E-admissible act with respect to $\mathcal{M}$ and denote by $\mathcal{M}_a$ the set $\{\pi \in \mathcal{M} : a \in \mathbb{A}_\pi\}$. We define the (maximal) extent $ext_{\mathcal{M}}(a)$ of act a as the number*

$$ext_{\mathcal{M}}(a) := \sup_{\pi,\pi' \in \mathcal{M}_a} \| \pi - \pi' \|$$

i.e. as the maximum distance of probability measures $\pi, \pi' \in \mathcal{M}_a$ with respect to $\|\cdot\|$ for which act a maximizes expected utility.

Why is the measure $ext_{\mathcal{M}}(\cdot)$ sensible for the question motivating the section? To see that, first note that intuitively if $ext_{\mathcal{M}}(a)$ is large, then act a maximizes expected utility with respect to very different (in the sense of highly distant) probability measures from $\mathcal{M}$. To directly connect this observation to the size of the set $\mathcal{M}_a$, it is important to mention that $\mathcal{M}_a$ is a convex set and therefore all measure lying on the "line" between the two maximum distance measures again have to be contained in $\mathcal{M}_a$. Thus, $ext_{\mathcal{M}}(a)$ indeed can be viewed as a measure of the size of the set of probabilities for which act a is optimal and therefore is sensible for the above questions. The following proposition gives an algorithm for computing $ext_{\mathcal{M}}(a)$ by solving a series of linear programming problems for the case that $\|\cdot\| := \|\cdot\|_\infty$.

Proposition 3. *Let $(\mathbb{A}, \Theta, u)$ and $\mathcal{M}$ be defined as before and let $a_z \in \mathbb{A}_{\mathcal{M}}$. Consider, for every $j = 1, \ldots, m$, the linear programming problem*

$$\gamma_{1j} - \gamma_{2j} \longrightarrow \max_{(\gamma_{11},\ldots,\gamma_{1m},\gamma_{21},\ldots,\gamma_{2m})} \tag{P$_j$}$$

with constraints $(\gamma_{11}, \ldots, \gamma_{1m}, \gamma_{21}, \ldots, \gamma_{2m}) \geq 0$ and

- $\sum_{j=1}^m \gamma_{ij} = 1$ *for all* $i = 1, 2$
- $\underline{b}_s \leqslant \sum_{j=1}^m f_s(\theta_j) \cdot \gamma_{ij} \leqslant \overline{b}_s$ *for all* $s = 1, ..., r$ *and* $i = 1, 2$
- $\sum_{j=1}^m (u_{\ell j} - u_{zj}) \cdot \gamma_{ij} \leqslant 0$ *for all* $i = 1, 2$ *and* $\ell = 1, \ldots, n$

Denote by $g(j)$ the optimal objective of problem (P_j). Then the maximal extent of act a with respect to $\|\cdot\|_\infty$ is given by $ext_{\mathcal{M}}(a) = \max_j g(j)$.

Proof. For $j \in \{1,\dots,m\}$, let $(\gamma_{11}^j,\dots,\gamma_{1m}^j,\gamma_{21}^j,\dots,\gamma_{2m}^j)$ denote an optimal solution to problem (P_j).[11] Then the constraints guarantee that by setting $\pi_1^j(\{\theta_t\}) := \gamma_{1t}^j$ and $\pi_2^j(\{\theta_t\}) := \gamma_{2t}^j$ for all $t = 1,\dots,m$ we define two measures $\pi_1^j, \pi_2^j \in \mathcal{M}_a$ with the property

i) $g(j) = |\pi_1^j(\{\theta_j\}) - \pi_2^j(\{\theta_j\})| \geq |\pi_1(\{\theta_j\}) - \pi_2(\{\theta_j\})|$ for all $\pi_1, \pi_2 \in \mathcal{M}_a$

Let $j^* \in \{1,\dots,m\}$ with $g(j^*) = |\pi_1^{j^*}(\{\theta_{j^*}\}) - \pi_2^{j^*}(\{\theta_{j^*}\})| = \max_j g(j)$. Due to i), for all $j \in \{1,\dots,m\}$ arbitrary, it then holds that:

ii) $|\pi_1^{j^*}(\{\theta_{j^*}\}) - \pi_2^{j^*}(\{\theta_{j^*}\})| \geq |\pi_1(\{\theta_j\}) - \pi_2(\{\theta_j\})|$ for all $\pi_1, \pi_2 \in \mathcal{M}_a$

This implies that $\| \pi_1^{j^*} - \pi_2^{j^*} \|_\infty \geq \| \pi_1 - \pi_2 \|_\infty$ for all $\pi_1, \pi_2 \in \mathcal{M}_a$, which then implies that $ext_{\mathcal{M}}(a_z) = g(j^*) = \max_j g(j)$. □

Remark 3. *Note that, instead of solving m linear programming problems for computing the value $ext_{\mathcal{M}}(a)$ as proposed in Proposition 3, one alternatively could solve one bilinear programming problem with objective function*

$$\sum_{j=1}^{m} \xi_j(\gamma_{1j} - \gamma_{2j}) + \sum_{j=1}^{m} \xi_{m+j}(\gamma_{2j} - \gamma_{1j}) \longrightarrow \max_{(\gamma_{11},\dots,\gamma_{2m},\xi_1,\dots\xi_{2m})} \tag{3}$$

with the same constraints as above and additional constraints $\xi_1,\dots\xi_{2m} \geq 0$ and $\sum_{j=1}^{2m} \xi_j = 1$. This approach has the advantage that the value $ext_{\mathcal{M}}(a)$ can also be computed with respect to $\|\cdot\|_1$ instead of $\|\cdot\|_\infty$. To see that, simply replace the constraint $\sum_{j=1}^{2m} \xi_j = 1$ by the set of constraints $\sum_{j=1}^{2m} \xi_j = m$ and $\xi_j + \xi_{m+j} = 1$ for all $j = 1,\dots,m$. However, note that the resulting bilinear programming problem then no longer is decomposable into m linear programming problems, since the solutions of the single problems can no longer be treated independently of each other as in the case of the $\|\cdot\|_\infty$-norm.

Despite its intuitiveness, the extent $ext_{\mathcal{M}}(\cdot)$ of an E-admissible act has a drawback in certain situations: It measure the size of set $\mathcal{M}_a$ only in one direction, namely the most extreme one. Therefore, the maximal extent alone might be not capable of distinguishing situations that definitely are worth to be distinguished in this context. This drawback is most easily explained by the schematic picture in Figure 1.

In order to react the problem that might arise when only considering the extent $ext_{\mathcal{M}}(a)$ of an E-admissible act $a \in \mathbb{A}_{\mathcal{M}}$ for measuring the size of the set $\mathcal{M}_a$, we now introduce another concept for addressing this question, and call it uniform extent. This measure relates to the diameter of the largest barycentric ε-cube that can be inscribed into $\mathcal{M}_a$.

Definition 3. *Let $(\mathbb{A}, \Theta, u)$ and $\mathcal{M}$ be defined as before and let $a \in \mathbb{A}_{\mathcal{M}}$. We define the uniform extent $uxt_{\mathcal{M}}(a)$ of act a with respect to $\mathcal{M}$ as the number*

$$uxt_{\mathcal{M}}(a) := \max\Big\{\varepsilon : \exists \pi \in \mathcal{M}_a \text{ s.t. } B_\varepsilon(\pi) \subset \mathcal{M}_a\Big\} \tag{4}$$

[11] Note that, due to standard results from linear programming theory, such an optimal solution always exists since the constraint set is bounded and there always exists an admissible solution since $a_z \in \mathbb{A}_{\mathcal{M}}$.

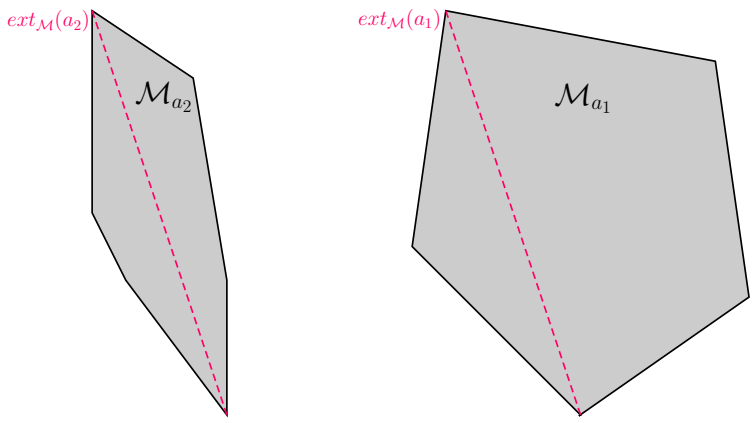

Figure 1: Two sets $\mathcal{M}_{a_1}$ and $\mathcal{M}_{a_2}$ with the same extent, however, quite different size.

where $B_\varepsilon(\pi) := \mathrm{conv}(\{\pi_\varepsilon^{1+}, \pi_\varepsilon^{1-}, \pi_\varepsilon^{2+}, \pi_\varepsilon^{2-}, \ldots, \pi_\varepsilon^{m+}, \pi_\varepsilon^{m-}\})$ *with*

$$\pi_\varepsilon^{j^*+}(\{\theta_j\}) = \begin{cases} \pi(\{\theta_j\}) + \varepsilon & \text{if } j = j^* \\ \pi(\{\theta_j\}) - \frac{\varepsilon}{m-1} & \text{if } j \neq j^* \end{cases}$$

$$\pi_\varepsilon^{j^*-}(\{\theta_j\}) = \begin{cases} \pi(\{\theta_j\}) - \varepsilon & \text{if } j = j^* \\ \pi(\{\theta_j\}) + \frac{\varepsilon}{m-1} & \text{if } j \neq j^* \end{cases}$$

is the barycentric ε-cube around π. Thus, the uniform extent of act a is half the diameter of the largest barycentric ε-cube that can be inscribed into $\mathcal{M}_a$.

The uniform extent for the schematic picture discussed earlier is illustrated in Figure 2. We see that we now can distinguish between the two situations. However, as easy to imagine, also the uniform extent might sometimes be too pessimistic just as the maximal extent is too optimistic. Hence, a good approach is to consider both measures *ext* and *uxt* simultaneously. Together, they will give a pretty good impression of the extent of E-admissibility.

We now propose an algorithm for computing the uniform extent of some fixed E-admissible act with respect to $\mathcal{M}$ under consideration. Again, it shows that this can be done by solving one single, relatively simple, linear programming problem. Here, the main idea is to explicitly model the distributions $\pi_\varepsilon^{1+}, \pi_\varepsilon^{1-}, \ldots, \pi_\varepsilon^{m+}, \pi_\varepsilon^{m-}$ from Definition 3 by decision variables and utilizing the fact that $\mathcal{M}_a$ is a convex set. The uniform extent is then computed by maximizing over the value of ε. Precisely, we arrive at the following proposition.

Proposition 4. *Let $(\mathbb{A}, \Theta, u)$ and $\mathcal{M}$ be defined as before and let $a_z \in \mathbb{A}_\mathcal{M}$. Consider the linear programming problem*

$$\varepsilon \longrightarrow \max_{(\gamma_1, \ldots, \gamma_m, \varepsilon)} \tag{5}$$

with constraints $(\gamma_1, \ldots, \gamma_m, \varepsilon) \geq 0$ and

- $\sum_{j=1}^m \gamma_j = 1$
- $\gamma_j \geq \varepsilon \quad$ *for all* $j = 1, \ldots, m$

- $\underline{b}_s \leqslant \sum_{j=1}^m f_s(\theta_j) \cdot \gamma_j + \varepsilon \cdot d(j^*, s) \leqslant \overline{b}_s \quad$ *for all* $s = 1, ..., r, j^* = 1, \ldots, m$
- $\underline{b}_s \leqslant \sum_{j=1}^m f_s(\theta_j) \cdot \gamma_j - \varepsilon \cdot d(j^*, s) \leqslant \overline{b}_s \quad$ *for all* $s = 1, ..., r, j^* = 1, \ldots, m$
- $\sum_{j=1}^m (u_{\ell j} - u_{zj}) \cdot \gamma_j + \varepsilon \cdot c(j^*, \ell) \leqslant 0 \quad$ *for* $\ell = 1, \ldots, n, j^* = 1, \ldots, m$
- $\sum_{j=1}^m (u_{\ell j} - u_{zj}) \cdot \gamma_j - \varepsilon \cdot c(j^*, \ell) \leqslant 0 \quad$ *for* $\ell = 1, \ldots, n, j^* = 1, \ldots, m$

where

$$c(j^*, \ell) = (u_{\ell j^*} - u_{zj^*}) - \frac{1}{(m-1)} \sum_{j \neq j^*} (u_{\ell j} - u_{zj})$$

and

$$d(j^*, s) = f_s(\theta_{j^*}) - \frac{1}{(m-1)} \sum_{j \neq j^*} f_s(\theta_j)$$

Then the uniform extent $uxt_{\mathcal{M}}(a_z)$ *of* a_z *is given by the optimal value of problem (5).*

Proof. First, note that every pair $(\pi, \varepsilon) \in \mathcal{M}_{a_z} \times \mathbb{R}_0^+$ with $B_\varepsilon(\pi) \subset \mathcal{M}_{a_z}$ induces an admissible solution to (5) with objective value ε by setting $\gamma_j := \pi(\{\theta_j\})$, since we have

$$\sum_{j=1}^m f_s(\theta_j) \cdot \gamma_j + \varepsilon \cdot d(j^*, s) = \mathbb{E}_{\pi_\varepsilon^{j^*+}}(f_s) \in (\underline{b}_s, \overline{b}_s)$$

$$\sum_{j=1}^m f_s(\theta_j) \cdot \gamma_j - \varepsilon \cdot d(j^*, s) = \mathbb{E}_{\pi_\varepsilon^{j^*-}}(f_s) \in (\underline{b}_s, \overline{b}_s)$$

for all $s = 1, ..., r, j^* = 1, \ldots, m$ and it then holds that

$$\sum_{j=1}^m (u_{\ell j} - u_{zj}) \cdot \gamma_j + \varepsilon \cdot c(j^*, \ell) = \mathbb{E}_{\pi_\varepsilon^{j^*+}}(u_\ell - u_z) \leqslant 0$$

$$\sum_{j=1}^m (u_{\ell j} - u_{zj}) \cdot \gamma_j - \varepsilon \cdot c(j^*, \ell) = \mathbb{E}_{\pi_\varepsilon^{j^*-}}(u_\ell - u_z) \leqslant 0$$

for $\ell = 1, \ldots, n, j^* = 1, \ldots, m$ due to the constraints. Since $a_z \in \mathbb{A}_{\mathcal{M}}$ and, therefore, there exists $\pi_0 \in \mathcal{M}_{a_z}$, it is then guaranteed that problem (5) always possesses an admissible solution (just take the one induced by $(\pi_0, 0)$). Since the set of admissible solutions is obviously bounded, it also possesses an optimal solution by standard results on linear programming theory.

Let $(\gamma_1^*, \ldots, \gamma_m^*, \varepsilon^*)$ denote such an optimal solution to (5). Utilizing again the above identities (in the opposite way), we see that setting $\pi^*(\{\theta_j\}) := \gamma_j^*$ defines a probability measure $\pi^* \in \mathcal{M}_{a_z}$ such that $B_{\varepsilon^*}(\pi^*) \subset \mathcal{M}_{a_z}$ with $\varepsilon^* = uxt_{\mathcal{M}}(a_z)$. □

Remark 4. *The linear programming problem (5) possesses* $m+1$ *decision variables and* $1+2mr+2mn$ *constraints. It therefore might become computationally expensive for very large problems.*

We conclude the section by applying the proposed measures of the extent of E-admissibility to the toy example that was already introduced at the end of Section 3.1.

Example 2. *Consider again the situation of Example 1. We want to compute the extent* $ext_{\mathcal{M}}(\cdot)$ *of both E-admissible acts* a_1 *and* a_2*. Solving the series of linear programming problems from Proposition 3 for both acts gives* $ext_{\mathcal{M}}(a_1) = 0.3$ *and* $ext_{\mathcal{M}}(a_2) = 0.2$ *with respect to the* $\|\cdot\|_\infty$*-norm. Therefore, it could be argued that* a_1 *is the most preferable among the E-admissible acts with respect to* $\mathcal{M}$*. Additionally, we are interested in the uniform extent* $uxt_{\mathcal{M}}(\cdot)$ *of the acts* a_1 *and* a_2*. Solving the linear programming problem introduced in Proposition 4 gives the results* $uxt_{\mathcal{M}}(a_1) = 0.15$ *and* $uxt_{\mathcal{M}}(a_2) = 0.1$*, even strengthening the argument that act* a_1 *is the most preferable among the E-admissible acts.*

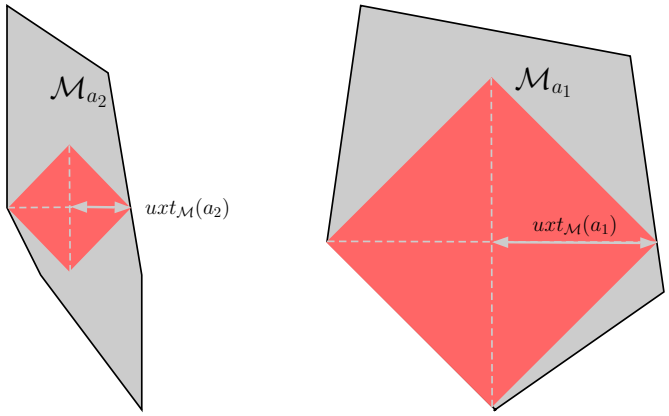

Figure 2: The measure $uxt_{\mathcal{M}}(\cdot)$ indeed gives different values to the sets $\mathcal{M}_{a_1}$ and $\mathcal{M}_{a_2}$ and therefore resolves the drawback of the measure $ext_{\mathcal{M}}(\cdot)$.

4. The Ordinal Case

Up to this point, all decision criteria discussed, with the exception of Wald's maximin principle, made explicit use of the cardinality of the utility function u involved in the basic decision problem $(\mathbb{A}, \Theta, u)$. However, as widely known, assuming cardinal utility implicitly demands the decision maker's preferences to satisfy pretty strong axiomatic assumptions which are often not met in practice. If the deviation from these axioms is too strong, it often makes sense to work with decision criteria that can cope with purely ordinal preferences.[12] For this reason, in this section the utility function u in the decision problem $(\mathbb{A}, \Theta, u)$ is solely interpreted as an ordinal utility representation. Particularly, utility differences with respect to u have no meaningful interpretation apart from their sign in what follows.

We again start by briefly summarizing some criteria that still make sense in the presence of purely ordinal preferences. If, additional to the ordinal utility information, a precise probability measure π on the state space is available, again several different criteria appear natural:

(C_1) *Pairwise Stochastic Dominance:* Label any act $a_0 \in \mathbb{A}$ optimal for which there does not exist another act $a_1 \in \mathbb{A} \setminus \{a_0\}$ such that $\mathbb{E}_\pi(t \circ u_{a_1}) \geq \mathbb{E}_\pi(t \circ u_{a_0})$ for every non-decreasing function $t : \mathbb{R} \to \mathbb{R}$. If, contrarily, it is the case that $\mathbb{E}_\pi(t \circ u_{a_1}) \geq \mathbb{E}_\pi(t \circ u_{a_0})$ for every non-decreasing function $t : \mathbb{R} \to \mathbb{R}$, we say that a_1 stochastically dominates a_0 (cf., e.g., Lehmann (1955); Mosler and Scarsini (1991)).

Clearly, pairwise stochastic dominance can rather be viewed as a local decision criterion, since the preference between two acts $a_0, a_1 \in \mathbb{A}$ does not depend on which other acts from $\mathbb{A} \setminus \{a_0, a_1\}$ are also available to the decision maker. Moreover, it also possesses a very natural interpretation: Act a_1 is preferred to act a_0 if every expectation maximizing decision maker with the same ordinal utility function would have the same preference between the two acts. Note that often acts will be incomparable with respect to stochastic dominance, since it will hold $\mathbb{E}_\pi(t_1 \circ u_{a_1}) > \mathbb{E}_\pi(t_1 \circ u_{a_0})$

[12] Another, very prominent, way for proceeding in such situations is working with partially cardinal preference relations as done in Seidenfeld et al. (1995).

for one function t_1 and $\mathbb{E}_\pi(t_2 \circ u_{a_1}) < \mathbb{E}_\pi(t_2 \circ u_{a_0})$ for another function t_2.

(C_2) *Joint Stochastic Dominance:* Label every act $a_0 \in \mathbb{A}$ optimal for which there exists a strictly increasing function $t^* : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}_\pi(t^* \circ u_{a_0}) \geq \mathbb{E}_\pi(t^* \circ u_a)$ for all $a \in \mathbb{A}$, i.e. if there exists one expectation maximizing agent with the same ordinal utility function for which a_0 maximizes expected utility among all other available acts (cf. Jansen et al. (2017b)).

Obviously, this is an example for a global criterion: If there exists a function t^* with the desired properties for all $a \in \mathbb{A}$, this does not necessarily imply the existence of such a function for $\mathbb{A}^* := \mathbb{A} \cup \{a^*\}$ (simply choose a^* to have higher utility that every act in $\mathbb{A}$ in every state of nature).

(C_3) *Pairwise Statistical Preference:* Label every act $a_0 \in \mathbb{A}$ optimal for which there exists no other act $a_1 \in \mathbb{A} \setminus \{a_0\}$ such that

$$\pi(\{\theta : u_{a_1}(\theta) \geq u_{a_0}(\theta)\}) > \pi(\{\theta : u_{a_0}(\theta) \geq u_{a_1}(\theta)\})$$

i.e. if there is no other act a_1 which has higher probability of yielding a higher utility value than a_0. If contrarily there exists such an act a_1, we say that a_1 statistically dominates a_0 (cf., e.g., Montes (2014, Section 2.2.1)).

Clearly, statistical preference can rather be viewed as a local decision criterion, since the preference between two acts a_0 and a_1 does not depend on acts from $\mathbb{A} \setminus \{a_0, a_1\}$.

(C_4) *Joint Statistical Preference:* Label every act $a_0 \in \mathbb{A}$ optimal for which it holds that $D_\pi(a_0) \geq D_\pi(a)$ for all $a \in \mathbb{A}$, where

$$D_\pi(a) := \pi(\{\theta : u(a, \theta) \geq u(a', \theta) \text{ for all } a' \in \mathbb{A}\})$$

that is if a_0 has the highest probability to be utility dominant among all other available acts.

This criterion is clearly global: Enlarging the set of acts $\mathbb{A}$ to a new set of acts $\mathbb{A}^* := \mathbb{A} \cup \{a^*\}$ might completely change the preference between two acts acts $a_0, a_1 \in \mathbb{A}$ in the sense that $D_\pi^{\mathbb{A}}(a_0) > D_\pi^{\mathbb{A}}(a_1)$ but $D_\pi^{\mathbb{A}^*}(a_0) < D_\pi^{\mathbb{A}^*}(a_1)$.[13]

If no precise probability measure π is available and the uncertainty on the state space is again characterized by a credal set $\mathcal{M}$ of the form defined in (1), then there are several possibilities to generalize the decision criteria (C_1), (C_2). (C_3) and (C_4). A detailed study of these different possibilities as well as algorithmic approaches that are capable to deal with the resulting criteria is given in Montes (2014, Sections 4.1.1 and 4.1.2) and Jansen et al. (2017b). An algorithm for detecting stochastic dominance for the case that the different decision consequences are only partially ordered is introduced in Schollmeyer et al. (2017). Here, we only give a small selection of

[13] For a simple example consider the decision problems $(\mathbb{A}, \Theta, u)$ and $(\mathbb{A} \cup \{a^*\}, \Theta, \tilde{u})$ given by

	θ_3	θ_2	θ_3
a_1	2	2	5
a_2	3	3	3

	θ_3	θ_2	θ_3
a_1	2	2	5
a_2	3	3	3
a^*	1	2	6

and the prior π on Θ induced by $(\pi(\{\theta_1\}), \pi(\{\theta_2\}), \pi(\{\theta_3\})) = (0.2, 0.2, 0.6)$. Here we have that $D_\pi^{\mathbb{A}}(a_1) = 0.6 > 0.4 = D_\pi^{\mathbb{A}}(a_2)$ but $D_\pi^{\mathbb{A}^*}(a_2) = 0.4 > 0 = D_\pi^{\mathbb{A}^*}(a_1)$.

the criteria:

(D_1) *Joint Statistical Preference (Imprecise Version)*: Label any act $a_0 \in \mathbb{A}$ optimal for which it holds that $\min_{\pi \in \mathcal{M}} D_\pi(a_0) \geq \min_{\pi \in \mathcal{M}} D_\pi(a)$ for all other $a \in \mathbb{A}$, i.e. which maximizes the lower probability of the act to be dominant to all other available acts.

(D_2) *Joint Stochastic Dominance (Imprecise Version):* Label act $a_0 \in \mathbb{A}$ optimal if there exists a strictly increasing function $t^* : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}_\pi(t^* \circ u_{a_0}) \geq \mathbb{E}_\pi(t^* \circ u_a)$ for all $a \in \mathbb{A}$ and all $\pi \in \mathcal{M}$.

All the ordinal decision criteria just discussed can be handled either by hand or by utilizing linear programming techniques similar as seen in detail for the criteria discussed in Section 3 (see Jansen et al. (2017b) for details). Here, we only give an impression of how this could be done for the example of the imprecise version of joint stochastic dominance: To check whether an act $a_z \in \mathbb{A}$ is optimal in the sense of joint stochastic dominance in the imprecise version, we explicitly model the transformation function t^* by decision variables. Additionally, we require the extreme points $\pi^{(1)}, \ldots, \pi^{(T)}$ of the underlying credal set $\mathcal{M}$. We then consider the linear programming problem with the objective function

$$\varepsilon \longrightarrow \max_{(\varepsilon, t_{11}, \ldots, t_{nm})} \tag{6}$$

and constraints $(\varepsilon, t_{11}, \ldots, t_{nm}) \geq 0$ and

- $t_{11}, \ldots, t_{nm} \leq 1$
- $\sum_{j=1}^{m} (u_{zj} t_{zj} - u_{ij} t_{ij}) \cdot \pi^{(t)}(\{\theta_j\}) \geq 0$ for all $t = 1, \ldots, T,\ i = 1, \ldots, n$
- For $i, i^{'} \in \{1, \ldots, n\}$, $j, j^{'} \in \{1, \ldots, m\}$: $u_{ij} = u_{i^{'}j^{'}} \Rightarrow t_{ij} = t_{i^{'}j^{'}}$
- For $i, i^{'} \in \{1, \ldots, n\}$, $j, j^{'} \in \{1, \ldots, m\}$: $u_{ij} < u_{i^{'}j^{'}} \Rightarrow t_{ij} + \varepsilon \leq t_{i^{'}j^{'}}$

One then can show that act a_z is optimal in the sense of joint stochastic dominance in the imprecise version if and only if the optimal objective of the above program is strictly greater than 0. The idea here is that if there exists an optimal solution $(\varepsilon^*, t_{11}^*, \ldots, t_{nm}^*)$ with $\varepsilon^* > 0$, then the solution t_{ij}^* describes the necessary strictly increasing transformation of u, and we receive a desired function by choosing any increasing function $t^* : \mathbb{R} \to \mathbb{R}$ satisfying that $t^*(u_{ij}) = t_{ij}^* \cdot u_{ij}$ for all $i = 1, \ldots n$ and $j = 1, \ldots, m$.

Of course solving this linear program might become computationally very expensive and cumbersome as the number of extreme points of the set $\mathcal{M}$ might become as large as $m!$ (cf., Derks and Kuipers (2002); Wallner (2007)). However, convenient classes of credal sets exist where furthermore efficient enumeration procedures are available (such special cases include ordinal probabilities (cf., Kofler (1989, p. 26)), comparative probabilities (cf., Miranda and Destercke (2015)), necessity measures (cf., Schollmeyer (2015)), p-boxes (cf., Montes and Destercke (2017)), probability intervals (cf., Weichselberger and Pöhlmann (1990, Chapter 2) or de Campos et al. (1994)) or pari-mutual models (cf., Montes et al. (2017))).

Another easy to handle case appears if the credal set $\mathcal{M}$ under consideration directly arises as the convex hull of a finite number of precise probability estimates $\pi_{E_1}, \ldots, \pi_{E_K}$ of a committee of experts $E_1, \ldots, E_K$. In such cases the extreme points of the credal set $\mathcal{M}$ are always among the

experts guesses $\pi_{E_1}, \ldots, \pi_{E_K}$ for the probabilities, and the algorithm described above can directly be applied without the need for any previous computation of the extreme points. We conclude the section by a small example that continues Examples 1 and 2.

Example 3. *Consider again the situation of Examples 1 and 2. Here, the unique optimal act with respect to joint statistical preference in the imprecise version is* a_1 *with a value of* 0.3*. If we consider joint stochastic dominance in the imprecise version, we first need to compute the extreme points of* $\mathcal{M}$*, which are here given by the measures* $\pi^{(1)}, \pi^{(2)}$ *induced by* $\pi^{(1)}(\{\theta_1\}) = 0.3$ *and* $\pi^{(2)}(\{\theta_1\}) = 0.8$*. Solving the above linear programming problem (6) for all acts gives that acts* a_3 *and* a_4 *are optimal in terms of joint stochastic dominance in the imprecise version whereas acts* a_1 *and* a_2 *are not.*

5. A Stylized Application Example

In this section, we discuss a more realistic, yet stylized, application example in some more detail: Consider the situation where the decision maker wants to invest money in stocks of some company. The acts then correspond to the stocks of the different companies. Say the agent compares ten different stocks collected in $\mathbb{A} = \{a_1 \ldots, a_{10}\}$. Moreover, the states of nature then correspond to different economic scenarios which might or might not occur and which, each differently, would influence the payoffs of the stocks of the different companies. Say the agent incorporates the scenarios collected in $\Theta = \{\theta_1, \ldots, \theta_5\}$ in her market analysis. She summarizes the payoffs of the different stocks under the different scenarios in the following utility table:

$u(a_i, \theta_j)$	θ_1	θ_2	θ_3	θ_4	θ_5
$\mathbf{a_1}$	37	25	23	73	91
$\mathbf{a_2}$	50	67	2	44	94
$\mathbf{a_3}$	60	4	96	1	83
$\mathbf{a_4}$	16	24	31	26	100
$\mathbf{a_5}$	3	86	76	85	11
$\mathbf{a_6}$	12	49	66	56	14
$\mathbf{a_7}$	39	10	92	88	57
$\mathbf{a_8}$	62	52	80	71	42
$\mathbf{a_9}$	90	8	74	70	38
$\mathbf{a_{10}}$	63	68	36	69	9

Moreover, suppose the decision maker has observed the market development for quite a while, so that she can specify bounds for the probabilities of the different economic scenarios to occur (alternatively, the bounds for the scenario probabilities could also come from opinions of different expert the agent has consulted). Precisely, she specifies the uncertainty underlying the situation by the credal set

$$\mathcal{M} = \Big\{\pi : \underline{b}_s \leq \mathbb{E}_\pi[f_s] \leq \overline{b}_s \text{ for } s = 1, \ldots, 5\Big\}$$

where

- $f_s : \Theta \to \mathbb{R}$ is given by $f_s(\theta) := \mathbb{1}_{\{\theta_s\}}(\theta)$ for $s = 1, \ldots, 5$ and

- $\begin{pmatrix} \underline{b}_1 & \overline{b}_1 \\ \underline{b}_2 & \overline{b}_2 \\ \underline{b}_3 & \overline{b}_3 \\ \underline{b}_4 & \overline{b}_4 \\ \underline{b}_5 & \overline{b}_5 \end{pmatrix} = \begin{pmatrix} 0.1 & 0.3 \\ 0.05 & 0.1 \\ 0.1 & 0.2 \\ 0.2 & 0.4 \\ 0.15 & 0.2 \end{pmatrix}$

Applying the different decision criteria and the other concepts discussed in the paper, the decision maker arrives at the following results:

- Stock a_8 is the unique non-randomized $\mathcal{M}$-maximin act, i.e. $\mathbb{A}_{\underline{\mathcal{M}}} = \{a_8\}$. Thus, for a very pessimistic and ambiguity averse agent, act a_8 is the appropriate investment.

- Solving the programming problem from Proposition 2 for ε set to 0 for each act, we find the set of E-admissible acts with respect to $\mathcal{M}$ is given by $\mathbb{A}_{\mathcal{M}} = \{a_7, a_9\}$ (since the optimal value of the program is 10 for both acts). Hence, the $\mathcal{M}$-maximin act is not E-admissible with respect to $\mathcal{M}$. In order to further compare the E-admissible acts a_7 and a_9, we first compute the extent $ext_{\mathcal{M}}(\cdot)$ from Definition 2 for both of them. Solving the series of linear programming problems described in Proposition 3 gives the results $ext_{\mathcal{M}}(a_7) \approx 0.152$ and $ext_{\mathcal{M}}(a_9) = 0.2$, for which reason it could be argued that a_9 is the most preferable among the E-admissible acts. To see how informative the extent of the acts is, we are additionally interested in their uniform extents $uxt_{\mathcal{M}}(\cdot)$ in the sense of Definition 3. Solving the linear programming problem introduced in Proposition 4 gives $uxt_{\mathcal{M}}(a_7) = 0.025$ as well as $uxt_{\mathcal{M}}(a_9) = 0.025$. Thus, if we consider the uniform extent in order to measure the amount of E-admissibility of acts, it could be argued that the decision maker should be indifferent between the E-admissible acts a_7 and a_9.

- Solving the programming problem from Proposition 2 for ε set to 100 for each act, we find the set of $\mathcal{M}$-maximal acts is given by $\mathbb{A}_{\max} = \{a_7, a_8, a_9\}$. In order to make a decision between the $\mathcal{M}$-maximin act a_8 and the E-admissible acts a_7 and a_9, it is of interest how far a_8 is from being E-admissible. Solving the linear program from Proposition 2 for varying value of ε gives that a_8 is E_ε-admissible in the sense of Definition 1 already for a value of $\varepsilon = 0.01$. Hence, a_8 is very close to being E-admissible and, therefore, could be argued to be preferable to a_7 and a_9.

- The unique optimal act with respect to joint statistical preference in the imprecise version is a_7 with a value of 0.2. In order to see which of the acts are optimal in the sense of joint stochastic dominance in the imprecise version as discussed in the previous section, we first need to compute the extreme points of $\mathcal{M}$. There are 15 such extreme points.[14] They are given in the Table 2:

  Having obtained the extreme points, we can use algorithm (6) from Section 4 for every act in $\mathbb{A} = \{a_1, \dots a_{10}\}$. We find that the acts a_6 and a_{10} are not optimal in the sense of joint stochastic dominance in the imprecise version, whereas the acts in $\mathbb{A} \setminus \{a_6, a_{10}\}$ are.

[14]The calculation was performed with the `rcdd package` (see Geyer and Meeden (2017)), which provides an interface for using Fukuda (2017)'s `cdd library` in the R statistical computing environment (see R Core Team (2017)).

	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_4\}$	$\{\theta_5\}$
$\pi^{(1)}(\cdot)$	0.30	0.10	0.20	0.20	0.20
$\pi^{(2)}(\cdot)$	0.30	0.05	0.20	0.25	0.20
$\pi^{(3)}(\cdot)$	0.30	0.10	0.20	0.25	0.15
$\pi^{(4)}(\cdot)$	0.30	0.05	0.20	0.30	0.15
$\pi^{(5)}(\cdot)$	0.30	0.10	0.10	0.30	0.20
$\pi^{(6)}(\cdot)$	0.30	0.05	0.10	0.35	0.20
$\pi^{(7)}(\cdot)$	0.30	0.10	0.10	0.35	0.15
$\pi^{(8)}(\cdot)$	0.15	0.05	0.20	0.40	0.20
$\pi^{(9)}(\cdot)$	0.15	0.10	0.20	0.40	0.15
$\pi^{(10)}(\cdot)$	0.20	0.05	0.20	0.40	0.15
$\pi^{(11)}(\cdot)$	0.20	0.10	0.10	0.40	0.20
$\pi^{(12)}(\cdot)$	0.25	0.05	0.10	0.40	0.20
$\pi^{(13)}(\cdot)$	0.25	0.10	0.10	0.40	0.15
$\pi^{(14)}(\cdot)$	0.10	0.10	0.20	0.40	0.20
$\pi^{(15)}(\cdot)$	0.30	0.05	0.10	0.40	0.15

Table 2: Extreme points in the application example.

6. Summary and Concluding Remarks

In this paper we introduced and discussed some ideas in the context of decision theory using imprecise probabilistic model. Here, we first introduced a new decision criterion, E_ε-admissibility, that selects acts that are not too far from E-admissibility, where the accepted deviation from E-admissibility can be explicitly controlled by an additional parameter ε. Subsequently, we investigated how to measure the extent of E-admissibility of an E-admissible act of interest. Precisely, we introduced two different measures for this purpose: the maximal extent $ext_{\mathcal{M}}(a)$ and the uniform extent $uxt_{\mathcal{M}}(a)$ of an E-admissible act a. While the former corresponds to the maximal diameter of the set $\mathcal{M}_a$, the latter is related to the side length of the maximal barycentric ε-cube that can be inscribed into $\mathcal{M}_a$. For all concepts discussed we proposed (bi-)linear programming driven algorithms for computation. In the second part of the paper we recalled some concepts for decision making if a cardinal utility function is no longer available and there is (potentially) only imprecise probabilistic information. For the concept of imprecise joint stochastic dominance, we also discussed some details about computation.

There are several interesting directions that could be followed in future research of which we only want to briefly mention one: Consider again the viewpoint that the credal set $\mathcal{M}$ arises from the opinions of a committee of experts. In the discussion directly following Definition 1, we argued in favor of the concept of E_ε-admissibility, since it allows to take into account more than only one expert opinion while simultaneously allowing to control how far the involved experts may differ in opinion. This idea could easily be extended: Instead of only controlling how far the involved experts may differ from each other in terms of opinion, one could also control how far their opinions differ from some externally given criterion. If we take again our example of some politician with an advisory body of experts, the external criterion could for instance be the opinion of the politician herself, so that she only takes expert opinions into account that do not differ too much from her own one. Of course other examples for external criteria are imaginable.

Acknowledgments

We want to express our gratitude to Jean Baccelli, Seamus Bradley, Arthur Merin, Jürgen Landes, Gregory Wheeler and many other members of the Munich Center for Mathematical Philosophy for stimulating talks and discussion on decision theory and imprecise probabilities during the last few years, and to Teddy Seidenfeld for inspiring lectures and discussions not only during his visits to Munich.

References

Augustin, T. (2003). On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view: A cautionary note on updating imprecise priors. In Bernard, J.-M., Seidenfeld, T., and Zaffalon, M., editors, *Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '03, Lugano)*, pages 31–45. Carleton Scientific.

Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors (2014). *Introduction to Imprecise Probabilities*. Wiley, Chichester.

Bradley, S. (2015). How to choose among choice functions. In Augustin, T., Doria, S., Miranda, E., and Quaeghebeur, E., editors, *Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '15, Pescara)*, pages 57–66. Aracne.

de Campos, L., Huete, J., and Moral, S. (1994). Probability interval: A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196.

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

Etner, J., Jeleva, M., and Tallon, J.-M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, 26:234–270.

Fukuda, K. (2017). The cdd and cddplus Homepage. `https://www.inf.ethz.ch/personal/fukudak/cdd_home/` Last accessed Dec 22, 2017.

Geyer, C. J. and Meeden, G. D. (2017). *Package rcdd*. `https://CRAN.R-project.org/package=rcdd` and `www.stat.umn.edu/geyer/rcdd/`, Last accessed Dec 22, 2017.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153.

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

Herron, T., Seidenfeld, T., and Wasserman, L. (1997). Divisive conditioning: Further results on dilation. *Philosophy of Science*, 64:411–444.

Hodges, J. and Lehmann, E. (1952). The use of previous experience in reaching statistical decisions. *Annals of Mathematical Statistics*, 23:396–407.

Huntley, N., Hable, R., and Troffaes, M. (2014). Decision making. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors, *Introduction to Imprecise Probabilities*, pages 190–206. Wiley, Chichester.

Jansen, C., Augustin, T., and Schollmeyer, G. (2017a). Decision theory meets linear optimization beyond computation. In Antonucci, A., Cholvy, L., and Papini, O., editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty. ECSQARU 2017. Lecture Notes in Computer Science, vol 10369*, pages 329–339. Springer, Cham.

Jansen, C., Schollmeyer, G., and Augustin, T. (2017b). Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., 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.

Kikuti, D., Cozman, F., and Filho, R. (2011). Sequential decision making with partially ordered preferences. *Artificial Intelligence*, 175:1346–1365.

Kofler, E. (1989). *Prognosen und Stabilität bei unvollständiger Information [Forecasts and Stability under Incomplete Information]*. Campus, Frankfurt.

Kofler, E. and Menges, G. (1976). *Entscheiden bei unvollständiger Information [Decision Making under Incomplete Information]*. Springer, Berlin.

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

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, 71:391–418.

Liu, H. (2015). Common knowledge, ambiguity, and the value of information in games. In Augustin, T., Doria, S., Miranda, E., and Quaeghebeur, E., editors, *Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '15, Pescara)*, pages 167–176. Aracne.

Majo-Wilson, C. and Wheeler, G. (2016). Scoring imprecise credences: A mildly immodest proposal. *Philosophy and Phenomenological Research*, 93:55–78.

Miranda, E. and Destercke, S. (2015). Extreme points of the credal sets generated by comparative probabilities. *Journal of Mathematical Psychology*, 64:44–57.

Montes, I. (2014). *Comparison of Alternatives under Uncertainty and Imprecision*. PhD thesis, Oviedo. `http://digibuo.uniovi.es/dspace/bitstream/10651/28953/6/TD_IgnacioMontesGutierrez.pdf`. Last accessed December 22, 2017.

Montes, I. and Destercke, S. (2017). On extreme points of p-Boxes and belief functions. In Ferraro, M., Giordani, P., Vantaggi, B., Gagolewski, M., Gil, M., Grzegorzewski, P., and Hryniewicz, O., editors, *Soft Methods for Data Science*, pages 363–371. Springer.

Montes, I., Miranda, E., and Destercke, S. (2017). A study of the pari-mutuel model from the point of view of imprecise probabilities. In Antonucci, A., Corani, G., Couso, I., and Destercke, S., editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '17, Lugano)*, volume 62 of *Proceedings of Machine Learning Research*, pages 229–240. PMLR.

Mosler, K. and Scarsini, M. (1991). Some theory of stochastic dominance. In Mosler, K. and Scarsini, M., editors, *Stochastic Orders and Decision under Risk*, pages 203–212. Institute of Mathematical Statistics, Hayward, CA.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schervish, M., Seidenfeld, T., Kadane, J., and Levi, I. (2003). Extensions of expected utility theory and some limitations of pairwise comparisons. In Bernard, J.-M., Seidenfeld, T., and Zaffalon, M., editors, *Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '03, Lugano)*, pages 496–510. Carleton Scientific.

Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (2013). The effect of exchange rates on statistical decisions. *Philosophy of Science*, 80:504–532.

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

Schollmeyer, G., Jansen, C., and Augustin, T. (2017). 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)*. Available under: `https://epub.ub.uni-muenchen.de/40450/`. Last accessed December 22, 2017.

Seidenfeld, T. (1988). Decision theory without independence or without ordering, What is the difference? (with discussion). *Economics and Philosophy*, 4:267–315.

Seidenfeld, T. (1994). When normal form and extensive form solutions differ. In Prawitz, D., Skyrms, B., and Westerstahl, D., editors, *Logic, Methodology and Philosophy of Science IX (Uppsala, 1991)*, pages 451–463. Elsevier.

Seidenfeld, T. (2004a). A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus E-admissibility. *Synthese*, 140:69–88.

Seidenfeld, T. (2004b). Some decision theoretic issues for imprecise probability theory. Slides of a lecture given on the *First SIPTA School on Imprecise Probabilities*. Available under: `http://www.sipta.org/ssipta04/topics.htm`. Last accessed December 22, 2017.

Seidenfeld, T., Kadane, J., and Schervish, M. (1995). A representation of partially ordered preferences. *Annals of Statistics*, 23:2168–2217.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2010). Coherent choice functions under uncertainty. *Synthese*, 172:157–176.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2012). Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning*, 53:1248–1261.

Seidenfeld, T. and Wassermann, L. (1993). Dilation for sets of probabilities. *Annals of Statistics*, 21:1139–1154.

Seidenfeld, T. and Wassermann, L. (1996). Discussion of "Inferences from multinomial data: Learning about a bag of marbles" by P. Walley. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:49.

Troffaes, M. (2007). Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45:17–29.

Utkin, L. and Augustin, T. (2005). Powerful algorithms for decision making under partial prior information and

general ambiguity attitudes. In Cozman, F., Nau, R., and Seidenfeld, T., editors, *Proceedings of the Forth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '05, Pittsburgh)*, pages 349–358. SIPTA.

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

Vicig, P. and Seidenfeld, T. (2012). Bruno de Finetti and imprecision: Imprecise probability does not exist! *International Journal of Approximate Reasoning*, 53:1115–1123.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:3–57.

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

Wassermann, L. and Seidenfeld, T. (1994). The dilation phenomenon in robust Bayesian inference. (with discussion). *Journal of Statistical Planning and Inference*, 40:345–356.

Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept [Elementary Foundations of a More General Calculus of Probability I: Interval Probability as a Comprehensive Concept]*. Physica, Heidelberg.

Weichselberger, K. and Pöhlmann, S. (1990). *A Methodology for Uncertainty in Knowledge-based Systems*. Springer, Heidelberg.

Wheeler, G. (2012). Objective Bayesianism and the problem of non-convex evidence. *British Journal for the Philosophy of Science*, 63:841–850.

Contribution 3:

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

The original publication is available at Elsevier via

`https://doi.org/10.1016/j.ijar.2018.04.011`

International Journal of Approximate Reasoning 98 (2018) 112–131

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences ☆,☆☆

C. Jansen *, G. Schollmeyer, T. Augustin

Department of Statistics, LMU Munich, Ludwigsstr. 33, 80539 Munich, Germany

ARTICLE INFO

Article history:
Received 12 October 2017
Received in revised form 25 April 2018
Accepted 27 April 2018
Available online 27 April 2018

Keywords:
Partial preferences
Ordinality and cardinality
Utility representation
Imprecise probabilities
Stochastic dominance
Linear programming

ABSTRACT

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

1. Introduction

One of the constantly recurring topics discussed in the community of researchers working with imprecise probabilities (and on ISIPTA conferences in particular) is defining meaningful criteria for decision making under complex uncertainty, finding persuading axiomatic justifications for these criteria and providing efficient algorithms capable to deal with them. Examples for such works are ranging from rather early IJAR and ISIPTA contributions by, e.g., [23,1,48,53] to more recent ones by, e.g., [55,58,24,36,4].

However, in the vast majority of works in this field, the complexity underlying the decision situation is assumed to solely arise from the fact that the decision maker's beliefs on the mechanism generating the states of nature are expressed by an imprecise probabilistic model. In contrast, the cardinal utility function adequately describing the decision maker's preference

☆ A preliminary version of this paper was presented at the *Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA)*, 10–14 July 2017 in Lugano, see [26].

☆☆ This paper is part of the Virtual special issue on Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'17), edited by Alessandro Antonucci, Giorgio Corani, Inés Couso and Sébastien Destercke.

* Corresponding author.
E-mail address: christoph.jansen@stat.uni-muenchen.de (C. Jansen).

https://doi.org/10.1016/j.ijar.2018.04.011
0888-613X/

structure is often unquestioned and assumed to be precisely given in advance.[1] Unfortunately, also this can be problematic. Wrongfully pretending to have perfect information on the level of utilities might lead to bad decision making just as doing the same on the level of beliefs: What's worth a decision that is derived on the basis of an inadequate utility function?

For this reason, our paper generalizes both the classical setting of decision making under risk as well as the generalized setting of decision making under ambiguity to situations in which the assumption of a known cardinal utility structure is no longer justified. Particularly, we consider the case that the (information on the) decision maker's preference structure is both of partially ordinal and of partially cardinal scale and, therefore, no longer can be characterized by (a set of positive linear transformations of) *one* cardinal utility function. Instead, we model the decision maker's utility by the set of all utility representations that are compatible with both the ordinal and the cardinal information concerning her preferences.

The paper is structured as follows: In Section 2, we give a brief overview on the background of our work and show how our approach naturally fits into this picture. Moreover, we discuss related literature and the connections to our work. In Section 3, we introduce the crucial concept of a *preference system* over a set of alternatives that allows for modeling partially ordinal and partially cardinal preference structures. Section 4 introduces three different approaches for decision making with acts taking values in a preference system by proposing decision criteria based on generalized expectation intervals (Section 4.2), on global comparisons of acts (Section 4.3) and on pairwise comparisons of acts (Section 4.4). For certain criteria, we give linear programming driven algorithms for checking feasibility of acts in finite decision settings. Section 5 is devoted to an application of the theory. There, we illustrate all the concepts developed in the paper in an example and thereby also show a class of situations in which our approach seems natural: The case where the consequences that acts can attain belong to some product space with both ordinal and cardinal dimensions. Section 6 concludes the paper.

2. Fundamentals underlying our approach and related literature

In classical subjective expected utility theory (SEUT), the decision maker (synonymously called agent in the following) is assumed to be able to specify (I) a real-valued *cardinal* utility function u (unique up to a positive linear transformation) representing her preferences on a set A of alternatives and (II) a unique and *precise* subjective probability measure π on the space S of states of nature adequately specifying her beliefs on the occurrence of the different states $s \in S$. Once these two ingredients are specified, according to SEUT, the decision maker should choose any act $X : S \to A$ that maximizes the expected utility $\mathbb{E}_\pi(u \circ X)$ with respect to her utility function u and her subjective probability measure π among all other available acts.

However, as is well known, in practice both assumptions (I) and (II) often turn out to be systematically too restrictive. In particular, (I) demands the decision maker to act in accordance with the axioms of von Neumann and Morgenstern, i.e. to be able to specify a complete preference ranking of all simple lotteries on the set A that is both independent and continuous (see, e.g., [17, Ch. 8] for details), whereas (II) requires that the decision maker can completely order the resulting utility-valued acts by preference in accordance with the axioms of de Finetti, i.e. continuous, additive and monotone (see, e.g., [19, Ch. 9] for details).

Consequently, there exists plenty of literature relaxing these assumptions. If *only* (II) is violated in the sense that there is only *partial* probabilistic information on the occurrence of the states of nature together with a perfectly cardinal preference structure (represented by a cardinal utility function u), the common relaxation is to allow for *imprecise* probabilistic models for representing the probabilistic information (for instance one could use the credal set $\mathcal{M}$ of all probability measures that are compatible with the given probability constraints). In this case, one can define optimality of acts, for instance depending on the attitude of the decision maker towards the ambiguity underlying the situation, in terms of some imprecise decision criterion such as:

- Γ-*maximin* (Γ-*maximax*): Choose any arbitrary act X yielding maximal expected utility with respect to the worst (best) compatible probability measure, i.e. that maximizes the value $\inf_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u \circ X)$ (the value $\sup_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u \circ X)$) among all available acts.
- *Maximality*: Dismiss each act X for which there is available another act Y that dominates it in expectation with respect to all compatible probability measures, i.e. for which it holds that $\mathbb{E}_\pi(u \circ X) < \mathbb{E}_\pi(u \circ Y)$ for all $\pi \in \mathcal{M}$.
- *E-admissibility*: Dismiss each act X that does not maximize expected utility $\mathbb{E}_\pi(u \circ X)$ among the available acts with respect to at least one compatible probability measure $\pi \in \mathcal{M}$, i.e. where for all $\pi \in \mathcal{M}$ there exists an act Y_π with $\mathbb{E}_\pi(u \circ X) < \mathbb{E}_\pi(u \circ Y_\pi)$.

The original sources of the criteria just discussed are given in [30,34,35,20,60]. Further criteria for the case of cardinal utility and imprecise probabilities, each in its own way taking into account the whole set $\mathcal{M}$ of compatible probability measures, are reviewed in, e.g., [22]. Additionally, there exists a variety of efficient and powerful algorithms to deal with this kind of violation of the classical assumptions (see, e.g., [57,28,21,25]). However, note that the assumption of a cardinal utility

[1] Exceptions include Montes [39, Section 4.2.1], who uses set-valued utility functions, Landes [32] who axiomatically characterizes preferences over utility intervals and Troffaes and Sahlin [56], who propose elicitation procedures for partially specified utility functions. These references, among others, are discussed in some more detail at the end of Section 2.

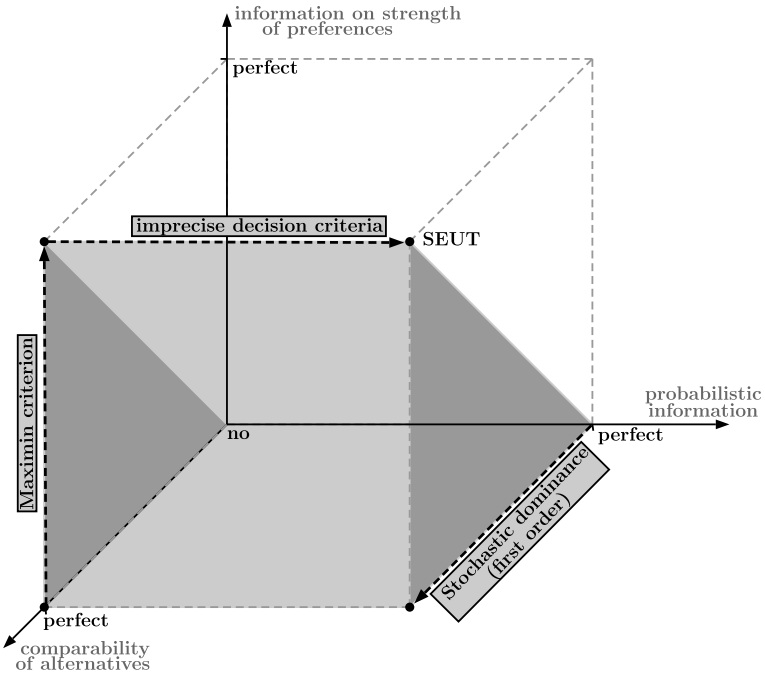

Fig. 1. Schematic illustration of the approaches recalled in Section 2: The points on the black dotted arrows indicate situations in which classical criteria exist, whereas the points in the filled gray-shaded prism indicate situations that are captured by our approach.

function u is essential for all these criteria in order to be applicable. The situations where these imprecise decision criteria are appropriate are illustrated as one of the horizontal edges in the schematic cube given in Fig. 1.

If it is the case that (I) is violated in the sense that the decision maker has complete *ordinal* preferences but no cardinal information (for instance in form of a complete and transitive binary relation R on A) and (II) is violated in the sense that there is no probabilistic information at all, it is nearly unanimously favored to define optimality of acts based on Wald's classical maximin criterion (see [59] for details), which reads here as

- *Maximin criterion*: Choose any act X receiving highest possible rank under the worst possible state of nature, i.e. with $(\inf_{s\in S} X(s), \inf_{s\in S} Y(s)) \in R$ for every other available act Y.

However, note that the completeness of the involved ordinal ranking is essential, since, otherwise, the worst consequences of two distinct acts might be incomparable and, therefore, an optimal act with respect to the maximin criterion simply does not exist. Even more severe, also the vacuousness assumption concerning the information on probabilities is crucial: Applying the maximin criterion in the presence of (partial) probabilistic information means willingly ignoring information. This seems not reasonable at all (cf. also Example 1 for an illustration). Situations in which Wald's maximin criterion appears to be appropriate are illustrated as one of the vertical edges in the schematic cube given in Fig. 1.

Finally, if *only* (I) is violated in the sense that there is no cardinal information at all and the available ordinal information is possibly incomplete (meaning that the relation R from above might not be complete), however, (II) holds true in the sense that beliefs about the states can be described by a precise probability measure π, one common criterion to be applied is the concept of

- *(First oder) Stochastic dominance*: An act Y is said to weakly stochastically dominate another act X, if it holds that $\mathbb{E}_\pi(u \circ Y) \geq \mathbb{E}_\pi(u \circ X)$ for every measurable utility function u that is monotone with respect to R (meaning $(a, b) \in R$ implies $u(a) \geq u(b)$ for arbitrary $a, b \in A$). We denote the stochastic dominance relation by $\geq_{SD}$. This induces the following choice rule: Dismiss an act X if it is strictly stochastically dominated by another available act, i.e. if there exists an act Y such that $\mathbb{E}_\pi(u \circ Y) \geq \mathbb{E}_\pi(u \circ X)$ for every measurable utility function u that is monotone with respect to R and such that $\mathbb{E}_\pi(u_0 \circ Y) > \mathbb{E}_\pi(u_0 \circ X)$ for at least one such function u_0.

Also for this case there exist well-established theory as well as efficient algorithms for computation (see, e.g. [33,27,44,54, 47,50]). Situations for which first order stochastic dominance should be the decision criterion of choice are indicated as one of the horizontal edges in the schematic cube drawn in Fig. 1.

Further, note that there exists a fair amount of work on generalizing the notion of first order stochastic dominance to situations where the underlying uncertainty is characterized by a credal set of probability measures or situations where

the utility function is only partially specified in terms of a multi-valued mapping: In [42, Section 5] the authors introduce and study a generalization where an act X is said to dominate another act Y whenever it stochastically dominates it with respect to every distribution from the underlying credal set. Moreover, in [42, Sections 3 and 4] the authors study different possibilities to extend the notion of stochastic dominance (with respect to a precise probability measure) to suitable binary relations on sets of random variables. These approaches can be viewed as generalizations of stochastic dominance to imprecise utilities. In [43] this framework is further investigated and applied to a real world example. The interrelations of these works and the present one are discussed in some more detail at the end of Section 4.4 when the required concepts are formulated.

Beyond the connections to the literature already mentioned, other related work exists: In [13] the author also studies generalizations of stochastic dominance to imprecise probabilistic models, however, for the special case of belief functions on the real line. More precisely, the paper studies how different orderings between intervals on the real line induce different orderings between belief functions and the mass function associated with them. Afterwards, the paper studies how these orderings relate to the notion of stochastic dominance. Some more details on the connection between [13] and the present work are provided in the discussion directly following the proof of Proposition 6 in Section 4.4.

Of course, there is also related work on non-fully specified utilities: In [32] the author considers the situation of decision making under complete uncertainty (i.e. with a credal set $\mathcal{M}$ containing all possible probability distributions on the state space) with acts taking values in some *linearly* ordered space. To each such act it is then associated an utility interval. Afterwards, the author axiomatically characterizes desirable properties of binary relation on such utility intervals and shows that these axioms uniquely determine a particular binary relation, the so-called Min–Max Relation (see [32, Corollary 3] in particular).

In [56] the authors consider the case of decision making with acts taking values in some multi-attribute space. For such problems they propose a two-step elicitation procedure for utility functions, where the first step consists in precisely eliciting the marginal utility functions on the different attributes and the second step consists in imprecisely eliciting the weights with respect to which these marginal utilities are extended to an utility function on the whole multi-attribute space.

Finally, in [9] and [10] and [8] the authors study decision making problems where both the utility values assigned to the consequences and the probability values assigned to the states are allowed to be only imprecisely known. In particular, they investigate situations where the statements about probability and utility values can be formed by one of three types of sentences, so-called vague sentences, interval sentences and comparative sentences. The set of available acts together with the sets of probability and utility sentences then forms the so-called information frame. For such decision problems they then propose a decision criterion, the so-called t-admissibility, which relies on the idea to prefer an act X to another act Y whenever it holds that $\mathbb{E}_\pi(u \circ X) - \mathbb{E}_\pi(u \circ Y) \geq t$ for all pairs (π, u) that are compatible with both the set of utility sentences and the set of probability sentences, that is with the information frame (see [9, Section 2.2]). For evaluating this criterion the authors then propose bilinear as well as linear optimization approaches and apply their theory to a real world problem.

Furthermore, in [8] the framework developed in [9] is generalized to the case where the pairwise comparison between the acts X and Y is no longer made by considering solely the differences of the expected values, but where it can be made by arbitrary functionals f depending besides of the acts under consideration also on π and u as well as additional parameters (see [8, Section 3]). Their work is also implemented in the decision user interface *DecideIt*, which is introduced and described in [10]. Note that our relation $R_{\forall\forall}$ that is introduced in Equation (9) and our concept of local $R_{\forall\forall}$-admissibility from Definition 8 that is based on it are closely related to the concept of t-admissibility from the works by Danielson et al.. This is discussed in some more detail at the end of Section 4.4.

In the following sections of the paper, we introduce and discuss different concepts for decision making in situations in which simultaneously both assumptions (I) and (II) are violated (i.e. situations corresponding to inner points of the gray-shaded prism from Fig. 1) and thus none of the concepts just recalled can be applied. Therefore, the contribution of the present paper consists in filling up the gray-shaded prism in the schematic cube drawn in Fig. 1.

3. Preference systems

In this section we start by defining the concept of a preference system, which is essential for what follows throughout the rest of the paper. The intuition behind this concept is very simple: In many practically relevant decision problems, the (available information on the) agent's preferences are (is) incomplete. More precisely, it often will be the case that certain pairs of possible decision outcomes are incomparable for the agent, whereas others can be ordered by preference. Additionally, for some pairs there might even be an idea of the strength of the preference, that is an idea of how much the one outcome is preferred to the other.

There are several circumstances that could give rise to this type of incomplete preferences. For example, if a company wants to analyze the choice behavior of their (potential) customers, the information on the customer's preferences will often be given in form of observed binary choices and/or survey data. Obviously, usually such data won't be sufficient to specify the full preference structure of the customer, since this require too many observations. In this case, incompleteness is a missing data problem and originates in lacking information about the choice behavior.

However, also the agent herself might have incomplete preferences. Suppose she knows (e.g. from earlier choice experience) certain potential decision outcomes better than others. Then for pairs involving better known outcomes, she might be able to specify a preference ranking and even some intuition for the strength of the preference, whereas for pairs involving unfamiliar outcomes, she might be able to specify only a ranking or even can't make a comparison at all. The following definition captures the intuition just described.

Definition 1. Let A be a non-empty set and let $R_1 \subseteq A \times A$ denote a pre-order (i.e. reflexive and transitive) on A. Moreover, let $R_2 \subseteq R_1 \times R_1$ denote a pre-order on R_1. Then the triplet $\mathcal{A} = [A, R_1, R_2]$ is called a **preference system** on A.

The interpretation of the relations R_1 and R_2 contained in a preference system $\mathcal{A}$ in the sense of Definition 1 is in perfect accordance with the intuition that should be captured by it: If some pair $(a, b) \in R_1$, we interpret that as *a is at least as desirable as b*, that is a and b can be ordered by preference. If both $(a, b) \notin R_1$ and $(b, a) \notin R_1$, then a and b are incomparable. Moreover, if a pair of pairs $((a, b), (c, d)) \in R_2$, we interpret this as *exchanging alternative b by alternative a is at least as desirable as exchanging alternative d by alternative c*, that is a is more strongly preferred over b than c is over d. Again, if both $((a, b), (c, d)) \notin R_2$ and $((c, d), (a, b)) \notin R_2$, then the exchange of b by a is incomparable with the exchange of d by c.

Except from transitivity, Definition 1 makes no rationality and/or compatibility assumption regarding the relations R_1 and R_2. Accordingly, a preference system in the sense of the above Definition 1 needs by no means to be reasonable or rational. In [31, Chapter 4], an axiomatic approach for characterizing consistent preference systems is provided for the case that the involved relations are complete. The corresponding axioms then imply the existence of a real valued function representing both relations simultaneously that is unique up to a positive linear transformation. Another axiomatization that uses quaternary relations instead of pairs of relations is established in [46], where it is shown that under some quite strong conditions (like, e.g., solvability) there exists a multi-utility characterization of the corresponding quaternary relation.

A weaker consistency condition that still applies to settings in which conditions like solvability no longer can be expected is given in the following definition, for which we need some further notation: If R is a pre-order on A, we denote by I_R and P_R its indifference and its strict part, respectively. More precisely, for $(a, b) \in A \times A$, we have $(a, b) \in I_R \Leftrightarrow ((a, b) \in R \wedge (b, a) \in R)$ and $(a, b) \in P_R \Leftrightarrow ((a, b) \in R \wedge (b, a) \notin R)$.

Definition 2. Let $\mathcal{A} = [A, R_1, R_2]$ be a preference system. Then $\mathcal{A}$ is said to be **consistent** if there exists a function $u : A \to [0, 1]$ such that for all $a, b, c, d \in A$ the following two properties hold:

i) If $(a, b) \in R_1$, then $u(a) \geq u(b)$ with equality iff $(a, b) \in I_{R_1}$.
ii) If $((a, b), (c, d)) \in R_2$, then $u(a) - u(b) \geq u(c) - u(d)$ with equality iff $((a, b), (c, d)) \in I_{R_2}$.

Every such function u is then said to **(weakly[2]) represent** the preference system $\mathcal{A}$. The set of all (weak) representations u of $\mathcal{A}$ is denoted by $\mathcal{U}_\mathcal{A}$. The set of all $u \in \mathcal{U}_\mathcal{A}$ satisfying $\inf_{a \in A} u(a) = 0$ and $\sup_{a \in A} u(a) = 1$ is denoted by $\mathcal{N}_\mathcal{A}$.

We will call a preference system *non-trivial* if there exists a pair $(a, b) \in P_{R_1}$, that is if there is at least one alternative that is strictly preferred to another one. In the rest of the paper we will throughout consider non-trivial preference systems and, therefore, drop the prefix non-trivial from now on. Note that trivial preference systems are represented by arbitrary maps $c : A \to [0, 1]$.

The idea behind the set $\mathcal{N}_\mathcal{A}$ of normalized representations in the above definition is the following: For the special case that the preference system $\mathcal{A}$ is in accordance with the axioms in [31, Chapter 4], the representation is unique up to a positive linear transformation. Hence, the conditions $\inf_a u(a) = 0$ and $\sup_a u(a) = 1$ guarantee a unique representation for that special case. For the general case of a consistent preference system $\mathcal{A}$ with non complete relations R_1 and R_2, restricting analysis to the set $\mathcal{N}_\mathcal{A}$ ensures that comparison will not be made with respect to equivalent representations which only measure utility on a different scale. Specifically, if $u \in \mathcal{U}_\mathcal{A}$, we have also $\lambda \cdot u \in \mathcal{U}_\mathcal{A}$ for arbitrary $\lambda \in (0, 1)$. This means that both functions u and $\lambda \cdot u$ represent the preference system $\mathcal{A}$, however, they measure utility on different scales, namely $[\min_{a \in A} u(a), \max_{a \in A} u(a)]$ and $[\min_{a \in A} \lambda \cdot u(a), \max_{a \in A} \lambda \cdot u(a)]$. By restricting analysis to the set $\mathcal{N}_\mathcal{A}$ we therefore ensure that all considered representations measure utility on a $[0, 1]$-scale. The restriction on $\mathcal{N}_\mathcal{A}$, together with the concept of *granularity* from Definition 3, will prove crucial when comparing acts by means of the numerical representation in Section 4.2.

Further, note that for finite A, the boundedness condition on the utility function in Definition 2 implies the existence of alternatives in A that attain a greatest and a lowest utility value, but not necessarily of worst and best alternatives in A with respect to the relation R_1: An element of A attaining the highest utility value for a certain representation $u : A \to [0, 1]$ might indeed be incomparable to all other elements of A.

Obviously, for a preference system $\mathcal{A} = [A, R_1, R_2]$ to be consistent, certain compatibility criteria between the relations R_1 and R_2 have to be satisfied. For example it cannot be the case that, for some elements $a, b, c \in A$, it simultaneously holds

[2] Here, the term *weakly* refers to the fact that the representation is meant in the *if* and not in the *if and only if (short: iff)* sense.

that $(c,a) \in P_{R_1}$ and $((a,b),(c,b)) \in R_2$, since any element $u \in \mathcal{U}_{\mathcal{A}}$ would have to satisfy $u(c) > u(a)$ and $u(a) - u(b) \geq u(c) - u(b)$. We now provide an algorithm for checking the consistency of a finite preference system.

Proposition 1. *Let $\mathcal{A} = [A, R_1, R_2]$ be a preference system, where $A = \{a_1, \ldots, a_n\}$ is a finite and non-empty set. Consider the linear optimization problem*

$$\varepsilon = \langle (0, \ldots, 0, 1)', (u_1, \ldots, u_n, \varepsilon)' \rangle \longrightarrow \max_{(u_1,\ldots,u_n,\varepsilon) \in \mathbb{R}^{n+1}} \tag{1}$$

with constraints $0 \leq (u_1, \ldots, u_n, \varepsilon) \leq 1$ and

i) $u_p = u_q$ *for all* $(a_p, a_q) \in I_{R_1} \setminus \{(a,a) : a \in A\}$
ii) $u_q + \varepsilon \leq u_p$ *for all* $(a_p, a_q) \in P_{R_1}$
iii) $u_p - u_q = u_r - u_s$ *for all* $((a_p, a_q), (a_r, a_s)) \in I_{R_2} \setminus \{((a,b),(a,b)) : (a,b) \in R_1\}$
iv) $u_r - u_s + \varepsilon \leq u_p - u_q$ *for all* $((a_p, a_q), (a_r, a_s)) \in P_{R_2}$

Then $\mathcal{A}$ is consistent if and only if the optimal outcome of (1) is strictly positive.

Proof. First, note that $(0, \ldots, 0) \in \mathbb{R}^{n+1}$ defines an admissible solution of (1). Thus, the set of admissible solutions of (1) is non-empty. Since it is also bounded due to $0 \leq (u_1, \ldots, u_n, \varepsilon) \leq 1$, we can deduce the existence of an optimal solution of (1) by utilizing that linear programming problems with a bounded and non-empty set of admissible solutions always possess an optimal solution. Let $(u_1^*, \ldots, u_n^*, \varepsilon^*)$ denote such an optimal solution.

If: Assume $\varepsilon^* > 0$. Define $u : A \to [0,1]$ by setting $u(a_i) := u_i^*$ for all $i \in \underline{n} := \{1, \ldots, n\}$. One then straightforwardly verifies that conditions i) to iv) imply that $u \in \mathcal{U}_{\mathcal{A}}$. Hence, $\mathcal{A}$ is consistent.

Only if: Assume $\varepsilon^* = 0$ and, for contradiction, that there exists $v \in \mathcal{U}_{\mathcal{A}}$. Define the values $v_i := v(a_i)$ for all $i \in \underline{n}$ and $\delta := \min\{m_1, m_2\}$, where $m_1 := \min\{v_i - v_j : a_i P_{R_1} a_j\}$ and $m_2 := \min\{(v_i - v_j) - (v_k - v_l) : (a_i, a_j) P_{R_2} (a_k, a_l)\}$. Then, since $v \in \mathcal{U}_{\mathcal{A}}$ and it therefore holds that $m_1 > 0$ due to Definition 2 i) and that $m_2 > 0$ due to Definition 2 ii), we have $\delta > 0$. One then straightforwardly verifies that $(v_1, \ldots, v_n, \delta)$ is an admissible solution to (1) with $\delta > \varepsilon^*$, contradicting the optimality of $(u_1^*, \ldots, u_n^*, \varepsilon^*)$. □

The linear programming problem (1) possesses $|R_2| + n + 2$ constraints. Thus, the number of constraints increases with the preciseness of the available information on the agent's preferences. In applications, typically the relation R_2 will be rather sparse (i.e. contain few comparable pairs of pairs), whereas the relation R_1 will be rather dense (i.e. contain many comparable pairs). This is intuitive: While R_1 is directly observable in the choice behavior of the agent, edges in R_2 need to be gained by hypothetical comparisons in interviews and polls by asking questions like: "Imagine you have objects a and b. Would you rather be willing to accept the exchange of a by c or the exchange of b by d?"

In order to reduce the number of constraints of the problem, note that (weak) representability of a preference system $\mathcal{A} = [A, R_1, R_2]$ automatically implies transitivity of the represented relations. Therefore, in the constraints of the above optimization problem (1) it actually suffices to quantify only over (the corresponding indifference parts $I_{R_1^*}$, $I_{R_2^*}$ and strict parts $P_{R_1^*}$, $P_{R_2^*}$ of) some transitive reductions R_1^*, R_2^* of the relations R_1 and R_2. However, note that this makes necessary to compute the corresponding transitive reductions which, again, raises the complexity of the problem to some extent.

Before turning to decision theory with preference system valued acts, we need one further concept, which will be of particular relevance in Section 4.2.

Definition 3. Let $\mathcal{A} = [A, R_1, R_2]$ be a consistent preference system. Moreover, for $\delta \in (0,1)$, let $\mathcal{N}_{\mathcal{A}}^{\delta}$ denote the set of all $u \in \mathcal{N}_{\mathcal{A}}$ satisfying $u(a) - u(b) \geq \delta$ for all $(a,b) \in P_{R_1}$ and $u(a) - u(b) - u(c) + u(d) \geq \delta$ for all $((a,b),(c,d)) \in P_{R_2}$. Then, $\mathcal{N}_{\mathcal{A}}^{\delta}$ is called the **(weak) representation set** of **granularity** (at least) δ. Moreover, the decision system $\mathcal{A}$ is called **δ-consistent** if $\mathcal{N}_{\mathcal{A}}^{\delta} \neq \emptyset$.

On the one hand, the granularity δ from Definition 3 can be given a similar interpretation as the just noticeable difference in the context of psychophysics (see [37] for details): It is the minimal difference in utility that the specific decision maker under consideration is able to notice given that utility is measured on a $[0,1]$-scale. On the other hand, the granularity can also be given a more constructive interpretation, namely as a controlling device for the specific decision maker: Choosing a granularity parameter $\delta > 0$ ensures that an act will not be labeled superior to another based solely on some utility function that involves utility differences that are practically not noticeable at all and, accordingly, should not influence the decision to be made.

The restriction of the analysis to utility functions that reflect the fact that utility differences below some threshold are not distinguishable empirically will play a crucial role when it comes to defining generalized expectations (and the decision criteria based on these) in Section 4.2. For now, it is sufficient to note that the algorithm given in Proposition 1 straightforwardly extends to checking whether a preference system $\mathcal{A} = [A, R_1, R_2]$ is δ-consistent. This is the statement of the following proposition.

Proposition 2. *Let $\mathcal{A} = [A, R_1, R_2]$ be a preference system, where $A = \{a_1, \dots, a_n\}$ is a finite and non-empty set and let $\delta \in (0, 1)$. Then $\mathcal{A}$ is δ-consistent if and only if the optimal outcome of (1) is at least δ.*

Proof. Let $(u_1^*, \dots, u_n^*, \varepsilon^*)$ denote an optimal solution to problem (1) and define $u : A \to [0, 1]$ by setting $u(a_i) := u_i^*$ for all $i \in \underline{n}$. First, note there have to exist elements $a^+, a^- \in A$ such that $u(a^-) = 0$ and $u(a^+) = 1$, since otherwise normalizing u to $[0, 1]$ would induce a solution to (1) with objective value strictly greater than ε^*. Thus, $u \in \mathcal{N}_{\mathcal{A}}$. If $\varepsilon^* \geq \delta$, then the constraints of (1) guarantee that $u(a) - u(b) \geq \delta$ for all $(a, b) \in P_{R_1}$ and $u(a) - u(b) - u(c) + u(d) \geq \delta$ for all $((a, b), (c, d)) \in P_{R_2}$. Hence, $u \in \mathcal{N}_{\mathcal{A}}^{\delta}$ and $\mathcal{A}$ is δ-consistent. If conversely $\mathcal{A}$ is δ-consistent, we can choose $u^+ \in \mathcal{N}_{\mathcal{A}}^{\delta}$ and we know from the proof of Proposition 1 that $(u_1^+, \dots, u_n^+, \delta)$ with $u_i^+ := u^+(a_i)$ for $i \in \underline{n}$ defines an admissible solution to (1). Since ε^* is the optimal outcome of (1), we know that $\varepsilon^* \geq \delta$. □

4. Decision theory with ps-valued acts

Differently from axiomatic approaches followed in, e.g., [51,45,18], where (multi-)utility and (imprecise) probability representations are obtained by different axiomatic characterizations of preferences over acts, the aim of the present paper is to go the opposite direction and to obtain preferences on acts given a preference system and some additional, commonly partial, probabilistic information about the occurrence of the states of nature.

As already discussed in more detail in Section 2, most existing criteria for decision making under uncertainty are not applicable in such situations, since they require either a perfectly cardinal preference structure (like, e.g., maximizing expected utility or Γ-maximin) or, complementary, a precise probability measure representing the beliefs on the states of nature (like, e.g., first order stochastic dominance[3]). Therefore, we now propose and discuss three different approaches for decision making under uncertainty when the considered acts take values in some arbitrary preference system (abbreviated by *ps-valued* acts in the following) and when there is partial probabilistic information on the occurrence of the states available.

4.1. Basic setting

We start by defining the central concepts of the theory for the most general case. Let S denote some non-empty set equipped with some suitable σ-algebra $\sigma(S)$. The elements of S are interpreted as all possible states of nature about whose occurrence the decision maker under consideration is uncertain. Moreover, let $\mathcal{M}$ denote some credal set on the measurable space $(S, \sigma(S))$, which is interpreted as the set of all probability measures on $(S, \sigma(S))$ that are compatible with the available (partial) probabilistic information and thus describing the uncertainty about the occurrence of the states. For a given consistent preference system $\mathcal{A}$, a state space S and a credal set $\mathcal{M}$, a ps-valued act is a mapping $X : S \to A$ assigning states of nature to elements of the preference system.

Given this, define the set $\mathcal{F}_{(\mathcal{A},\mathcal{M},S)} \subseteq A^S := \{X | X : S \to A\}$ by setting

$$\mathcal{F}_{(\mathcal{A},\mathcal{M},S)} := \left\{ X \in A^S : u \circ X \text{ is } \sigma(S)\text{-}\mathcal{B}_{\mathbb{R}}\text{-measurable for all } u \in \mathcal{U}_{\mathcal{A}} \right\} \tag{2}$$

where $\mathcal{B}_{\mathbb{R}}$ denotes the Borel sigma field on $\mathbb{R}$. By construction, the space $\mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ consists of exactly those acts $X : S \to A$ whose expectation exists with respect to all pairs $(u, \pi) \in \mathcal{U}_{\mathcal{A}} \times \mathcal{M}$ of compatible probability measure and utility representation (since bounded and measurable random variables have finite expectation). We can now define our main object of study:

Definition 4. In the situation above, call every subset $\mathcal{G} \subseteq \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ a **decision system** (with information base $(\mathcal{A}, \mathcal{M})$). Moreover, call a decision system $\mathcal{G}$ finite, if both $|\mathcal{G}| < \infty$ and $|S| < \infty$, that is if both the set of states and the set of available acts are finite.

The elements of a decision system $\mathcal{G}$ are interpreted as those elements of the space $\mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ that are available in the specific choice situation under consideration. Given a decision system $\mathcal{G}$, we are interested in the following question: How can we utilize the information base $(\mathcal{A}, \mathcal{M})$ best possibly in order to define meaningful and reasonable choice criteria on the set $\mathcal{G}$? In the following sections, we propose three different approaches that address exactly this question.

4.2. Criteria based on generalized expectation intervals

We start by introducing and studying decision criteria that are based on the analysis of generalized expectation intervals of the available acts. Depending on the attitude of the agent of interest towards the ambiguity underlying the situation (for instance she could be ambiguity *seeking* or ambiguity *averse* or something inbetween), such intervals can give rise to a variety of different optimality criteria for decision making. Specifically, for a ps-valued act X and a decision maker with a granularity parameter $\delta > 0$, the corresponding interval will range from the lowest to the highest possible expected value

[3] For approaches directly generalizing stochastic dominance to credal sets, see [13,7,43,6].

that choosing the act X can lead to under some pair $(u,\pi) \in \mathcal{N}_{\mathcal{A}}^{\delta} \times \mathcal{M}$ that is compatible with the preference system $\mathcal{A}$ and the probabilistic information $\mathcal{M}$. This leads to the definition of the basic concept of this section.

Definition 5. Let $X \in \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ and $\delta \in (0,1)$. With $\mathcal{D}_\delta := \mathcal{N}_{\mathcal{A}}^{\delta} \times \mathcal{M}$, we call

$$\mathbb{E}_{\mathcal{D}_\delta}(X) := \Big[\underline{\mathbb{E}}_{\mathcal{D}_\delta}(X), \overline{\mathbb{E}}_{\mathcal{D}_\delta}(X)\Big] := \Big[\inf_{(u,\pi)\in\mathcal{D}_\delta} \mathbb{E}_\pi(u \circ X), \sup_{(u,\pi)\in\mathcal{D}_\delta} \mathbb{E}_\pi(u \circ X)\Big] \tag{3}$$

the **generalized interval expectation** of X with respect to $\mathcal{A}, \mathcal{M}$ and granularity δ.

Examples for how the generalized interval expectation is computed in concrete decision situations are given in the second part of Example 1 in Section 4.3 and in the application example in Section 5.2. Note that, in the spirit of the theory of imprecise probabilities, the set $\mathbb{E}_{\mathcal{D}_\delta}(X)$ can be given an epistemic or an ontologic interpretation (see [61] or [2, p. 143]): If the imprecision/ambiguity in the sets arises from lack of information in the sense of e.g. partially observed choice behavior and/or partially known precise probabilities, the set $\mathbb{E}_{\mathcal{D}_\delta}(X)$ is the set of all expectations arising in at least one situation that is compatible with the data. In contrast, if both sets $\mathcal{N}_{\mathcal{A}}^{\delta}$ and $\mathcal{M}$ have an ontologic interpretation, i.e. are interpreted as holistic entities of their own, the same holds true for the set of expectations $\mathbb{E}_{\mathcal{D}_\delta}(X)$.

Of course, all decision theory that is based on comparisons of the set $\mathbb{E}_{\mathcal{D}_\delta}(X_i)$ of different acts X_i should reflect the underlying interpretation. The following definition gives three criteria rather relying on an ontologic interpretation of the set $\mathcal{D}_\delta$. Note that all of them are straightforward generalizations of the (complete order inducing) decision criteria commonly used in the theory of imprecise probabilities and reviewed, e.g., in [22].

Definition 6. Let $\mathcal{G} \subseteq \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ be a decision system and $\delta \in (0,1)$. An act $X \in \mathcal{G}$ is called

i) $\mathcal{D}_\delta$**-maximin** iff $\forall Y \in \mathcal{G}: \ \underline{\mathbb{E}}_{\mathcal{D}_\delta}(X) \geq \underline{\mathbb{E}}_{\mathcal{D}_\delta}(Y)$

ii) $\mathcal{D}_\delta$**-maximax** iff $\forall Y \in \mathcal{G}: \ \overline{\mathbb{E}}_{\mathcal{D}_\delta}(X) \geq \overline{\mathbb{E}}_{\mathcal{D}_\delta}(Y)$

iii) $\mathcal{D}_\delta^\alpha$**-maximix** iff

$$\forall Y \in \mathcal{G}: \ \alpha\underline{\mathbb{E}}_{\mathcal{D}_\delta}(X) + (1-\alpha)\overline{\mathbb{E}}_{\mathcal{D}_\delta}(X) \geq \alpha\underline{\mathbb{E}}_{\mathcal{D}_\delta}(Y) + (1-\alpha)\overline{\mathbb{E}}_{\mathcal{D}_\delta}(Y)$$

where $\alpha \in [0,1]$ is some fixed parameter.

We denote by $\underline{\mathcal{G}}_\delta$, $\overline{\mathcal{G}}_\delta$ and $\mathcal{G}_\delta^\alpha$ the sets of $\mathcal{D}_\delta$-maximin, $\mathcal{D}_\delta$-maximax and $\mathcal{D}_\delta^\alpha$-maximix acts in $\mathcal{G}$.

Independent of its interpretation, we need ways for computing the set $\mathbb{E}_{\mathcal{D}_\delta}(X)$ in concrete situations. The following proposition gives a linear programming based algorithm for doing so in finite decision systems. However, note that applying the proposition requires the extreme points of the underlying credal set $\mathcal{M}$ and, therefore, is ideal for situations where the number of extreme points is moderate and where closed formulas for computing the extreme points are available. For credal sets induced by 2-monotone lower/ 2-alternating upper probabilities such formulas exist (cf., [52, Theorem 3, p. 19]). While generally the number of extreme points could be very high (maximally $|S|!$ for lower probabilities, cf. [15] and [62]), convenient cases exist where furthermore efficient enumeration procedures are available (such special cases include ordinal probabilities (cf., [29, p. 26]), comparative probabilities (cf., [38]), necessity measures (cf., [49]), p-boxes (cf., [40]), probability intervals (cf., [63, Chapter 2] or [11]) or pari-mutuel models (cf., [41])).

Proposition 3. *Let $\mathcal{A} = [A, R_1, R_2]$ be a consistent preference system, where $A = \{a_1, \ldots, a_n\}$ such that $(a_1, b), (b, a_n) \in R_1$ for all $b \in A$ and let ε^* denote the optimal outcome of problem (1). Moreover, let $S = \{s_1, \ldots, s_m\}$ be finite, $\mathcal{M}$ be some polyhedral credal set on $(S, 2^S)$ with extreme points $\mathcal{E}(\mathcal{M}) := \{\pi^{(1)}, \ldots, \pi^{(T)}\}$ and let $X \in \mathcal{G}$. For $\varepsilon^* \geq \delta > 0$, consider the collection of linear programs $LP_1^\delta, \ldots, LP_T^\delta$ given by:*

$$\sum_{i=1}^{n} u_i \cdot \pi^{(t)}(X^{-1}(\{a_i\})) \longrightarrow \min_{(u_1,\ldots,u_n)\in\mathbb{R}^n} / \max_{(u_1,\ldots,u_n)\in\mathbb{R}^n} \tag{LP$_t^\delta$}$$

with constraints $0 \leq (u_1, \ldots, u_n) \leq 1$, $u_1 = 1$, $u_n = 0$ and i) to iv) as given in Proposition 1 (with $\varepsilon := \delta$ fixed). Let $\underline{v}(t,\delta)$ and $\overline{v}(t,\delta)$ denote the optimal outcomes of problem LP_t^δ in minimum and maximum form. Then, we have $\mathbb{E}_{\mathcal{D}_\delta}(X) = [\min_t \underline{v}(t,\delta), \max_t \overline{v}(t,\delta)]$.

Proof. Let $X \in \mathcal{G}$ and $\varepsilon^* \geq \delta > 0$. Then, the set $\mathcal{N}_{\mathcal{A}}^{\delta}$ is non-empty and we can define the function $f: \mathcal{D}_\delta \to \mathbb{R}, (u,\pi) \mapsto \mathbb{E}_\pi(u \circ X)$. For any representation $u \in \mathcal{N}_{\mathcal{A}}^{\delta}$ fixed, the function $\pi \mapsto f(u,\pi)$ is linear in π and, therefore, both convex and concave. By utilizing the facts that the pointwise infimum of any family of concave functions is a concave function and that the pointwise supremum of any family of convex functions is a convex function, we know that the functions $\pi \mapsto \inf_u f(u,\pi)$ and $\pi \mapsto \sup_u f(u,\pi)$ have to be concave and convex, respectively. But concave functions on polyhedral

set attain their minimum and convex functions on polyhedral set attain their maximum on the set of extreme points. Hence, in order to find global maximum and minimum of the function f, it suffices to check for it on the set $\mathcal{N}_{\mathcal{A}}^{\delta} \times \mathcal{E}(\mathcal{M})$.

Now, let $(u_1^*, \dots, u_n^*)$ denote an optimal solution to problem LP_t^{δ} in maximum form for fixed $t \in \{1, \dots, T\}$. One then easily verifies that the constraints imply $u^* \in \mathcal{N}_{\mathcal{A}}^{\delta}$, where $u^* : A \to [0,1]$, $u^*(a_i) := u_i^*$ and $\overline{v}(t,\delta) = \mathbb{E}_{\pi^{(t)}}(u^* \circ X) = \sup\{\mathbb{E}_{\pi^{(t)}}(u \circ X) : u \in \mathcal{N}_{\mathcal{A}}^{\delta}\}$. Analogous reasoning for the problem in minimum form yields $\underline{v}(t,\delta) = \inf_{u \in \mathcal{N}_{\mathcal{A}}^{\delta}} \mathbb{E}_{\pi^{(t)}}(u \circ X)$. Thus, applying our considerations from before yields $\mathbb{E}_{\mathcal{D}_\delta}(X) = [\min_t \underline{v}(t,\delta), \max_t \overline{v}(t,\delta)]$. □

Another way to compute the bounds in (3) in the case of 2-alternating upper probabilities (2-monotone lower probabilities) on a finite space A is to use the Choquet representation of the upper (lower) expectation (cf., e.g., [12, Proposition 10.3, p. 126]): For a fixed utility u and a 2-alternating upper probability ν with associated credal set $\mathcal{M}_\nu$ the corresponding expected upper utility can be written as $\overline{\mathbb{E}}_{\{u\}\times\mathcal{M}_\nu}(X) = \sum_{i=1}^{n} \left(u_{(i)} - u_{(i-1)}\right) \cdot \nu(\{s \in S \mid u(X(s)) \geq u_{(i)}\})$, where $u_{(i)}$ denotes the i-th value of the increasingly ordered involved utility values $u_1, \dots, u_n$.

If R_1 is complete then the expectation is a linear form in the utility u and the maximization $\max_{u \in \mathcal{N}_{\mathcal{A}}^{\delta}} \mathbb{E}_{\{u\}\times\mathcal{M}_\nu}(X)$ translates to a simple linear program. If the relation R_1 is not complete then the ordering of the utility values u_i can change as u ranges in $\mathcal{N}_{\mathcal{A}}^{\delta}$ and one has to compute the expectation separately for every possible ordering of the utility values and then take the maximum. If there are totally comparable values u_i meaning that for every u_j either $u_i \leq u_j$ or $u_i > u_j$, independently from the concrete $u \in \mathcal{N}_{\mathcal{A}}^{\delta}$ then one can split the sum in a part containing all utility values below u_i and a part containing all utility values above u_i and then analyze every sub-sum independently which would help in reducing the combinatorial complexity.

The criteria from Definition 6 allow for comparing acts *given* the granularity δ of the specific decision maker of interest. However, note that knowing the granularity might be a strong assumption if R_1 and R_2 are partial orderings, since experimental settings in which this additional parameter could precisely be elicited are not as straightforward as in the case of complete orderings. A natural way for addressing this issue in practical problems is to compute the generalized interval expectation along varying values of δ. Clearly, it holds that $\mathbb{E}_{\mathcal{D}_{\delta_1}}(X) \subseteq \mathbb{E}_{\mathcal{D}_{\delta_2}}(X)$ whenever $\delta_1 \geq \delta_2$, since it holds $\mathcal{D}_{\delta_1} \subseteq \mathcal{D}_{\delta_2}$ and, thus, the inf and the sup in (3) are taken for a smaller set for δ_1 than for δ_2. That is, the generalized interval expectation of an act X becomes narrower (or more precisely, not gets wider) as the value of δ increases. Utilizing this fact, in order to decide between two competing acts X and Y, one could proceed as follows: Once having decided for one of the criteria from Definition 6, one can compute the general interval expectation for increasing values of δ until the chosen criterion can discriminate between the acts X and Y for the first time, for instance in favor of X (say this happens for the value δ^*). Afterwards, the decision maker is asked whether it is acceptable for her that utility differences below δ^* are not taken into account by the decision procedure. If the answer is yes, the decision maker should rank act X before Y, otherwise no decision can be made.

Further possibilities to deal with these issues are treated in the next two sections, where we propose two approaches completely overcoming the choice of a granularity parameter.

4.3. Criteria based on global comparisons

The decision criteria defined in Section 4.2 all construct complete rankings on the set $\mathcal{G}$ by comparing numerical representations of parts of the decision system and by somehow ignoring the inherent utility and probability structure. Therefore, when defining optimality of acts in terms of one of the criteria from Definition 6, it makes no difference if the ranking is constructed by pairwise or global comparisons. In the next sections, we turn to two approaches that explicitly take into account a global and local viewpoint for defining optimality of acts, respectively.[4]

We start with defining criteria taking the global perspective: For an act X in order to be labeled optimal, it is necessary that there exists (depending on the concrete approach at least) one fixed pair $(u,\pi) \in \mathcal{U}_{\mathcal{A}} \times \mathcal{M}$ for which this act maximizes the expected utility among all other available acts $Y \in \mathcal{G}$. In particular, the pair (u,π) for which X dominates the other acts in expectation must not depend on the concrete competing act under consideration, but has to be constant for all acts from $\mathcal{G}$. Optimality criteria for which the pair (u,π) may depend on the concrete competing act are considered in Definition 8 in Section 4.4.

Moreover, note that the concepts from the following definition take a global point of view also from another perspective: The utility component in the required pair(s) (u,π) should not depend on its probability component and also its probability component should no depend on the utility component. This is reflected in the fact that in the admissibility concepts of Definition 7 a ∀ quantifier can follow an ∃ quantifier but not vice versa.

Definition 7. Let $\mathcal{G} \subseteq \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ denote a decision system. We call an act $X \in \mathcal{G}$

i) $\mathcal{A}|\mathcal{M}$**-admissible** iff $\exists u \in \mathcal{U}_{\mathcal{A}}\ \exists \pi \in \mathcal{M}\ \forall Y \in \mathcal{G}:\ \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$

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

[4] Note that in the context of IP decision theory, fundamental differences between global criteria and criteria based on pairwise comparisons have already been discussed in [48].

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

iv) $\mathcal{A}|\mathcal{M}$**-dominant** iff $\forall u \in \mathcal{U}_{\mathcal{A}}\ \forall \pi \in \mathcal{M}\ \forall Y \in \mathcal{G}:\ \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$

Denote by $\mathcal{G}_{\mathcal{A}|\mathcal{M}}$, $\mathcal{G}_{\mathcal{A}}$, $\mathcal{G}_{\mathcal{M}}$ and $\mathcal{G}^d_{\mathcal{A}|\mathcal{M}}$ the sets of such acts, respectively.

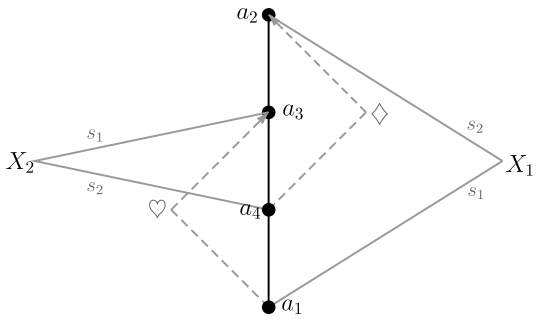

Fig. 2. The black part shows the Hasse graph of the relation R_1, the gray dashed lines show the edges $\heartsuit = (a_3, a_1)$ and $\diamondsuit = (a_2, a_4)$ with $\diamondsuit$ being preferred to $\heartsuit$. The solid gray lines show which elements of A are attained by the acts X_1 and X_2 under the different states of nature.

All four act properties just defined rely on the idea that, if there *was* perfect information on both the state probabilities (i.e. $\mathcal{M} = \{\pi\}$ is a singleton) and the utility values (i.e. the utility representation u is unique up to a positive linear transformation), then an act X should be labeled optimal iff X has greater or equal expected utility than every other act $Y \in \mathcal{G}$ with respect to (u, π). However, they differ in the way they handle the ambiguity underlying the involved sets $\mathcal{M}$ and $\mathcal{U}_{\mathcal{A}}$: While $\mathcal{A}|\mathcal{M}$-admissibility only demands the existence of at least *one* compatible combination (u, π) with respect to which X maximizes expected utility, $\mathcal{A}|\mathcal{M}$-dominance requires this for *all* compatible combinations. $\mathcal{M}$- and $\mathcal{A}$-admissibility relax the $\forall$-assumption on probability and utility level, respectively. Clearly, it holds that $\mathcal{G}_{\mathcal{A}}, \mathcal{G}_{\mathcal{M}}, \mathcal{G}^d_{\mathcal{A}|\mathcal{M}} \subseteq \mathcal{G}_{\mathcal{A}|\mathcal{M}}$ and $\mathcal{G}^d_{\mathcal{A}|\mathcal{M}} \subseteq \mathcal{G}_{\mathcal{A}}$ and $\mathcal{G}^d_{\mathcal{A}|\mathcal{M}} \subseteq \mathcal{G}_{\mathcal{M}}$, but in general neither $\mathcal{G}_{\mathcal{A}} \subseteq \mathcal{G}_{\mathcal{M}}$ nor $\mathcal{G}_{\mathcal{M}} \subseteq \mathcal{G}_{\mathcal{A}}$.

Note that [8, Section 4] also proposes a decision rule that rather could be viewed from a global perspective. Precisely, in that paper an act X is labeled simultaneously superior to the remaining acts from $\mathcal{G} \setminus \{X\}$, whenever the expectation $\mathbb{E}_\pi(u \circ X)$ of X is greater or equal as the mean of the expectations $\frac{1}{|\mathcal{G}|-1} \sum_{Y \in \mathcal{G} \setminus \{X\}} \mathbb{E}_\pi(u \circ Y)$ of the remaining acts for every compatible pair (u, π) of utility and probability representation. The author then proves that this criterion induces the same ranking of the acts than a similar criterion applied solely for pairwise comparisons of acts. This is an important difference to the concepts introduced in the present paper: Here, in general, the global admissibility concepts introduced in Definition 7 do indeed induce different orderings of the acts than the concepts of local admissibility introduced in Definition 8 from Section 4.4.

Note also that, if the involved set of utility representations $\mathcal{U}_{\mathcal{A}}$ is a class of positive linear transformations, i.e. belongs to a perfectly cardinal preference structure, then both $\mathcal{A}|\mathcal{M}$-admissibility and $\mathcal{M}$-admissibility reduce to E-admissibility as recalled in Section 2. The following example demonstrates that ignoring the available information base and applying the maximin criterion instead leads to counter-intuitive decisions even in very simple situations.

Example 1. Let $A = \{a_1, a_2, a_3, a_4\}$, the (complete) relation R_1 induced by $a_2 P_{R_1} a_3 P_{R_1} a_4 P_{R_1} a_1$ and $P_{R_2} = \{((a_2, a_4), (a_3, a_1))\}$ consists of one single edge. Consider the decision system $\mathcal{G} = \{X_1, X_2\}$, where the acts $X_1, X_2 : \{s_1, s_2\} \to A$ are defined by $(X_1(s_1), X_1(s_2)) = (a_1, a_2)$ and $(X_2(s_1), X_2(s_2)) = (a_3, a_4)$. An illustration of the decision system is given in Fig. 2.

Moreover, suppose there is additional probabilistic information available which is given by the credal set $\mathcal{M} := \{\pi : \pi(\{s_1\}) \leq 0.5\}$. In this case, act X_1 is $\mathcal{A}|\mathcal{M}$-dominant, since it maximizes expected utility with respect to every pair $(u, \pi) \in \mathcal{U}_{\mathcal{A}} \times \mathcal{M}$. In contrast, X_2 is not even $\mathcal{A}|\mathcal{M}$-admissible, although it is the unique optimal act with respect to Wald's maximin criterion (since it holds that $\inf_{s \in S} X_2(s) = a_4 P_{R_1} a_1 = \inf_{s \in S} X_1(s)$). Moreover, we can go on computing the generalized interval expectations of the acts X_1 and X_2 (in the sense of Definition 5) for varying degrees of granularity, say $\delta = 0, 0.1, 0.15, 0.25$. The resulting expectation intervals for the acts are visualized in Fig. 3.

To complete the section, we give a proposition containing a linear programming based approach for checking whether an act X is $\mathcal{A}$-admissible in finite decision settings.

Proposition 4. *Let $\mathcal{A} = [A, R_1, R_2]$ be a consistent preference system, where $A = \{a_1, \ldots, a_n\}$. Moreover, let $S = \{s_1, \ldots, s_m\}$ be finite, $\mathcal{M}$ be some polyhedral credal set on $(S, 2^S)$ with extreme points $\mathcal{E}(\mathcal{M}) := \{\pi^{(1)}, \ldots, \pi^{(T)}\}$ and let $\mathcal{G} := \{X_1, \ldots, X_k\} \subseteq \mathcal{F}_{(\mathcal{A}, \mathcal{M}, S)}$ denote a finite decision system with $X_z \in \mathcal{G}$. Consider again the linear optimization problem (1) with additional constraints*

$$\sum_{i=1}^{n} u_i \cdot \pi^{(t)}(X_z^{-1}(\{a_i\})) \geq \sum_{i=1}^{n} u_i \cdot \pi^{(t)}(X_l^{-1}(\{a_i\})) \ \text{ for all } \ l = 1, \ldots, k \tag{C_t}$$

for every $t = 1, \dots, T$. Then X_z is $\mathcal{A}$-admissible if and only if the optimal outcome of this optimization problem is strictly greater than 0.

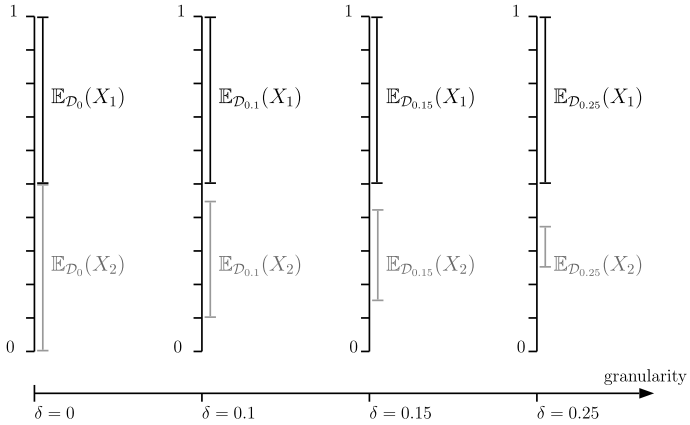

Fig. 3. Development of the generalized interval expectations of the acts X_1 and X_2 along increasing granularity. For the case $\delta = 0$ the lower expectation of act X_1 and the upper expectation of act X_2 coincide with a value of 0.5.

Proof. A similar argument as in the proof of Proposition 1 guarantees the existence of an optimal solution $(u_1^*, \dots, u_n^*, \varepsilon^*)$ to the optimization problem. If $\varepsilon^* = 0$, then there exists no vector $(u_1, \dots, u_n, \varepsilon)$ with $\varepsilon > 0$ satisfying the constraints of the optimization problem. Since every function $u \in \mathcal{U}_{\mathcal{A}}$ with $\mathbb{E}_{\pi^{(t)}}(u \circ X_z) \geq \mathbb{E}_{\pi^{(t)}}(u \circ X_l)$ for all $l = 1, \dots, k$ and $t = 1, \dots, T$ induces such a vector, we conclude that such u cannot exist. Since $\mathcal{E}(\mathcal{M}) \subseteq \mathcal{M}$, we conclude that X_z is not $\mathcal{A}$-admissible.

If $\varepsilon^* > 0$, constraints i) to iv) guarantee that $u : A \to \mathbb{R}$, $u(a_i) := u_i^*$ for all $i \in \underline{n}$ (weakly) represents the preference system $\mathcal{A}$. Now, let $\pi \in \mathcal{M}$ be arbitrary. Choose $\alpha \in \Delta_{T-1}$ such that $\pi(\cdot) = \sum_{t=1}^{T} \alpha_t \cdot \pi^{(t)}(\cdot)$. Then, condition ($C_t$) additionally guarantees that for all $l = 1, \dots, k$ it holds

$$\begin{aligned}
\mathbb{E}_{\pi}(u \circ X_z) &= \sum_{i=1}^{n} u_i^* \cdot \pi(X_z^{-1}(\{a_i\})) \\
&= \sum_{i=1}^{n} u_i^* \cdot \Big(\sum_{t=1}^{T} \alpha_t \cdot \pi^{(t)}(X_z^{-1}(\{a_i\})) \Big) \\
&= \sum_{t=1}^{T} \alpha_t \Big(\sum_{i=1}^{n} u_i^* \cdot \pi^{(t)}(X_z^{-1}(\{a_i\})) \Big) \\
&\geq \sum_{t=1}^{T} \alpha_t \Big(\sum_{i=1}^{n} u_i^* \cdot \pi^{(t)}(X_l^{-1}(\{a_i\})) \Big) \\
&= \sum_{i=1}^{n} u_i^* \cdot \Big(\sum_{t=1}^{T} \alpha_t \cdot \pi^{(t)}(X_l^{-1}(\{a_i\})) \Big) \\
&= \mathbb{E}_{\pi}(u \circ X_l)
\end{aligned}$$

Hence, X_z maximizes expected utility with respect to (u, π). Since $\pi \in \mathcal{M}$ was chosen arbitrarily, this implies that X_z is $\mathcal{A}$-admissible. □

Note that a similar algorithm as given in Proposition 4 could be used for checking $\mathcal{M}$-admissibility of acts. However, this would require the set $\mathcal{E}(\mathcal{U}_{\mathcal{A}})$ of extreme points of the representation set to be known, which is way less straightforward than assuming $\mathcal{E}(\mathcal{M})$ to be known.

4.4. Criteria based on pairwise comparisons

While the criteria defined in Section 4.3 rather relied on global comparisons of acts in the sense that an act, in order to be labeled admissible, has to dominate all other available acts from $\mathcal{G}$ in expectation for (at least one) *fixed* pair (π, u) simultaneously, we now turn to criteria induced by pairwise expectation comparisons of acts (i.e. binary relations on the set of acts). There, roughly spoken, the idea is to first compare the expectation of a fixed act X of interest to the expectation

of every other available act $Y \in \mathcal{G}$ *separately* and, afterwards, to label this act admissible if and only if none of the other available acts from $\mathcal{G}$ dominates it. In particular, the pair (u_Y, π_Y) for which the expectation of act X is compared to the expectation of act Y might now depend on Y, for which reason the following criteria could rather be viewed from a local perspective.

Similarly as already seen in the global case, there are several different ways to define such relations each of which reflecting a different attitude towards the underlying ambiguity between the different compatible probability measures and/or the indeterminacy on the utility level. In particular, we define six binary relations $R_{\exists\exists}$, $R^1_{\exists\forall}$, $R^2_{\exists\forall}$, $R^1_{\forall\exists}$, $R^2_{\forall\exists}$ and $R_{\forall\forall}$ on $\mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ by setting for all $X, Y \in \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$:

$$(X,Y) \in R_{\exists\exists} \text{ iff } \exists u \in \mathcal{U}_{\mathcal{A}}\ \exists \pi \in \mathcal{M} : \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) \tag{4}$$

$$(X,Y) \in R^1_{\exists\forall} \text{ iff } \exists u \in \mathcal{U}_{\mathcal{A}}\ \forall \pi \in \mathcal{M} : \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) \tag{5}$$

$$(X,Y) \in R^2_{\exists\forall} \text{ iff } \exists \pi \in \mathcal{M}\ \forall u \in \mathcal{U}_{\mathcal{A}} : \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) \tag{6}$$

$$(X,Y) \in R^1_{\forall\exists} \text{ iff } \forall u \in \mathcal{U}_{\mathcal{A}}\ \exists \pi \in \mathcal{M} : \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) \tag{7}$$

$$(X,Y) \in R^2_{\forall\exists} \text{ iff } \forall \pi \in \mathcal{M}\ \exists u \in \mathcal{U}_{\mathcal{A}} : \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) \tag{8}$$

$$(X,Y) \in R_{\forall\forall} \text{ iff } \forall \pi \in \mathcal{M}\ \forall u \in \mathcal{U}_{\mathcal{A}} : \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) \tag{9}$$

Obviously, it holds that $R_{\forall\forall}$ is subset of all other relation, whereas $R_{\exists\exists}$ is a superset of them. For the remaining relations, in general, no sub- or superset relation has to be satisfied. Furthermore, transitivity is only guaranteed for $R_{\forall\forall}$ in general. Similarly as already discussed in the global case, each of the desirability relations just defined relies on the idea that, given perfect information on utilities and probabilities, maximizing expected utility should be the criterion of choice. Again, the relations differ only in the way they handle the ambiguity on the involved sets $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{M}$. Naturally, each of the relations defined above induces a different criterion of (local) admissibility. These criteria are summarized in the following definition.

Definition 8. Let $R \in \{R_{\exists\exists}, R^1_{\exists\forall}, R^2_{\exists\forall}, R^1_{\forall\exists}, R^2_{\forall\exists}, R_{\forall\forall}\} =: \mathcal{R}_p$. We call an act $X \in \mathcal{G}$ **locally admissible** with respect to R, if it is an element of the set

$$\max_R(\mathcal{G}) := \{Y \in \mathcal{G} : \nexists Z \in \mathcal{G} \text{ s.t. } (Z,Y) \in P_R\}$$

that is if it is a maximal element in $\mathcal{G}$ with respect to the relation $R \cap (\mathcal{G} \times \mathcal{G})$.

So, which of the relations in $\mathcal{R}_p$ defined above are most important in our context? To address this question, it certainly makes sense to start by discussing some special cases of them: If the credal set $\mathcal{M}$ is a singleton $\mathcal{M} = \{\pi\}$, that is a precise probability available, and if the set of compatible utility representations $\mathcal{U}_{\mathcal{A}} = \{a \cdot u_0 + b \mid a > 0, b \in \mathbb{R}\}$ is unique up to a positive linear transformation of one utility function u_0, that is a perfectly cardinal utility, then all relations $R \in \mathcal{R}_p$ coincide with the classical expected utility criterion, i.e. with choosing an act that maximizes the expectation with respect to π and one arbitrary chosen utility representation from $\mathcal{U}_{\mathcal{A}}$.

If $\mathcal{M}$ still is a singleton, however, $\mathcal{U}_{\mathcal{A}}$ is the class of all non-decreasing functions with respect to R_1 (this essentially corresponds to the case where the relation R_2 of the underlying preference system is empty), then the relations $R^2_{\exists\forall}$, $R^1_{\forall\exists}$ and $R_{\forall\forall}$ essentially coincide with the classical concept of first order stochastic dominance (cf., e.g., [33,27,44]), while second order stochastic dominance is obtained if $\mathcal{U}_{\mathcal{A}}$ is the set of all continuous concave non-decreasing utility functions that are related to the concept of decreasing returns to scale. An intermediate case would arise if one has information about decreasing returns to scale only for parts of the preference system.

Finally, if the involved credal set $\mathcal{M}$ is no longer a singleton and utility is given perfectly cardinal again, then the relations $R_{\exists\exists}$, $R^2_{\exists\forall}$ and $R^1_{\forall\exists}$ all coincide and are exactly the ones corresponding to the criterion of maximality as recalled in Section 2. More precisely, the acts that are locally admissible with respect to one of the relations $R_{\exists\exists}$, $R^2_{\exists\forall}$ and $R^1_{\forall\exists}$ in that special case, are exactly the acts that are not dismissed when applying maximality. Additionally, the relations $R^1_{\exists\forall}$, $R^2_{\forall\exists}$ and $R_{\forall\forall}$ for that case reduce to Bewley's structural dominance (see, e.g., [3] or [16, p. 243]).

To check whether an act X dominates another acts Y with respect to one of the relations $R_{\exists\exists}$ and $R_{\forall\forall}$ in the general (yet finite) case, one can apply a similar technique as described in Proposition 3 by noting that $\mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$ is equivalent to $\mathbb{E}_\pi(u \circ X - u \circ Y) \geq 0$. Utilizing this fact leads us to the following proposition.

Proposition 5. *Let $\mathcal{A} = [A, R_1, R_2]$ be a consistent preference system, where $A = \{a_1, \ldots, a_n\}$. Moreover, let $S = \{s_1, \ldots, s_m\}$ be finite, $\mathcal{M}$ be some polyhedral credal set on $(S, 2^S)$ with extreme points $\mathcal{E}(\mathcal{M}) := \{\pi^{(1)}, \ldots, \pi^{(T)}\}$ and let $X, Y \in \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$. Consider the collection of linear programs $LO_1, \ldots, LO_T$ given by:*

$$\sum_{i=1}^{n} u_i \cdot [\pi^{(t)}(X^{-1}(\{a_i\})) - \pi^{(t)}(Y^{-1}(\{a_i\}))] \longrightarrow \min_{(u_1,\ldots,u_n)\in\mathbb{R}^n} / \max_{(u_1,\ldots,u_n)\in\mathbb{R}^n} \tag{LO$_t$}$$

with constraint i)–iv) from Proposition 1 where ε is set to 0. For $t = 1, \dots, T$, denote by $\underline{v}(t)$ and $\overline{v}(t)$ the optimal value of problem LO_t in minimum and maximum form, respectively. Then, the following holds:

i) $(X, Y) \in R_{\forall\forall}$ *if and only if* $\min_t \underline{v}(t) \geq 0$
ii) $(X, Y) \in R_{\exists\exists}$ *if* $\max_t \overline{v}(t) > 0$

Proof. Define the function $g : \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M} \to \mathbb{R}$ by setting $g(u, \pi) = \mathbb{E}_\pi(u \circ X - u \circ Y)$, where $\tilde{\mathcal{U}}_{\mathcal{A}}$ denotes the set of all functions $u : A \to [0, 1]$ which are monotone (but not necessarily strictly monotone) with respect to the relations R_1 and R_2. Note that $\mathcal{U}_{\mathcal{A}} \subseteq \tilde{\mathcal{U}}_{\mathcal{A}}$. A similar argument as performed in the proof of Proposition 3 shows that $\min_t \underline{v}(t) = \min\{g(u, \pi) : (u, \pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}\}$ and $\max_t \overline{v}(t) = \max\{g(u, \pi) : (u, \pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}\}$.

Part i): If $\min_t \underline{v}(t) \geq 0$ (note that this actually means $\min_t \underline{v}(t) = 0$ since the vector $(0, \dots, 0)$ is an admissible solution of LO_t for all $t = 1, \dots, T$), then, according to the above identity, it holds that $\mathbb{E}_\pi(u \circ X - u \circ Y) \geq 0$ for all $(u, \pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}$. Since $\mathcal{U}_{\mathcal{A}} \subseteq \tilde{\mathcal{U}}_{\mathcal{A}}$, this implies $\mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$ for all $(u, \pi) \in \mathcal{U}_{\mathcal{A}} \times \mathcal{M}$. If contrarily $\min_t \underline{v}(t) < 0$, let $(u_1^*, \dots, u_n^*)$ denote a solution yielding $\min_t \underline{v}(t)$ and define $u^* \in \tilde{\mathcal{U}}_{\mathcal{A}}$ by setting $u^*(a_i) := u_i^*$. If $u^* \in \mathcal{U}_{\mathcal{A}}$ we are done. If $u^* \in \tilde{\mathcal{U}}_{\mathcal{A}} \setminus \mathcal{U}_{\mathcal{A}}$, choose $u_0 \in \mathcal{U}_{\mathcal{A}} \neq \emptyset$ (this is possible since $\mathcal{A}$ is assumed to be consistent) such that $\mathbb{E}_\pi(u_0 \circ X - u_0 \circ Y) < |\min_t \underline{v}(t)|$ for all $\pi \in \mathcal{M}$ (this is possible since, with any $u \in \mathcal{U}_{\mathcal{A}}$, we have also $\lambda \cdot u \in \mathcal{U}_{\mathcal{A}}$ for arbitrary $\lambda \in (0, 1)$). One then easily verifies that $u^+ := \frac{u^* + u_0}{2} \in \mathcal{U}_{\mathcal{A}}$. One also easily verifies that, if $\pi^+ \in \mathcal{M}$ is chosen to be a credal element yielding outcome $\min_t \underline{v}(t)$ in combination with u^*, then it holds $\mathbb{E}_{\pi^+}(u^+ \circ X) < \mathbb{E}_{\pi^+}(u^+ \circ Y)$. This completes the proof of i).

Part ii): If $\max_t \overline{v}(t) > 0$, then, due to $\max_t \overline{v}(t) = \max\{g(u, \pi) : (u, \pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}\}$, there exists a pair $(u^*, \pi^+) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}$ such that $\mathbb{E}_{\pi^+}(u^* \circ X) \geq \mathbb{E}_{\pi^+}(u^* \circ Y)$. If $u^* \in \mathcal{U}_{\mathcal{A}}$ we are done. If $u^* \in \tilde{\mathcal{U}}_{\mathcal{A}} \setminus \mathcal{U}_{\mathcal{A}}$, choose $u_0 \in \mathcal{U}_{\mathcal{A}} \neq \emptyset$ (again utilizing the consistency of the preference system $\mathcal{A}$) such that $\mathbb{E}_\pi(u_0 \circ X - u_0 \circ Y) > -\max_t \overline{v}(t)$ (again utilizing the fact that, with any $u \in \mathcal{U}_{\mathcal{A}}$, we have also $\lambda \cdot u \in \mathcal{U}_{\mathcal{A}}$ for arbitrary $\lambda \in (0, 1)$). Analogously as in part i), we have that $u^+ := \frac{u^* + u_0}{2} \in \mathcal{U}_{\mathcal{A}}$. Moreover, one easily verifies that it holds that $\mathbb{E}_{\pi^+}(u^+ \circ X) > \mathbb{E}_{\pi^+}(u^+ \circ Y)$. Thus, there exists a pair $(u^+, \pi^+) \in \mathcal{U}_{\mathcal{A}} \times \mathcal{M}$ with the desired property and, therefore, it holds that $(X, Y) \in R_{\exists\exists}$. This completes the proof of part ii). □

Note that the converse implication in part ii) of Proposition 5 is not necessarily true (for a trivial example consider the pair $(X, X) \in R_{\exists\exists}$). A non-trivial situation where the opposite direction fails to hold is illustrated by the following toy example:

Example 2. Let $A = \{a_1, a_2, a_3, a_4\}$, the (complete) relation R_1 induced by $a_1 I_{R_1} a_4 P_{R_1} a_2 I_{R_1} a_3$ and the relation $R_2 = \emptyset$. Consider the decision system $\mathcal{G} = \{X_1, X_2\}$ consisting of two acts $X_1, X_2 : \{s_1, s_2\} \to A$ defined by $(X_1(s_1), X_1(s_2)) = (a_1, a_2)$ and $(X_2(s_1), X_2(s_2)) = (a_3, a_4)$. Suppose the uncertainty about the states is characterized by the credal set $\mathcal{M} = \{\pi : \pi(\{s_1\}) \leq 0.5\}$. Then we have $(X, Y) \in R_{\exists\exists}$, since for π the uniform distribution and u defined by $u(a_1) = u(a_4) = 0.75$ and $u(a_2) = u(a_3) = 0.25$ we have $\mathbb{E}_\pi(u \circ X_1) = \mathbb{E}_\pi(u \circ X_2)$. However, it holds that

$$\begin{aligned}
\max_t \overline{v}(t) &= \sup_{(u,\pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}} \mathbb{E}_\pi(u \circ X_1 - u \circ X_2) \\
&= \sup_{(u,\pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}} u(a_1)(\pi(\{s_1\}) - \pi(\{s_2\})) + u(a_2)(\pi(\{s_2\}) - \pi(\{s_1\})) \\
&= \sup_{(u,\pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}} (u(a_1) - u(a_2))(\pi(\{s_1\}) - \pi(\{s_2\})) \\
&= 0
\end{aligned}$$

where the last equality holds since $u(a_1) - u(a_2) \geq 0$ due to $a_1 P_{R_1} a_2$ and, therefore, the product is maximal when $\pi(\{s_1\})$ is, which is the case for $\pi(\{s_1\}) = 0.5$. Hence, we have constructed a situation where $(X, Y) \in R_{\exists\exists}$ but not $\max_t \overline{v}(t) > 0$.

In the following proposition, we formulate an additional assumption under which also the opposite implication of Proposition 5 ii) is valid.

Proposition 6. *Consider again the situation of Proposition 5. Additionally, assume that both R_1 and R_2 are antisymmetric relations and that there exists an element $a_0 \in A$ such that*

$$\pi^{(t)}(X^{-1}(\{a_0\})) - \pi^{(t)}(Y^{-1}(\{a_0\})) > 0$$

for all $t = 1, \dots, T$. Then, we have that $(X, Y) \in R_{\exists\exists}$ if and only if $\max_t \overline{v}(t) > 0$.

Proof. It follows from Proposition 5 ii) that $\max_t \overline{v}(t) > 0$ implies $(X, Y) \in R_{\exists\exists}$. For the converse implication assume that $(X, Y) \in R_{\exists\exists}$. We then can choose $u^* \in \mathcal{U}_{\mathcal{A}}$ such that $\mathbb{E}_{\pi^+}(u^* \circ X) \geq \mathbb{E}_{\pi^+}(u^* \circ Y)$ for some $\pi^+ \in \mathcal{M}$. If the inequality holds strictly we are done, since then setting $u_i^* := u^*(a_i)$ for all $i \in \underline{n}$ induces a solution with an objective value strictly greater

than 0. Thus, assume $\mathbb{E}_{\pi^+}(u^* \circ X) = \mathbb{E}_{\pi^+}(u^* \circ Y)$. Choose $\alpha \in \Delta_{T-1}$ such that $\pi^+(\cdot) = \sum_{t=1}^{T} \alpha_t \cdot \pi^{(t)}(\cdot)$ and let $\delta > 0$ be defined as in the proof of Proposition 1, i.e. as the minimal difference with respect to u over all elements of P_{R_1} and P_{R_2}. Define the function $u^+ : A \to [0,1]$ by $u^+(a) := u^*(a)$ for $a \neq a_0$ and $u^+(a_0) := u^*(a_0) + \frac{\delta}{4}$. Then one can show that $u^+ \in \mathcal{U}_{\mathcal{A}}$ (utilizing that both relations R_1 and R_2 are antisymmetric).[5] Set $z(a,t) := \pi^{(t)}(X^{-1}(\{a\})) - \pi^{(t)}(Y^{-1}(\{a\}))$ for $t = 1, \dots T$ and $a \in A$. Then, we can compute

$$
\begin{aligned}
\max_t \overline{v}(t) &\geq \mathbb{E}_{\pi^+}(u^+ \circ X - u^+ \circ Y) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \mathbb{E}_{\pi^{(t)}}(u^+ \circ X - u^+ \circ Y) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \Big(\sum_{a \in A} u^+(a) \cdot z(a,t) \Big) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \Big(\Big(\sum_{a \in A \setminus \{a_0\}} u^+(a) \cdot z(a,t) \Big) + u^+(a_0) \cdot z(a_0,t) \Big) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \Big(\Big(\sum_{a \in A \setminus \{a_0\}} u^*(a) \cdot z(a,t) \Big) + (u^*(a_0) + \tfrac{\delta}{4}) \cdot z(a_0,t) \Big) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \Big(\Big(\sum_{a \in A} u^*(a) \cdot z(a,t) \Big) + \tfrac{\delta}{4} \cdot z(a_0,t) \Big) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \sum_{a \in A} u^*(a) \cdot z(a,t) + \tfrac{\delta}{4} \cdot \sum_{t=1}^{T} \alpha_t \cdot z(a_0,t) \\
&= \sum_{t=1}^{T} \alpha_t \cdot \mathbb{E}_{\pi^{(t)}}(u^* \circ X - u^* \circ Y) + \tfrac{\delta}{4} \cdot \sum_{t=1}^{T} \alpha_t \cdot z(a_0,t) \\
&= \mathbb{E}_{\pi^+}(u^* \circ X - u^* \circ Y) + \tfrac{\delta}{4} \cdot \sum_{t=1}^{T} \alpha_t \cdot z(a_0,t) \\
&= 0 + \tfrac{\delta}{4} \cdot \sum_{t=1}^{T} \alpha_t \cdot (\pi^{(t)}(X^{-1}(\{a_0\})) - \pi^{(t)}(Y^{-1}(\{a_0\}))) \\
&> 0
\end{aligned}
$$

where the first inequality sign is valid since, as seen in the proof of Proposition 5, we have that $\max_t \overline{v}(t) = \max\{g(u,\pi) : (u,\pi) \in \tilde{\mathcal{U}}_{\mathcal{A}} \times \mathcal{M}\}$ and where the last strict inequality holds since we have $\pi^{(t)}(X^{-1}(\{a_0\})) - \pi^{(t)}(Y^{-1}(\{a_0\})) > 0$ for all $t = 1, \dots T$ by assumption. This gives $\max_t \overline{v}(t) > 0$ and completes the proof of the proposition. □

Note that the other relations $R \in \mathcal{R}_p \setminus \{R_{\forall\forall}, R_{\exists\exists}\}$ do not appear to be manageable in such a straightforward manner. However, in the special case that $\mathcal{M}$ is the core of a belief function, all $\pi \in \mathcal{M}$ can be understood as obtained from a mass transfer of probability mass to singleton sets of S (cf., e.g., [5, Corollary 3, p. 273] or [14, Theorem 2, p. 29] in the context of game theory). Since classical first order stochastic dominance can be alternatively checked via the solution of a suitable mass transportation problem (cf., [44, p. 269]), the computation of $R^2_{\exists\forall}$ can be done by solving a composite mass transportation problem.

On the other hand, if R_1 is totally ordered and R_2 is empty, then first order stochastic dominance can be characterized for a precise probability π as $Y \geq_{SD} X \iff \forall c \in A : \pi(Y \geq c) \geq \pi(X \geq c)$. This can be generalized to imprecise probabilities

[5] Antisymmetry is required since otherwise changing u^* only on the element a_0 would mean that u^+ cannot represent the relations I_{R_1} and I_{R_2} on pairs of the form $(a_0, a_1) \in I_{R_1}$ with $a_0 \neq a_1$ and pairs of the form $((a_0,a_1),(a_2,a_3)) \in I_{R_2}$ with $a_1, a_2, a_3 \in A \setminus \{a_0\}$.

Given antisymmetry, proving that $u^+ \in \mathcal{U}_{\mathcal{A}}$ is then straightforward, however, involves some tedious arithmetic exercises. One has to show that u^+ represents both relations R_1 and R_2. Therefore, one first has to note that by definition of δ it holds that $u^*(a) - u^*(b) > \frac{\delta}{2}$ for all $(a,b) \in P_{R_1}$ and that $u^*(a) - u^*(b) - (u^*(c) - u^*(d)) > \frac{\delta}{2}$ for all $((a,b),(c,d)) \in P_{R_2}$. It is then immediate that u^+ represents R_1 and R_2 for pairs not containing a_0, since for such pairs u^+ equals u^* and $u^* \in \mathcal{U}_{\mathcal{A}}$. Thus, we need only care about pairs containing a_0. There are several cases to distinguish. We only show the most complicated one. Assume for $a_1, a_2 \in A \setminus \{a_0\}$ it holds that $((a_1,a_0),(a_0,a_2)) \in P_{R_2}$. Then, due to the second of the above identities, it holds that $u^*(a_1) - u^*(a_0) - \frac{\delta}{2} > u^*(a_0) - u^*(a_2)$. This implies $u^*(a_1) - (u^*(a_0) + \frac{\delta}{4}) > u^*(a_0) + \frac{\delta}{4} - u^*(a_2)$, which implies $u^+(a_1) - u^+(a_0) > u^+(a_0) - u^+(a_2)$. Thus, u^+ represents R_2 on pairs of this form. The remaining cases are similar.

by replacing the probability of the events $X \geq c$ and $Y \geq c$, respectively by the lower or the upper probabilities associated with the credal set $\mathcal{M}$. This would lead to four other generalizations of stochastic dominance for imprecise probabilities and was studied in detail in [13] for the case of belief functions. In the special case of some notion of "independence" of X and Y the relation $Y \geq X \iff \forall c \in A : \underline{P}(Y \geq c) \geq \overline{P}(X \geq c)$ would be equivalent to our relation $R_{\forall\forall}$. Here, we used the notations $\underline{P}(B) = \inf_{\pi \in \mathcal{M}} \pi(B)$ and $\overline{P}(B) = \sup_{\pi \in \mathcal{M}} \pi(B)$ for $B \subseteq A$ and the term "independence" means that there always exists some $\pi \in \mathcal{M}$ that attains at the same time $\pi(X \geq c) = \overline{P}(X \geq c)$ and $\pi(Y \geq c) = \underline{P}(Y \geq c)$.

Note that the characterization of stochastic dominance via the probability of the events of the form $X \geq c$ (and $Y \geq c$) becomes far more complicated when dealing with a relation R_1 that is only partial. In this case one has to consider the probabilities of all events of the form $X \in U$ (and $Y \in U$), where U is an arbitrary upset,[6] and the number of upsets can become extremely large such that explicitly checking all upsets becomes intractable. (But note that for the case of a precise probability, checking all upsets can be done by using linear programming techniques described in [50].)

It should be further mentioned that in [42, Section 3] and in [43, Section 3] the authors introduce six binary relations $\succeq_1, \ldots, \succeq_6$ relying on a quite similar idea as the relations in $\mathcal{R}_p$, however, in a slightly different context. Here, the authors explore six ways of extending a binary relation $\succeq$ between random variables to binary relations $\succeq_1, \ldots, \succeq_6$ between sets of random variables that are based on the same construction principle as the ones collected in the set $\mathcal{R}_p$ (i.e. considering all variants of placing the $\exists$ and the $\forall$ quantifier). The authors then propose to apply these relations for decision making with acts attaining uncertain rewards: By considering more general acts $\Gamma : S \to 2^A \setminus \{\emptyset\}$ yielding set-valued outcomes (i.e. acts that are random sets), and their associated sets of random variables $S(\Gamma) = \{X : S \to A : X(s) \in \Gamma(s)\}$, they propose to prefer act Γ_1 before Γ_2 whenever $S(\Gamma_1) \succeq_{i_0} S(\Gamma_2)$, where $\succeq_{i_0} \in \{\succeq_1, \ldots, \succeq_6\}$ is the extension of choice. The main difference to the relations proposed in the present paper is that we do not consider acts with uncertain reward, but certain rewards with uncertain utility assignment. More precisely, we do exactly know which consequence from A is attained by which act under which state of nature, however, we do not know which is the concrete utility assignment. Consequently, we find ourselves in a more structured setting than the authors in [42,43] and we would ignore information by solely considering the relations $\succeq_1, \ldots, \succeq_6$.

Furthermore, in [42, Section 5] the authors propose two binary relation $\succeq_s^{\mathcal{M}}$ and $\succeq_w^{\mathcal{M}}$ between acts if the uncertainty on the states is characterized by a credal set $\mathcal{M}$. Here, they first assume a family $(\succeq_\pi)_{\pi \in \mathcal{M}}$ of relations on the states each representing the ordering of the acts given π was the true distribution and, afterwards, define act X to be preferable to act Y, that is $(X, Y) \in \succeq_s^{\mathcal{M}}$ or $(X, Y) \in \succeq_w^{\mathcal{M}}$, if it holds that $(X, Y) \in \succeq_\pi$ for all $\pi \in \mathcal{M}$ or it holds that $(X, Y) \in \succeq_\pi$ for some $\pi \in \mathcal{M}$, respectively. Here, there are some close connections to the relations from the set $\mathcal{R}_p$: If we assume a preference system $\mathcal{A} = [A, R_1, R_2]$ with a complete ordinal relation R_1 and an empty cardinal relation R_2 and we additionally choose $\succeq_\pi$ to be defined as first order stochastic dominance with respect to π for every $\pi \in \mathcal{M}$, then it holds that $(X, Y) \in R_{\forall\forall}$ if and only if $(X, Y) \in \succeq_s^{\mathcal{M}}$ as well as $(X, Y) \in R^2_{\exists\forall}$ if and only if $(X, Y) \in \succeq_w^{\mathcal{M}}$.

Finally, note that the relation $R_{\forall\forall}$ is also discussed in [9, Section 2]: If one considers the concept of t-admissibility proposed in that paper for the special case that $t = 0$ and one additionally assumes the sets V and P from [9] to consist of exactly those utility and probability sentences that characterize the sets $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{M}$ respectively, then the set of t-admissible acts coincides with the set of acts that are locally admissible with respect to $R_{\forall\forall}$ in the sense of Definition 8. However, in general, it will not always be possible to describe the utility constraints induced by the relation R_2 and the probability constraints induced by the credal set $\mathcal{M}$ by one of the three types of utility and probability sentences that are considered in that paper. Of course, these constraints could straightforwardly incorporated in the framework developed in the paper by allowing for larger classes of such sentences, since they are still linear in the corresponding values.

Clearly, the set of locally $R_{\forall\forall}$-admissible acts coincides with the set of $\mathcal{A}|\mathcal{M}$-dominant acts. Thus, part i) of Proposition 5 can also be used for checking whether an act X is $\mathcal{A}|\mathcal{M}$-dominant by solving the problems $(\mathrm{LO}_t)_{t=1,\ldots,T}$ for every pair $((X, Y))_{Y \in \mathcal{G}}$, where $\mathcal{G} \subseteq \mathcal{F}_{(\mathcal{A},\mathcal{M},S)}$ once again denotes a finite set of available concurring acts. This is a unique feature offered by $R_{\forall\forall}$: In general, the other global concepts of admissibility from Definition 4.3 cannot be expressed as induced by one of the local criteria from Definition 4.4 (for the special case of a cardinal u this is discussed and shown in [48]).

5. A real world toy example

In this section, we apply certain aspects of the proposed framework for decision making under uncertainty by computing selected decision criteria for a prototypical toy example. Particularly, we thereby demonstrate, firstly, that our framework is computationally feasible and, secondly, show a class of situations in which ps-valued acts naturally appear in practical applications, namely situations where the orderings R_1 and R_2 arise from the fact that the acts map into some bivariate product space with one cardinal and one (potentially partial) ordinal dimension.

5.1. Setup of the example

The example reads as follows: Suppose the agent under consideration is currently looking for a new job. As she has very high qualification in her field, she immediately receives three different job offers, say J_1, J_2 and J_3, each of which

[6] A set $U \subseteq A$ is called an upset if $\forall x, y \in A : x \in U \;\; \& \;\; y \geq x \Longrightarrow y \in U$.

appears to be of high interest for her at first sight. In order to come to a decision between the different job offers, she decides to address the situation systematically by comparing the jobs with respect to the offered monthly salary after tax and the offered additional benefits. She specifies the additional benefits collected in $\mathcal{X} = \{b_1, \dots, b_5\}$ to be important for her, where

b_1	b_2	b_3	b_4	b_5
overtime premium	child care	advanced training	promotion prospects	flexible hours

Under the assumption that the different additional benefits are incomparable for our agent, the situation just described very naturally induces the following preference system $\mathcal{A}' = [A', R_1', R_2']$, where $A' = \mathbb{R}^+ \times 2^{\mathcal{X}}$ is the set of possible decision outcomes (each of which consists of a potential salary offer and a set of additional benefits), the relation $R_1' \subseteq A' \times A'$ is the component-wise ordering given by

$$R_1' = \{((y_1, B_1), (y_2, B_2)) : y_1 \geq y_2 \wedge B_1 \supseteq B_2\} \tag{10}$$

and the relation $R_2' \subseteq R_1' \times R_1'$ partially specifying the strength of preferences is given by

$$R_2' = \Big\{(((y_1, B_1), (y_2, B_2)), ((y_3, B_3), (y_4, B_4))) : y_1 - y_2 \geq y_3 - y_4 \wedge B_1 \supseteq B_3 \supseteq B_4 \supseteq B_2\Big\} \tag{11}$$

The relation R_1' is interpretable in a pretty straightforward manner: An element (y_1, B_1) of A' is preferred to another element (y_2, B_2), whenever it is preferable in terms of salary, i.e. $y_1 \geq y_2$ and offers a super-set of additional benefits, i.e. $B_1 \supseteq B_2$. Otherwise, the elements are incomparable with respect to R_1'. Moreover, also the relation R_2' possesses a very natural interpretation: Whenever, for elements $((y_1, B_1), (y_2, B_2)), ((y_3, B_3), (y_4, B_4)) \in R_1'$, it is clear that exchanging B_2 by B_1 is preferable to exchanging B_4 by B_3 since it holds that $B_1 \supseteq B_3 \supseteq B_4 \supseteq B_2$, one can compare the exchanges of elements from A by simply checking whether the difference $y_1 - y_2$ is greater than the difference $y_3 - y_4$ in the salaries.

Finally, the agent specifies a set $S = \{s_1, \dots, s_4\}$ of four different economic scenarios which might affect the offers of the companies in different ways (for example, here, $\{s_4\}$ might be some event having very negative influence on the stock price of the company offering job J_1, whereas $\{s_1\}$ might be an event causing the opposite). Particularly, the agent can specify the following decision system describing her situation:

	s_1	s_2	s_3	s_4
J_1	$\underbrace{(5000, \mathcal{X})}_{=:a_1}$	$\underbrace{(2700, \{b_1, b_2\})}_{=:a_2}$	$\underbrace{(2300, \{b_1, b_2, b_3\})}_{=:a_3}$	$\underbrace{(1000, \emptyset)}_{=:a_4}$
J_2	$\underbrace{(3500, \{b_1, b_5\})}_{=:a_5}$	$\underbrace{(2400, \{b_1, b_2\})}_{=:a_6}$	$\underbrace{(1700, \{b_1, b_2\})}_{=:a_7}$	$\underbrace{(2500, \{b_1\})}_{=:a_8}$
J_3	$\underbrace{(3000, \{b_1, b_2, b_3\})}_{=:a_9}$	$\underbrace{(1000, \{b_1\})}_{=:a_{10}}$	$\underbrace{(2000, \{b_1\})}_{=:a_{11}}$	$\underbrace{(3000, \{b_1, b_4, b_5\})}_{=:a_{12}}$

Once having set up the decision system, the agent can also determine the relevant preference system $\mathcal{A} = [A, R_1, R_2]$ by setting $A := \{a_1, \dots, a_{12}\}$, $R_1 := R_1' \cap (A \times A)$ and $R_2 := R_2' \cap (R_1 \times R_1)$, i.e. by restricting all sets contained in the triplet $\mathcal{A}'$ to the relevant ones. The Hasse graph of the order R_1 (which is clearly anti-symmetric here, since all elements of A are distinct) then can be visualized as in Fig. 4. Note that R_2 is not anti-symmetric, since for instance it holds that $((a_3, a_7), (a_9, a_6)) \in I_{R_2}$ while $(a_3, a_7) \neq (a_9, a_6)$ and therefore distinct equivalent elements with respect to R_2 exist.

5.2. Checking consistency and applying the decision criteria

First, we want to check whether the preference system $\mathcal{A} = [A, R_1, R_2]$ of the considered agent is consistent in the sense of Definition 2. Therefore, we apply the algorithm described in Proposition 1, whose objective function translates as

$$\varepsilon = \langle (0, \dots, 0, 1)', (u_1, \dots, u_{12}, \varepsilon)' \rangle \longrightarrow \max_{(u_1, \dots, u_{12}, \varepsilon) \in \mathbb{R}^{13}} \tag{12}$$

and whose constraints are determined by the relations R_1 and R_2 from the preference system under consideration as described in Proposition 1. For example, since $(a_9, a_2) \in P_{R_1}$, we add the constraint $u_2 + \varepsilon \leq u_9$ and, since $((a_7, a_{10}), (a_8, a_{11})) \in P_{R_2}$ (the edges (a_7, a_{10}) and (a_8, a_{11}) are indicated with ♠ and ♣ in the figure), we add the constraint $u_8 - u_{11} + \varepsilon \leq u_7 - u_{10}$. Solving the resulting linear programming problem (12) gives an optimal objective of $0.\overline{037}$. Hence, according to Proposition 1, the preference system $\mathcal{A}$ is consistent. An optimal solution to the problem is given by $(u_1^*, \dots, u_{12}^*, \varepsilon^*) \approx (1, 0.\overline{4}, 0.\overline{370}, 0, 0.\overline{629}, 0.\overline{370}, 0.\overline{2}, 0.\overline{370}, 0.\overline{518}, 0.\overline{037}, 0.\overline{259}, 0.\overline{5}, 0.\overline{037})$, which induces an element $u^* \in \mathcal{U}_{\mathcal{A}}$ by setting $u^*(a_i) := u_i^*$ for $i = 1, \dots, 12$.

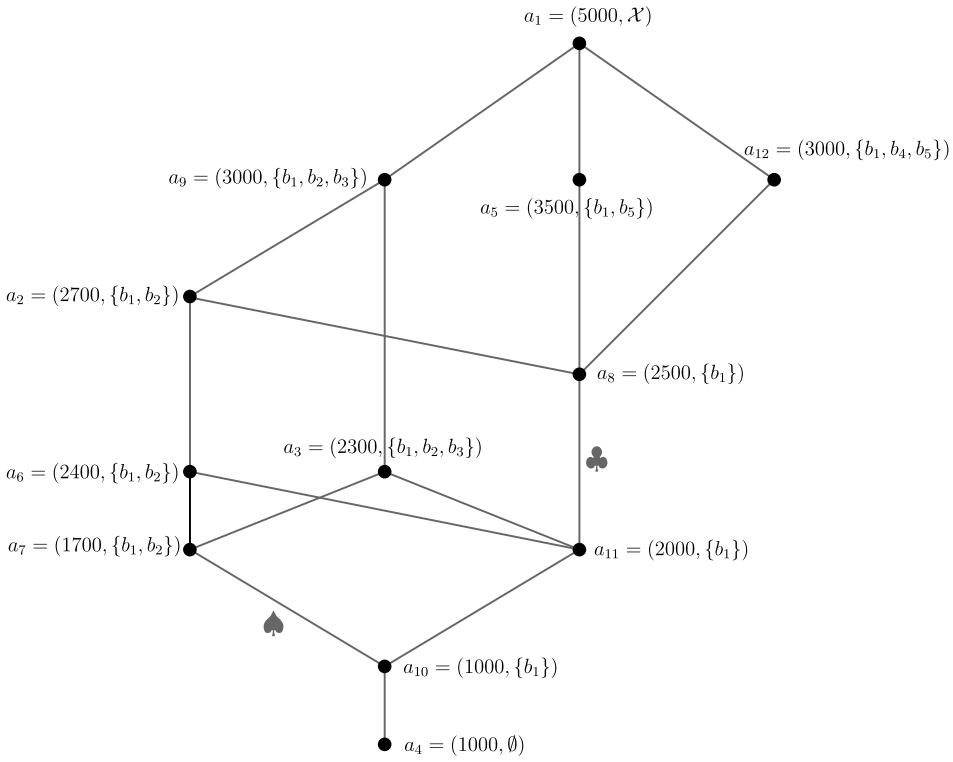

Fig. 4. Hasse graph of the relation R_1 of the example. The symbols ♠ and ♣ mark examples of elements of $R_1 \times R_1$ that are comparable with respect to the strict relation P_{R_2}.

Now, suppose our agent collects some more information that allows her to order the different economic scenarios $s_1, \ldots, s_4$ by their probability to occur, i.e. by an ordinal probability specified by the credal set $\mathcal{M} = \{\pi : \pi(\{s_1\}) \geq \pi(\{s_2\}) \geq \pi(\{s_3\}) \geq \pi(\{s_4\})\}$. In this situation, the set of extreme point of $\mathcal{M}$ possesses exactly four elements and is given by $\mathcal{E}(\mathcal{M}) = \{\pi^{(1)}, \ldots, \pi^{(4)}\}$, where we have $\pi^{(t)}(\{s_j\}) = \mathbb{1}_{\{1,\ldots,t\}}(j) \cdot \frac{1}{t}$ for $j, t \in \{1, 2, 3, 4\}$ (cf., [29, p. 26] or [38, Proposition 5 and Algorithm 1]).

Then, we want to check which of the jobs J_1, J_2, J_3 are $\mathcal{A}$-admissible in the sense of Definition 7, part ii). The linear optimization problem described in Proposition 4, for instance applied for job J_1, then possesses the same objective function as the program for checking consistency, namely (12). Moreover, it also includes all the constraints of problem (12), however, additionally involves the constraints (as described in Proposition 4)

$$
\begin{aligned}
(C_1)\ & u_1 \geq u_5 \\
& u_1 \geq u_9 \\
(C_2)\ & \tfrac{1}{2}(u_1 + u_2) \geq \tfrac{1}{2}(u_5 + u_6) \\
& \tfrac{1}{2}(u_1 + u_2) \geq \tfrac{1}{2}(u_9 + u_{10}) \\
(C_3)\ & \tfrac{1}{3}(u_1 + u_2 + u_3) \geq \tfrac{1}{3}(u_5 + u_6 + u_7) \\
& \tfrac{1}{3}(u_1 + u_2 + u_3) \geq \tfrac{1}{3}(u_9 + u_{10} + u_{11}) \\
(C_4)\ & \tfrac{1}{4}(u_1 + u_2 + u_3 + u_4) \geq \tfrac{1}{4}(u_5 + u_6 + u_7 + u_8) \\
& \tfrac{1}{4}(u_1 + u_2 + u_3 + u_4) \geq \tfrac{1}{4}(u_9 + u_{10} + u_{11} + u_{12})
\end{aligned}
$$

that are due to the information about the uncertainty that is given by the credal set $\mathcal{M}$ (where C_i, for $i = 1, \ldots, 4$, here describes the constraint induced by the ith extreme point). Again, solving the resulting linear programming problem gives an optimal objective of $0.\overline{037}$ and, again, an optimal solution to the problem is given by $(u_1^*, \ldots, u_{12}^*, \varepsilon^*)$ from above. However, the interpretation of the optimal solution is quite different: If we define $u^* \in \mathcal{U}_{\mathcal{A}}$ as above, then job J_1 maximizes expected utility with respect to (u^*, π) for every $\pi \in \mathcal{M}$ compatible with the agent's probabilistic information. In contrast, solving the same linear programming problem from Proposition 4 for the jobs J_2 and J_3 gives an optimal objective of 0 each time indicating that both jobs are not $\mathcal{A}$-admissible. According to $\mathcal{A}$-admissibility, therefore, our agent should decide for job J_1.

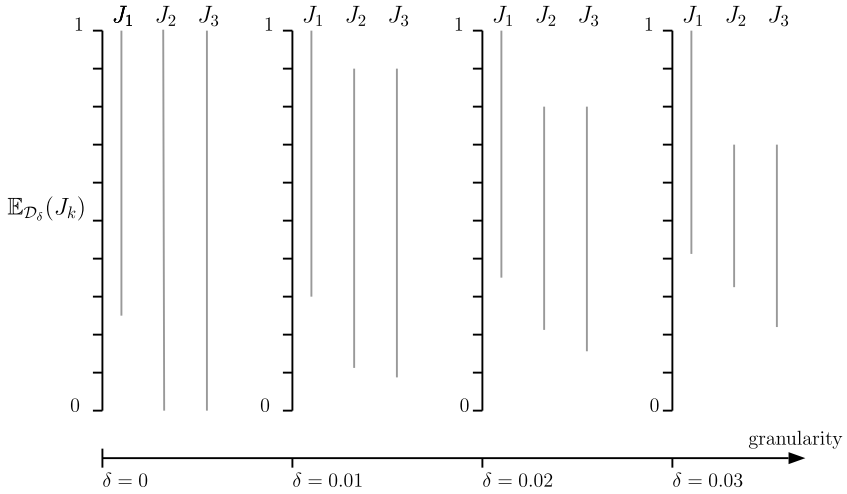

Fig. 5. Generalized expectation intervals of the different jobs along increasing value of granularity.

Next, we want to compute the generalized interval expectations from Definition 5 of choosing one of the three jobs under consideration along varying granularity value $\delta = 0, 0.01, 0.02, 0.03$.[7] According to Proposition 3, for job J_k ($k = 1, 2, 3$) and fixed value of the granularity parameter δ, this makes necessary solving the optimization problems

$$\sum_{i=1}^{12} u_i \cdot \pi^{(t)}(J_k^{-1}(\{a_i\})) \longrightarrow \min_{(u_1,\ldots,u_{12})\in\mathbb{R}^{12}} / \max_{(u_1,\ldots,u_{12})\in\mathbb{R}^{12}} \tag{13}$$

for $t = 1, \ldots, 4$, with constraints as described in Proposition 1, with the difference that ε here is not one of the variables, but is set to δ. Thus, eight linear programs have to be solved for each choice of job J_k and δ and the generalized expectation interval can be computed as described in the proposition (i.e. by taking for each job the minimum of the minima as lower bound and the maximum of the maxima as upper bound). Solving the corresponding optimization problems gives the following results:

	$\delta = 0$	$\delta = 0.01$	$\delta = 0.02$	$\delta = 0.03$
$\mathbb{E}_{\mathcal{D}_\delta}(J_1)$	[0.25, 1]	[0.305, 1]	[0.36, 1]	[0.415, 1]
$\mathbb{E}_{\mathcal{D}_\delta}(J_2)$	[0, 1]	[0.1075, 0.9]	[0.215, 0.8]	[0.3225, 0.7]
$\mathbb{E}_{\mathcal{D}_\delta}(J_3)$	[0, 1]	$[0.07\bar{3}, 0.9]$	$[0.14\bar{6}, 0.8]$	[0.22, 0.7]

Since both lower and upper bound of the interval of J_1 are greater than the respective bounds of the intervals of J_2 and J_3 (independent of which granularity value is chosen), job J_1 is also optimal with respect to all criteria introduced in Definition 6. The generalized expectation intervals of the different jobs along increasing value of granularity are visualized in Fig. 5.

Finally, we apply Proposition 5 in order to investigate how the different job offers J_1, J_2 and J_3 relate to each other with respect to the relations $R_{\exists\exists}$ and $R_{\forall\forall}$. The results are summarized in the following table:

	(J_1, J_2)	(J_1, J_3)	(J_2, J_1)	(J_2, J_3)	(J_3, J_1)	(J_3, J_2)
$R_{\exists\exists}$	$\in$	$\in$	$\in$	$\in$	$\in$	$\in$
$R_{\forall\forall}$	$\notin$	$\notin$	$\notin$	$\notin$	$\notin$	$\notin$

As discussed in the second paragraph after the proof of Proposition 5, from the fact that $(J_1, J_2) \notin R_{\forall\forall}$ one can also conclude that job J_1 is not $\mathcal{A}|\mathcal{M}$-dominant in the sense of Definition 7.

Concluding the example, we have seen that the agent under consideration should most likely decide for job J_1, since it is the only offer which is $\mathcal{A}$-admissible in the sense of Definition 7. Moreover, J_1 is the unique optimal offer with respect to the criteria based on generalized expectation intervals as introduced in Definition 6. In contrast, when preferring pairwise comparison of the different job offers with respect to the binary relations $R_{\exists\exists}$ and $R_{\forall\forall}$, no clear decision can be made: While every job offer J_i dominates any other job offer J_j, where $i, j \in \{1, 2, 3\}$, with respect to $R_{\exists\exists}$, none of the job offers are comparable with respect to $R_{\forall\forall}$.

[7] Note that, since the optimal objective value of program (12) equals $0.\overline{037}$, it makes no sense to consider values of δ any greater than that.

6. Summary and outlook

In this paper, we proposed three approaches for decision making under severe uncertainty if the acts under consideration take values in some preference system, i.e. can be understood as partial cardinal and partial ordinal valued. Our first approach is based on comparing granularity-dependent expectation intervals. Specifically, we proposed three decision criteria based on these intervals that are direct generalizations of the decision criteria known from the theory of imprecise probabilities. The other two approaches for decision making discussed in the paper rely on local and global comparisons of specific compatible expectations of the considered acts, respectively: For the former approach one searches for compatible pairs (u, π) of utility and probability representations with respect to which the act X of interest simultaneously dominates all the other available acts $Y \in \mathcal{G}$ in expectation. For the latter approach, it suffices if for each other available act $Y \in \mathcal{G}$ there exists a pair (u_Y, π_Y) such that X dominates Y with respect to this specific pair. At several points, we discussed how special cases of our criteria relate to concepts from the classical theory like for instance stochastic dominance or the criteria from decision theory using imprecise probabilities. For certain decision criteria proposed in the paper, we moreover provided linear programming algorithms to evaluate them. Finally, we illustrated a class of situations where our framework appears natural by means of a prototypical toy examples.

There are, of course, several challenges that could be addressed in future research. Clearly, further algorithms for evaluating the remaining criteria that were proposed in the paper need to be explored in order to make the theory computationally more tractable and, therefore, applicable in practice (compare, in particular, the discussion directly following the proof of Proposition 6 of Section 4.4). Further, it is certainly worth investigating in more detail how the criteria from the different approaches relate to each other and what can be learned about them by considering special cases of imprecise probabilistic models. Finally, designing experimental settings for eliciting the parameter δ could help to receive a more canonical interpretation of granularity (compare, in particular, the discussions directly following Definition 3 of Section 3 as well as the discussions at the end of Section 4.2).

Acknowledgements

We are most grateful to the anonymous referees, from whose valuable and insightful comments as well as hints to further relevant references the paper benefited very much. We also thank the referees of the ISIPTA'17 contribution underlying this paper (see [26]) and participants of Progic 2017 and ISIPTA'17 for helpful comments. Further, we want to thank Jean Baccelli for ongoing stimulating discussions on the topic.

References

[1] T. Augustin, On decision making under ambiguous prior and sampling information, in: G. de Cooman, T. Fine, S. Moral, T. Seidenfeld (Eds.), Proceedings of the Second International Symposium on Imprecise Probability: Theories and Applications, Shaker, 2001, pp. 9–16.
[2] T. Augustin, G. Walter, F. Coolen, Statistical inference, in: T. Augustin, Frank Coolen, G. de Cooman, Matthias Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley, Chichester, 2014, pp. 143–189.
[3] T. Bewley, Knightian decision theory. Part I, Decis. Econ. Finance 25 (2002) 79–110.
[4] S. Bradley, How to choose among choice functions, in: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur (Eds.), Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications, Aracne, 2015, pp. 57–66.
[5] A. Chateauneuf, J.-Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Math. Soc. Sci. 17 (3) (1989) 263–283.
[6] I. Couso, S. Destercke, Ranking of fuzzy intervals seen through the imprecise probabilistic lens, Fuzzy Sets Syst. 278 (2015) 20–39.
[7] I. Couso, D. Dubois, An imprecise probability approach to joint extensions of stochastic and interval orderings, Commun. Comput. Inf. Sci. 299 (2012) 388–399.
[8] M. Danielson, Generalized evaluation in decision analysis, Eur. J. Oper. Res. 162 (2005) 442–449.
[9] M. Danielson, L. Ekenberg, A framework for analyzing decision under risk, Eur. J. Oper. Res. 104 (1998) 474–484.
[10] M. Danielson, L. Ekenberg, J. Johansson, A. Larsson, The DecideIT decision tool, in: J.-M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications, Carleton Scientific, 2003, pp. 204–217.
[11] L. de Campos, J. Huete, S. Moral, Probability interval: a tool for uncertain reasoning, Int. J. Uncertain. Fuzziness 2 (1994) 167–196.
[12] D. Denneberg, Non-additive Measure and Integral, Kluwer Academic Publishers, Dordrecht, Boston and London, 1994.
[13] T. Denoeux, Extending stochastic ordering to belief functions on the real line, Inf. Sci. 179 (2009) 1362–1376.
[14] J. Derks, H. Haller, H. Peters, The selectope for cooperative games, Int. J. Game Theory 29 (1) (2000) 23–38.
[15] J. Derks, J. Kuipers, On the number of extreme points of the core of a transferable utility game, in: P. Borm, H. Peters (Eds.), Chapters in Game Theory in Honour of Stef Tijs, Kluwer, 2002, pp. 83–97.
[16] J. Etner, M. Jeleva, J.-M. Tallon, Decision theory under ambiguity, J. Econ. Surv. 26 (2012) 234–270.
[17] P. Fishburn, Utility Theory for Decision Making, Wiley, London and New York, 1970.
[18] T. Galaabaatar, E. Karni, Subjective expected utility with incomplete preferences, Econometrica 81 (2013) 255–284.
[19] I. Gilboa, Theory of Decision Under Uncertainty, Cambridge University Press, New York, 2009.
[20] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique prior, J. Math. Econ. 18 (1989) 141–153.
[21] R. Hable, M. Troffaes, Computation, in: T. Augustin, F. Coolen, G. de Cooman, M. Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley, Chichester, 2014, pp. 329–337.
[22] N. Huntley, R. Hable, M. Troffaes, Decision making, in: T. Augustin, Frank Coolen, G. de Cooman, Matthias Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley, Chichester, 2014, pp. 190–206.
[23] J.-Y. Jaffray, Rational decision making with imprecise probabilities, in: G. de Cooman, F. Cozman, S. Moral, P. Walley (Eds.), Proceedings of the First International Symposium on Imprecise Probability: Theories and Applications, IPP, 1999, pp. 183–188.
[24] J.-Y. Jaffray, M. Jeleva, Information processing under imprecise risk with an insurance demand illustration, Int. J. Approx. Reason. 49 (2008) 117–129.

[25] C. Jansen, T. Augustin, G. Schollmeyer, Decision theory meets linear optimization beyond computation, in: A. Antonucci, L. Cholvy, O. Papini (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2017, in: Lecture Notes in Computer Science, vol. 10369, Springer, Cham, 2017, pp. 329–339.
[26] C. Jansen, G. Schollmeyer, T. Augustin, Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences, in: A. Antonucci, G. Corani, I. Couso, S. Destercke (Eds.), Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications, in: Proceedings of Machine Learning Research, vol. 62, PMLR, 2017, pp. 181–192.
[27] T. Kamae, U. Krengel, G. O'Brien, Stochastic inequalities on partially ordered spaces, Ann. Probab. (1977) 899–912.
[28] D. Kikuti, F. Cozman, R. Filho, Sequential decision making with partially ordered preferences, Artif. Intell. 175 (2011) 1346–1365.
[29] E. Kofler, Prognosen und Stabilität bei unvollständiger Information, Campus, Frankfurt, 1989.
[30] E. Kofler, G. Menges, Entscheiden bei unvollständiger Information, Springer, Berlin, 1976.
[31] D. Krantz, R. Luce, P. Suppes, A. Tversky, Foundations of Measurement. Volume I: Additive and Polynomial Representations, Academic Press, San Diego and London, 1971.
[32] J. Landes, Min–max decision rules for choice under complete uncertainty: axiomatic characterizations for preferences over utility intervals, Int. J. Approx. Reason. 55 (2014) 1301–1317.
[33] E. Lehmann, Ordered families of distributions, Ann. Math. Stat. 26 (1955) 399–419.
[34] I. Levi, On indeterminate probabilities, J. Philos. 71 (1974) 391–418.
[35] I. Levi, The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance, MIT Press, Cambridge, Mass., 1983.
[36] K. Loquin, Significance of a decision making problem under uncertainty, in: F. Cozman, T. Denoeux, S. Destercke, T. Seidenfeld (Eds.), Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, 2013, pp. 203–212.
[37] R. Luce, Semiorders and a theory of utility discrimination, Econometrica 24 (1956) 178–191.
[38] E. Miranda, S. Destercke, Extreme points of the credal sets generated by comparative probabilities, J. Math. Psychol. 64 (2015) 44–57.
[39] I. Montes, Comparison of Alternatives Under Uncertainty and Imprecision, PhD thesis, Oviedo, 2014, digibuo.uniovi.es/dspace/bitstream/10651/28953/6/TD_IgnacioMontesGutierrez.pdf.
[40] I. Montes, S. Destercke, Extreme points of p-boxes and belief functions, Ann. Math. Artif. Intell. 81 (2017) 405–428.
[41] I. Montes, E. Miranda, S. Destercke, A study of the pari-mutuel model from the point of view of imprecise probabilities, in: A. Antonucci, G. Corani, I. Couso, S. Destercke (Eds.), Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications, in: Proceedings of Machine Learning Research, vol. 62, PMLR, 2017, pp. 229–240.
[42] I. Montes, E. Miranda, S. Montes, Decision making with imprecise probabilities and utilities by means of statistical preference and stochastic dominance, Eur. J. Oper. Res. 234 (2014) 209–220.
[43] I. Montes, E. Miranda, S. Montes, Stochastic dominance with imprecise information, Comput. Stat. Data Anal. 71 (2014) 867–885.
[44] K. Mosler, M. Scarsini, Some theory of stochastic dominance, in: K. Mosler, M. Scarsini (Eds.), Stochastic Orders and Decision Under Risk, Institute of Mathematical Statistics, Hayward, CA, 1991, pp. 203–212.
[45] R. Nau, The shape of incomplete preferences, Ann. Stat. 34 (2006) 2430–2448.
[46] M. Pivato, Multiutility representations for incomplete difference preorders, Math. Soc. Sci. 66 (2013) 196–220.
[47] T. Range, L. Osterdal, Checking Bivariate First Order Dominance, Technical report, Discussion Papers on Business and Economics, 2013, Available under, http://static.sdu.dk/mediafiles/0/C/2/%7B0C290F51-8643-421F-A9EF-F38D17A7F570%7Ddpbe9_2013.pdf, 2013.
[48] M. Schervish, T. Seidenfeld, J. Kadane, I. Levi, Extensions of expected utility theory and some limitations of pairwise comparisons, in: J.-M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), Proceedings of the Third International Symposium on Imprecise Probability: Theories and Applications, Carleton Scientific, 2003, pp. 496–510.
[49] 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 (Eds.), Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications, Aracne, 2015, pp. 277–286.
[50] G. Schollmeyer, C. Jansen, 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, available under https://epub.ub.uni-muenchen.de/40416/.
[51] T. Seidenfeld, J. Kadane, M. Schervish, A representation of partially ordered preferences, Ann. Stat. 23 (1995) 2168–2217.
[52] L. Shapley, Cores of convex games, Int. J. Game Theory 1 (1971) 11–26.
[53] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, Int. J. Approx. Reason. 38 (2005) 133–147.
[54] F. Tarp, L. Osterdal, Multivariate Discrete First Order Stochastic Dominance, Discussion Papers 07-23, University of Copenhagen, Department of Economics, 2007, available under http://www.economics.ku.dk/research/publications/wp/2007/0723.pdf.
[55] M. Troffaes, Decision making under uncertainty using imprecise probabilities, Int. J. Approx. Reason. 45 (2007) 17–29.
[56] M. Troffaes, U. Sahlin, Imprecise swing weighting for multi-attribute utility elicitation based on partial preferences, in: A. Antonucci, G. Corani, I. Couso, S. Destercke (Eds.), Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications, in: Proceedings of Machine Learning Research, vol. 62, PMLR, 2017, pp. 333–345.
[57] L. Utkin, T. Augustin, Powerful algorithms for decision making under partial prior information and general ambiguity attitudes, in: F. Cozman, R. Nau, T. Seidenfeld (Eds.), Proceedings of the Fourth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, 2005, pp. 349–358.
[58] L. Utkin, T. Augustin, Decision making under incomplete data using the imprecise Dirichlet model, Int. J. Approx. Reason. 44 (2007) 322–338.
[59] A. Wald, Statistical decision functions, Ann. Math. Stat. 20 (1949) 165–205.
[60] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[61] P. Walley, T. Fine, Towards a frequentist theory of upper and lower probability, Ann. Stat. 10 (1982) 741–761.
[62] A. Wallner, Extreme points of coherent probabilities on finite spaces, Int. J. Approx. Reason. 44 (2007) 339–357.
[63] K. Weichselberger, S. Pöhlmann, A Methodology for Uncertainty in Knowledge-Based Systems, Springer, Heidelberg, 1990.

Contribution 4:

Jansen, C.; Schollmeyer, G.; Augustin, T. (2018c): A probabilistic evaluation framework for preference aggregation reflecting group homogeneity. Revision under review for *Mathematical Social Sciences*.

As the paper is still under review, the following technical report that is identical to the submitted revision is included in Chapter 3 of this thesis instead:

Jansen, C.; Schollmeyer, G.; Augustin, T.. (2016/2018): A Probabilistic Evaluation Framework for Preference Aggregation Reflecting Group Homogeneity. Department of Statistics (LMU München): *Technical Reports*, Nr. 193, last updated March 2018.

Available under:

`https://epub.ub.uni-muenchen.de/42751/`

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK

Christoph Jansen, Georg Schollmeyer and Thomas Augustin

A Probabilistic Evaluation Framework for Preference Aggregation Reflecting Group Homogeneity

Technical Report Number 193, 2018
Department of Statistics
University of Munich

http://www.statistik.uni-muenchen.de

A Probabilistic Evaluation Framework for Preference Aggregation Reflecting Group Homogeneity

C. Jansen[a,*], G. Schollmeyer[a], T. Augustin[a]

[a] *Department of Statistics, LMU Munich, Ludwigsstr. 33, 80539 Munich, Germany*

Abstract

Groups differ in the homogeneity of their members' preferences. Reflecting this, we propose a probabilistic criterion for evaluating and comparing the adequateness of preference aggregation procedures that takes into account information on the considered group's homogeneity structure. Further, we discuss two approaches for approximating our criterion if information is only imperfectly given and show how to estimate these approximations from data. As a preparation, we elaborate some general minimal requirements for measuring homogeneity and discuss a specific proposal for a homogeneity measure. Finally, we investigate our framework by comparing aggregation rules in a simulation study.

Keywords: Aggregation procedure, preference profile, voting theory, imprecise probabilities, maximum entropy, homogeneity measure, group decision making. JEL classification: C1, C6

1. Introduction

One of the fundamental tasks in social choice theory is to define adequately justified rules for aggregating the preferences of a group of individuals into one global consensus order. Due to the generality of this problem, it is hardly surprising that many different rules have been proposed since the pioneering works by de Borda (1781); de Condorcet (1785); Hare (1857) (see Brams and Fishburn (2002) for a survey). More generally, the question of aggregating collections of binary relations in a meaningful way does not exclusively concern social choice theory, but also appears in classification problems in statistics (see, e.g., Maniqueta and Mongin (2016)), benchmarking of algorithms in the computer sciences (see, e.g., Mersmann et al. (2015)) or problems of judgment aggregation in philosophy (see, e.g., Hartmann and Sprenger (2012)) to name only a few examples.

Given the diversity of aggregation rules, criteria for evaluating and comparing their quality need to be established. Many different criteria have been proposed, and comparisons of aggregation rules with respect to them have been studied intensively (see, e.g., Grofman and Feld (2004)). However, almost all these criteria are non-group-specific: They are intended to be valid independently of the group whose members' preferences are to be aggregated. But what is a perfectly adequate aggregation procedure for one group may not be as appropriate for another one. The adequateness of an aggregation procedure may, beyond compatibility with non-group-specific criteria, additionally

*Corresponding author
Email address: christoph.jansen@stat.uni-muenchen.de (C. Jansen)

depend on certain characteristics of the specific group under consideration. One such characteristic is the homogeneity of the group members (see Section 2.2 for a discussion of the literature on homogeneity). In this paper we propose a group-specific quality criterion for aggregation rules that takes into account information on the homogeneity of group members' preference structure. Moreover, we show different ways to approximate our criterion under partial probabilistic information and discuss how to estimate these approximations in the presence of data or expert knowledge.

More precisely, the paper is structured as follows: In Section 2, we discuss measures for quantifying the homogeneity of a group that is represented by a fixed profile $(R_1, \ldots, R_n)$ of strict weak orders. Specifically, in Section 2.3, we elaborate a list of three minimal requirements that every reasonable measure should satisfy. In Section 2.4, we then propose a concrete measure, the maximum consensus homogeneity, and discuss why it is reasonable beyond its mere compatibility with these minimal requirements. Section 3, after reviewing some basics on Bayesian theory in Section 3.1, introduces a framework for evaluating and comparing aggregation procedures in the presence of probabilistic information on the considered group. This involves three steps: In Section 3.2, we introduce an optimality criterion that requires perfect knowledge of the probabilities with respect to which the group constitutes different profiles $(R_1, \ldots, R_n)$. Section 3.3 discusses approaches for approximating this criterion if the probabilistic information on the group is partial in the sense that only the probability distribution of some homogeneity measure is given. Finally, Section 3.4 discusses several statistical approaches for estimating this distribution in the presence of data, expert knowledge, or both. Section 4 starts by briefly reviewing some common aggregation procedures relevant to our context (Section 4.1). Afterwards, Section 4.2 summarizes an aggregation procedure recently proposed in Schollmeyer (2017).[1] In Section 5, we investigate the aggregation procedures reviewed, in respect to our criterion in a simulation study. Section 6 is reserved for concluding remarks as well as an outlook on future research questions.

2. Measuring Homogeneity of Preference Profiles

We begin the section by introducing our notation and terminology (Section 2.1) and surveying some related work on the topic (Section 2.2). Subsequently, we establish and discuss a weak set of conditions (Section 2.3) as well as a concrete proposal (Section 2.4) for measuring the homogeneity of a fixed collection $(R_1, \ldots, R_n)$ of strict weak orders each of which representing the opinion of a member of a group of size n.

2.1. Notation and Terminology

Throughout the paper, C denotes a finite set of at least two consequences. The elements of C have to be ranked by the members of a specific group G_n of fixed size $n \geq 2$, where certain requirements of rationality regarding the individual orders involved are imposed. Specifically, we work with the following spaces of binary relations on C:

$$\mathcal{R} := \{R \subset C^2 : R \text{ asymmetric, negatively transitive}\} \tag{1}$$

$$\mathcal{Q} := \{Q \subset C^2 : Q \text{ asymmetric}\} \tag{2}$$

In the sequel, every $R \in \mathcal{R}$ is termed a strict weak order on C. For every $R \in \mathcal{R}$, define the usual equivalence relation $\sim_R$ on C by setting $a \sim_R b$ if and only if $(a, b) \notin R \wedge (b, a) \notin R$. Given this,

[1] For an explanation of the procedure and a discussion see Section 4.2.

interpret $(a, b) \in R$ as a is strictly preferred to b and $(a, b) \in \sim_R$ as indifference between a and b. The elements of $\mathcal{R}$ are associated with the individual orders of the group members. Hence, the group members are assumed to have asymmetric and negatively transitive preferences. Importantly, note that our model of the individual preferences excludes incomparability of consequences: For alternatives $a, b \in C$ chosen arbitrarily, every group member is thus assumed to be able to decide if she strictly prefers a to b, or b to a, or if she ranks them equally desirable. Thus, we explicitly assume that incomparability with respect to $R \in \mathcal{R}$ is interpreted as indifference (see, e.g., Kreps (1988, Chapter 2) for a discussion of this convention).[2] For $n \geq 2$, an element $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$ is called a preference profile on C and each component of $\underline{R}$ is interpreted as the opinion of a member of G_n about how the consequences in C should be ranked.

Contrarily, every element $Q \in \mathcal{Q}$ is called a consensus order (or group preference). Except for asymmetry, we do not impose any further restrictions on the consensus order. This allows for also investigating aggregation procedures for which the group preference is not always as well-behaved as the individual orders (this includes, e.g., Condorcet's method, see Section 4.1, which might yield intransitive consensus orders). In this context, every mapping $S : \mathcal{R}^n \to \mathcal{Q}$ is called a preference aggregation function. Particularly, for every preference profile $\underline{R} \in \mathcal{R}^n$, the image $S(\underline{R}) \in \mathcal{Q}$ is the consensus order of the group represented by $\underline{R}$ with respect to the aggregation procedure described by S.

2.2. Preference Homogeneity in Related Work

In literature on social choice theory at least two different lines of how to establish a notion of homogeneity of groups can be identified. One line (see, e.g., Niemi (1969); Jamison and Luce (1972); Berg (1985); Gehrlein and Lepelley (2010); Lepelley and Valognes (2003)), which could be called "model-based", builds up stochastic models that govern the constitution of profiles and have specific parameters implicitly regulating the group's homogeneity. One prominent example is the multivariate Pólya-Eggenberger urn model (see, e.g., Johnson and Kotz (1977)), which has been used for instance in Berg (1985); Gehrlein and Lepelley (2010); Lepelley and Valognes (2003) in order to analyze the relationship between group homogeneity and the probability of the voting paradox or the manipulability of different aggregation functions. The Pólya-Eggenberger model contains two other well-established models as special cases: impartial culture and impartial anonymous culture, which are also often presumed in studies of the voting paradox and the manipulability of aggregation procedures (see, e.g., Aleskerov et al. (2012); Diss et al. (2012); Pritchard and Slinko (2006)). Other model-based approaches, in which the orders in the profile are assumed to be randomly drawn with replacement, measure the homogeneity of the generating process by the probabilities p_i $(i = 1, \ldots, |C|!)$ with respect to which the order R_i is drawn: Natural measures of homogeneity are then the variance of the p_is used for instance in Abrams (1976) or the Herfindahl index $\sum_{i=1}^{|C|!} p_i^2$ used, for instance, in Gehrlein (1981). Measures that only rely on the values of the p_i's and not on the concrete associated orders R_i are called non-profile specific measures (see Gehrlein (1981)). Since they are related to the probabilities p_i, they are also called population specific homogeneity measures in Gehrlein and Lepelley (2010, p. 191).

[2] An alternative approach would be to directly model the individual preferences by weak orders, i.e. complete and transitive binary relations $P \subset C^2$. To every such relation we then can associate its strict part $R_P \subset C^2$ by setting $(a, b) \in R_P$ if and only if $(a, b) \in P \wedge (b, a) \notin P$ for all $a, b \in C$. The relation R_P is then asymmetric and negatively transitive. Our model thus explicitly assumes that the individual orders $R \in \mathcal{R}$ arise as strict parts of a weak order.

A second line of establishing a notion of homogeneity, which can be called "data-based", relates homogeneity not to a probabilistic model but to the actually observed data in a profile. For example, in the above approaches, one can replace the probability p_i of observing the order R_i in a profile with the relative frequency of the associated order in the actually observed profile. Then one arrives at a notion of homogeneity that is no related to a generating process, but instead related to the observed profile. Such measures are called situation specific homogeneity measures in Gehrlein and Lepelley (2010, p. 192). A further type of such data-based measures are distance-based measures, which additionally utilize the information in the orders of the profile. These measures, arising not only in social choice theory but also in statistics and computer sciences (see, e.g., Fligner and Verducci (1986); Dwork et al. (2001)), rather rely on a geometric understanding and first introduce a distance between pairs of orders. Based on this distance, one defines a measure of heterogeneity by computing the average distance of all pairs of orders in the profile. Homogeneity of the profile is then measured by comparing the maximal distance to this average distance. This type of measures is local in the sense that not the whole group is examined simultaneously, as only pairs are considered. Another data-based measure of homogeneity, especially used in social choice theory (see, e.g., Fishburn (1973)), is the W coefficient introduced in Kendall and Smith (1939). This measure intends to analyze the whole population simultaneously by looking at the variance of the vector of the summarized ranks of each consequence. However, note that also this measure, as shown by Kendall and Smith (1939), could be alternatively represented as the average Spearman correlation coefficient of pairs of rank-vectors and is thus also local in the above sense.

Beyond concrete proposals for data-based homogeneity measures, axiomatic approaches have also been studied (see, e.g., Bosch (2006); Alcalde-Unzu and Vorsatz (2013)). Here, the terms consensus and cohesiveness are used instead of homogeneity. For the concept of polarization, a concept very similar, but not identical to the concept of heterogeneity,[3] an axiomatic characterization of a measure of polarization of profiles is given in Can et al. (2015).

2.3. Minimal Requirements for Measuring Homogeneity

Before introducing a concrete non-local and profile-specific homogeneity measure in the next section, we first set out to agree on some minimal requirements that, in our eyes, every reasonable candidate for such a measure should necessarily satisfy. We list these requirements in the following definition. Afterwards, a discussion of each is given.

Definition 1. *A preference homogeneity measure (for a group of size n) is a map $A_n : \mathcal{R}^n \to [0,1]$ satisfying the following three properties:*

(S1) ***Consensus sensitivity****: $A_n(\underline{R}) = 1$ if and only if $\underline{R} = (R^*, \dots, R^*)$ for some $R^* \in \mathcal{R}$.*

(S2) ***Anonymity****: Let $\phi : \{1, \dots, n\} \to \{1, \dots, n\}$ be a bijective map. Then $A_n(R_1, \dots, R_n) = A_n(R_{\phi(1)}, \dots, R_{\phi(n)})$ for all $(R_1, \dots, R_n) \in \mathcal{R}^n$.*

(S3) ***Majority strengthening****: Let $\underline{R} \in \mathcal{R}^n$. Define $k(j) := \{i : R_i = R_j\}$. If there exists $j_0 \in \{1, \dots, n\}$ such that $n > |k(j_0)| \geq \lfloor \frac{n}{2} \rfloor$, choose $j_1 \in \{1, \dots, n\} \setminus k(j_0)$ and define*

[3] While the notion of heterogeneity refers here to the diversity of the orders in the profile, polarization means that the orders in the profile are clustered in two or more "opposite" subgroups. A clear cut rigorous disambiguation between polarization and heterogeneity for the case of preference profiles is, as far as the authors are aware, not yet established. For a more elaborate disambiguation between polarization and heterogeneity/inequality in the context of, for instance, poverty measurement, see, e.g., Esteban and Ray (1994); Duclos et al. (2004).

$\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *by* $\phi(j) = j_0$*, if* $j \in k(j_0) \cup \{j_1\}$ *and* $\phi(j) = j$ *else. Then we have* $A_n(R_1, \ldots, R_n) \leq A_n(R_{\phi(1)}, \ldots, R_{\phi(n)})$.

So, why is it reasonable to require (S1), (S2), and (S3) as a minimal basis for measuring homogeneity of preference profiles? Consensus sensitivity states that every measure should be capable of identifying perfect consensus by attaining its maximal value 1 if and only if all group members share identical preferences. This is certainly reasonable since it reflects the fact that we know exactly that the identical profiles are ideal and superior to the non-identical ones with respect to homogeneity. This knowledge should not get lost by the construction of the measure. Anonymity ensures that the homogeneity value of a profile does not depend on the order in which the individuals state their preferences, as no individual has greater influence. Finally, majority strengthening can be interpreted as a weak demand for monotonicity: If a subgroup consisting of at least $\lfloor \frac{n}{2} \rfloor$ group members shares identical preferences and one member from outside this subgroup changes her mind towards this subgroup, then the homogeneity value of the modified profile should not decrease.

Clearly, all three conditions rely solely on the categorical and not the ordinal scale of measurement of the orders in the profile, i.e. one conceptually only distinguishes between equal and non-equal orders and does not make use of for example a notion of how similar different orders are by e.g. counting edges that two different orders have in common. Of course, one could also establish a notion of (S3) that uses the ordinal structure by stating for instance that if one order R in the profile is changed towards another order R' that is more similar to the order of the majority, then the homogeneity should not decrease. However, this would require a notion of what the terms "majority" and "more similar order" then exactly mean. Note further that adequately axiomatizing more subtle aspects like the difference between heterogeneity and polarization seems to be not possible if one only relies on the categorical scale of measurement of the orders in the profile. In this sense, the conditions (S1), (S2), and (S3) should indeed be understood as minimal requirements for a notion of homogeneity that leaves much space for content matter considerations in the final choice of the measure.[4]

2.4. The Maximum Consensus Homogeneity

We now introduce a specific homogeneity measure, the so-called maximum consensus homogeneity, show that this measure satisfies the minimal requirements given in Definition 1, and discuss why it is a reasonable choice for our purposes beyond its mere compatibility with the minimal requirements. The basic idea of the measure is to compare, for each pair (a, b) separately, the maximal number of coinciding opinions about that pair in the profile to the maximal possible number n.[5]

Some additional notation is needed: Let $n \geq 2$ and let $R_0 \in \mathcal{R}$ with $\sim_{R_0} = \{(c, c) : c \in C\}$ be fixed, such that R_0 always contains exactly one of the pairs (a, b) or (b, a) for all distinct $a, b, \in C$. For a fixed preference profile $\underline{R} \in \mathcal{R}^n$ and a fixed pair of distinct consequences $(a, b) \in C^2$, we define the expressions $c_{\underline{R}}(a, b) := |\{i : (a, b) \in R_i\}|$ and $e_{\underline{R}}(a, b) := |\{i : (a, b) \in \sim_{R_i}\}|$ to be, respectively, the number of individuals in $\underline{R}$ that prefer a to b and the number of individuals that are indifferent between these options.

[4] An (in parts) similar axiomatization, however stronger, is given in Alcalde-Unzu and Vorsatz (2013) in the context of measuring cohesiveness of preferences profiles.

[5] A similar measure is introduced in Can et al. (2015): There, the authors first list a set of axioms for measures of *polarization* that uniquely characterize a measure that is closely related to the one used in this work.

Definition 2. *The mapping* $\delta_n : \mathcal{R}^n \to [0,1]$ *defined by*

$$\delta_n(\underline{R}) := \frac{\sum_{(a,b)\in R_0} \max\{c_{\underline{R}}(a,b), c_{\underline{R}}(b,a), e_{\underline{R}}(a,b)\}}{n \cdot \binom{|C|}{2}} \tag{3}$$

for all $\underline{R} \in \mathcal{R}^n$ *is called maximum consensus homogeneity.*

Importantly, note that the definition of δ_n does not depend on the choice of $R_0 \in \mathcal{R}$ with the desired properties (see Appendix A1). As a first step in our discussion of the proposed maximum consensus measure δ_n, we show that it does indeed satisfy the conditions (S1), (S2), and (S3). Therefore, we consider it compatible with the minimal requirements a measure of homogeneity should satisfy. This is the assertion made in the following proposition. The proof consists in straightforwardly verifying (S1), (S2), and (S3) from Definition 1 and is given in Appendix A1.

Proposition 1. *The maximum consensus homogeneity* δ_n *satisfies (S1), (S2), and (S3).*

So, why is the maximum consensus measure δ_n a reasonable candidate for measuring homogeneity and what makes it preferable to the measures discussed in Section 2.2 for our purposes? First, δ_n utilizes the information encoded in the orders collected in the inserted profiles and does not solely rely on the shares p_i of identical orders. Accordingly, δ_n is profile-specific (in contrast to, e.g., Herfidahl's index). This certainly is a desirable property, since any measure of homogeneity should be capable of distinguishing between profiles of very similar yet not identical orders and profiles of completely opposed orders. Second, δ_n is not a local measure in the sense of being only based on pairwise comparisons of individual orders (see the discussion in Section 2.2): For computing the value $\delta_n(\underline{R})$ the whole profile needs to be examined simultaneously. This is a very desirable property conceptually, since group homogeneity should depend on the group as a whole rather than on comparisons of pairs of individuals only. Note that, in addition to this argument, distance-based homogeneity measures also satisfy the minimal conditions (S1) to (S3) from Definition 1.[6] Finally, note that the classical measure W of Kendall and Smith mentioned in Section 2.2 does not satisfy majority strengthening: For a counterexample, take $C = \{a, b, c, d, e\}$ and consider the profiles $\underline{R} = (R_1, R_1, R_1, R_2, R_3)$ and $\underline{R'} = (R_1, R_1, R_1, R_1, R_3)$ where relation R_1 ranks $a\ b\ c\ d\ e$, relation R_2 ranks $a\ b\ c\ e\ d$ and relation R_3 ranks $e\ b\ c\ a\ d$. Clearly, the majority strengthening condition (S3) requires assigning higher homogeneity to the profile $\underline{R}'$, but simple calculations yield $W(\underline{R'}) = 0.584 < 0.592 = W(\underline{R})$.

3. A Probabilistic Evaluation Framework for Preference Aggregation Functions

Section 3.1 recalls required concepts from Bayesian statistics (see, e.g., Berger (1980); Gelman et al. (2004) for monographs). Subsequently, based on the concept of preference homogeneity measures from the previous section, we propose a probabilistic criterion for evaluating the adequateness of a preference aggregation function S for a fixed group G_n of size n. Specifically, this will involve three steps: Firstly, in Section 3.2, we introduce a theoretical criterion $m^u_{G_n}(S)$ that measures the expected similarity that S yields given the true probability measure $\mathbb{P}_{G_n}$ with respect to which the group constitutes different profiles $\underline{R} \in \mathcal{R}^n$. Secondly, reflecting the fact that in reality the

[6] A formal justification of the non-locality and a discussion of distance-based homogeneity measures in the light of Definition 1 are given in Appendix 3.

measure $\mathbb{P}_{G_n}$ will typically be unknown, Section 3.3 shows how to construct approximations for it if only the distribution of some homogeneity measure is available instead. Finally, in Section 3.4, we discuss different methods for estimating the distribution of a homogeneity measure.

3.1. Required Concepts of Bayesian Statistics

Roughly stated, Bayesian theory addresses two fundamental questions: (Q1) *How to model an agent's beliefs in the light of uncertainty* and (Q2) *How to update the model once new information is gained.* In classical Bayesian theory as pioneered by de Finetti's concept of subjective probability (see, in particular, de Finetti (1974)), Question (Q1) is addressed by the assumption that an agent's beliefs/information about any uncertain phenomenon (independent of whether that phenomenon is random or not) can be perfectly characterized by a unique subjective probability measure π on the space of potential outcomes of the phenomenon. All reasoning should then be based on this unique probability measure π.

However, apart from classic Bayesian theory, this uniqueness is often strongly doubted for being too demanding in regard to the consistency of the agent's beliefs. Instead, beliefs are assumed to be only partial in that they specify a whole set $\mathcal{M}$ of probability measures compatible with them. Then, two main approaches are followed: The first one establishes criteria for choosing one particular distribution from the set $\mathcal{M}$ and, subsequently, bases all further analyses on the chosen representative. The most common choice for such a criterion is Jayne's maximum entropy principle (see Jaynes (1957)): Among all measures compatible with the beliefs, choose the one that is least informative and thus best captures the complete ignorance among the compatible measures (see Rosenkrantz (1977, Section 3.5)). Particularly, the informativeness of a distribution used in the definition of the principle is measured by means of Shannon's entropy (see Shannon (1949)). For a more recent justification of the maximum entropy principle see, e.g., Landes and Williamson (2013). The second approach treats the set $\mathcal{M}$ of all compatible distributions, also called credal set[7] in this context, as an entity of its own: The agent's beliefs are represented by all members of $\mathcal{M}$, not just by one single representative. Clearly, an argument supporting this approach is that it avoids any selection: Even a well-justified criterion might select a rather bad representative in certain situations and therefore could yield misleading reasoning. Contrarily, reasoning based solely on the credal set obviously produces less informative results. For a detailed discussion of the advantages and disadvantages of the two directions and a decision-theoretical justification of maximum entropy see Walley (1991, Section 5.12). From a practical point of view, it often makes sense to consider both approaches simultaneously: Use a well-established selection criterion (such as maximum entropy) and analyze the credibility and robustness of the derived inferences by additionally considering the set of inferences drawn by the credal set.

Let us turn now to Question (Q2): Suppose (new) information x about the uncertain phenomenon is gained (e.g. in the form of data). The agent then updates the unique[8] measure π describing her beliefs, also called prior distribution in this context, to a posterior measure $\pi|x$ according to Bayes' rule. This posterior measure $\pi|x$ is then assumed to appropriately express the updated beliefs about the uncertain phenomenon given the data x. In the specific context

[7]The name credal set is attributed to Isaac Levi (see Levi (1980)). For the general framework of imprecise probabilities, working with sets of probabilities or interval-valued assignments, see Walley (1991) and Weichselberger (2001), or, for a recent introduction, see Augustin et al. (2014).

[8]For Question (Q2), for the sake of brevity, we restrict presentation to classic Bayesian theory. For discussions on how to adequately update credal sets in light of new information see, e.g., the discussions in Walley (1991).

of Bayesian statistics as used here, this translates as follows: Suppose some random variable $X : \Omega \to \mathcal{X}$, mapping from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to a measurable space $(\mathcal{X}, \sigma(\mathcal{X}))$, for which we can specify its distribution up to a parameter θ from a parameter space Θ, i.e. we know $X \sim \mathbb{P}_\theta$ given $\theta \in \Theta$ is the true parameter. Following the ideas of Bayesian theory described above, we can then describe the uncertainty about the true parameter θ by a random variable $V : \Omega \to \Theta$ taking values in the measurable space $(\Theta, \sigma(\Theta))$ with prior distribution π, i.e. $V \sim \pi$, and we know that $\mathbb{P}_\theta(A) = \mathbb{P}(X \in A | V = \theta)$ for all $\theta \in \Theta$ and $A \in \sigma(\mathcal{X})$. After having observed a sample x of X, one then computes the posterior distribution $\pi|x$ by setting $\pi|x(B) := \mathbb{P}(V \in B | X = x)$ for all $B \in \sigma(\Theta)$ and utilizing Bayes' rule.

If an estimate for the parameter rather than the posterior measure itself is of interest, popular choices are to use the expectation, the median or the mode of the posterior distribution. For sake of computational convenience, so-called conjugate families of distributions are often used: A family of distributions $\mathcal{D}_1(\Xi)$ with parameter space Ξ is called conjugate to another family $\mathcal{D}_2(\Theta)$ with parameter space Θ if, whenever $\pi \in \mathcal{D}_1(\Xi)$ and $X \sim \mathbb{P}_\theta \in \mathcal{D}_2(\Theta)$, it holds that $\pi|x \in \mathcal{D}_1(\Xi)$, where x is an observation of X. Hence, such models guarantee that the posterior belongs to the same distribution family as the prior and, therefore, that the posterior as well as its moments can basically be computed by updating only the parameter of the prior distribution. An example for such a conjugate model is the Dirichlet-categorical model, which will be used in Section 3.4.

3.2. A Probabilistic Criterion for Evaluating Preference Aggregation Functions

We now turn to the first step of the construction of our criterion. Therefore, for the moment, we assume the probabilities according to which the members of the group G_n constitute the different profiles contained in $\mathcal{R}^n$ are known. More formally, we consider the measurable space $(\mathcal{R}^n, 2^{\mathcal{R}^n})$ together with a known group-specific probability measure $\mathbb{P}_{G_n}$, and we interpret the value $\mathbb{P}_{G_n}(\{\underline{R}\})$ as the probability that the members of G_n constitute the preference profile $\underline{R}$. In order to utilize the probabilistic information (given by $\mathbb{P}_{G_n}$) in the construction of our criterion, we want to compute the similarity of the individual orders collected in $\underline{R}$ and the consensus order $S(\underline{R})$ that an aggregation rule S yields in expectation. However, before such a criterion can be defined, we need to be more precise about what we mean by the similarity of a consensus relation to a profile of relations. Specifically, we will consider similarity measures of the following kind:

Definition 3. *Let $u : \mathbb{R}^+ \to \mathbb{R}^+$ be a monotone increasing function and let $S : \mathcal{R}^n \to \mathcal{Q}$ be a preference aggregation function. The mapping*

$$Y_S^u : \mathcal{R}^n \to \mathbb{R} \quad , \quad \underline{R} \mapsto \sum_{i=1}^{n} u(|R_i \cap S(\underline{R})|) \tag{4}$$

is called the similarity measure for S with respect to u.

The basic idea underlying a similarity measure in the above sense is to quantify similarity of pairs of relations R_1 and R_2 by computing the cardinality of their intersection $|R_1 \cap R_2|$ or, in other words, by counting the edges shared by both relations. An axiomatic justification is given in Kemeny and Snell (1962), were the authors show that the distance measure $d(R_1, R_2) := |R_1 \triangle R_2| = |(R_1 \cup R_2) \setminus (R_1 \cap R_2)| = |R_1| + |R_2| - 2|R_1 \cap R_2|$ is unique in satisfying four desirable conditions (including the properties of a metric). Together with the assumption that similarity of R_1 and R_2 should be high whenever their uniquely determined distance $d(R_1, R_2)$ is low and vice versa, using $|R_1 \cap R_2| = \frac{1}{2}(|R_1| + |R_2| - d(R_1, R_2)) = \frac{1}{2}(|C|(|C| + 1) - d(R_1, R_2))$ is a natural

choice. For a fixed profile $\underline{R} = (R_1, \ldots, R_n)$, the idea of a similarity measure Y_S^u is then to compute the pairwise similarity $|R_i \cap S(\underline{R})|$ between every individual order R_i and the group order $S(\underline{R})$ separately and, afterwards, sum up monotone transformations $u(|R_i \cap S(\underline{R})|)$ of these values. The role of u is to control the influence of high similarity values $|R_{i_0} \cap S(\underline{R})|$ for certain orders R_{i_0} on the global similarity $Y_S^u(\underline{R})$: If u is chosen to be a convex function, high pairwise similarity values will have a strong influence on the global similarity, whereas if u is concave, then increasing the similarity of an inadequately represented group member contributes more to global similarity than doing the same for a group member that is already appropriately represented.

Once having decided which specific similarity measure Y_S^u to use, one can go on to construct a quality criterion for the aggregation function S: We evaluate S by computing the expectation of the chosen similarity function with respect to the group specific probability $\mathbb{P}_{G_n}$.

Definition 4. *Let G_n be a group consisting of n members and let $\mathbb{P}_{G_n}$ denote its group specific probability measure on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$. For a preference aggregation function S and a monotone increasing function $u : \mathbb{R}^+ \to \mathbb{R}^+$ with associated similarity measure Y_S^u, we define the value*

$$m_{G_n}^u(S) := \mathbb{E}_{\mathbb{P}_{G_n}}(Y_S^u) = \sum_{\underline{R} \in \mathcal{R}^n} Y_S^u(\underline{R}) \cdot \mathbb{P}_{G_n}(\{\underline{R}\}) \tag{5}$$

Then $m_{G_n}^u(S)$ is called the expected similarity of the aggregation function S with respect to Y_S^u.

The criterion $m_{G_n}^u(\cdot)$ is intended to be applied as follows: Given two aggregation functions S_1 and S_2 and a group G_n that agrees to measure similarity by Y_s^u, the group should prefer aggregation rule S_1 whenever $m_{G_n}^u(S_1) \geq m_{G_n}^u(S_2)$, i.e. if S_1 yields higher expected similarity than S_2.

In practice, this criterion will often not be directly applicable, since $\mathbb{P}_{G_n}$ cannot be fully specified. However, in many applications there will be at least some information about the homogeneity structure of the preferences of the group under investigation. In the following Section 3.3, we demonstrate how to construct approximations for the true group-specific measure $\mathbb{P}_{G_n}$ if this information is given in the form of the probability distribution of some homogeneity measure A_n and how these approximations can be utilized for estimating expected similarity.

3.3. Constructing Approximations for Expected Similarity

This leads us to the second step of our construction: Let $A_n : \mathcal{R}^n \to [0,1]$ denote a fixed preference homogeneity measure attaining exactly the values $k_1 < k_2 < \cdots < k_\xi \in [0,1]$. We assume that the available information on the homogeneity of G_n can be specified as the probability distribution of this homogeneity measure A_n. More formally, we (for the moment) assume to know $\alpha := (\alpha_1, \ldots, \alpha_\xi) \in \Delta^{\xi-1} := \{x \in [0,1]^\xi : \sum_{i=1}^{\xi} x_i = 1\}$ such that $\mathbb{P}_{G_n}(A_n = k_j) = \alpha_j$ for all $j = 1 \ldots, \xi$. Substantially, this relates to the assumption that, even if the full group-specific measure $\mathbb{P}_{G_n}$ is unknown, we still know the probabilities α that the group G_n constitutes a certain degree of homogeneity, which is characterized by the chosen preference homogeneity measure A_n. Given this, our goal is to approximate the true underlying group-specific probability measure $\mathbb{P}_{G_n}$ such that the available knowledge on the distribution of A_n is utilized in the best possible way.

To reach this goal, first note that our assumption naturally characterizes a set of probability measures on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$, namely the credal set $\mathcal{M}_\alpha$ of all probability measures that are compatible with the available information on the distribution of A_n. Formally, we have

$$\mathcal{M}_\alpha := \big\{\pi \in \mathcal{P}(\mathcal{R}^n) : \ \pi(A_n^{-1}(k_j)) = \alpha_j \text{ for all } j = 1, \ldots, \xi\big\} \tag{6}$$

where $\mathcal{P}(\mathcal{R}^n)$ denotes the set of all probability measures on the space of profiles $(\mathcal{R}^n, 2^{\mathcal{R}^n})$. Consequently, any element of $\mathcal{M}_\alpha$ is a candidate for the true group specific measure and, therefore, a plausible candidate for approximating it. As discussed in Section 3.1, (at least) two different approaches exist for dealing with the ambiguity between the compatible measures in $\mathcal{M}_\alpha$ in such situations: Applying the maximum entropy principle to $\mathcal{M}_\alpha$ in order to specify the least informative measure or directly working with the set $\mathcal{M}_\alpha$ as an entity of its own. We contrast both approaches and the approximations for expected similarity obtained by them in the sequel:

Maximum entropy approach: Given the distribution of A_n, we know that the probability of the homogeneity class $A_n^{-1}(k_j) := \{\underline{R} \in \mathcal{R}^n : A_n(\underline{R}) = k_j\}$ equals α_j. Contrarily, there is complete ignorance between all measures fixing the probabilities of these classes. Applying the maximum entropy principle, we choose the representative among the compatible measures in $\mathcal{M}_\alpha$ that maximizes Shannon's entropy and that therefore can be viewed as the least informative one. The measure satisfying the desired property is induced by the assignment

$$\mathbb{P}^*_\alpha(\{\underline{R}\}) := \frac{\alpha_{\phi(A_n(\underline{R}))}}{|A_n^{-1}(A_n(\underline{R}))|} \tag{7}$$

for all $\underline{R} \in \mathcal{R}^n$, where $\phi(k_j) := j$ for $j = 1, \ldots, \xi$. Among all measures fixing the probability values of the homogeneity classes $A_n^{-1}(k_j)$ to α_j, the resulting probability $\mathbb{P}^*_\alpha$ is exactly the one giving equal probability mass to all profiles belonging to the same class. Therefore, beyond maximizing entropy, the measure $\mathbb{P}^*_\alpha$ is also intuitively appealing: Why should two profiles with coinciding homogeneity value be assumed to have different probability?

Credal set approach: Directly approximating $\mathbb{P}_{G_n}$ with the credal set $\mathcal{M}_\alpha$ protects against possibly misleading inferences based on an unlucky selection of a representative. Obviously, the set $\mathcal{M}_\alpha$ contains exactly these probability measures that are compatible with the probabilities α_j of the homogeneity classes $A_n^{-1}(k_j)$. By construction, we therefore have $\mathbb{P}_{G_n} \in \mathcal{M}_\alpha$ and $\mathbb{P}^*_\alpha \in \mathcal{M}_\alpha$, i.e. the true measure and the maximum entropy measure are contained in the credal set.

By using the two approaches just described, the expectation in (5) can now straightforwardly be approximated by replacing the true measure $\mathbb{P}_{G_n}$ in the expression with the corresponding approximation $\mathbb{P}^*_\alpha$ or $\mathcal{M}_\alpha$. Note that for the credal set approach this will lead to a set-valued approximation, each element of the set representing the expected similarity with respect to a different distribution from $\mathcal{M}_\alpha$. Formally, this leads to the following two approaches for approximation.

Maximum entropy approximation: Compute the expected similarity with respect to the maximum entropy measure $\mathbb{P}^*_\alpha$. We then arrive at the following real-valued approximation:

$$m^{u*}_{G_n}(S) := \mathbb{E}_{\mathbb{P}^*_\alpha}(Y^u_S) = \sum\nolimits_{\underline{R} \in \mathcal{R}^n} Y^u_S(\underline{R}) \cdot \mathbb{P}^*_\alpha(\{\underline{R}\}) \tag{8}$$

The maximum entropy is represented by a single real number and, therefore, allows for easy comparisons of different aggregation functions S_1 and S_2. However, the maximum entropy approximation $\mathbb{P}^*_\alpha$ might differ from the true underlying measure $\mathbb{P}_{G_n}$ in a way yielding $m^{u*}_{G_n}(S_1) > m^{u*}_{G_n}(S_2)$ but $m^u_{G_n}(S_2) > m^u_{G_n}(S_1)$ and, thus, might produce misleading comparisons.

Credal approximation: Compute the expectation with respect to the set $\mathcal{M}_\alpha$, i.e. the interval ranging from the lowest to the highest expected similarity value compatible with a measure from $\mathcal{M}_\alpha$. We arrive at the following interval-valued approximation:

$$M^u_{G_n}(S) := [\underline{M}^u_{G_n}(S), \overline{M}^u_{G_n}(S)] := \Big[\inf_{\pi \in \mathcal{M}_\alpha} \mathbb{E}_\pi(Y^u_S), \sup_{\pi \in \mathcal{M}_\alpha} \mathbb{E}_\pi(Y^u_S)\Big] \tag{9}$$

Again, by construction, it holds that $m^u_{G_n}(S) \in M^u_{G_n}(S)$ and $m^{u*}_{G_n}(S) \in M^u_{G_n}(S)$ and thus both the true expected similarity value and its maximum entropy approximation are contained in the interval given by the credal approximation. The smaller the width of the credal interval, the less ambiguity underlies the situation. Consequently, analyses based on the maximum entropy approximation are then more reliable.

For the computation of the approximations (8) and (9), we give a proposition showing that, once the preimages of the homogeneity values are computed, one only has to compute the scalar products of the weight vector α with corresponding fixed vectors associated to the previously computed preimages. This will prove very valuable also to our study in Section 5. Checking the validity of the proposition is straightforward and therefore omitted.

Proposition 2. *For the maximum entropy approximation and the credal approximation defined in (8) and (9), the following equations hold, respectively:*

i) $m^{u*}_{G_n}(S) = \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \frac{1}{|A_n^{-1}(k_j)|}\sum_{\underline{R}\in A_n^{-1}(k_j)} Y^u_S(\underline{R})\Big)$

ii) $\underline{M}^u_{G_n}(S) = \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \min_{\underline{R}\in A_n^{-1}(k_j)} Y^u_S(\underline{R})\Big)$

iii) $\overline{M}^u_{G_n}(S) = \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \max_{\underline{R}\in A_n^{-1}(k_j)} Y^u_S(\underline{R})\Big)$

3.4. Estimation of Homogeneity Class Probabilities

Finally, we turn to the last step of the construction described at the beginning of Section 3: In real-world applications, not only the group-specific probability $\mathbb{P}_{G_n}$, but also the precise homogeneity class probabilities α_j will typically be unknown. Accordingly, an estimate $\hat{\alpha} := (\hat{\alpha}_1, \ldots, \hat{\alpha}_\xi)$ for these probabilities has to be obtained. In principle, different ways of addressing this estimation task are conceivable. Firstly, one can draw on expert knowledge, i.e. ask experts from the investigated field for their probability estimates of the homogeneity classes. If more than one expert is involved, an estimate could be received by using either an weighted average of the experts' estimates or by directly working with the credal set containing all of them.

Secondly, one can collect data. For this purpose, a questionnaire can be designed consisting of d items (covering a relevant topic), each of which demanding the participant to order $q := |C|$ alternatives by preference. Each of the n group members participates in the survey such that, after combining the questionnaires, each item produces a preference profile of the relevant group and, therefore, a collection of d preference profiles $\underline{R}_1, \ldots, \underline{R}_d$ is received. For each of these profiles we compute the homogeneity measure and receive data $\underline{x} := (x_1, \ldots, x_d)$, where $x_s := A_n(\underline{R}_s)$ for $s = 1, \ldots, d$. We then estimate α_j by computing relative frequencies

$$\hat{\alpha}_j := \frac{1}{d} \cdot \sum_{s=1}^{d} \mathbb{1}_{\{k_j\}}(x_s) \tag{10}$$

Finally, expert knowledge and available data can be combined by following a Bayesian approach: A preference homogeneity measure $A_n : \mathcal{R}^n \to [0,1]$ defines a categorically distributed random variable[9] taking values in $\{k_1, \ldots, k_\xi\}$. Specifically, since $\alpha_j = \mathbb{P}(A_n = k_j)$, we have $A_n \sim \mathrm{Cat}(\alpha)$.

[9] A random variable X with possible values $\{x_1, \ldots, x_k\}$ is called categorically distributed with parameter vector $\lambda = (\lambda_1, \ldots, \lambda_k) \in \Delta_{k-1}$, formally $X \sim \mathrm{Cat}(\lambda)$, if the probability that X attains value x_i equals λ_i for all $i = 1, \ldots, k$. The categorical distribution is that special case of the multinomial distribution (see, e.g., Berger (1980, p. 562)) where the sample size $n = 1$.

If we, as described in Section 3.1, interpret the parameter $\alpha = (\alpha_1, \ldots, \alpha_\xi)$ as a random quantity with a Dirichlet distribution with (hyper-)parameter $\gamma := (\gamma_1, \ldots, \gamma_\xi) \in \mathbb{R}^\xi_+$ as a prior[10], we can use the data $(x_1, \ldots, x_d)$ from above and compute the posterior distribution of α given $\underline{x}$. For specifying the parameter γ of the prior distribution drawing on expert knowledge seems to be reasonable. As the family of Dirichlet distributions is conjugate to the family of categorical distributions (see Section 3.1), the posterior is again a Dirichlet distribution with updated posterior-parameter $\gamma|\underline{x} := (\gamma_1|\underline{x}, \ldots, \gamma_\xi|\underline{x})$, where $\gamma_j|\underline{x} := \gamma_j + \sum_{s=1}^{q} \mathbb{1}_{\{k_j\}}(x_s)$ for $j = 1, \ldots, \xi$.[11] The common choice for estimating α is then the posterior expectation given by

$$\hat{\alpha}_j := \frac{\gamma_j|\underline{x}}{\sum_{l=1}^{\xi} \gamma_l|\underline{x}} \tag{11}$$

Which approach to follow also depends on the situation: If q is large and the homogeneity measure can attain many different values, taking the relative frequencies will often fail, since doing so requires too many data points to be stable. Particularly, in such cases the Bayesian approach has certain advantages. However, this approach needs to specify a hyper-parameter γ.[12] Note that, when it comes to eliciting experts, the advantages of the proposed framework become perfectly clear: Instead of directly asking experts for their probability estimates on the space of profiles $\mathcal{R}^n$, which contains $(q!)^n$ different elements, one could let them specify a distribution α on the much smaller space $\{k_1, \ldots, k_\xi\}$. Due to its very intuitive interpretation as a relatively small homogeneity scale, the distribution α is much easier to enquire about: How homogeneous do you think the considered group is in probability?

4. Aggregation Rules investigated in the Study

In Section 4.1, we briefly survey some common preference aggregation procedures and demonstrate how they straightforwardly extend to our definition of preference aggregation functions. Importantly, it should be noted that all preference aggregation procedures listed in the following section 4.1 are adaptations of the classic rules from literature to the framework that is used in the present paper. In Section 4.2, we shortly describe a new aggregation method, recently proposed in Schollmeyer (2017), which is based on a generalized concept of quantiles on complete lattices.

4.1. Adaptations of some common Aggregation Procedures

Mean rank (Borda count): For $R \in \mathcal{R}$ and $a \in C$, let $\text{rank}_R(a)$ denote the rank of alternative a with respect to R.[13] The mean rank aggregation function $\text{MR} : \mathcal{R}^n \to \mathcal{Q}$ is defined by $(a, b) \in$

[10] The Dirichlet distribution with parameter vector $\mu = (\mu_1, \ldots \mu_k) \in \mathbb{R}^k_+$ is a probability distribution on the unit simplex Δ_{k-1}. It can therefore be used as a prior distribution for the parameter of a categorical distribution with k categories. For details, see, e.g., Berger (1980, p. 561).

[11] For further details concerning the Dirichlet-Categorical Model see, e.g., Gelman et al. (2004).

[12] If no expert knowledge for specifying the hyper-parameter is available, a so-called near-vacuous prior model, such as the imprecise Dirichlet model (IDM), can be chosen (see Walley (1996) for the original work or Bernard (2005) for further interesting properties).

[13] Formally, we have $\text{rank}_R(a) := |\{b \in C : (a, b) \in R\}| + \frac{1}{2}|\{b \in C : (a, b) \in \sim_R \wedge a \neq b\}| + 1$. Note that this definition of the rank, as common in statistics, assigns the mean value of all possible ranks to the members of the equivalence classes of $\sim_R$ (see for instance, Yule and Kendall (1924)). Other, less common, choices are to assign to these consequences the minimum or the maximum rank.

$\mathrm{MR}(\underline{R})$ if and only if $\sum_{i=1}^{n}\big(\mathrm{rank}_{R_i}(a) - \mathrm{rank}_{R_i}(b)\big) > 0$, where $\underline{R} \in \mathcal{R}^n$. Specifically, the group assigns each alternative its average rank and prefers alternative a to alternative b iff the latter achieves a strictly lower average rank. a and b are equivalent with respect to $\sim_{\mathrm{MR}(\underline{R})}$ if and only if they have equal average rank in the profile $\underline{R}$.

Condorcet's method: The Condorcet aggregation function $\mathrm{CO} : \mathcal{R}^n \to \mathcal{Q}$ is defined by $(a,b) \in \mathrm{CO}(\underline{R})$ if and only if $(c_{\underline{R}}(a,b) > c_{\underline{R}}(b,a) \wedge c_{\underline{R}}(a,b) > e_{\underline{R}}(a,b))$, where $\underline{R} \in \mathcal{R}^n$. Hence, for each pair (a,b) we decide if the majority of the group prefers a to b or vice versa or if the majority of the group is indifferent between a and b. The consequences a and b are considered equivalent with respect to $\sim_{\mathrm{CO}(\underline{R})}$ if and only if either at least half of the group is indifferent between them or an equal number of persons prefer a to b and vice versa. Importantly, note that this is one adaptation of Condorcet's method to our framework: In principle there may be other plausible ways of defining the group's indifference relation.

Instant runoff (Hare's method): Instant runoff is a sequential aggregation procedure: In the first step, all alternatives with the fewest number of first-place votes are excluded from C. These form the alternatives that are least preferred by the group, and between them the group is indifferent. Afterwards, we exclude the alternatives with the fewest first-place votes in the profile on the reduced space of alternatives and receive a set of alternatives that the group prefers second least. Again, between these alternatives the group is indifferent, but each of them is preferred to every alternative excluded in the first step. Successively repeating this procedure, we end up with a set of optimal options with the same number of first-place votes. This describes a preference aggregation function $\mathrm{IR} : \mathcal{R}^n \to \mathcal{Q}$ defined by $(a,b) \in \mathrm{IR}(\underline{R})$ if and only if a is excluded at a later stage than b, where $\underline{R} \in \mathcal{R}^n$. The consequences a and b are equivalent with respect to $\sim_{\mathrm{IR}(\underline{R})}$ if and only if they are excluded at the same stage of the procedure.

Coombs' rule:[14] The basic idea of Coombs' rule is very similar to that of instant runoff voting, as it is also based on sequential exclusion of alternatives. However, in contrast to instant runoff voting, we exclude the alternatives with the maximal number of last-place votes in every step. The corresponding aggregation function $\mathrm{CM} : \mathcal{R}^n \to \mathcal{Q}$ is defined by $(a,b) \in \mathrm{CM}(\underline{R})$ if and only if a is excluded at a later stage than b, where $\underline{R} \in \mathcal{R}^n$. Moreover, a and b are equivalent with respect to $\sim_{\mathrm{CM}(\underline{R})}$ if and only if they are excluded at the same stage.

Kemeny's rule:[15] Given a profile $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$, the idea of Kemeny's rule is to choose that consensus order $Q \in \mathcal{R}$ that minimizes the sum of the distances $d(R_i, Q)$ to the individual orders (see Section 3.2). Formally, we say Q^* is a Kemeny consensus ranking for $\mathcal{R}$ whenever it holds that $Q^* \in \mathrm{argmin}_{Q \in \mathcal{R}} \sum_{i=1}^{n} d(R_i, Q) =: \mathrm{KE}(\underline{R})$. Since such Q^* will generally not be unique, in order to define an aggregation function from this rule, we need to choose a choice function $f : 2^{\mathcal{R}} \setminus \{\emptyset\} \to \mathcal{R}$ satisfying $f(\mathcal{C}) \in \mathcal{C}$ for all $\mathcal{C} \in 2^{\mathcal{R}} \setminus \{\emptyset\}$. The Kemeny aggregation function $\mathrm{KE}_f : \mathcal{R}^n \to \mathcal{R}$ with respect to f is then given by $\mathrm{KE}_f(\underline{R}) := f(\mathrm{KE}(\underline{R}))$ for all $\underline{R} \in \mathcal{R}^n$.[16] In

[14] Cf. Coombs and Cohen (1984) for a discussion or Grofman and Feld (2004) for comparisons with Hare's method.

[15] Originally proposed in Kemeny (1959); some nice properties of the method are shown in Young and Levenglick (1978). Generally, note that determining a Kemeny consensus ranking for $\underline{R}$ is NP-hard (see Bartholdi et al. (1989). Also, compare Ali and Meila (2012) for a comparison of algorithms alleviating the NP-hardness.

[16] Note that, since $\sum_{i=1}^{n} d(R_i, Q)$ is minimal iff $\sum_{i=1}^{n} |R_i \cap Q|$ is maximal (see Section 3.2), this implies $Y^{id}_{\mathrm{KE}_f}(\underline{R}) \geq Y^{id}_{S}(\underline{R})$ for all $\underline{R} \in \mathcal{R}^n$ for every other preference aggregation function S, where $id(x) := x$ for all $x \in \mathbb{R}$. Consequently, this implies $\mathbb{E}_{\mathbb{P}}(Y^{id}_{\mathrm{KE}_f}) \geq \mathbb{E}_{\mathbb{P}}(Y^{id}_{S})$ for any probability measure $\mathbb{P}$ and, thus, Kemeny's rule is superior to every other method if similarity is measured based on the function $u = id$.

the study in Section 5, we enumerated the orders and defined $f(\mathcal{C}) = R_{j_0}$, where j_0 is the smallest index of an order belonging to $\mathcal{C}$.

Dictatorship: For $i_0 \in \{1, \ldots, n\}$, the dictatorship aggregation function $\mathrm{DI}_{i_0} : \mathcal{R}^n \to \mathcal{Q}$ is defined by $(a, b) \in \mathrm{DI}_{i_0}(\underline{R})$ iff $(a, b) \in R_{i_0}$, where $\underline{R} \in \mathcal{R}^n$. Hence, the group prefers a to b whenever individual i_0 does. Even if this does not seem like the fairest way of aggregating preferences, it might be worth investigating how the dictatorship function performs in comparison to the others.

4.2. An aggregation rule based on quantiles on complete lattices: commonality sharing

We now briefly sketch the idea of the commonality sharing aggregation rule that was recently proposed in Schollmeyer (2017) and that initially arose in a different context, namely through attempts to generalize concepts of centrality and outlyingness of observations to partially ordered data. Opposed to the other methods investigated here, this aggregation procedure does not locally look at different alternatives or pairs of alternatives, but takes into account the full order of all individuals and embeds these orders into the complete lattice of binary relations on C equipped with the set intersection and set union as meets and joins, respectively. Then, a notion of outlyingness in this space, described in Schollmeyer (2017), is used to select one or more orders of individuals who are most centered in the population. Concretely, the following procedure can be applied:

For a given minimum size k, one looks at every possible sub-population M_k^i consisting of at least k individuals. Then, one considers the set Q_k of all individuals q_k^j who share with every sub-population M_k^i all commonalities of this sub-population (i.e., all edges (a, b) that the population M_k^i has in common should also be edges of every order q_k^j in Q_k). The set Q_k of individuals who share with every sub-population of size $\geq k$ its commonalities is to some extent representative for every such sub-population. If k is too small, then Q_k is empty. In contrast, for $k = n$ the set Q_k is the whole population. Now, for a given order q, the smaller the smallest k such that Q_k still contains q, the more central is the order q, since then q is a representative for a bigger collection of sub-populations including smaller sub-populations with bigger and thus more specific commonalities. Finally, to select a consensus order, choose k as small as possible under the restriction that $Q_k \neq \emptyset$ and choose the arising Q_k as the set of candidates for the consensus order. If Q_k has more than one element then for a unique consensus order choose arbitrarily from the set Q_k or apply some further aggregation rule to the orders in Q_k. In the study of Section 5 we apply the first approach and, like in the procedure for Kemeny's rule, choose that most central order with the smallest index.

Note that the commonality sharing consensus rule is in fact a non-local rule in the sense that if for example two individuals in a profile both prefer all alternatives in the set $\{a, b, c\}$ over the alternatives in the set $\{d, e, f\}$, but with different orders within these sets, then the consensus order could possibly change if the individuals swap their orders within the set $\{a, b, c\}$ with each other while maintaining their orders within the set $\{d, e, f\}$. Thus, the orders do in fact play a role as a whole. This is the main difference from the aggregation rules of Section 4.1 (except dictatorship): There, for example, it does not matter which individual an alternative gets its score from in the mean rank aggregation or where edges for pairs of alternatives in Condorcet's method are counted without differentiating between edges belonging to the same individual and edges belonging to different individuals. Note that commonality sharing can be computed in $\mathcal{O}(n \cdot |C|^2)$ time, much simpler than one would expect from the conceptual description.

5. A Simulation Study for the Case $n = 8$ and $|C| = 4$

In this section, we apply the evaluation framework developed in Sections 2 and 3 to the aggregation functions discussed in Section 4. Specifically, we design a study allowing for the comparison of the appropriateness of these aggregation functions for groups of varying degrees of homogeneity in Section 5.1 and then discuss the results in Section 5.2.

5.1. Setup of the Study

Throughout this study, we consider groups G_8 consisting of $n = 8$ group members, each of which has ranked $|C| = 4$ consequences. We further assume that within the individual preferences there is no indifference, i.e., we restrict analysis to the set $\mathcal{H}$ of all strict weak orders $R \in \mathcal{R}$ satisfying $\sim_R = \{(c, c) : c \in C\}$. In order to analyze the appropriateness of different aggregation functions S for groups of varying degrees of homogeneity, we compute and compare the maximum entropy approximation from (8) and the credal set approximation from (9) for different choices of the homogeneity class probabilities α. Therefore, we need to specify three things: A preference homogeneity measure A_n in the sense of Definition 1, a similarity measure Y_S^u in the sense of Definition 3, and a sequence of probability vectors $\alpha^{[i]}$, each determining a distribution of A_n while representing a different degree of group homogeneity.

Firstly, for A_n, we choose the maximum consensus homogeneity δ_n from Definition 2 restricted to the domain $\mathcal{H}^8$, which in this case reduces to $\delta_8(\underline{R}) = \frac{1}{48} \sum_{(a,b)\in R_0} \max\{c_{\underline{R}}(a, b), c_{\underline{R}}(b, a)\}$ for all $\underline{R} \in \mathcal{H}^8$. Secondly, for measuring similarity between a profile and its consensus order with respect to an aggregation function S, we consider three different choices for the function u in the similarity measure Y_S^u, taking into account the discussion after Definition 3: The linear function $u_1(x) = x$ giving equal influence to all group members, the convex function $u_2(x) = x^2$ giving higher influence to appropriately represented group members, and the concave function $u_3(x) = 36 - (6 - x)^2$ strengthening the influence of inadequately represented members. Note that u_1 and u_3 were also proposed in Kemeny (1959).

Finally, in order to model varying degrees of group homogeneity of the considered groups G_8, first note that δ_8 attains the values $k_1 = \frac{24}{48} < k_2 = \frac{25}{48} < \cdots < k_{25} = \frac{48}{48}$, where $\frac{1}{2}$ indicates minimal homogeneity and 1 indicates perfect homogeneity of the inserted profile. Accordingly, assuming δ_8 to be categorically distributed with parameter vector $\alpha_{min} = (1, 0, \ldots, 0)$ obviously represents a group G_8 with a lower degree of homogeneity than assuming the same for parameter vector $\alpha_{max} = (0, \ldots, 0, 1)$. Generalizing from this idea, we can construct a sequence $\alpha^{[0]}, \ldots, \alpha^{[50]} \in \Delta^{24}$ of parameter vectors representing groups G_8 of increasing degrees of homogeneity by setting

$$\alpha_j^{[i]} := \mathrm{Bin}\big(24, \tfrac{i}{50}\big)(\{j\}) = \tbinom{24}{j} \cdot (\tfrac{i}{50})^j \cdot (1 - \tfrac{i}{50})^{24-j} \tag{12}$$

with $i = 0, \ldots, 50$ and $j = 0, \ldots, 24$, where $\alpha_j^{[i]}$ denotes the jth component of the ith parameter vector and $\mathrm{Bin}(n, p)$ denotes the binomial distribution with parameters n and p. Using the constructed sequence then allows for analyzing the performance of different aggregation function for varying degrees $\alpha^{[i]}$ of homogeneity. Note that, due to Proposition 2, computation of the approximations is possible without computing the whole assessment for every single $\alpha^{[i]}$.

5.2. Discussion of the Results

The results of the study described in Section 5.1 for the similarity measure $Y_S^{u_2}$ are visualized in Figures 1 and 2, while the results for the similarity measures $Y_S^{u_1}$ and $Y_S^{u_3}$ can be found in Figures 3 and 4 in Appendix A4. We will refer to these in our discussion at several points.

Figure 1 shows the credal approximation $M^{u_2}_{G_n}(S)$ (gray shaded region) and the maximum entropy approximation $m^{u_2*}_{G_n}(S)$ (black line) for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \ldots, 50\}$. Specifically, an abscissa value of i indicates that the information on the group G_8 is given by $\delta_8 \sim \text{Cat}(\alpha^{[i]})$. The ordinate ranges from 0 to 288, where 0 is the minimal and 288 is the maximal attainable expected similarity value.

Clearly, for all investigated aggregation functions the width of the intervals $M^{u_2}_{G_n}(S)$ does depend on the homogeneity of the group, where very homogeneous and very inhomogeneous groups tend to produce more narrow intervals than groups of medium homogeneity. Moreover, for all functions except instant-runoff, both the upper expected similarity $\overline{M}^{u}_{G_n}(S)$ and the maximum entropy approximation $m^{u_2*}_{G_n}(S)$ strictly increase along increasing homogeneity. Contrarily, the lower expected similarity $\underline{M}^{u}_{G_n}(S)$ does not strictly increase along homogeneity for mean rank, dictatorship, Kemeny's rule, and commonality sharing, while it does for the other functions (the same is true if analysis is based on $Y_S^{u_1}$ or $Y_S^{u_3}$). Finally, the results show that Kemeny's rule, commonality sharing, and Condorcet's method produce rather narrow credal approximations compared to the other functions, whereas dictatorship, Coombs' rule, and instant-runoff lead to rather wide intervals. Note that the comparison of the width of the credal intervals provides highly relevant information as it indicates how sensitively the evaluation of an aggregation procedure reacts to choosing one (possibly wrong) approximating measure from the credal set: If we mistakenly evaluate an aggregation function by using maximum entropy, the average error we make will be higher for aggregation functions whose evaluation reacts very sensitively to the choice of a representative.

Figure 2 consists of two different graphs: The upper graph compares the lower expected similarity $\underline{M}^{u_2}_{G_n}(S)$ along an increasing degree of homogeneity $\alpha^{[i]}$ for all aggregation procedures. With the exception of very inhomogeneous groups ($i \leq 2$) where dictatorship exceeds Kemeny's rule, it shows that Kemeny's rule and commonality sharing outperform all other methods independent of the underlying $\alpha^{[i]}$. Comparing commonality sharing with Kemeny's rule, we see that the latter exceeds the former for groups with $i \geq 6$, while the opposite is the case for inhomogeneous groups with $i \leq 5$. Further, it turns out that for rather inhomogeneous groups ($i \leq 18$), choosing a dictatorship performs better than all other methods except commonality sharing and Kemeny's rule. For medium to perfect homogeneity ($i > 18$), however, Condorcet's rule shows the third best performance, for high homogeneity ($i \geq 35$), very closely followed by mean rank and Coombs' rule. Moreover, it is interesting to note that instant-runoff is outperformed by all other aggregation functions independent of the underlying $\alpha^{[i]}$. For $Y_S^{u_1}$ and $Y_S^{u_3}$ (see Appendix A4), the main difference is that the lower expected similarity values $\underline{M}^{u_1}_{G_n}(S)$ and $\underline{M}^{u_3}_{G_n}(S)$ of Kemeny's rule exceed those of the other functions, independent of $\alpha^{[i]}$.

Comparing the maximum entropy approximations $m^{u*}_{G_n}(S)$ in the lower graph gives a similar picture. Again, commonality sharing and Kemeny's rule, whose (numerically) coinciding approximations cannot be distinguished, are superior to all the other aggregation methods. However, mean rank aggregation now outperforms Condorcet's method independent of homogeneity and is already superior to a dictatorship for groups with a rather low homogeneity value ($i \geq 10$). For homogeneous groups ($i \geq 30$), all methods except dictatorship and instant-runoff show a very similar performance. Again, instant-runoff is outperformed by all other functions.

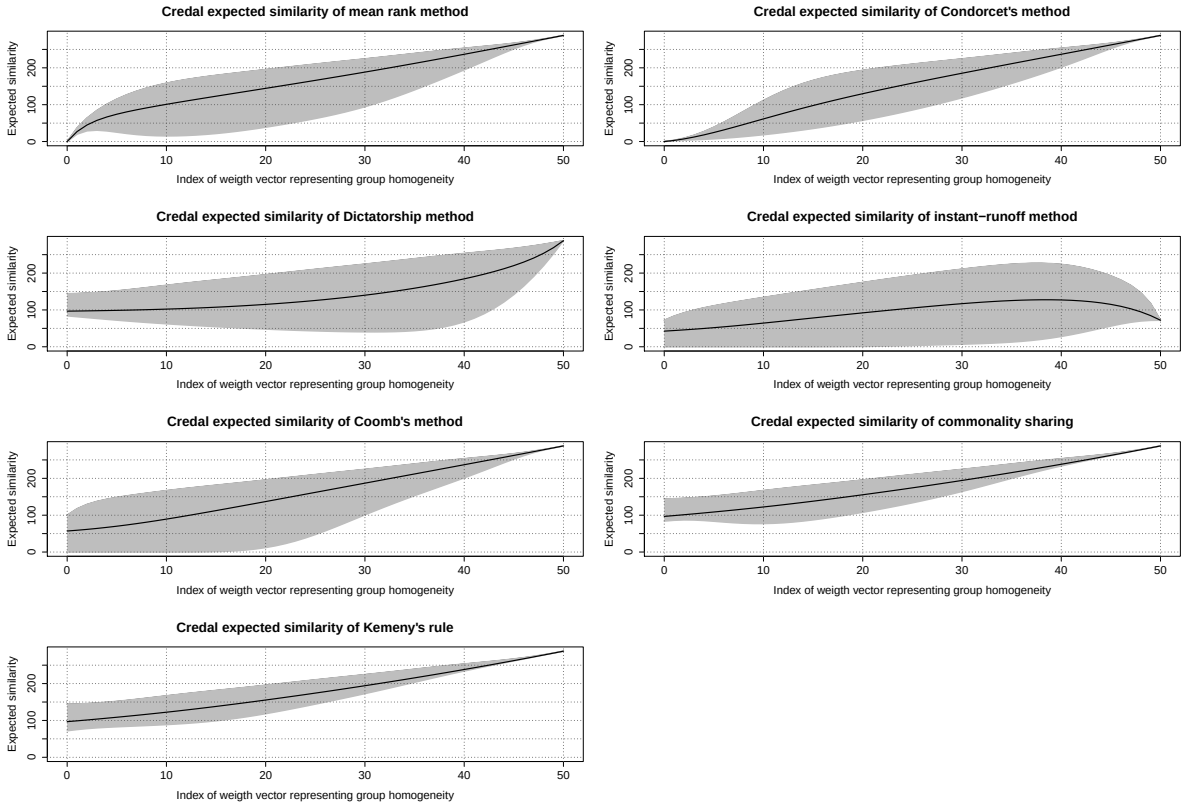

Figure 1: The figures shows the credal approximation $M^{u_2}_{G_n}(S)$ (gray shaded region) and the maximum entropy approximation $m^{u_2*}_{G_n}(S)$ (black line) for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \ldots, 50\}$ and $u_2(x) = x^2$.

6. Summary, concluding remarks, and discussion

In this paper, we introduced the expected similarity $m^u_{G_n}(S)$ as a theoretical criterion for evaluating and comparing the performance of different preference aggregation functions S if perfect probabilistic information on the homogeneity structure of the group members' preferences is available. Approaches for approximating the true value of $m^u_{G_n}(S)$ under imperfect probabilistic information are fundamentally based on the concept of preference homogeneity measures, for which we gave both a set of minimal requirements and a concrete proposal. Specifically, we studied two different ways to approximate $m^u_{G_n}(S)$: the maximum entropy approximation and the credal set approximation. Finally, by comparing these approximations, we investigated the performance of six common aggregation procedures as well as the recently proposed commonality sharing rule by means of a simulation study for groups along varying degrees of homogeneity. Specifically, we were able to show that the adequateness of a preference aggregation function for a fixed group indeed depends on the group's homogeneity structure.

In future research this framework needs to be applied to real-world data. Particularly, we plan to apply the proposed estimation procedures and evaluation framework to survey data on political opinions and investigate whether groups of significantly differing degrees of homogeneity can be

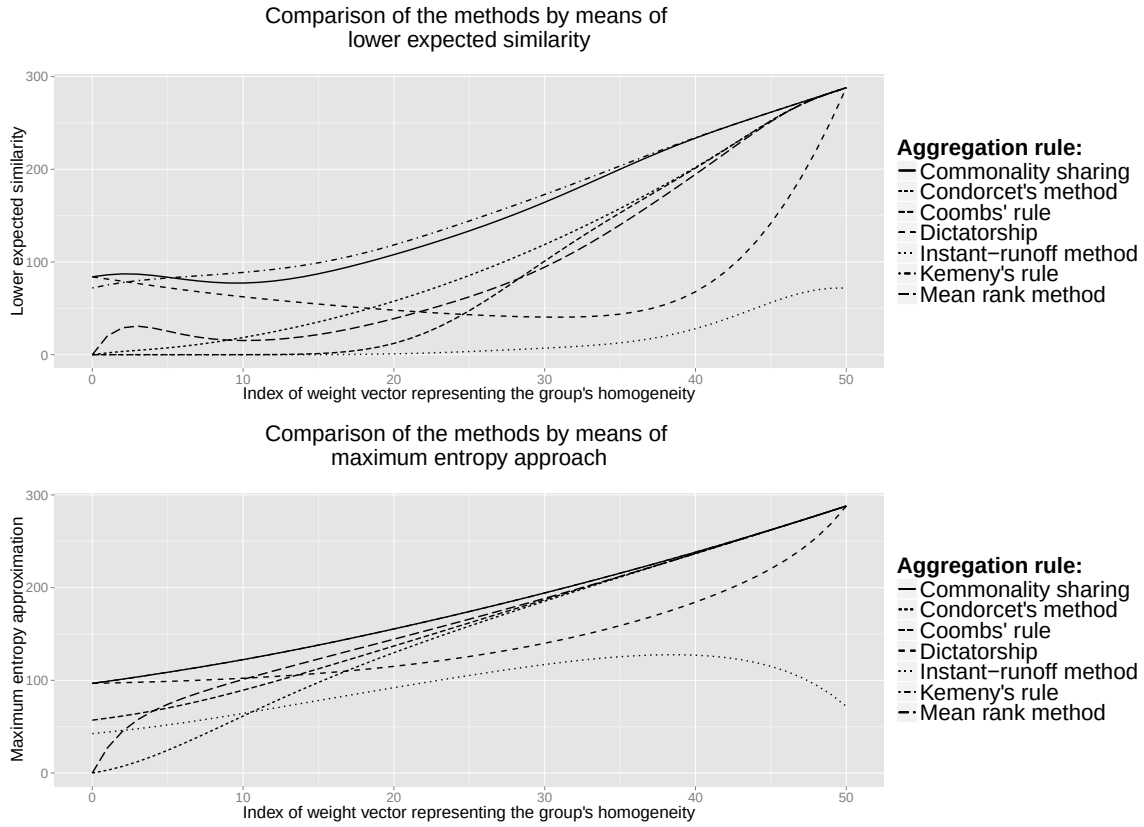

Figure 2: The upper graph shows the lower expected similarity $\underline{M}^{u_2}_{G_n}(S)$ for different choices of the aggregation function S along increasing degrees of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \ldots, 50\}$ and $u_2(x) = x^2$. The lower graph shows the same for the maximum entropy approximation $m^{u_2*}_{G_n}(S)$.

identified in empirical studies. Beyond this, three further extensions seem particularly promising:

Partial individual preferences: The preferences of the group members are modeled by asymmetric and negatively transitive relations $R \in \mathcal{R}$. This explicitly excludes the case of group members judging certain consequences in C incomparable (since incomparability with respect to R is treated as indifference, see Section 2.1 and Footnote 2 in particular). Allowing also for incomparability of consequences could lead to a more realistic model in certain situations. Of course, this requires appropriate adaptations of the aggregation rules from Section 4.

Axiomatic foundations: The conditions that have been listed in Definition 1 are to be understood as minimal requirements for measures of preference homogeneity. However, they are rather weak, as they only look at the profile on a categorical scale. Going beyond the categorical scale in the spirit of Bosch (2006) and Alcalde-Unzu and Vorsatz (2013) could give a more detailed picture of what is actually meant by homogeneity on an axiomatic level.

Efficient algorithms for simulation: In Section 5, we presented a study for a group of $n = 8$ members ranking $|C| = 4$ alternatives. For this setting, the approximations for expected similarity were able to be computed analytically. However, for larger settings this becomes computer-intensive, and computation using simulations has to be applied instead. A proposal for simulation designs

is given in Appendix A2. More efficient designs, comparable to the MCMC-driven approaches already used in the statistical analysis of networks (see, e.g., Hunter et al. (2012)), are planned to be investigated in future research.

Acknowledgement

The authors would like to thank the reviewers and the associate editor for very supportive and helpful comments which improved the paper considerably.

References

Abrams, R. (1976). The voter's paradox and the homogeneity of individual preference orders. *Public Choice*, 26(1):19–27.

Alcalde-Unzu, J. and Vorsatz, M. (2013). Measuring the cohesiveness of preferences: an axiomatic analysis. *Social Choice and Welfare*, 41(4):965–988.

Aleskerov, F., Karabekyan, D., Sanver, M. R., and Vyacheslav, Y. (2012). On the manipulability of voting rules: the case of 4 and 5 alternatives. *Mathematical Social Sciences*, 64(1):67–73.

Ali, A. and Meila, M. (2012). Experiments with Kemeny ranking: what works when? *Mathematical Social Sciences*, 64(1):28–40.

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

Bartholdi, J., Tovey, C., and Trick, M. (1989). Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(2):157–165.

Berg, S. (1985). Paradox of voting under an urn model: the effect of homogeneity. *Public Choice*, 47(2):377–387.

Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis*. Springer.

Bernard, J. (2005). An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2–3):123–150.

Bosch, R. (2006). *Characterizations on Voting Rules and Consensus Measures (Ph.D. dissertation)*. Tilburg University, The Netherlands.

Brams, S. and Fishburn, P. (2002). Voting procedures. In Arrow, K., Sen, A., and Suzumura, K., editors, *Handbook of Social Choice and Welfare, Vol. 1*, pages 173–236. North-Holland.

Can, B., Ozkes, A. I., and Storcken, T. (2015). Measuring polarization in preferences. *Mathematical Social Sciences*, 78:76–79.

Coombs, C. H. and Cohen, J. L. (1984). Social choice observed: five presidential elections of the American Psychological Association. *The Journal of Politics*, 46(2):479–502.

de Borda, J. C. (1781). *Memoire sur les elections au scrutin*. Historie de l'Academie Royale des Sciences.

de Condorcet, N. (1785). *Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix*. Paris.

de Finetti, B. (1974). *Theory of Probability*. Wiley.

Diss, M., Louichi, A., Merlin, V., and Smaoui, H. (2012). An example of probability computations under the IAC assumption: The stability of scoring rules. *Mathematical Social Sciences*, 64(1):57–66.

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

Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. (2001). Rank aggregation methods for the web. In Shen, V., Saito, N., Lyu, M. R., and Zurko, M. E., editors, *Proceedings of the 10th International Conference on World Wide Web*, pages 613–622. ACM.

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

Fishburn, P. C. (1973). Voter concordance, simple majorities, and group decision methods. *Behavioral Science*, 18(5):364–376.

Fligner, M. A. and Verducci, J. S. (1986). Distance-based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):359–369.

Gehrlein, W. V. (1981). Non-profile specific measures of social homogeneity. *Proceedings of the Southeast American Institute for Decision Sciences*, pages 167–169.

Gehrlein, W. V. and Lepelley, D. (2010). *Voting Paradoxes and Group Coherence: The Condorcet Efficiency of Voting Rules*. Springer.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall.

Grofman, B. and Feld, S. (2004). If you like the alternative vote (a.k.a. the instant runoff), then you ought to know about the Coombs rule. *Electoral Studies*, 23:641–659.

Hare, T. (1857). *The Machinery of Representation*. Maxwell.

Hartmann, S. and Sprenger, J. (2012). Judgment aggregation and the problem of tracking the truth. *Synthese*, 187(1):209–221.

Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21(4):856–882.

Jamison, D. and Luce, E. (1972). Social homogeneity and the probability of intransitive majority rule. *Journal of Economic Theory*, 5(1):79–87.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. Wiley.

Kemeny, J. and Snell, J. (1962). *Mathematical Models in the Social Sciences*. Ginn and Company.

Kemeny, J. G. (1959). Mathematics without numbers. *Quantity and Quality*, 88(4):577–591.

Kendall, M. G. and Smith, B. B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, 10(3):275–287.

Kreps, D. M. (1988). *Notes on the theory of choice (Underground Classics in Economics)*. Westview Press Incorporated.

Landes, J. and Williamson, J. (2013). Objective Bayesianism and the maximum entropy principle. *Entropy*, 13:3528–3591.

Lepelley, D. and Valognes, F. (2003). Voting rules, manipulability and social homogeneity. *Public Choice*, 116(1-2):165–184.

Levi, I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT press.

Maniqueta, F. and Mongin, P. (2016). A theorem on aggregating classifications. *Mathematical Social Sciences*, 79:6–10.

Mersmann, O., Preuss, M., Trautmann, H., Bischl, B., and Weihs, C. (2015). Analyzing the BBOB results by means of benchmarking concepts. *Evolutionary Computation Journal*, 23(1):161–185.

Niemi, R. G. (1969). Majority decision-making with partial unidimensionality. *American Political Science Review*, 63(2):488–497.

Pritchard, G. and Slinko, A. (2006). On the average minimum size of a manipulating coalition. *Social Choice and Welfare*, 27(2):263–277.

Rosenkrantz, R. D. (1977). *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*. Springer.

Schollmeyer, G. (2017). Application of lower quantiles for complete lattices to ranking data: analyzing outlyingness of preference orderings. *Technical Report 208, Department of Statistics, LMU Munich (2017)*. Available under: `https://epub.ub.uni-muenchen.de/40452/`.

Shannon, C. E. (1949). A mathematical theory of communication data. *The Bell System Technical Journal*, 27:379–423.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):3–57.

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

Young, H. P. and Levenglick, A. (1978). A consistent extension of Condorcet's election principle. *SIAM Journal of Applied Mathematics*, 35:285–300.

Yule, G. U. and Kendall, M. G. (1924). *An introduction to the theory of statistics*. C. Griffin.

Appendix

A1: *Proof of Proposition 1*

First, note that the definition of δ_n does not depend on the choice of $R_0 \in \mathcal{R}$ with $\sim_{R_0} = \mathrm{diag}(C^2)$, since every such relation contains exactly one of the pairs (a,b) and (b,a) for all $a,b \in C$ with $a \neq b$ and summation is commutative. Moreover, one easily verifies that $\mathrm{Im}(\delta_n) \subset [0,1]$. Hence, δ_n is well-defined.

(S1): We have to show that 1 is attained iff the profile consists of identical orders. Obviously, δ_n equals 1 for identical profiles by construction. In contrast, if $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$ is a non-identical profile, a pair $(a,b) \in R_0$ exists such that $\max\{c_{\underline{R}}(a,b), c_{\underline{R}}(b,a), e_{\underline{R}}(a,b)\} < n$. This gives $\delta_n(\underline{R}) < 1$.

(S2): Let $\phi : \{1,\ldots,n\} \to \{1,\ldots,n\}$ be a bijective map and $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$. By definition, we have $c_{(R_1,\ldots,R_n)}(a,b) = c_{(R_{\phi(1)},\ldots,R_{\phi(n)})}(a,b)$ and $e_{(R_1,\ldots,R_n)}(a,b) = e_{(R_{\phi(1)},\ldots,R_{\phi(n)})}(a,b)$. This implies $\delta_n(R_1,\ldots,R_n) = \delta_n(R_{\phi(1)},\ldots,R_{\phi(n)})$.

(S3): Let $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$ be a preference profile such that exactly $k \in \{\lfloor \frac{n}{2} \rfloor, \ldots, n-1\}$ group members share identical preferences. Without loss of generality, we assume it holds that $R_1 = \cdots = R_k =: R^*$ (otherwise we can rearrange the profile in this way due to (S2)). For all distinct pairs $(a,b) \in C^2$ it then holds that $f_{\underline{R}}(a,b) := \max\{c_{\underline{R}}(a,b), c_{\underline{R}}(b,a), e_{\underline{R}}(a,b)\} \geq k$, since each pair (a,b) is identically ranked within the orders $R_1,\ldots,R_k$. Choose an arbitrary index $j_0 \in \{k+1,\ldots,n\}$ and define $\underline{Q} := (Q_1,\ldots,Q_n)$ to be the profile that arises from $\underline{R}$ by exchanging order R_{j_0} with order R^*. We show that $\delta_n(\underline{R}) \leq \delta_n(\underline{Q})$. Therefore, let $a_0, b_0 \in C$, $a_0 \neq b_0$ be arbitrary but fixed. We distinguish two cases:

Case 1: $f_{\underline{R}}(a_0,b_0) = k$. Clearly, this implies $f_{\underline{Q}}(a_0,b_0) = k+1 > k = f_{\underline{R}}(a_0,b_0)$, since (a_0,b_0) is then identically ranked by exactly $Q_1,\ldots,Q_k$ and Q_{j_0}.

Case 2: $f_{\underline{R}}(a_0,b_0) > k$. For arbitrary but fixed $R \in \mathcal{R}$ and $a,b \in C$, $a \neq b$, define the expression $R_{\{a,b\}} := \{(x,y) : x,y \in \{a,b\} \ \wedge \ (x,y) \in R\}$. We then distinguish two sub-cases:

Sub-case 1: $\forall\, j \in \{k+1,\ldots,n\} : R^*_{\{a_0,b_0\}} \neq (R_j)_{\{a_0,b_0\}}$.
This implies that $(R_{j_1})_{\{a_0,b_0\}} = (R_{j_2})_{\{a_0,b_0\}}$ for all $j_1, j_2 \in \{k+1,\ldots,n\}$ (and that $k = \lfloor \frac{n}{2} \rfloor$ and n is odd), since otherwise $f_{\underline{R}}(a_0,b_0) > k$ would not be possible. Hence, the pair (a_0,b_0) is ranked identically by $\lfloor \frac{n}{2} \rfloor + 1$ members and, therefore, we have $f_{\underline{R}}(a_0,b_0) = \lfloor \frac{n}{2} \rfloor + 1$. However, it also holds that $f_{\underline{Q}}(a_0,b_0) = \lfloor \frac{n}{2} \rfloor + 1$, since (a_0,b_0) is identically ranked by exactly $Q_1,\ldots,Q_k$ and Q_{j_0}.

Sub-case 2: $\exists\, j \in \{k+1,\ldots,n\} : R^*_{\{a_0,b_0\}} = (R_j)_{\{a_0,b_0\}}$.
Then, if $R^*_{\{a_0,b_0\}} = (R_{j_0})_{\{a_0,b_0\}}$ we have $f_{\underline{R}}(a_0,b_0) = f_{\underline{Q}}(a_0,b_0)$, and if $R^*_{\{a_0,b_0\}} \neq (R_{j_0})_{\{a_0,b_0\}}$ we have $f_{\underline{R}}(a_0,b_0) < f_{\underline{Q}}(a_0,b_0)$. In either case, we have $f_{\underline{R}}(a_0,b_0) \leq f_{\underline{Q}}(a_0,b_0)$.

Thus, we have shown that, in every case, it holds that $f_{\underline{R}}(a_0,b_0) \leq f_{\underline{Q}}(a_0,b_0)$. Since the pair (a_0,b_0) was chosen arbitrarily, this implies $\delta_n(\underline{R}) \leq \delta_n(\underline{Q})$, completing the proof. □

A2: *A possible simulation design for larger settings*

The simulation can be done in the following way: Draw N random samples $\underline{R}_1,\ldots,\underline{R}_N$ from the space of profiles $\mathcal{R}^n$. For all $j = 1,\ldots,\xi$, define the set $N_j := \{\underline{R}_i : A_n(\underline{R}_i) = k_j\}$ of all samples mapped to homogeneity class k_j. For given weights $\alpha := (\alpha_1,\ldots,\alpha_\xi)$ and aggregation function S, we use the characterization of the assessments given in Proposition 2 and receive

$$
\begin{aligned}
m^{u*}_{G_n}(S) &= \textstyle\sum_{j=1}^{\xi}\Big(\alpha_j \cdot \frac{1}{|A_n^{-1}(k_j)|}\sum_{\underline{R}\in A_n^{-1}(k_j)} Y^u_S(\underline{R})\Big) \;\approx\; \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \frac{1}{|N_j|}\sum_{\underline{R}\in N_j} Y^u_S(\underline{R})\Big)\\
\underline{M}^u_{G_n}(S) &= \textstyle\sum_{j=1}^{\xi}\Big(\alpha_j \cdot \min_{\underline{R}\in A_n^{-1}(k_j)} Y^u_S(\underline{R})\Big) \;\approx\; \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \min_{\underline{R}\in N_j} Y^u_S(\underline{R})\Big)\\
\overline{M}^u_{G_n}(S) &= \textstyle\sum_{j=1}^{\xi}\Big(\alpha_j \cdot \max_{\underline{R}\in A_n^{-1}(k_j)} Y^u_S(\underline{R})\Big) \;\approx\; \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \max_{\underline{R}\in N_j} Y^u_S(\underline{R})\Big)
\end{aligned}
$$

However, note that this simulation design requires a sample satisfying the condition $N_j \neq \emptyset$ for all $j = 1, \ldots, \xi$, i.e., the sample needs to be rich enough that every homogeneity class has been met at least once. Consequently, such a design becomes computationally intensive as n and $|C|$ increase.

A simulation design producing fewer computational costs can be realized by taking advantage of the fact that the maps A_n and Y_S are invariant under permutations of the inserted profile. Let Φ denote the set of all bijective maps $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$. For $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$ and $\phi \in \Phi$, we set $\underline{R}_\phi := (R_{\phi(1)}, \ldots, R_{\phi(n)})$ and define an equivalence relation $\sim_\Phi$ on $\mathcal{R}^n$ by setting

$$\underline{R} \sim_\Phi \underline{Q} \quad :\Leftrightarrow \quad \exists \phi \in \Phi : \; \underline{R} = \underline{Q}_\phi$$

Moreover, let $\mathcal{R}^n_{\sim_\Phi}$ denote the quotient space produced by $\sim_\Phi$ and let $f : \mathcal{R}^n_{\sim_\Phi} \to \mathcal{R}^n$ be any choice function satisfying $f(\mathcal{C}) \in \mathcal{C}$ for all $\mathcal{C} \in \mathcal{R}^n_{\sim_\Phi}$. Further, for every possible homogeneity value k_j, where $j = 1, \ldots, \xi$, we define the set $\mathcal{L}_j := \{\mathcal{C} \in \mathcal{R}^n_{\sim_\Phi} : \; A_n(f(\mathcal{C})) = k_j\}$ of all equivalence classes with members that are mapped to k_j. Due to Proposition 2 and the fact that both A_n and Y^u_S are constant on every $\mathcal{C} \in \mathcal{R}^n_{\sim_\Phi}$ (as they are invariant under permutations of the inserted profile), one easily verifies the following identities:

$$
\begin{aligned}
m^{u*}_{G_n}(S) &= \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \frac{\sum_{\mathcal{C}\in\mathcal{L}_j} Y^u_S(f(\mathcal{C}))\cdot|\mathcal{C}|}{\sum_{\mathcal{C}\in\mathcal{L}_j}|\mathcal{C}|}\Big)\\
\underline{M}^u_{G_n}(S) &= \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \min_{\mathcal{C}\in\mathcal{L}_j} Y^u_S(f(\mathcal{C}))\Big)\\
\overline{M}^u_{G_n}(S) &= \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \max_{\mathcal{C}\in\mathcal{L}_j} Y^u_S(f(\mathcal{C}))\Big)
\end{aligned}
$$

Using the above identities allows the application of a similar simulation design as proposed above, however, instead of drawing samples from the space $\mathcal{R}^n$, we can now sample from the smaller space $\mathcal{R}^n_{\sim_\Phi}$. In our context, this means we can sample from the space of all n-combinations of $\mathcal{R}$ instead of sampling from the space of n-permutations of $\mathcal{R}$.

A3: *Non-locality of the maximum consensus homogeneity δ_n*

The measure δ_n is not local as it cannot be represented as an average similarity of pairs of orders: For a counterexample, consider the profile $\underline{R} = (R_1, R_1, R_1, R_2, R_2)$ on $C = \{a, b, c\}$, where R_1 ranks $a\ b\ c$ and R_2 ranks $c\ b\ a$. An arbitrary homogeneity measure h_n based on average pairwise similarities would satisfy $h_5(\underline{R}) = \frac{1}{10}[3 \cdot h_2(R_1, R_1) + 6 \cdot h_2(R_1, R_2) + h_2(R_2, R_2)]$. However, the maximum homogeneity measure δ_n does not satisfy this identity, since we have $\delta_5(\underline{R}) = 0.6$, but at the same time $\frac{1}{10}[3 \cdot \delta_2(R_1, R_1) + 6 \cdot \delta_2(R_1, R_2) + \delta_2(R_2, R_2)] = 0.7$.

Note that, despite their locality, homogeneity measures based on pairwise distances of relations satisfy the minimal requirements listed in Definition 1. Conditions (S1) and (S2) are trivially true. For (S3), the triangle inequality can be used: If an order R changes from a non-majority order to the majority order R^* then the distances $d(R, R^*)$ change to $d(R^*, R^*) = 0$ and the distances $d(R, R_j)$ from R to non-majority orders R_j change to $d(R^*, R_j)$. With $d(R^*, R_j) \leq d(R^*, R) + d(R, R_j)$ we get $d(R, R^*) \geq d(R^*, R_j) - d(R, R_j)$. Since in addition to the order R that changes to R^* there are as least as many majority orders as non-majority orders, we can match every increase in distance associated with a pair (R, R_j) to a decrease associated to $d(R, R^*)$ that is greater or equal, so the overall change in the sum of all distances can only be a decrease or zero.

A4: *Results for the linear function* $u_1(x) = x$ *and the concave function* $u_3(x) = 36 - (6 - x)^2$

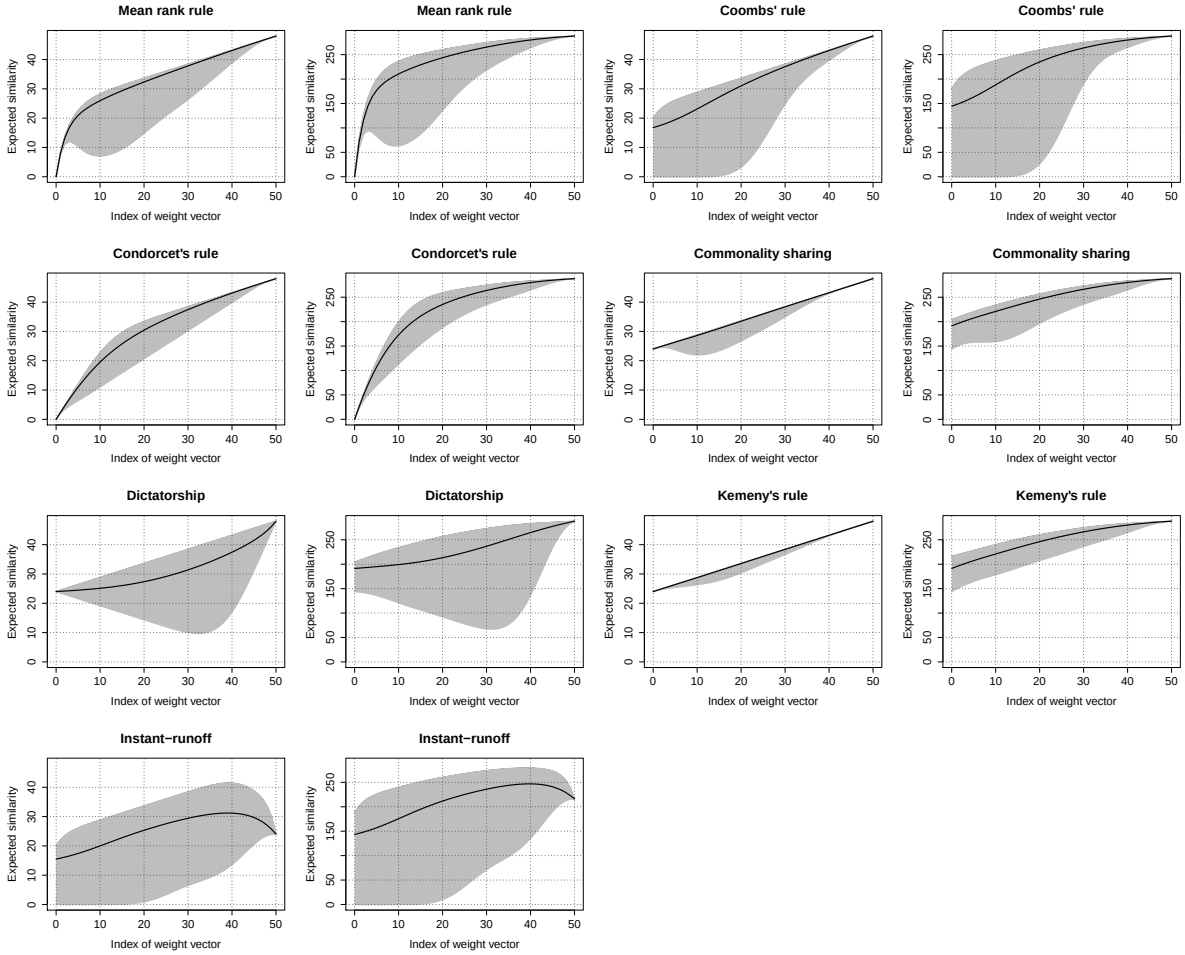

Figure 3: The graphs show the credal approximation $M^u_{G_n}(S)$ (gray shaded region) and the maximum entropy approximation $m^{u*}_{G_n}(S)$ (black line) for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \ldots, 50\}$. For the same S, the left graph corresponds to the choice $u(x) = u_1(x) = x$, whereas the right graph corresponds to the choice $u(x) = u_3(x) = 36 - (6 - x)^2$.

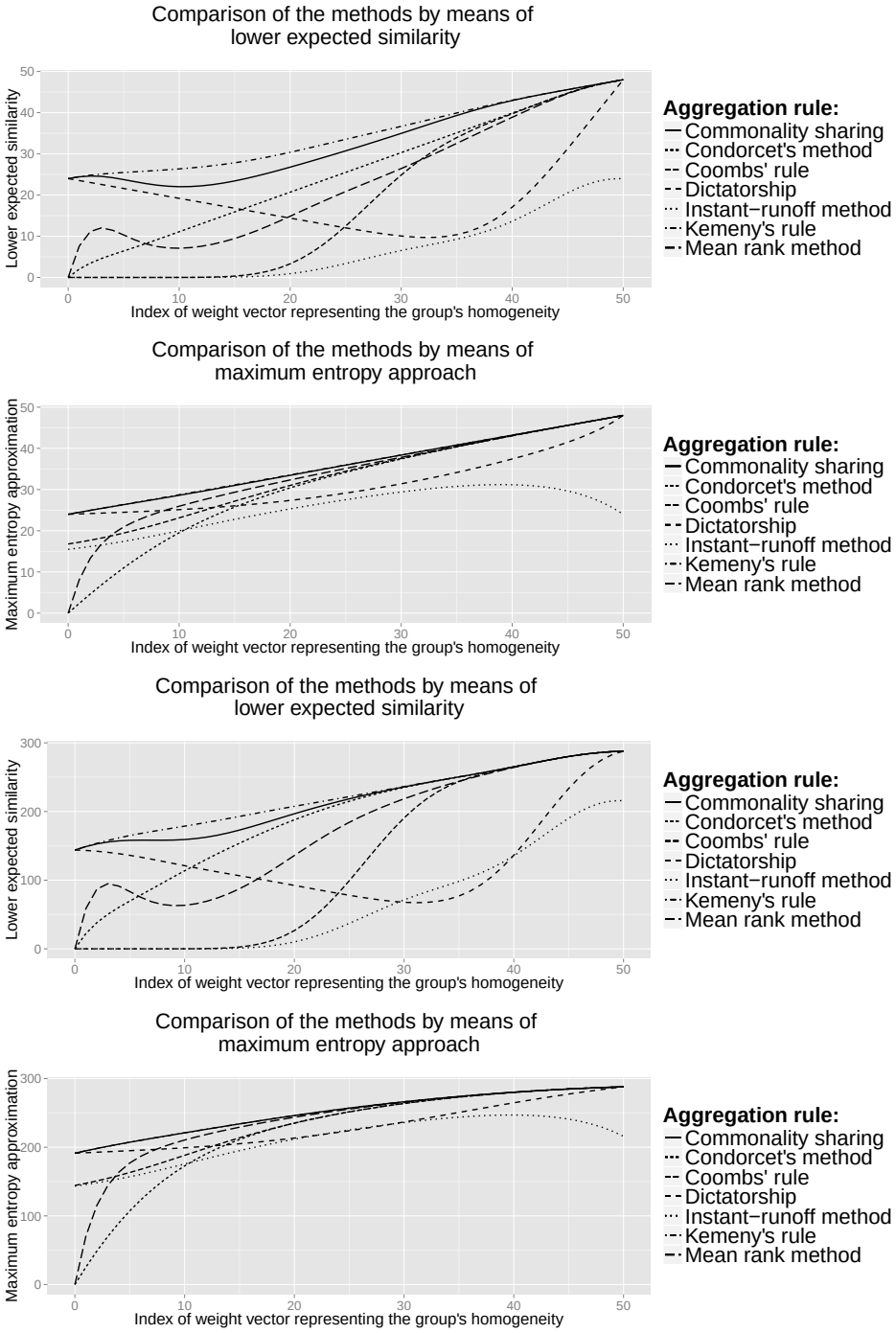

Figure 4: Graphs 1 and 3 show the lower expected similarity $\underline{M}^u_{G_n}(S)$ for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \ldots, 50\}$ for the choices $u(x) = u_1(x) = x$ and $u(x) = u_3(x) = 36 - (6 - x)^2$. Graphs 2 and 4 show the same for the maximum entropy approximation $m^{u*}_{G_n}(S)$.

Contribution 5:

Schollmeyer, G.; *Jansen, C.*; Augustin, T. (2017): Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems. Department of Statistics (LMU München): *Technical Reports*, Nr. 209, August 2017.

Available under:

`https://epub.ub.uni-muenchen.de/40416/`

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK

Georg Schollmeyer, Christoph Jansen, Thomas Augustin

Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems

Technical Report Number 209, 2017
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Detecting stochastic dominance for poset-valued random variables as an example of linear programming on closure systems

Georg Schollmeyer Christoph Jansen Thomas Augustin

Abstract

In this paper we develop a linear programming method for detecting stochastic dominance for random variables with values in a partially ordered set (poset) based on the upset-characterization of stochastic dominance. The proposed detection-procedure is based on a descriptively interpretable statistic, namely the maximal probability-difference of an upset. We show how our method is related to the general task of maximizing a linear function on a closure system. Since closure systems are describable via their valid formal implications, we can use here ingredients of formal concept analysis. We also address the question of inference via resampling and via conservative bounds given by the application of Vapnik-Chervonenkis theory, which also allows for an adequate pruning of the envisaged closure system that allows for the regularization of the test statistic (by paying a price of less conceptual rigor). We illustrate the developed methods by applying them to a variety of data examples, concretely to multivariate inequality analysis, item impact and differential item functioning in item response theory and to the analysis of distributional differences in spatial statistics. The power of regularization is illustrated with a data example in the context of cognitive diagnosis models.

Keywords: stochastic dominance, multivariate stochastic order, linear programming, closure system, formal concept analysis, formal implication, Vapnik-Chervonenkis theory, regularization.

1 Introduction

Stochastic (first order) dominance plays an important role in a huge 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 (IRT, 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 cite just a few. Most treatments of stochastic dominance are devoted to the univariate case with emphasis also on higher order stochastic dominance

or to the classical multivariate case where one has $\mathbb{R}^d$-valued random variables with the natural ordering $\leq = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mid \forall i \in \{1, \ldots, d\} : x_i \leq y_i\}$. In this paper we treat the general case of random variables that have values in a partially ordered set[1] (poset) $\mathbb{V} = (V, \leq)$.

Detecting stochastic dominance in this general case is especially interesting in the context of multivariate inequality or poverty analysis (cf., [Alkire et al., 2015]) in the situation where one has more dimensions of inequality that are additionally possibly only of a partial ordinal scale of measurement. One thinkable dimension with an only partially ordered scale of measurement is the dimension education, because different highest educational achievements may be incomparable due to different specifics of different courses of education. In this paper, the example of multivariate inequality analysis will serve as a prototypic example of multivariate stochastic dominance analysis.

In contrast to the simple univariate case, for random variables with values in a partially ordered set the notion of stochastic dominance cannot be simply described with the distribution function, anymore[2]. For two random variables $X : \Omega \longrightarrow V$ and $Y : \Omega \longrightarrow V$ with values in a partially ordered set $(V, \leq)$, one says that X is (weakly) stochastically smaller than Y, denoted by $X \leq_{SD} Y$, if there exist random variables X' and Y' on a further probability space $(\Omega', \mathcal{F}', P')$ with $X \stackrel{d}{=} X', \quad Y \stackrel{d}{=} Y'$ and $P'(X' \leq Y') = 1$. The property of stochastic dominance can be characterized by three essentially equivalent, more constructive statements: The random variable X is stochastically smaller than the random variables Y if one of the three following conditions is satisfied[3]:

i) $P(X \in A) \leq P(Y \in A)$ for every (measurable) upset $A \subseteq V$

ii) $\mathbb{E}(u \circ X) \leq \mathbb{E}(u \circ Y)$ for every bounded non-decreasing Borel-measurable[4] function $u : V \longrightarrow \mathbb{R}$

iii) It is possible to obtain the density[5] f_X from the density f_Y by transporting probability mass from values v to smaller values $v' \leq v$.

In this paper we will deal with the problem of detecting stochastic dominance between two random variables X and Y for which one has observed an i.i.d. sample $(x_1, \ldots, x_{n_x})$

[1]This includes especially the multivariate case of $\mathbb{R}^d$ where the natural order $x \leq y \iff \forall i \in \{1, \ldots, d\} : x_i \leq y_i$ 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].

[2]If one would rely on the distribution function in the multivariate case, then one would get another order, the so-called lower orthant or upper orthant order, cf., e.g., [Müller and Stoyan, 2002].

[3]The equivalence between (ii) and (i) was shown by Lehmann [1955] and independently proved by Levhari et al. [1975]. The equivalence between (iii) and (i) is a consequence of Strassen's Theorem ([Strassen, 1965]), see Kamae et al. [1977].

[4]Here, we have to assume that $(V, \leq)$ can be equipped with an appropriate topology that makes it a partially ordered polish space.

[5]This statement is of course only equivalent if the densities f_X and f_Y actually exist.

of the unknown random variable X and an i.i.d. sample $(y_1, \ldots, y_{n_y})$ of the unknown random variable Y. Actually, one would be interested in detecting $X \leq_{SD} Y$, but one does not exactly know the true law of X and Y. So, here we will deal with detecting empirical stochastic dominance between X and Y, denoted by $X \leq_{\hat{SD}} Y$, where the true laws of X and Y are replaced by the corresponding empirical laws. The problem of statistical inference that is concerned with the question of how stochastic dominance w.r.t. the empirical laws can be translated to stochastic dominance w.r.t. the true laws will also be discussed in this paper. The typical situation in this paper will be the analysis of differences between two subpopulations of some population. The typical subpopulations analyzed in this paper will be subpopulations of male and female persons. Here, we think of the random variable X as the outcome of a random sample from the subpopulation of the male, and Y as a random sample from the subpopulation of the female persons. Note that in the formal definition of stochastic dominance one compares random variables on the same probability space $(\Omega, \mathcal{F}, P)$. In our case of comparing subpopulations we can ensure that X and Y are random variables on the same underlying probability space by thinking of jointly sampling from the male and the female subpopulation. Note that the notion of stochastic dominance does not rely on the possible dependencies between X and Y, because all terms involved in the characterizing properties $i) - iii)$ of stochastic dominance only rely on the marginal distribution of X and Y. Note further that the definition of stochastic dominance could thus be simply extended to random variables living on different probability spaces. Thus, also for the replacement of the true laws by empirical laws, different sample sizes for the male and the female samples would not introduce any problem, here.

For detecting stochastic dominance in the above sense, we will make substantial use of the upset-characterization $i)$. The characterization via a mass transfer can also be used to check for stochastic dominance, see, e.g., Mosler and Scarsini [1991] (for empirical applications see, e.g., [Arndt et al., 2012, 2013]), while an alternative approach would be to make use of a network flow formulation of the problem, as outlined in Preston [1974] or Hansel and Troallic [1978] and then check for dominance via computation of the maximum flow. The main reason for putting emphasis on the upset approach is that with this approach we could not only check for stochastic dominance, but we will also get additionally some well-interpretable statistic for free, upon which we can also base an attempt to do inference. Beyond this, the family of all upsets of a given poset is a well understood closure system[6] and a natural question is then, how the linear programming approach outlined here, can be generalized to the case of arbitrary closure systems.

The paper is structured as follows: In Section 2 we briefly introduce basic mathematical concepts of partially ordered sets, complete lattices and formal concept analysis needed in the paper. Section 3 develops and analyses a linear program for detecting first order sto-

[6]A closure system $\mathcal{S}$ is a family of subsets of a space Ω that contains Ω and is closed under arbitrary intersections.

chastic dominance for random variables with values in a poset. In Section 4 we generalize the linear programming approach to optimization on closure systems. Statistical inference for the developed methods, especially the application of Vapnik-Chervonenkis theory, possible regularization and characterizations of the Vapnik-Chervonenkis dimension of selected closure systems (as well as concretely computing the Vapnik-Chervonenkis dimension) are treated in Section 5. Examples of application, ranging from inequality analysis based on stochastic dominance to a geometrical generalization of the Kolmogorov-Smirnov test for spatial statistics are given in Section 6, while Section 7 concludes.

2 Mathematical preliminaries

In this section, we very briefly introduce elementary basics of partially ordered sets and of formal concept analysis. A far more detailed introduction to partially ordered sets can be found in Davey and Priestley [2002], which also gives a short introduction to formal concept analysis. An introduction into formal concept analysis can be found in Ganter and Wille [2012]. The concepts of formal concept analysis are actually only needed for the optimization problems on general closure systems indicated in Section 4, the reader only interested in the problem of detecting first order stochastic dominance can skip Section 2.2.

2.1 Ordered sets and lattices

Definition 1 (posets and lattices)**.** *A* **partially ordered set (poset)** $\mathbb{V} = (V, \leq)$ *is a pair of a set* V *and a binary relation* $\leq$ *on* V *that is reflexive transitive and antisymmetric. A poset* $(V, \leq)$ *is called* **linearly ordered***, if every two elements* x, y *of* V *are comparable (meaning that* $x \leq y$ *or* $y \leq x$*). For two different elements* x, y *of a poset* $\mathbb{V}$ *we say that* y *is an upper neighbor of* x *(or that* x *is a lower neighbor of* y*), and denote this by* $x \lessdot y$*, if* $x \leq y$ *and if there is no further element* $z \in V$ *(different from* x *and* y*) with* $x \leq z \leq y$*.*

A **lattice** $\mathbb{L} = (L, \leq)$ *is a poset such that every set* $\{x, y\}$ *of two elements* $x, y \in L$ *has a least upper bound and a greatest lower bound. A lattice is called* **complete***, if every arbitrary set* M *has a least upper bound and a greatest lower bound. The least upper bound of a set* M *is called* **supremum** *or* **join** *of* M *and it is denoted with* $\bigvee M$*. The greatest lower bound of a set* M *is called* **infimum** *or* **meet** *of* M *and it is denoted with* $\bigwedge M$*. An element* x *of a complete lattice* $(L, \leq)$ *is called* **join-irreducible** *if for arbitrary subsets* $B \subseteq L$ *from* $x = \bigvee B$ *it follows* $x = b$ *for some* $b \in B$*. The set of all join-irreducible elements of a poset* $\mathbb{V}$ *is denoted with* $\mathcal{J}(\mathbb{V})$*.*

Definition 2 (upset and downset, principal ideal and principal filter)**.** *Let* $(V, \leq)$ *be a poset. A set* $V \subseteq M$ *is called an* ***upset*** *(or filter) if we have* $\forall x, y \in V : x \leq y \ \& \ x \in M \Longrightarrow y \in M$*. A subset* $M \subseteq V$ *is called* ***downset*** *(or ideal) if* $\forall x, y \in V : x \leq y \ \& \ y \in M \Longrightarrow x \in M$*. The set of all upsets of a poset* $(V, \leq)$ *is denoted with* $\mathcal{U}((V, \leq))$ *and the set of all downsets is denoted with* $\mathcal{D}((V, \leq))$*. An upset of the form* $\uparrow x := \{y \in V \mid y \geq x\}$

with $x \in V$ is called a **principal filter**. *A downset of the form $\downarrow x := \{y \in V \mid y \leq x\}$ with $x \in V$ is called a* **principal ideal.**

Remark 1. *The complement of an upset is a downset and the complement of a downset is an upset.*

Definition 3 (chain, antichain and width)**.** *Let $(V, \leq)$ be a poset. A set $M \subseteq V$ is called a* ***chain*** *if every two arbitrary elements x and y of M are comparable (meaning that $x \leq y$ or $y \leq x$). A subset M of a poset $(V, \leq)$ is called an* ***antichain*** *if every two arbitrary different elements x and y of M are incomparable (meaning that neither $x \leq y$ nor $y \leq x$). The* **width** *of a finite poset $(V, \leq)$ is the maximal cardinality of an antichain of $(V, \leq)$.*

Remark 2. *For every upset M the set $\min M$ of all minimal elements of M is an antichain. Furthermore, every finite upset M can be characterized by its minimal elements as $M = \uparrow \min M := \{x \in V \mid \exists y \in \min M : y \leq x\}$. Analogous statements are valid for downsets.*

Definition 4 (order dimension)**.** *The* **order dimension** *of a poset $(V, \leq)$ is the smallest number k such that there exist k linearly ordered sets $(V, L_1), \ldots, (V, L_k)$ that represent $(V, \leq)$ via $\leq = \bigcap_{i=1}^{k} L_i$.*

2.2 Formal concept analysis

Formal concept analysis (FCA) is an applied mathematical theory rooted in an attempt to mathematically formalize the notion of a *concept*. In its origins initially motivated by some philosophical attempt to restructure lattice theory (cf., [Wille, 1982]) it nowadays also has very broad applications in computer science, for example in data mining, text mining, machine learning or knowledge management, to name just a few.

Concretely, in formal concept analysis one starts with a so-called **formal context** $\mathbb{K} = (G, M, I)$ where G is a set of objects, M is a set of attributes and $I \subseteq G \times M$ is a binary relation between the objects and the attributes with the interpretation $(g, m) \in I$ iff object g has attribute m. If $(g, m) \in I$ we also use infix notation and write gIm. In the context of statistical data analysis, the objects are often the data points, for example the persons that participated in a survey.The attributes are the observed values of the interesting variables, for example the answer *yes* or *no* to the posed questions and gIm means that person g answered the question m with *yes*. (If the answers to the questions in a survey are not binary, then one can transform them into binary attributes with the method of conceptual scaling, see below.) A **formal concept** of the context $\mathbb{K}$ is a pair (A, B) of a set $A \subseteq G$ of objects, called **extent**, and a set $B \subseteq M$ of attributes, called **intent**, with the following properties:

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 $m \in M \backslash A$ that is also shared by all objects $g \in A$ (i.e. $\forall m \in M : (\forall g \in A : gIm) \Longrightarrow m \in B$).

Conceptually, the concept extent describes, which objects belong to the formal concept and the intent describes, which attributes characterize the concept. The property of being a formal concept can be characterized with the following operators

$$\begin{aligned} \Phi : 2^M \longrightarrow 2^G : &\quad B \mapsto \{g \in G \mid \forall m \in B : gIm\} \\ \Psi : 2^G \longrightarrow 2^M : &\quad A \mapsto \{m \in M \mid \forall g \in A : gIm\} \end{aligned}$$

as

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

This can be verbalized as: "The pair (A, B) is a formal concept iff B is exactly the set of all common attributes of the objects of A and A is exactly the set of all objects having all attributes of B." In the sequel, we will abbreviate both Ψ and Φ with $'$. (Which of the two operators is meant will be always clear from the context.) Furthermore, for singleton sets $\{g\} \subseteq G$ or $\{m\} \subseteq M$ we abbreviate $\{g\}'$ by g' and $\{m\}'$ by m'.

On the set of all formal concepts we can define a subconcept relation as

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

(Actually, for formal concepts the equivalence $A \subseteq C \iff B \supseteq D$ holds.) If the concept (A, B) is a subconcept of (C, D) then it is a more specific concept containing less objects that have more attributes in common. The set of all formal concepts of a context $\mathbb{K}$ together with the subconcept relation is called the **concept lattice** of $\mathbb{K}$ and it is denoted with $\mathfrak{B}(\mathbb{K})$. The concept lattice is in fact a complete lattice. The set of the concept extents of all formal concepts of $\mathfrak{B}(\mathbb{K})$ is denoted with $\mathfrak{B}_1(\mathbb{K})$ and the set of all concept intents is denoted with $\mathfrak{B}_2(\mathbb{K})$. The family of sets $\mathfrak{B}_1(\mathbb{K})$ is a closure system on G and the family $\mathfrak{B}_2(\mathbb{K})$ is a closures system on M: A (set-theoretic) **closure system** $\mathcal{S}$ on a space Ω is a family $\mathcal{S} \subseteq 2^\Omega$ of subsets of Ω that contains the space Ω and is closed under arbitrary intersections. If a family $\mathcal{F}$ of subsets of a space Ω is not a closure system, one can compute its **closure** $\mathrm{cl}(\mathcal{F}) := \bigcap\{\mathcal{S} \mid \mathcal{S} \supseteq \mathcal{F} \ \& \ \mathcal{S} \text{ is a closure system on } \Omega\}$ that is the smallest closure system containing all sets of $\mathcal{F}$.

Every closure system $\mathcal{S}$ on Ω can be described by all valid formal implications of $\mathcal{S}$: A **formal implication** is a pair (Y, Z) of subsets of Ω, also denoted by $Y \longrightarrow Z$. We say that an implication $Y \longrightarrow Z$ is **valid** in a family $\mathcal{S}$ of subsets of Ω (which needs not to be a closure system) if every set of $\mathcal{S}$ that contains all elements of Y also contains all elements of Z. In this case we also say that the family $\mathcal{S}$ **respects** the implication $Y \longrightarrow Z$. Similarly, we say that a subset of Ω respects an implication $Y \longrightarrow Z$ if it either is not a superset of Y or if it is a superset of Z. The first component of a formal implication

is also called the **premise** or the **antecedent** and the second component is also called **conclusion** or the **consequent** of the formal implication. A formal implication is called **simple** if its premise is a singleton.

A closure system $\mathcal{S}$ can be characterized by formal implications as follows: Define for $\mathcal{S}$ the set $\mathfrak{I}(\mathcal{S})$ of all formal implications that are valid in $\mathcal{S}$. Then, given the set $\mathfrak{I}(\mathcal{S})$, the closure system $\mathcal{S}$ can be rediscovered from $\mathfrak{I}(\mathcal{S})$ as the set of all subsets of Ω that respect all formal implications of $\mathfrak{I}(\mathcal{S})$. The set $\mathfrak{I}(\mathcal{S})$ of all valid implications of a closure system $\mathcal{S}$ is usually very large. To efficiently describe a closure system, it suffices to look at a so-called implication base of $\mathfrak{I}(\mathcal{S})$: Given an arbitrary set $\mathfrak{I}$ of formal implications, we say that a further set $\mathfrak{J}$ of implications is a **base** of $\mathfrak{I}$, if we have

$$\forall M \subseteq \Omega: \quad M \text{ respects all implications of } \mathfrak{I} \iff M \text{ respects all implications of } \mathfrak{J} \qquad (1)$$

and if furthermore $\mathfrak{J}$ is minimal w.r.t. this property, i.e. for every other subset $\mathfrak{J}' \subsetneq \mathfrak{J}$ the equivalence (1) is not valid anymore. In the sequel, we will mainly deal with formal implications of the closure system of the concept intents of a given formal context $\mathbb{K}$. Such implications are sometimes also called attribute-implications to indicate that one is speaking about implications between attributes and not between objects of a context. Here, we will always use the short term implications and will also say that an implication is valid in a context $\mathbb{K}$ instead of saying that an implication is valid in the closure system of all concept intents of $\mathbb{K}$.

In the context of statistical data analysis one often has data that are not binary but for example categorical with more than two possible values. To analyze such data with methods of formal concept analysis one can use the technique of **conceptual scaling** (cf. [Ganter and Wille, 2012, p.36-45]) to fit the categorical data into a binary setting: For a categorical variable with the possible values in $\{1, \ldots, K\}$ one can introduce the K attributes "$= 1$", ..., "$= K$" and say that an object g has attribute "$= i$" if the value of K equals i. In a similar way, for an ordinal variable with possible values $\{1 < 2 < \ldots < K\}$ we can introduce the attributes "≤ 1","≤ 2",..., "$\leq K$" and say that object g has attribute "$\leq i$" if the value of object g is lower than or equal to i. One can also additionally introduce the attributes "≥ 1", ..., "$\geq K$". This concrete way of conceptually scaling an ordinal variable is called **interordinal scaling** and will be used in one example of application given in Section 6.2.

3 Detecting stochastic dominance

We now turn to the development of a technique for detecting stochastic dominance for poset-valued random variables based on linear programming and the upset-characterization of stochastic dominance.

3.1 Characterizing stochastic dominance via linear programming

Let $(V, \leq) = (\{v_1, \ldots, v_k\}, \leq)$ be a finite poset[7], let $x = (x_1, \ldots, x_{n_x})$ be an i.i.d. sample of a random variable X and let $y = (y_1, \ldots, y_{n_y})$ be an i.i.d. sample of Y. Let $w^x = (w_1^x, \ldots, w_k^x)$ where w_i^x denotes the number of observed samples of X with value v_i, divided by n_x. Analogously, let $w^y = (w_1^y, \ldots, w_k^y)$ where w_i^y denotes the number of samples of Y with value v_i, divided by n_y. With $\mathcal{U}((V, \leq))$ we denote the set of all upsets of $(V, \leq)$. We identify an upset $M \in (V, \leq)$ with its characteristic vector $m \in \{0, 1\}^k$ via $m_i = 1 \iff v_i \in M$. Additionally, we also identify the relation $\leq$ with the relation $\{(i, j) \mid i, j \in \{1, \ldots, k\}, v_i \leq v_j\}$, the same for the covering relation $\lessdot$. To the samples x and y we associate the empirical analogue $\hat{P}$ of the true law P via $\hat{P}(X = v_i) = w_i^x$ and $\hat{P}(Y = v_i) = w_i^y$. To check if $X \leq_{\hat{SD}} Y$ we have to check

$$\forall M \in \mathcal{U}((V, \leq)) : \hat{P}(X \in M) \leq \hat{P}(Y \in M). \tag{2}$$

Obviously, $\hat{P}(X \in M) = \langle w^x, m \rangle$ and $\hat{P}(X \in M) = \langle w^x, m \rangle$, so (2) is equivalently characterizable as

$$\begin{aligned}
& \forall M \in \mathcal{U}((V, \leq)) : && \hat{P}(X \in M) \leq \hat{P}(Y \in M) \\
\iff & \forall M \in \mathcal{U}((V, \leq)) : && \langle w^x, m \rangle \leq \langle w^y, m \rangle \\
\iff & \forall M \in \mathcal{U}((V, \leq)) : && \langle w^x, m \rangle - \langle w^y, m \rangle \leq 0 \\
\iff & \forall M \in \mathcal{U}((V, \leq)) : && \langle w^x - w^y, m \rangle \leq 0 \\
\iff & \sup_{M \in \mathcal{U}((V, \leq))} && \langle w^x - w^y, m \rangle \leq 0.
\end{aligned}$$

This means that the problem is characterizable as a linear program over the family $\mathcal{S} := \mathcal{U}((V, \leq))$ of subsets of V. To solve this program we can look at the concrete structure of the family $\mathcal{S}$. The family $\mathcal{S}$ consists of all upsets of $(V, \leq)$, i.e., of all sets M satisfying

$$\forall i, j \in \{1, \ldots, k\} : \quad v_i \in M \ \& \ v_i \leq v_j \Longrightarrow v_j \in M$$

which is equivalent to

$$\forall i, j \in \{1, \ldots, k\} : \quad v_i \leq v_j \Longrightarrow m_j \geq m_i$$

and this set of inequalities can be easily implemented in a linear program:

We have $X \leq_{\hat{SD}} Y$ if and only if the linear binary program

$$\begin{aligned}
\langle w^x - w^y, m \rangle \longrightarrow \max & \qquad (3) \\
w.r.t. & \\
m \in \{0, 1\}^k & \\
\forall (i, j) \in \leq : \quad m_j \geq m_i &
\end{aligned}$$

[7]This is actually no restriction because we are in the first place interested in detecting stochastic dominance for samples that are always finite.

has a maximal value of zero. (Note that the maximal value of (3) is always at least 0, because for $M = \emptyset$ we have $\langle w^x - w^y, (0, \ldots, 0)\rangle = 0$.) If one analyzes this above binary program further (see the last paragraph of Section 4.1 at page 21), one sees that it is not necessary to take the m_i as binary variables, one can relax the integrality conditions and solve instead the far more simple classical linear program

$$\begin{aligned} \langle w^x - w^y, m\rangle &\longrightarrow \max \\ w.r.t. & \\ m &\in [0,1]^k \\ \forall (i,j) \in \leq: \quad m_j &\geq m_i \end{aligned} \tag{4}$$

which could be further simplified to

$$\begin{aligned} \langle w^x - w^y, m\rangle &\longrightarrow \max \\ w.r.t. & \\ m &\in [0,1]^k \\ \forall (i,j) \in \lessdot: \quad m_j &\geq m_i. \end{aligned} \tag{5}$$

In the sequel we will denote the maximal value of (5) with D^+ and the optimal value one would get if one would replace maximization by minimization in (5) with D^-.

3.2 Some analysis of the linear programming approach for detecting stochastic dominance

The obtained linear program for checking dominance involves k decision variables and $|\lessdot| + k$ inequalities, where $|\lessdot|$ can be shown to be bounded by $\lfloor \frac{k}{2} \rfloor \cdot \lceil \frac{k}{2} \rceil$, which indicates that the linear program is practically manageable for real data sets. One interesting question in this context is how the feasible set of the linear program looks like in special situations and what for example the simplex-algorithm would actually do. In applied situations, the poset $(V, \leq)$ is often of the form $V = \mathbb{R}^d$ or $\{0, \ldots, K\}^d$ and $x \leq y \iff \forall i \in \{1, \ldots, d\} : x_i \leq y_i$. For checking stochastic dominance only the actually observed $x \in V$ are of interest.This helps in reducing the effective size of the poset V but at the same times makes the structure of V only implicitly given. Thus, a general analysis seems to be difficult and we therefore restrict the analysis in Section 3.2.1 to some simple examples.

3.2.1 Some examples

In this section we exemplarily discuss some examples for posets $(V, \leq)$ of the form $V = \{0, \ldots, K\}^d$ and $x \leq y \iff \forall i \in \{1, \ldots, d\} : x_i \leq y_i$. We start with the simplest example

where $d = 1$ which corresponds to a linearly ordered set $V = \{0 < 1 < \ldots < K\}$. Then the linear program (5) would translate to

$$\begin{aligned} \langle w^x - w^y, m\rangle &\longrightarrow \max \\ &w.r.t. \\ m &\in [0,1]^k \end{aligned}$$

$$\underbrace{\begin{pmatrix} 1 & -1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 \\ & & \vdots & & & \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix}}_{=:A} \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

In this case the extreme points of the feasible set are simply the vectors of the form $m^l = (0, \ldots, \underbrace{0}_{\text{l-th entry}}, 1, 1, \ldots)$, where $l \in \{0, \ldots, k\}$. For $l, l' \in \{1, \ldots, k-1\}$ it is easy to see that every two different extreme points m^l and $m^{l'}$ are adjacent because A has full rank and for m^l the inequality constraint associated to the $l-th$ row of A is strict where the other inequalities are actually equalities and to "switch" from m^l to $m^{l'}$ one simply has to switch the $l'-th$ variable from basis to non-basis and the $l-th$ variable form non-basic to basic. A similar argumentation shows that also for arbitrary $l, l' \in \{0, \ldots, k\}$ every two different extreme points are adjacent which means that applying the simplex algorithm would in this case exactly mean that one scans every extreme point, i.e. every upset, so the simplex algorithm is not better than a naive inspection of every upset. However in the case of a linearly ordered set the number of upsets is $|V|$ and thus no problem from a computational point of view.

Now we come to the more difficult cases of $d > 1$. In these situations the feasible set of the linear program appears to be not so easily describable, there seems to be no simple rule that says which extreme points are adjacent. Table 1 gives lower and upper bounds[8] for the size u of the closure system of all upsets of $\{0, \ldots, K\}^d$ for different values of K and d. One can see that for high enough K or d the closure system is very big and explicitly checking all upsets is clearly not applicable. Compared to this, in Table 2 one can see the

[8]The upper bounds were computed with the help of the Sauer-Shelah lemma ([Sauer, 1972, Shelah, 1972]). The Sauer-Shelah lemma is also closely related to Vapnik-Chervonenkis theory which we use in Section 5.2, see also Bottou [2013] or `http://leon.bottou.org/_media/papers/vapnik-symposium-2011.pdf` for the curious history of the Sauer-Shelah lemma. The lower bounds were obtained by noting that for every $l \in \{1, \ldots, K\}$ the set $A_l := \{x \in \{0, 1, \ldots, K\}^d \mid \sum_{i=1}^{d} x_i = l\}$ of all K-bounded multisets of rank l is an antichain and thus for every non-empty set $B \subseteq A_l$ we get a different upset $\uparrow B$. Thus, $u \geq \sum_{l=1}^{K} (2^{|A_l|} - 1) + 2$, where the last $+2$ comes from noting that also the empty set and the whole set V are upsets, and the cardinality $|A_l|$ can be computed recursively.

number of iterations a dual simplex algorithm did take to get a solution. (For the objective function we simply took a standard normally distributed sample.) This indicates that with the linear programming approach the problem is still solvable for larger values of K and d.

		d						
		1	2	3	4	5	6	7	8
K=1	lower bound	2	5	16	95	2110	1.1e+06	6.9e+10	1.2e+21
	upper bound	2	10	9.2e+01	1.5e+04	1.1e+08	3.4e+16	5.1e+31	1.5e+64
K=2	lower bound	3	15	2.7e+02	6.6e+05	2.3e+15	2.8e+42	2.0e+118	
	upper bound	3	129	1.3+06	2.1e+18	1.4e+53	1.5e+154		
K=3	lower bound	4	37	1.0e+04	2.0e+13	9.1e+46	4.0e+174		
	upper bound	4	2516	4.2e+12	8.6e+49	4.6e+187			
K=4	lower bound	5	83	1.1e+06	4.1e+25	4.9e+114			
	upper bound	5	68405	1.6e+22	4.7e+106				
K=5	lower bound	6	177	3.4e+08	9.2e+43	1.3e+235			
	upper bound	6	2391495	2.1e+34	5.5e+196				
K=6	lower bound	7	367	2.9e+11	3.5e+69				
	upper bound	7	102022809	7.0e+49					
K=7	lower bound	8	749	7.1e+14	3.6e+103				
	upper bound	8	5130659560	1.0e+68					
K=8	lower bound	9	1515	4.9e+18	1.6e+147				
	upper bound	9	296881218693	6.9e+89					

Table 1: Upper and lower bounds for the size u of the closure system of all upsets of $\{1, \ldots, K\}^d$ for different values of K and d.

	d						
K	1	2	3	4	5	6	7
1	0	0	7	18	18	92	239
2	4	3	19	156	796	3861	23002
3	3	78	208	1901	4456	24628	27271
4	17	86	626	3518	23002	24173	24923
5	12	200	2380	10987			
6	29	353	2023	23002			
7	60	396	4959				
8	87	572	7698				

Table 2: Number of iterations for solving the linear program via dual simplex for detecting stochastic dominance for $V = \{0, \ldots, K\}^d$ for different values of K and d. The objective function was a standard normally distributed random sample.

3.2.2 Duality

In this section, we analyze the dual linear program of program (4) for detecting first order stochastic dominance. The most interesting inside will be that this dual program can be interpreted as a special kind of mass transportation problem. In order to determine the dual program of program (4), first note that the second class of constraints of problem (4) can equivalently be rewritten as

$$\forall i, j \in \{1, \ldots, k\} : m_i \geq I_{ij} \cdot m_j \tag{6}$$

where $I_{ij} := \mathbb{1}_{<}((v_j, v_i))$ and $<$ denotes the strict part of the partial order $\leq$. By defining for each $i \in \{1, \ldots, k\}$, the matrix $M^{(i)} \in \mathbb{R}^{k \times k}$ via

$$M^{(i)}_{pq} = \begin{cases} I_{ip} & \text{if } p = q \\ -1 & \text{if } q = i \wedge q \neq p \\ 0 & \text{else} \end{cases} \tag{7}$$

one then can reformulate the linear programming problem (4) by the equivalent linear programming problem

$$\begin{aligned} \langle w^x - w^y, m \rangle &\longrightarrow \max \\ &w.r.t. \\ m_1, \ldots, m_k &\geq 0 \\ \begin{pmatrix} E_k \\ M^{(1)} \\ \vdots \\ M^{(k)} \end{pmatrix} \cdot m &\leq (\underbrace{1, \ldots, 1}_{k-times}, \underbrace{0, \ldots, 0}_{k^2-times})^T =: b \end{aligned} \tag{8}$$

where E_k denotes the k-dimensional unit matrix. Define $w^{xy} := w^x - w^y$ and $z := (x_1, \ldots, x_k, z_{11}, \ldots, z_{1k}, \ldots, z_{k1}, \ldots, z_{kk})$ and let b be defined as in the constraints of the above linear program (8). Then the dual linear program of (8) is given by:

$$\begin{aligned} \sum_{l=1}^{k} x_l = \langle b, z \rangle &\longrightarrow \min \\ &w.r.t. \\ z &\in \mathbb{R}^{k+k^2}_{\geq 0} \\ \begin{pmatrix} E_k & M^{(1)T} & \ldots & M^{(k)T} \end{pmatrix} \cdot z &\geq \begin{pmatrix} w^{xy}_1 \\ \vdots \\ w^{xy}_k \end{pmatrix} \end{aligned} \tag{9}$$

In order to investigate what duality theory can teach us about our original problem, we rewrite the program (9) as:

$$\sum_{l=1}^{k} x_l \longrightarrow \min \tag{10}$$

$$w.r.t.$$

$$z \in {\mathbb{R}_+}^{k+k^2}$$

$$\forall i \in \{1,\ldots,k\} : x_i - \sum_{s\in\{1,\ldots,k\}\setminus\{i\}} z_{is} + \sum_{s\in\{1,\ldots,k\}\setminus\{i\}} I_{si}\cdot z_{si} \geq w_i^{xy}$$

For variables z_{is} with $I_{is} = 0$, for finding an optimal solution one can always set z_{is} to zero, because such z_{is} are not present in the objective function and do occur separated only in the $i-th$ inequality constraint with a negative sign. Thus, the program can be simplified to

$$\forall i \in \{1,\ldots,k\} : x_i - \sum_{s\in\{1,\ldots,k\}\setminus\{i\}} I_{is}\cdot z_{is} + \sum_{s\in\{1,\ldots,k\}\setminus\{i\}} I_{si}\cdot z_{si} \geq w_i^{xy},$$

which again can be simplified to

$$\forall i \in \{1,\ldots,k\} : x_i - \sum_{\{s:v_s<v_i\}} z_{is} + \sum_{\{s:v_i<v_s\}} z_{si} \geq w_i^{xy}. \tag{11}$$

Note that the resulting program (10) with the rewritten version (11) of the constraints is very similar, yet not identical to the mass transport algorithm for detecting stochastic dominance discussed in [Range and Østerdal, 2013, p. 5]: In case the optimal objective of the program equals 0, the values z_{ij}^* can be interpreted as the probability masses that need to be transported from strictly greater elements to strictly smaller elements w.r.t. $\leq$ in order to obtain the distribution of X from the distribution of Y (which exactly corresponds to characterization iii) of first order stochastic dominance that was recalled in the introduction). The main difference of our program (10) and the problem discussed in [Range and Østerdal, 2013, p. 5] is that, while there the authors have two classes of constraints, one class for the masses transported into each node and one class for the masses transported out of each node, our set of constraints considers the masses that are transported inside in- and out of each node simultaneously.

Note that there are also attempts to interpret the value of the sum $\sum_{ij} z_{ij}^*$ of the optimal z_{ij}^* values, or a weighted version of it in cardinal settings (see [Tarp and Østerdal, 2007, p.19-20]), as a measure for the extent of stochastic dominance that is given in the situation under consideration. However, as discussed in further detail in Section 3.3, in this paper we argue that in order to detect the extent of stochastic dominance using the optimal value of (10) might be a more sensible indicator for the extent of stochastic dominance since it avoids certain counter-intuitive characteristics.

In order to get a better impression of the structure of the above dual programming problem, we consider the following example: Suppose the poset V consists of seven elements, namely $V = \{v_1, \dots, v_7\}$. Moreover, suppose the partial order $\leq$ is specified by the following incidence matrix M:

$$M = \begin{array}{c|ccccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \\ \hline v_1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ v_2 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ v_3 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ v_4 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ v_5 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ v_6 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ v_7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

where we have that $M_{ij} = 1$ if and only if $v_i \leq v_j$. Finally, suppose we have observed samples of X and Y and computed the vectors w^x and w^y. Then, the dual linear programming problem from (10) takes the following form:

$$\begin{aligned} \sum_{l=1}^{7} x_l &\longrightarrow \min \\ w.r.t. & \\ (x_1, \dots, x_7, z_{11}, z_{12}, \dots, z_{76}, z_{77}) &\in {\mathbb{R}_+}^{56} \\ x_1 + (z_{21} + z_{31} + z_{41} + z_{51} + z_{61} + z_{71}) &\geq w_1^{xy} \\ x_2 - (z_{21}) + (z_{52} + z_{62} + z_{72}) &\geq w_2^{xy} \\ x_3 - (z_{31}) + (z_{53} + z_{63} + z_{73}) &\geq w_3^{xy} \\ x_4 - (z_{41}) + (z_{54} + z_{64} + z_{74}) &\geq w_4^{xy} \\ x_5 - (z_{51} + z_{52} + z_{53} + z_{54}) + (z_{75}) &\geq w_5^{xy} \\ x_6 - (z_{61} + z_{62} + z_{63} + z_{64}) + (z_{76}) &\geq w_6^{xy} \\ x_7 - (z_{71} + z_{72} + z_{73} + z_{74} + z_{75} + z_{76}) &\geq w_7^{xy} \end{aligned}$$

First, consider the observed samples lead to vectors $w^x = (\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7})$ and $w^y = (\frac{1}{7}, 0, 0, 0, \frac{2}{7}, \frac{2}{7}, \frac{2}{7})$. For that case, the optimal objective of the above programming problem is 0 (which, due to duality and Proposition 3.1, also indicates that Y first-order stochastic dominates X). A corresponding optimal solution vector is given by $(x_1^*, \dots, x_7^*, z_{11}^*, \dots, z_{77}^*)$, where every component equals 0 except of $z_{54}^* = z_{63}^* = z_{72}^* = \frac{1}{7}$. As discussed before, the z_{ij}^* variables exactly describe how the distribution of X can be obtained from the distribution of Y by a finite number of probability mass transfers to strictly smaller elements with respect to the partial order $\leq$. In our example, the distribution of X can be obtained from that of Y by transferring mass $\frac{1}{7}$ from node v_5 to v_4, mass $\frac{1}{7}$ from node v_6 to v_3 and mass $\frac{1}{7}$ from node v_7 to v_2. This is illustrated in Figure 1.

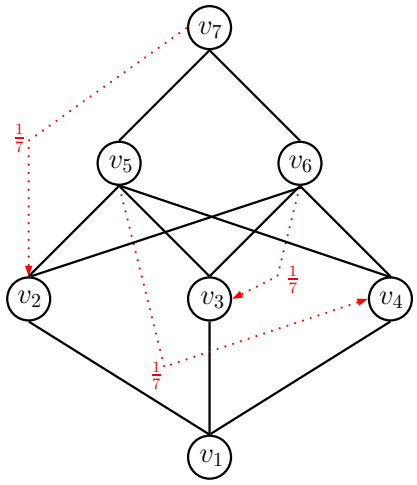

Figure 1: Mass transfer problem for $w^x = (\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7})$ and $w^y = (\frac{1}{7}, 0, 0, 0, \frac{2}{7}, \frac{2}{7}, \frac{2}{7})$.

A natural question is the following: Do optimal solutions of the programming problem (10) still possess a meaningful interpretation for the case that Y does not stochastically dominate X? To address this question, suppose we, instead of the previous situation, observed the vectors $w^x = \frac{1}{28} \cdot (4, 5, 6, 2, 1, 3, 7)$ and $w^y = \frac{1}{28} \cdot (4, 2, 5, 7, 6, 1, 3)$. In that case, the optimal objective of our example is $\frac{6}{28}$ (which indicates that Y does not stochastically dominate X by the same argument as given above) and an optimal solution vector is given by $(x_1^*, \ldots, x_7^*, z_{11}^*, \ldots, z_{77}^*)$, where all components equal 0 except $x_6^* = \frac{2}{28}$, $x_7^* = \frac{4}{28}$, $z_{52}^* = \frac{3}{28}$ and $z_{53}^* = \frac{1}{28}$. Indeed, also in the case of a non-dominant Y we receive a straightforward interpretation: Compared to the case of stochastic dominance, where the whole probability mass can be transported from higher values to lower values to obtain X from Y, in the case of non-dominance, not all mass can be transported and the optimal value of (10) could be understood as the amount of probability mass that cannot be transported and thus has to be externally introduced to supply X with enough probability mass. Again, the optimal solution is illustrated in Figure 2.

3.3 The minimal value as a measure of the extent and the argmin as an insight into the actual manifestation of dominance

With the linear program (5) we can detect stochastic dominance. However, as already betoken, generally one is not only interested in the presence or absence of stochastic dominance, one would also like to get some rough idea about the "extent" of dominance. In our very general setting of random variables/data with only a partially ordered scale of measurement, a reasonable definition of the term *extent of dominance* is not straight forward. Therefore, we will firstly go one step back and reconstruct, how the upset characterization of stochastic dominance, that was introduced only in purely

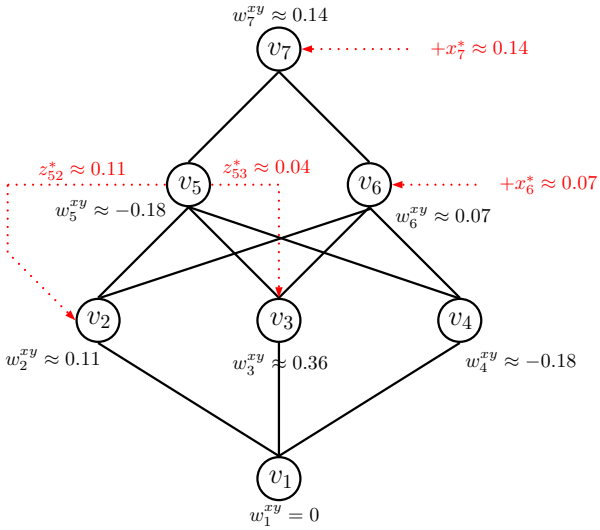

Figure 2: Mass transfer problem for $w^x = \frac{1}{28}\cdot(4,5,6,2,1,3,7)$ and $w^y = \frac{1}{28}\cdot(4,2,5,7,6,1,3)$.

mathematical terms until now, can be concretely interpreted in conceptual terms. We will do this by relying on one prototypic example of poverty/inequality analysis[9]. To make it simple, we will start with the notion of income poverty as a simple example of univariate poverty/inequality analysis. Consider for example that one is interested in the differences of income-poverty in two countries. One simple approach is here to firstly define a so-called poverty line c and to say that every person with income below the poverty line c can be termed *poor* whereas all persons with an income above the poverty line c can be termed *non-poor*. (The terms *poor* and *non-poor* are meant here in a purely descriptive sense free from value judgment). If the poverty line could be defined in a reasonable manner from a substance matter point of view, then for "measuring" the extent of inequality, one can compare the proportions of the *poor* persons in the two countries (also called head count ratio), for example by looking at the differences of the proportions in the wo countries. If it is difficult to specify the poverty line c, then one can get rid of the need for the specification of the poverty line by simultaneously looking at every reasonable poverty line c. If, independently from the choice of the poverty line c, the proportion of the poor is always greater for one country than for the other country, one can reasonably say that the income-poverty in one country is clearly greater than the income-poverty in the other country, which is exactly saying that one country is dominated by the other w.r.t. classical univariate first order dominance. In the situation of a given, fixed poverty line c, one can measure the extent of poverty for example with the income gap ratio, which is the relative difference between the income of the poor and the poverty line. The difference of the income gap ratios can then be used to measure

[9]Of course, in other concrete situations, the conceptual reconstruction done here could be less convincing.

the difference in the extent of poverty between the two countries. Compared to this, in the situation of multivariate inequality analysis, the involved dimensions of poverty like e.g. health or education are often only of ordinal scale of measurement. Of course, other dimensions like income have a higher scale of measurement, but this does not help in assessing, which amount of increase of income can compensate for which decrease in health or education. Of course, one can standardize every dimension in a reasonable way, but this would lead to a relative notion of inequality. Here, we go another way and use a notion of "extent" of dominance that is not related to units of the different dimensions but that is only based on the proportion of persons that are termed *non-poor* (or *poor*).

To do so, let us firstly think about the translation of the notion of a poverty line to the multivariate case: In the univariate case of income inequality we said that persons with income below the poverty line c could be termed *poor*, and the persons with income above the poverty line could be termed *non-poor*. In the multivariate setting, the way to term persons as *poor* or *non-poor* is only restricted by the underlying partial order $\leq$. If one terms one person i as *poor*, then one should also declare a person j as *poor* if the attributes x_j of person j are all lower than or equal to the attributes x_j of person x_j (i.e. $x_i \leq x_j$). This is exactly the concept of a downset of a partially ordered set: Every downset M of a poset $(V, \leq)$ is a reasonable concretion of the term *poor* in the sense that all $x \in M$ can be called *poor* and all $x \notin M$ could be called *non-poor*. The notion of a downset is the natural generalization of the notion of a poverty line to the multivariate case. In the sequel, we will deal with upsets instead of downsets. Dually to the notion of downstes, the notion of upsets[10] models the reasonable concretions of the term *non-poor* instead of the term *poor*. We can now interpret the maximal value D^+ and the minimal value D^- of the linear program (5) for detecting stochastic dominance: For the prototypic example of inequality analysis, if the maximal value D^+ is zero, then we know that X is stochastically dominated by Y, meaning that the proportion of the non-poor persons in subpopulation Y is greater than or equal to the proportion of the non-poor persons in subpopulation X, independently from the concretion of the term *non-poor*. Furthermore, in this case the minimal value D^- can be interpreted as some measure of the extent of stochastic dominance. The value $D^-$is exactly the difference between the proportions of non-poor persons of the two countries for that concretion of the term *non-poor* that is the most conservative in the sense that it maximizes the absolute value of the difference in proportions between the two countries. The extent of stochastic dominance can thus be measured to some extent with the minimal value of (5) with the following clear interpretation: The absolute value of D^- is exactly the proportion of poor persons in the poorer country that would have to be made non-poor to make the proportions of the poor (and thus also he proportions of the non-poor) in both countries the same, where the notion of *poor* is the most conservative, i.e., for every other reasonable notion of *poor* one would only have to make a smaller proportion of poor people of the poor country non-poor to make the proportions of the poor the same in each country.

[10]Note that the complements of downsets are upsets and that the complements of upsets are downsets.

As already mentioned, in [Tarp and Østerdal, 2007, p.19-20], a similar quantity for measuring the extent of dominance was proposed. There, the authors use the characterizing property (iii) of stochastic dominance and they measure the extent of dominance by the (weighted) amount of probability mass that is needed to obtain the density f_X from f_Y. However, the used measure seems to be not very sensitive in the following specific sense: Assume for simplicity real-valued random variables X and Y. Assume further that X is a simple transformation of Y, concretely $X := Y - \varepsilon$ with positive but very small ε. Then, one would have to transport the whole probability mass to obtain f_X from f_Y no matter how small the value ε is. This seems to be a very undesirable property of this measure of the extent of dominance. In such a situation, the measure D^- of extent proposed here behaves differently. For example if X and Y are normally distributed, then the maximal value D^- of (5) would be strictly increasing in ε and would furthermore converge to zero if ε converges to zero. For constant random variables X and Y the measure D^- would be either zero (if ε is zero) or one (if ε is grater than zero). This seems counter-intuitive at first glance, the measure is insensitive to the distance between X and Y. Actually this behavior is adequate, because one presumes only an ordinal scale of measurement for X and Y here, and thus one cannot reasonably measure the distance between X and Y. The fact that one could actually be sensitive to the distance between X and Y in the normally distributed case is due to the fact, that with our measure, we do not directly measure (non-existing) distances in the space of the values of X and Y, instead we measure indirectly the "distance" between X and Y by the amount of probability mass that has to be transported to compensate inequality for the most conservative choice of a poverty line.

While the value of D^- gives a quantitative insight in the extent of dominance, the upset U, where D^- is attained additionally gives a further more qualitative insight into how the worst possible concretion of the term *poor*, for which the maximal inequality in poverty is attained, looks like. This could be interesting for example if one is interested in the question, if the purely mathematical formalization of *poor* and *non-poor* via upsets is maybe too rigorous and if an extreme value of the test statistic is only due to an upset representing a very "skewed" concretion of the term *non-poor*, that could maybe be excluded as a reasonable concretion of the term *poor* because of substance matter considerations. In such a situation on may use the regularization techniques developed in Section 5.3.

3.4 Checking stochastic dominance as a linear program on a closure system

One important point to note is that the way we incorporated the property of being an upset was by introducing simple inequalities of the form $m_j \geq m_i$ for all pairs $(i, j) \in \preccurlyeq$. In the language of formal implications this means that we demanded that an upset should

contain with every v_i also every $v_j \succcurlyeq v_i$, which is exactly saying that the formal implication $\{v_i\} \longrightarrow \{v_j\}$ should be valid in the closure system of upsets. In fact, for the closure system $\mathcal{S}$ of upsets, the essential implications are the implications of the form $\{v_i\} \longrightarrow \{v_j\}$ with $v_i \prec v_j$, because all these implications are respected by $\mathcal{S}$ and they already describe $\mathcal{S}$ in the sense that they are a base of all valid implications. (Obviously there are further redundant implications like e.g. $\{v_i\} \longrightarrow \{v_i\}$ or $\{v_i, v_k\} \longrightarrow \{v_j\}$ with $v_i \leq v_j$ and v_k arbitrary.) For the case of upsets we were especially lucky, because all such essential implications had a simple premise (meaning that the premise A in $A \longrightarrow B$ is a singleton) and thus we could implement this implications via simple inequalities $m_i \leq m_j$ and could furthermore drop the integrality constraints. There are other situations that are such simple, too: Due to Birkhoffs theorem ([Birkhoff, 1937]), every (finite) closure system that is additionally closed under union[11] is describable via simple formal implications, and examples of such kinds of closure systems arise for example in the context of quasi-ordinal knowledge space theory (see, e.g., [Doignon and Falmagne, 2012, p.38-40], note also that there are neat connections between knowledge space theory and formal concept analysis, cf., [Rusch and Wille, 1996]). A natural question is now: Can we still solve the problem of maximizing/minimizing a linear function on an arbitrary closure system that is not describable via simple implications and could this have some application? The answer is simply yes: The next sections will give two examples of closure systems that are either explicitly given by an implication base or that are implicitly given as the concept extents of a given formal context.

4 Linear programming on general closure systems

4.1 The case of closure systems efficiently described by formal implications

In some situations, a closure system that is very big can still be efficiently described by an implication base of all valid implications. One example is the closure system $\mathcal{C}(\mathbb{R}^2)$ of all convex sets in $\mathbb{R}^2$ that could be of interest in the context of spatial statistics. The set of all valid implications of $\mathcal{C}(\mathbb{R}^2)$ is given by $\mathfrak{I} = \{A \longrightarrow \mathrm{co}(A) \mid A \subseteq \mathbb{R}^2\}$ where co is the operator that maps a set to its convex hull. Because of Carathéodory's theorem for convex hulls[12] the system $\mathcal{L} = \{A \longrightarrow \mathrm{co}(A) \mid A \subseteq \mathbb{R}^2, |A| \leq 3, \mathrm{co}(A) \supsetneq A\}$ is an implication base of $\mathfrak{I}$. In statistical applications in the context of spatial data analysis, for example in ecology, one is interested in differences in the spatial distribution of different species, for example male and female Pacific cods in the eastern Bering Sea analyzed in Syrjala [1996]. To describe differences in the spatial distributions of the two subpopulations, one

[11] The closure system of upsets is such a system.

[12] 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.

can use common test statistics[13]: Here, test statistics of many different statistical tests are available, for example generalizations of the Kolmogorv-Smirnov test or generalizations of the Cramér von Mises test (see [Syrjala, 1996]) could be used. For the Kolmogorov-Smirnov type generalization one determines for every rectangular area the difference in the proportion of male and female cods. Then one computes the maximal difference over all rectangular areas. This method needs a specification of a rectangular coordinate system and the results are dependent on the concrete choice of this coordinate system. Opposed to this, one could also simply look not only at all rectangular, but instead at all convex areas and then compute the maximal difference. This would be exactly an optimization of a linear function on a closure system. The result of the optimization on all convex sets instead of all rectangular areas would then be independent of the choice of a coordinate system, because for the definition of convexity, no specification of a coordinate system is needed at all. If one did observe cods at altogether k spatial points $(v_1, \ldots, v_k)$ then one actually does not need to look at the whole closure system $C(\mathbb{R}^2)$, it suffices to look only at the projected closure system $\mathcal{C}(\{v_1, \ldots, v_k\}) := \{A \cap \{v_1, \ldots, v_k\} \mid A \in \mathcal{C}(\mathbb{R}^2)\}$. To compute the test statistic one can solve a binary program, where all implications are implemented as inequality constraints. This method is generally applicable for arbitrary closure systems with a given implication base: For a given implication base $\mathcal{L}$ of an arbitrary finite closure system, one can compute the statistic

$$\sup_{A \in \subseteq V,\ A \text{ respects } \mathcal{L}} \langle w^x - w^y, \mathbb{1}_A \rangle$$

by solving the following binary program:

$$\begin{aligned} \langle w^x - w^y, m \rangle &\longrightarrow \max \\ w.r.t. & \\ m &\in \{0,1\}^k \\ \forall (Y, Z) \in \mathcal{L} : \sum_{i: v_i \in Y} m_i - \frac{1}{|Z|} \sum_{i: v_i \in Z} m_i &\leq |Y| - 1. \end{aligned}$$

Here, for any given implication $Y \longrightarrow Z$ of $\mathcal{L}$, the corresponding inequality constraint of the binary program is automatically satisfied if the premise of $Y \longrightarrow Z$ is not fulfilled, because then the left hand side is lower than $|Y| - 1$. If the premise Y is fulfilled, then the corresponding inequality translates to $-\frac{1}{Z} \sum_{i: v_i \in Z} m_i \leq -1$ or equivalently to $\frac{1}{Z} \sum_{i: v_i \in Z} m_i \geq 1$, thus demanding that all m_i with $i \in Z$ should be one, meaning that if the set described by the indicator function $(m_1, \ldots, m_k)$ contains all elements of Y, then it should also contain all elements of Z. In our concrete situation of convex sets we would have to solve a binary program with n decision variables and $\mathcal{O}(n^3)$ inequality constraints. Unfortunately,

[13]Here, in the first place we are mainly interested in the test statistic as a descriptive tool, the problem of inference will be discussed in a general setting in Section 5.

generally, the integrality constraints cannot be dropped, here. Thus, the program becomes very difficult to solve if n is large.

An important case where one can actually drop the integrality constraints is the case where one has only simple implications: In this case one can implement every simple implication $\{v_i\} \longrightarrow \{v_j\}$ as the inequality $m_i \leq m_j$. To see that one can drop the integrality constraints in this case, observe that any feasible vector $(m_1, \ldots, m_k)$ of the relaxed program with some $m_l \notin \{0, 1\}$ is not an extreme point of the feasible set of the relaxed program, since for $\varepsilon > 0$ chosen small enough (for example $\varepsilon = 1/2 \min\{|m_i - m_l| \mid i \in \{1, \ldots, n\}, m_i \neq m_l\}$) it can be represented as the convex combination of the two feasible vectors $m + x^\varepsilon$ and $m - x^\varepsilon$ where $x^\varepsilon \in \mathbb{R}^k$ is defined as $x_i^\varepsilon = \begin{cases} \varepsilon & \text{if } m_i = m_l \\ 0 & \text{else} \end{cases}$.

4.2 The case of closure systems efficiently described by a generating formal context

Closure systems also naturally arise in the theory of formal concept analysis: the family of all formal concept extents (as well as the family of all concept intents) of a concept lattice is a closure system. Furthermore, every arbitrary closure system can be represented as a closure system of extents (or intents) of an appropriately chosen formal context[14]. In statistical applications, it appears natural to take as objects the observed data points, for example persons in a social survey. As attributes one can take the values of different variables of interest, for example the answers of the persons to different questions. (If the questions are yes-no questions, then they can be incorporated directly, otherwise one can apply the method conceptual scaling to get binary data, cf. Section 2.2.) The formal concept lattice then gives valuable qualitative information about different subgroups of persons that supplied response patterns that belong to the same formal concept and thus share specific attributes. If one is interested in differences between different subgroups (e.g., male and female participants) w.r.t. the answers to the questions, one could look at every formal concept and analyze the differences between the subpopulations that belong to the given concept. Often, the concept lattice is very large and it becomes difficult to look at every formal concept. Then, one can look for example only on that concepts, for which the difference between the proportions of persons belonging to this concept in each subgroup is maximal or minimal. This is exactly the problem of maximizing a linear function on the closure system of concept extents. If the whole concept lattice can be computed explicitly, then one can simply explicitly compute for every extent the difference in proportions between both subpopulations. However, in many situations the concept lattice is so big that it is very hard to compute all extents/intents explicitly to perform the optimization. (In the worst case, a formal context can have $\min(2^{|G|}, 2^{|M|})$ associated

[14]For a closure system $\mathcal{S} \subseteq 2^V$ take the formal context $\mathbb{K} := (V, \mathcal{S}, \in)$, then the formal extents are exactly the sets of $\mathcal{S}$. Analogously, for the dual context $\mathbb{K} := (\mathcal{S}, V, \ni)$, the formal intents are exactly the sets of $\mathcal{S}$.

formal concepts.) In this situation one can use the fact that a pair (A, B) with $A \subseteq G$ and $B \subseteq M$ is a formal concept iff

$$A = B' \;\&\; B = A'$$

or equivalently iff

$$\forall g \in G, m \in M : \mathbb{1}_A(g) = \min_{m \in B} \mathbb{1}_{m'}(g) \;\&\; \mathbb{1}_B(m) = \min_{g \in A} \mathbb{1}_{g'}(m). \tag{12}$$

Characterization (12) can be used to describe the property of being a formal concept with the help of the following characterizing inequalities:

$$\forall g \in G, m \in B : \mathbb{1}_A(g) \leq \mathbb{1}_{m'}(g) \tag{13}$$

$$\forall g \in A, m \in M : \mathbb{1}_B(m) \leq \mathbb{1}_{g'}(m) \tag{14}$$

$$\forall g \in A, m \in B : \mathbb{1}_A(g) \geq \sum_{m \in B} \mathbb{1}_{m'}(g) - |B| + 1 \quad \& \tag{15}$$

$$\mathbb{1}_B(m) \geq \sum_{g \in A} \mathbb{1}_{g'}(m) - |A| + 1. \tag{16}$$

Equations (13) and (21) capture the fact that $\mathbb{1}_A(g) \leq \min_{m \in B} \mathbb{1}_{m'}(g)$ and $\mathbb{1}_B(m) \leq \min_{g \in A} \mathbb{1}_{g'}(m)$, respectively. Equations (15) and (16) say that $\mathbb{1}_A(g) \geq \min_{m \in B} \mathbb{1}_{m'}(g)$ and $\mathbb{1}_B(m) \geq \min_{g \in G} \mathbb{1}_{g'}(m)$, respectively, which is equivalent to the condition that if an object g has all attributes of B then it has to be in the extent A and that if an attribute m is shared by all objects of A, then it should be in the intent B. The characterization via inequality constraints can be used to optimize a linear function of the indicator function of the extents (or the intents, or both) with a binary program: Let $G = \{g_1, \ldots, g_m\}$ be the set of objects, $M = \{m_1, \ldots, m_n\}$ the set of attributes and let $A \in \{0,1\}^{m \times n}$ be a matrix describing the incidence I with the interpretation $A_{ij} = 1 \iff$ object number i has attribute number j. A formal concept can then be described by a binary vector $z = (z_1, \ldots, z_m, z_{m+1}, \ldots, z_{m+n}) \in \{0,1\}^{m+n}$, where the first m entries describe the extent via $z_i = 1$ iff object i belongs to the extent and the last n entries describe the intent as $z_{j+m} = 1$ iff attribute j belongs to the intent. The characterizing constraints (13) - (16) would then translate to the conditions

$$\forall (i,j) \text{ s.t. } A_{ij} = 0 : \qquad z_i \leq 1 - z_{j+m} \quad \& \quad z_{j+m} \leq 1 - z_i \tag{17}$$

$$\forall i \in \{1, \ldots, m\} : \qquad z_i \geq \sum_{k : A_{ik} = 1} z_{k+m} - \sum_{k=1,\ldots,n} z_{k+m} + 1 \tag{18}$$

$$\forall j \in \{1, \ldots, n\} : \qquad z_{j+m} \geq \sum_{k : A_{kj} = 1} z_k - \sum_{k=1,\ldots,m} z_k + 1. \tag{19}$$

$$\tag{20}$$

have to be satisfied. This could be simplified to the condition

$$\forall (i,j) \text{ s.t. } A_{ij} = 0 : \qquad z_i \leq 1 - z_{j+m} \tag{21}$$

$$\forall i \in \{1, \ldots, m\} : \qquad \sum_{k:A_{ik}=0} z_{k+m} \geq 1 - z_i \tag{22}$$

$$\forall j \in \{1, \ldots, n\} : \qquad \sum_{k:A_{kj}=0} z_k \geq 1 - z_{j+m}, \tag{23}$$

which has the following intuitive interpretation:

For every 0-entry in the i-th row and the j-th column of the matrix A we have:

1. if $A_{ij} = 0$ and if object g_i belongs to the extent, then necessarily attribute m_j cannot belong to the intent and vice versa.
2. If object g_i does not belong to the extent, then there exists at least one attribute m_k of the intent, that the object g_i does not have.
3. Dualy, if attribute m_j does not belong to the intent, then there exists at least one object g_k of the extent, that has not attribute m_j.

Thus, we can compute the maximum

$$\max_{(A,B)\in\mathfrak{B}(\mathbb{K})} \langle w^{ext}, \mathbb{1}_A \rangle + \langle w^{int}, \mathbb{1}_B \rangle$$

of an arbitrary linear objective function $(w_1^{ext}, \ldots, w_n^{ext}, w_1^{int}, \ldots, w_n^{int})$ of both the extents and the intents by solving the binary program

$$\begin{aligned} \langle (w_1^{ext}, \ldots, w_m^{ext}, w_1^{int}, \ldots, w_n^{int}), (z_1, \ldots, z_m, z_{m+1}, \ldots, z_{m+n}) \rangle &\longrightarrow \max \\ &w.r.t. \\ \forall (i,j) \text{ s.t. } A_{ij} = 0 : z_i &\leq 1 - z_{j+m} \\ \forall i \in \{1, \ldots, m\} : \sum_{k:A_{ik}=0} z_{k+m} &\geq 1 - z_i \\ \forall j \in \{1, \ldots, n\} : \sum_{k:A_{kj}=0} z_k &\geq 1 - z_{j+m} \end{aligned} \tag{24}$$

All in all, we would thus have to solve a **binary** program with $m + n$ variables and $|\{(i,j) \mid A_{ij} = 0)\}| + m + n$ constraints. This problem can become cumbersome if the formal context is big enough, especially because one cannot simply drop the integrality-constraints. However, in practical applications, often only the number of objects is large and the number of items is medium-sized. If one further analyzes the binary program, then one observes that the inequalities concerning the objects and the constraints concerning the attributes

are separated in the sense that if one relaxes only the integrality constraints of the variables $z_1, \ldots, z_m$ describing the extent, then the optimum of the associated relaxed mixed binary program is still (also) attained at a binary solution and thus one can relax the integrality constraints of the extent. The reason is that for fixed and binary $(z_{m+1}, \ldots, z_{m+n})$ the inequalities (21) and (22) are either redundant or reduce to equality constraints of the form $z_i = 0$ or $z_i = 1$ and inequality (23) is either redundant or demands that a sum of z_k's associated with the extent is greater or equal to 1. If one of the z_k's is not binary, than at least one other $z_{k'}$ has to be greater than zero. This means that for an appropriately chosen[15] $\varepsilon > 0$ the vectors $(z_1, \ldots, z_k + \varepsilon, \ldots, z_{k'} - \varepsilon, \ldots, z_{m+n})$ and $(z_1, \ldots, z_k - \varepsilon, \ldots, z_{k'} + \varepsilon, \ldots, z_{m+n})$ are still feasible with respect to the relaxed feasible set and this shows that non-integer points are no extreme-points of the restricted polytope where the binary variables describing the intent are fixed. Thus, the optimal value for the relaxed program is always also attained at a binary solution.

5 Statistical inference

We now treat the question of inference. Coming back to the example of detecting stochastic dominance, we were able to detect stochastic dominance in a sample. The natural question of inference is now: What can we reasonably infer about stochastic dominance in the population we sampled from? From a substance matter perspective, one would supposedly be interested for example in the hypotheses

$$\begin{aligned} H_0: &\quad X \text{ is not stochastically dominated by } Y \quad \text{vs} \\ H_1: &\quad X \text{ is stochastically dominated by } Y. \end{aligned}$$

However, a reasonable consistent classical statistical test of this pair of hypotheses is not reachable since already in the univariate case where the distribution function characterizes stochastic dominance, we have the problem that for every $X \leq_{SD} Y$, in every arbitrarily small neighborhood[16] of Y we can find some $\tilde{Y}$ with $X \not\leq_{SD} \tilde{Y}$. To circumvent this problem, one can modify the hypotheses, for example by switching the roles of H_0 and H_1 (for consistent statistical tests of this kind in the univariate case, see, e.g., [Barrett and Donald, 2003]). Here, we go a slightly different way. Since the value of D^+ characterizes $X \leq_{SD} Y$ via $X \leq_{SD} Y$ iff $D^+ = 0$ and D^- characterizes $Y \leq_{SD} X$ via $Y \leq_{SD} X$ iff $D^- = 0$ and furthermore $X \nsim_{SD} Y$ (where $X \nsim_{SD} Y$ means $X \not\leq_{SD} Y$ & $Y \not\leq_{SD} X$) iff $D^+ > 0$ & $D^- < 0$) we can simply test, if D^+ and D^- are significantly different from zero. (In the case of for example D^+ is significantly positive and D^- is not significantly negative,

[15] The choice of ε depends on all inequalities that involve z_k and $z_{k'}$ but since there are only finite many constraints, ε can in fact be chosen small enough and still greater than zero.

[16] This is meant w.r.t. e.g., the Kolmogorov-Smirnov distance. Note also, that in our situation, we have not much freedom of choice of other distances that induce other neighborhood concepts, since we can only make use of the partially ordered scale of measurement of X and Y.

we cannot directly conclude $X \leq_{SD} Y$ but at least $Y \not\leq_{SD} X$ and the possibility $X \not\leq_{SD} Y$ is only possible due to upsets A with $P(X \in A) \geq P(Y \in A)$ where the difference $P(X \in A) - P(Y \in A)$ is only slightly positive.) In the sequel we will put focus on D^+ and also do not explicitly correct for multiple testing if considering both D^+ and D^- simultaneously. Actually, conceptually, here we do not take the inference problem as the primitive and do not rigorously test a beforehand exactly stated hypothesis by doing a statistical test that provides us with a descriptively interpretable test statistic as a by-product. Instead, we see it a little bit the other way around: In the first place, we would like to get a good, conceptually rigorous descriptive insight into the data by not relying on traditional approaches based on somehow "arbitrarily chosen" location measures[17] summarizing the data by one number and then comparing the obtained numbers. Instead, by relying on stochastic dominance, we in a sense somehow look simultaneously at all reasonable location measures and if we know $X \leq_{SD} Y$, then we also know that every reasonable location measure[18] would give a lower (or equal) number to X than to Y. This is a conceptually much more reliable statement than simply comparing numbers (of course with the drawback of being less decisive). Only in a second step we think in statistical terms about to which extent the conceptually rigorous statement of stochastic dominance can be translated from the sample to the population.

5.1 Permutation-based tests

Now, let us come to the purely statistical aspects of inference for detecting stochastic dominance. (All considerations are similarly valid for linear optimization on general closure systems.) In the simple univariate case of real-valued, continuously distributed random variables X, Y, for the two-sample case under $H_0 : F_X = F_Y$, the distribution of the test statistic D^+ (and also D^- and $D := \max\{D^+, -D^-\}$) is independent of the true law F_X, has known asymptotics and can be furthermore computed exactly for identical sample sizes (see, e.g., [Pratt and Gibbons, 2012, Chapter 7]). Opposed to this, in the general multivariate situation, the statistic D^+ is not distribution free, anymore: Firstly, the distribution of D^+ depends on the concrete structure of the poset $(V, \leq)$: If the relation $\leq$ is very sparse, then the set of all upsets is very large and one would generally expect that D^+ would typically have higher values than for the case of a very dense relation $\leq$. Secondly, also the interplay between the structure of $(V, \leq)$ and the unknown true law is also of relevance: For example in a very large poset $(V, \leq)$ with a very sparse relation $\leq$ it could be still the case that the most probability mass is living on a much smaller subset $W \subseteq V$ on which the restricted relation $\leq \cap\, W \times W$ is actually very dense. This suggests that a rigorous analytic treatment of the distribution of D^+ seems to be only partially

[17] Note the non-classical scale of measurement we are dealing with, here.

[18] One of the few location measures that does not respect first order stochastic dominance is the mode. But note that the mode appears most naturally if we have a categorical or an interval scale of measurement, the mode seems to give no valuable information if we want to analyze inequality which is a genuinely ordinal concept.

possible[19]. Thus, a natural alternative is to apply a two sample observation-randomization test (permutation test, see, e.g., [Pratt and Gibbons, 2012, Chaper 6]), here. The procedure for evaluating the distribution of D^+ under $H_0 : P_X = P_Y$, which is the least favorable case of $\tilde{H}_0 : \quad D^+_{true} := \sup_{A \in \mathcal{U}((V,\leq))} P_X(A) - P_Y(A) = 0 \quad (\iff X \leq_{SD} Y))$ is straightforward:

1. Let a sample $x = (x_1, \ldots, x_{n_x})$ of size n_x for subpopulation X and a sample $y = (y_1, \ldots, y_{n_y})$ of size y of subpopulation Y be given.
2. Compute the statistic D^+ for the actually observed data.
3. Take the pooled sample $z = (x_1, \ldots, x_{n_x}, y_1, \ldots, y_{n_y})$.
4. Take all index sets $I \subseteq \{1, \ldots, n_x + n_y\}$ of size n_x and compute the test statistic D^+_I that would be obtained for a virtual sample $\tilde{x} = (z_i)_{i \in I}$ for population X and $\tilde{y} = (z_i)_{i \in \{1,\ldots,n_x+n_y\} \setminus I}$ for subpopulation Y.
5. Order all D^+_I in increasing order
6. Reject H_0 if the test statistic D^+ for the actually observed data is greater than the $\lceil \gamma \cdot |I| \rceil$-th value of the increasingly ordered values D^+_I, where γ is the envisaged confidence level.

In step 4 one has to compute the test statistic for a very huge number of resamples, thus one usually does not compute the test statistic for all resamples but only for a smaller number of randomly chosen resamples. In the context of linear programming on closure systems, the computation of the test statistic for one resample could be already computational demanding for very complex data sets, so the application of observation-randomization tests has some limitations, here.

5.2 Conservative bounds via Vapnik-Chervonenkis theory

Beyond applying resampling schemes for inference there is the further possibility to apply Vapnik-Chervonenkis theory (see, e.g., [Vapnik and Kotz, 1982]) to obtain conservative bounds for the test statistic: In Vapnik-Chervonenkis theory, among other things, one analyzes the distribution of

$$\sup_{A \in \mathcal{S}} |P_n(A) - P(A)|$$

or

$$\sup_{A \in \mathcal{S}} |P_n(A) - P'_n(A)|,$$

[19] Actually, there exists some literature on the asymptotic distribution of the optimal value of a random linear program (e.g., [Babbar, 1955, Sengupta et al., 1963, Prèkopa, 1966]). However, this literature seems to be not applicable in our situation, 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.

where P is an unknown probability law and P_n is the empirical law associated with an i.i.d.-sample of size n (and P'_n is the empirical law associated to a further independently drawn sample of the same size n). Here, the family $\mathcal{S}$ can be any arbitrary family of subsets of a given space Ω. In our situation, the family $\mathcal{S}$ is the underlying closure system of interest. The Vapnik-Chervonenkis inequalities (cf., [Vapnik and Kotz, 1982, p.170-172]) then state that

$$P\left(\sup_{A\in\mathcal{S}}|P_n(A)-P(A)|>\varepsilon\right)\leq 6\ m^{\mathcal{S}}(2n)\ \ e^{-n\varepsilon^2/4}\quad\text{and} \tag{25}$$

$$P\left(\sup_{A\in\mathcal{S}}|P_n(A)-P'_n(A)|>\varepsilon\right)\leq 3\ m^{\mathcal{S}}(2n)\ \ e^{-n\varepsilon^2}. \tag{26}$$

These inequalities[20] can be used to get conservative critical values for a one sample and a two sample test. (In the sequel, we will put focus on the two sample situation.) The crucial quantity involved in the right hand sides of these inequalities is the so-called **growth function**

$$\begin{aligned} m^{\mathcal{S}}(k) :=& \max_{A\subseteq\Omega,|A|=k}\Delta_{\mathcal{S}}(A),\ \text{where}\\ \Delta_{\mathcal{S}}(A) :=& |\{S\cap A\mid S\in\mathcal{S}\}| \end{aligned}$$

describes the cardinality of the projection of the family $\mathcal{S}$ on the set A. Obviously, if $\mathcal{S}$ is finite, then the growth function $m^{\mathcal{S}}(k)$ is always lower than or equal to the cardinality of $\mathcal{S}$ and thus $|\mathcal{S}|$ can be used to get a bound for the left hand sides of (25) and (26). Actually, in our setting, we will use as the underlying space always the subset V_{ess} of all actually observed values of the basic set V and an associated closure system $\mathcal{S}\subseteq 2^{V_{ess}}$ on the restricted space $\Omega := V_{ess}$. (Note that the projection of a closure system $\mathcal{S}'\subseteq 2^{\Omega'}$ on Ω' onto a subset $\Omega\subseteq\Omega'$ via $\mathcal{S}'_{|\Omega}=\{S\cap\Omega\mid S\in\mathcal{S}'\}$ is again a closure system on Ω.) Thus, with $A=V_{ess}$ we have $\Delta_{\mathcal{S}}(A)=|\mathcal{S}|$ and $m^{\mathcal{S}}(2n)=|\mathcal{S}|$, such that the bound $|\mathcal{S}|$ is a sharp bound for $m^{\mathcal{S}}(2n)$. If the family $\mathcal{S}$ is explicitly given, we could thus work with the computable bound $|\mathcal{S}|$. The far more interesting situation appears if the family $\mathcal{S}$ is very large and only implicitly given. Then there is another important bound (see [Vapnik and Kotz, 1982, p.167]) on the growth function that is related to the **Vapnik-Chervonenkis dimension (V.C.-dimension)** of the family[21] $\mathcal{S}$:

$$m^{\mathcal{S}}(k)\leq 1.5\frac{k^{VC-1}}{(VC-1)!},$$

[20] There is a bunch of similar Vapnik-Chervonenkis type inequalities that could be of help here, see, e.g., the summary given in Table 1 of [Vayatis and Azencott, 1999, p.4].

[21] Originally, Vapnik-Chervonenkis theory was mainly developed to be able to deal with infinite families $\mathcal{S}$. Here, we have finite families $\mathcal{S}$, and if we would know $|\mathcal{S}|$ then, in our setting, we would better bound $m^{\mathcal{S}}(k)$ by $|\mathcal{S}|$ instead of using the Vapnik-Chervonenkis dimension since in our setting of $\mathcal{S}\subseteq 2^{V_{ess}}$, this dimension essentially only provides an upper bound for the cardinality of $|\mathcal{S}|$. If the family $\mathcal{S}$ is very large and is only implicitly given, then the V.C.-dimension can still provide a good computable bound for $m^{\mathcal{S}}(k)$. Note further that sometimes also other bounds for $m^{\mathcal{S}}(2n)$ can be useful, for example for finite Ω wa have $m^{\mathcal{S}}(2n)\leq 2^{|\Omega|}$.

where VC is the Vapnik-Chervonenkis dimension of the family $\mathcal{S}$, that is defined as the cardinality of the largest possible subset A that can be shattered by $\mathcal{S}$: One says that a set A can be **shattered** by $\mathcal{S}$ (or alternatively that A is shatterable w.r.t. $\mathcal{S}$) if the projection of $\mathcal{S}$ on A contains all subsets of A, i.e. $2^A = \{S \cap A \mid S \in \mathcal{S}\}$ or equivalently $\Delta_{\mathcal{S}}(A) = 2^{|A|}$. In many cases, the V.C.-dimension cannot be computed explicitly. However, in our context it shows up that we can compute the V.C.-dimension either with the help of binary programs or with a sharp characterization of the V.C.-dimension. Of course, the Vapnik-Chervonenkis inequalities provide only very conservative bounds for inference. (Note that the right hand sights of (25) and (26) do not depend on the true law P.) If one is able to perform an observation randomization test, then one should do it instead of dealing with the conservative Vapnik-Chervonenkis inequalities. However, the Vapnik-Chervonenkis inequalities give us some guidance for dealing with situations where the closure system is so big that one would expect that the distribution of the test statistic is behaving too ugly to allow for a sensible statistical test with enough power. In such a situation, we can use Vapnik-Chervonenkis theory to appropriately reduce the closure system to hope for making the tail distribution of the test statistic more well-behaved[22]. If one appropriately reduces the cardinality of the closure system, then one could hope for a test statistic that has a better power for the detection of a "systematic" deviation[23] from H_0. This possibly increased power would then come along with a smaller and thus less fine-grained closure system $\mathcal{S}$ that is then not so sensitive to very specific alternatives. Note that the V.C.-inequality is essentially based on the effective size of $\mathcal{S}$. Thus, if $\mathcal{S}$ is explicitly given, one can simply drop some sets of $\mathcal{S}$ to make $\mathcal{S}$ smaller. However, in our situation, we often have a closure system that is implicitly given and only nicely describable because it is a closure system. A simple removal of some sets of $\mathcal{S}$ is thus not possible because sets of $\mathcal{S}$ are not explicitly given and an arbitrary removal of some sets could lead to a family $\mathcal{S}'$, that is not a closure system, and thus not easily describable, anymore. The beauty of V.C.-theory in this situation lies in the fact that if we can compute the V.C.-dimension by supplying a shatterable set A of maximal cardinality, then we also have a straightforward possibility to tame $\mathcal{S}$: Since big shatterable sets A make $\mathcal{S}$ very big, we can drop some or all elements of such sets A to tame $\mathcal{S}$ efficiently. Actually, one would not completely remove A (or a subset of A) for the whole data analysis, but only for the construction of the closure system under which the final data analysis takes place. Before explaining how this is exactly meant in different situations and how we ensure that the tamed system is still a closure system (cf. Section 5.3), we will now firstly characterize shatterable sets and the V.C.-dimension for different closure systems in the next section.

[22]Of course, V.C.-theory gives us only bounds on the tail behavior of the test statistic and no direct insight into the actual behavior of the tails, so a sharpening of the bound does not necessarily mean that the actual tail behavior will be getting better if we reduce the V.C.-dimension of the closure system.

[23]Of course, one cannot hope for a more powerful statistic w.r.t. every thinkable deviation from H_0 but only for a better power for detecting deviations that are not "too complex" w.r.t. V.C.-dimension.

5.2.1 The Vapnik-Chervonenkis dimension of several special closure systems

This section is actually very technical and not really necessary to understand the basic ideas in the following sections. The reader more interested in the basic concepts can thus skip this section and Section 5.2.2. For the reader interested in Vapnik-Chervonenkis theory and the reader interested in a detailed understanding, we would like to recommend the following sections, because, though looking a bit technical, there are no deep or cumbersome ideas involved in the following theorems. Contrarily, the relation between Vapnik-Chervonenkis theory and formal concept analysis seems to be very natural. Maybe somehow surprising, there seems to be not too much research that directly connects formal concept analysis and Vapnik-Chervonenkis theory, the only works in this direction, the authors are aware of, are the papers Anthony et al. [1990a,b], Albano and Chornomaz [2017], Chornomaz [2015], Albano [2017a,b], Makhalova and Kuznetsov [2017].

Definition & Proposition 1. *Let $\mathcal{S} \subseteq 2^\Omega$ be a closure system on Ω. Let furthermore $\mathfrak{I}(\mathcal{S})$ be the set of all formal implications the closure system $\mathcal{S}$ respects. A set $M \subseteq \Omega$ is called* **implication-free** *if there is no formal implication $(A, B) \in \mathfrak{I}$ where A and B are disjoint non-empty subsets of M. A set M is shatterable w.r.t. $\mathcal{S}$ if and only if it is implication-free and thus the Vapnik-Chervonenkis dimension of $\mathcal{S}$ is the maximal cardinality of an implication-free set $M \subseteq \Omega$.*

Definition 5 (Vapnik-Chervonenkis principal dimension (VCPI/VCPF))**.** *Let $(V, \leq)$ be a partially ordered sett. The* **Vapnik-Chervonenkis principal ideal dimension** *(VCPI) is the Vapnik-Chervonenkis dimension of the family*

$$\mathfrak{pi}((V, \leq)) := \{\downarrow x \mid x \in V\} = \{\{y \mid y \leq x\} \mid x \in V\}\}$$

of all principal ideals of $(V, \leq)$. If $(V, \leq)$ is a complete lattice, then $\mathfrak{pi}((V, \leq))$ is a closure system. In this case we also say that a set $M \subseteq V$ is **join-shatterable** *if it is shatterable w.r.t. the family $\mathfrak{pi}((V, \leq))$. Analogously, the* **Vapnik-Chervonenkis principal filter dimension** *(VCPF) is the Vapnik-Chervonenkis dimension of the family*

$$\mathfrak{pf}((V, \leq)) := \{\uparrow x \mid x \in V\} = \{\{y \mid y \geq x\} \mid x \in V\}\}$$

of all principal filters of $(V, \leq)$. If $(V, \leq)$ is a complete lattice, then $\mathfrak{pf}((V, \leq))$ is a closure system. In this case we also say that a set $M \subseteq V$ is **meet-shatterable** *if it is shatterable w.r.t. the family $\mathfrak{pf}((V, \leq))$.*

Theorem 1 (Motivating the notions *join-shatterable* and *meet-shatterable*)**.** *A subset M of a complete lattice $(V, \leq)$ is join-shatterable if and and only if we have for every $x \in M$:*

$$x \not\leq \bigvee M \backslash \{x\}. \tag{27}$$

Analogously, a subset M of a complete lattice $(V, \leq)$ is meet-shatterable if and and only if we have for every $x \in M$:

$$x \not\geq \bigwedge M \backslash \{x\}. \tag{28}$$

Proof. We only proof the first statement, the second statement can be proofed analogously.
if: Let $B \subseteq M$. Take $A :=\downarrow \bigvee B \in \mathfrak{pi}((V,\leq))$. Then $A \cap M \supseteq B$ since $\forall b \in B : b \leq \bigvee B$. Additionally, for every $x \in M\backslash B$ we have $B \subseteq M\backslash\{x\}$ and thus $x \notin A$, since if $x \in A$, because of $A \subseteq\downarrow \bigvee M\backslash\{x\}$ we would get $x \leq \bigvee M\backslash\{x\}$ which would be a contradiction to (27). Thus $A \cap M = B$ and because B was an arbitrary subset of M, we can conclude that M is shatterable.
only if: Let $x \leq \bigvee M\backslash\{x\}$ for some $x \in M$. Then the set $M\backslash\{x\}$ is not shatterable w.r.t. $\mathfrak{pi}((V,\leq))$, because every $a \in V$ with $\forall y \in M\backslash\{x\} : y \leq a$ is an upper bound of $M\backslash\{x\}$ and thus $a \geq \bigvee M\backslash\{x\} \geq x$. But this means, that every set $A =\downarrow a \in \mathfrak{pi}((V,\leq))$ that contains all elements of $M\backslash\{x\}$ necessarily also contains x which shows that in fact $M\backslash\{x\}$ is not shatterable. □

Theorem 2. *For every finite join-shatterable set M of a finite*[24] *complete lattice $(V,\leq)$ there exists another join-shatterable set J_M of join-irreducible elements of $(V,\leq)$ that has the same cardinality as M. This means that for determining the Vapnik-Chervonenkis principal ideal dimension it is enough to look at join-shatterable sets of join-irreducible elements.*

Proof. Let $M \subseteq V$ be a finite shatterable set. If all elements of M are join-irreducible then we are done. If there exists an $x \in M$ that is not join-irreducible we can find a join-irreducible element z such that the set $\tilde{M} := M\backslash\{x\} \cup \{z\}$ is still join shatterable. Since M is assumed to be finite, we can replace step by step every join-reducible element of M by a join-irreducible element and thus obtain a shatterable set of join-irreducible elements with the same cardinality: So let $x \in M\backslash\mathcal{J}(V)$. Then $x = \bigvee B$ for some set $B \subseteq \mathcal{J}(V)$. Furthermore, we have $z \not\leq \bigvee M\backslash\{x\}$ for at least one $z \in B$, because otherwise we would have $\bigvee M\backslash\{x\} \geq \bigvee B = x$ which is in contradiction with the assumption that the set M is join-shatterable. Now, take $\tilde{M} := M\backslash\{x\} \cup \{z\}$. Then, $\tilde{M}$ is join-shatterable. To see this, observe that M and $\tilde{M}$ only differ in the elements x and z and $z \leq x$. Thus $z \not\leq \bigvee M\backslash\{x\} = \bigvee \tilde{M}\backslash\{z\}$ and for every other $y \in \tilde{M}$ we have $y \not\leq \bigvee \tilde{M}\backslash\{y\}$ because otherwise we would have $y \leq \bigvee \tilde{M}\backslash\{y\} \leq \bigvee M\backslash\{y\}$ which is in contradiction with M being join-shatterable. □

Theorem 3. *The Vapnik-Chervonenkis principal ideal dimension VCPI (and also the Vapnik-Chervonenkis principal filter dimension VCPF) of a poset $(V,\leq)$ is bounded by its order dimension*[25] $\mathbf{odim}((V,\leq))$.

Proof. Let $d := \mathbf{odim}((V,\leq))$ and let $L_1,\ldots,L_d$ be d linear orders representing $\leq$ via $x \leq y \iff \forall i \in \{1,\ldots,d\} : xL_iy$. We show that every set M of more than d elements

[24]The finiteness assumption can be dropped if one only assumes that every element $x \in V$ can be written as a supremum of join-irreducible elements of V. This is for example the case if there are no infinite descending chains in V.

[25]Remember that the order dimension of a poset $(V,\leq)$ is the smallest number d of linear orders $L_1,\ldots,L_d \subseteq V \times V$ such that the relation $\leq$ can be represented as the intersection of these linear orders via $x \leq y \iff \forall i \in \{1,\ldots,d\} : xL_iy$.

of V is not join-shatterable: Take M with $|M| > d$ and take for every $i \in \{1, \ldots, d\}$ that element $x_i \in M$ that is the greatest element of M w.r.t. the linear order L_i. Then every principal ideal $\downarrow a$ that contains all the x_i necessarily also contains every further element $y \in M \backslash \{x_1, \ldots, x_d\} \neq \emptyset$ because for every $i \in \{1, \ldots, d\}$ we have $yL_ix_iL_ia$. □

Theorem 4. *Let $\mathbb{V} = (V, \leq)$ be a finite complete lattice. If $\mathbb{V}$ is distributive*[26]*, which can be characterized by saying that the condition*

$$\forall B \subseteq \mathcal{J}(\mathbb{V}) \forall x \in \mathcal{J}(\mathbb{V}) : \quad x \leq \bigvee B \Longrightarrow x \leq z \text{ for some } z \in B \tag{29}$$

is fulfilled, then the Vapnik-Chervonenkis principal dimension of $\mathbb{V}$ is exactly the width of $\mathcal{J}(\mathbb{V})$ and because of Birkhoffs theorem we have $\mathbb{V} \cong (\mathcal{D}(\mathcal{J}(\mathbb{V}))$ and the width of $(\mathcal{J}(\mathbb{V}))$ is exactly the order dimension of $\mathbb{V}$, so in this case we have $VCPI(\mathbb{V}) = \mathbf{odim}(\mathbb{V})$.

Proof. Because of Theorem 2 we only have to look at the set $\mathcal{J}(\mathbb{V})$ of the join-irreducible elements of $\mathbb{V}$. Let d denote the width of $\mathcal{J}(\mathbb{V})$. It is clear that a join-shatterable set $M \subseteq (\mathcal{J}(\mathbb{V}))$ necessarily is an antichain. Thus VCPI is lower than or equal to d. To see that $VCPI = d$ take an antichain A of size d. Then this antichain is obviously shatterable because for all $x \in A$ we have $x \not\leq \bigvee A \backslash \{x\}$ since if $x \leq \bigvee A \backslash \{x\}$ because of (29) we would have $x \leq z$ for some z in $A \backslash \{x\}$, but this would be in contradiction with A being an antichain. □

Definition & Proposition 2 (Vapnik-Chervonenkis upset dimension: Simply the width)**.** *Let $\mathbb{V} = (V, \leq)$ be a poset and $\mathcal{U}(\mathbb{V})$ be the set of all upsets of $\mathbb{V}$. Then the Vapnik-Chervonenkis dimension of $\mathcal{U}(\mathbb{V})$ is called the* **Vapnik-Chervonenkis upset dimension***. The Vapnik-Chervonenkis upset dimension is identical to the width of $\mathbb{V}$, because the shatterable sets are exactly the implication-free sets, which are in this case the antichains of $\mathbb{V}$. Analogously, the Vapnik-Chervonenkis dimension of all downsets is also equal to the width.*

Definition 6 (Vapnik-Chervonenkis formal context dimension (VCC))**.** *Let $\mathbb{K} := (G, M, I)$ be a formal context. Let*

$$\mathcal{S} := \mathfrak{B}_1((G, M, I)) = \{A \subseteq G \mid (A, B) \in \mathfrak{B}((G, M, I)) \text{ for some } B \subseteq M\}$$

be the closure system of all concept extents. The **Vapnik-Chervonenkis formal concept dimension** *(VCC) is defined as the Vapnik-Chervonenkis dimension of $\mathcal{S}$.*

Theorem 5 (cf. also [Albano and Chornomaz, 2017, Albano, 2017a,b])**.** *Let $\mathbb{K} := (G, M, I)$ be a formal context and let $\mathcal{S} := \mathfrak{B}_1((G, M, I))$. Then a set $\{g_1, \ldots, g_l\} \subseteq G$ of objects is shatterable w.r.t. $\mathcal{S}$ if and only if there exists a set $\{m_1, \ldots, m_l\} \subseteq M$ of attributes such that*

$$\forall i, j \in \{1, \ldots, l\} : (g_i, m_j) \in I \iff i \neq j. \tag{30}$$

[26] A lattice $\mathbb{L}$ is called distributive if we have $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ for arbitrary $x, y, z \in \mathbb{L}$.

Proof. **if:** Let $A \subseteq \{g_1, \ldots, g_l\}$. Take the formal concept (A'', A'). Then A'' contains all $g_i \in A$ and for all j with $g_j \notin A$ because of $m_j \in A'$ we have $g_j \notin A''$ and thus $A'' \cap \{g_1, \ldots, g_l\} = A$ which shows that $\{g_1, \ldots, g_l\}$ is shatterable w.r.t. $\mathcal{S}$.
only if: If $\{g_1, \ldots, g_l\}$ is shatterable then for every g_i there exists a formal concept (A_i, B_i) such that $g_i \notin A_i$ and $\forall j \in \{1, \ldots, l\} \backslash \{i\} : g_i \in A_i$. But this means that for every $i \in \{1, \ldots, l\}$ there exists an attribute m_i such that $(g_i, m_i) \notin I$ and $\forall j \in \{1, \ldots, l\} \backslash \{i\} : (g_i, m_j) \in I$. □

Corollary 1. *The Vapnik-Chervonenkis formal context dimension of a context (G, M, I) is equal to the Vapnik-Chervonenkis formal context dimension of the dual context (M, G, I^∂), where $I^\partial = \{(m, g) \mid g \in G, m \in M, gIm\}$.*

5.2.2 Computation of the Vapnik-Chervonenkis dimension

In this section we shortly propose some methods to actually compute the Vapnik-Chervonenkis dimension for different closure systems.

Computing the Vapnik-Chervonenkis dimension if the closure system is given via formal implications

If the closure system $\mathcal{S}$ is given by all valid formal implications, then computing the V.C.-dimension can be done by searching for an implication-free set A of maximal cardinality. To do this, one can solve the following binary program:

$$\sum_{i=1}^{k} m_i \longrightarrow max \tag{31}$$

$$w.r.t. \tag{32}$$

$$\forall (Y, Z) \in \mathfrak{I}(\mathcal{S}) : \sum_{i:v_i \in Y} m_i + \frac{1}{|Z|} \sum_{i:v_i \in Z} m_i \leq |Y| \tag{33}$$

$$m = (m_1, \ldots, m_k) \in \{0, 1\}^k \tag{34}$$

Here, condition (33) codifies the demand that for a valid implication $Y \longrightarrow Z$ a shatterable (implication-free) set A necessarily cannot contain any element of Z if it contains all elements of Y. Instead of the whole set of implications in (33) one can also use only that valid implications $Y \longrightarrow Z$ where Y is minimal (in the sense that $\tilde{Y} \longrightarrow Z$ is not valid anymore for every $\tilde{Y} \subsetneq Y$) and Z is maximal (in th sense that $Y \longrightarrow \tilde{Z}$ is not valid anymore for every $\tilde{Z} \supsetneq Z$). This set of implications is referred to as the **generic base** in formal concept analysis (cf., e.g., [Bastide et al., 2000], where also an algorithm for extracting the generic base is given). Note that in (33) one cannot use an arbitrary implication base: For example the implication base

$$\mathfrak{I} := \{\{v_1\} \longrightarrow \{v_2\}, \{v_2, v_3\} \longrightarrow \{v_4\}\}$$

induces the further implication $\{v_1, v_3\} \longrightarrow \{v_4\}$ and thus the set $A = \{v_1, v_3, v_4\}$ is not shatterable, but the "anti-implications"

$$\mathfrak{I} := \{\{v_1\} \longrightarrow \{\neg v_2\}, \{v_2, v_3\} \longrightarrow \{\neg v_4\}\}$$

obtained by demanding (33) only for the implication base $\mathfrak{I}$ would not exclude the set A although it is not shatterable.

If one wants to compute for example the Vapnik-Chervonenkis principal ideal dimension VCPI of an explicitly given complete lattice $\mathbb{L} = (L, \leq)$, one can firstly construct the formal context $\mathbb{K} := (V, V, \geq)$. Then, the closure system of all intents of this context is exactly the set of all principal ideals of $(L, \leq)$ and one can compute the generic base of all implications that are valid in this closure system. Finally, one can build and solve the binary program (31). Actually, due to Theorem 2 it suffices to look only at the reduced context where join-irreducible elements of $\mathbb{L}$ are removed.[27]

Computing the Vapnik-Chervonenkis upset dimension: Computing the width

If one wants to compute the Vapnik-Chervonenkis upset dimension, in principle one can use the binary program (31), but since the Vapnik-Chervonenkis upset dimension is simply the width, one can also use other more efficient algorithms to compute the width. One possibility is to reformulate the problem of computing the width of a poset as a matching problem in a bipartite graph: Define the bipartite graph $G = (V \times \{1\}, V \times \{2\}, E)$ where the set of vertices is the disjoint union of V and V and the two parts of G are essentially two copies of the poset V and an edge $e = ((v, 1), (w, 2))$ is in E iff $v < w$. Now one can compute a maximal matching in G. The maximum matching then corresponds to a minimum size chain partitioning of V where two elements v and w with $v < w$ are in the same partition iff the edge $((v, 1), (w, 2))$ is in the maximal matching. The number of partitions is then $|V| - m$ where m is the size of the maximal matching. This means that we have found a minimal chain partitioning of V with size $|V| - n$ which is due to Dilworth's theorem identical to the maximal cardinality of an antichain, i.e., the width. To actually compute the maximum matching one can use e.g. the algorithm of Hopcroft and Karp (Hopcroft and Karp [1971]), which would have time complexity $\mathcal{O}(|V|^{\frac{5}{2}})$ in our situation.

Computing the Vapnik-Chervonenkis formal context dimension VCC

[27]Note that for an explicitly given poset $(V, \leq)$ that is not a complete lattice, the family $\mathfrak{pi}((V, \leq))$ of all principal ideals is generally not a closure system, but one can look at the closure system that is generated by all principal ideals of $(V, \leq)$ (of course, without the need of explicitly computing it). Then, to make the computation more efficient, one can similarly remove all reducible attributes (and also all reducible objects) from the context $\mathbb{K} = (V, V, \geq)$, where an attribute a is called reducible if the formal concept $(\{a\}', \{a\}'')$ is meet-reducible and an object o is called reducible if the formal concept $(\{o\}'', \{o\}')$ is join-reducible.

To compute the Vapnik-Chervonenkis formal context dimension VCC one can simply make use of Theorem 5. One can equivalently express condition (30) of Theorem 5 by saying that A set $A = \{g_1, \ldots, g_l\}$ of objects is shatterable w.r.t. $\mathfrak{B}_1(\mathbb{K})$ if and only if there exists a set $B = \{m_1, \ldots, m_l\}$ such that for every object $g \in A$ there exists exactly one attribute $m \in B$ with $(g, m) \notin I$ and if furthermore, for every attribute $m \in B$ there exists also exactly one object $g \in A$ with $(g, m) \notin I$. These two conditions can also be incorporated via inequality constraints. Thus, we can compute the Vapnik-Chervonenkis formal context dimension of a context $\mathbb{K}$ by jointly analyzing pairs (A, B) of an object set A and a an intent set B satisfying (30). With the notation of Section 4.2 we have to solve the binary program

$$
\begin{aligned}
& & z_1 + \ldots + z_m \longrightarrow \max & \\
& & w.r.t. & \\
\forall i \in \{1, \ldots, m\}: & \quad (n-1) \cdot z_i + \sum_{j : A_{ij} = 0} z_{j+m} \leq n & & (35)\\
& & -z_i + \sum_{j : A_{ij} = 0} z_{j+m} \geq 0 & (36)\\
\forall j \in \{1, \ldots, n\}: & \quad (m-1) \cdot z_{j+m} + \sum_{i : A_{ij} = 0} z_i \leq m & & (37)\\
& & -z_{j+m} + \sum_{i : A_{ij} = 0} z_i \geq 0. & (38)
\end{aligned}
$$

Here, the constraints (35) and (36) are redundant if z_i is zero, e.g., if object g_i does not belong to the envisaged shatterable set A. If object g_i is in the envisaged shatterable set A, then (35) demands exactly that there is maximal one attribute m_j in the associated attribute set B with $(g_i, m_j) \notin I$ and constraint (36) further demands that there is also at least one such attribute. The constraints (37) and (38) analogously codify the dual statement where the roles of objects and attributes are exchanged. Here, unfortunately one generally cannot drop any integrality constraint, so the computation of the V.C. formal context dimension is generally very hard.

5.3 Taming the monster: pruning closure systems via Vapnik-Chervonenkis theory

The last section showed how to compute the V.C.-dimension for several closure systems and how to identify shatterable sets of maximal cardinality. The ability to identify such big shatterable sets supplies us with a simple possibility of effectively taming the closure system by removing such big shatterable sets to get a test statistic that is less crude in the sense that one gets better bounds in (25) and (26) due to a lower V.C.-dimension.

Concretely, for e.g. the closure system $\mathcal{U}((V,\leq))$ of upsets, every upset $M \in \mathcal{U}((V,\leq))$ can be characterized by the set $\min(M)$ of all minimal elements of M via $M = \uparrow \min(M)$. To tame $\mathcal{U}((V,\leq))$ one can compute an antichain[28] A of maximal cardinality and then remove this antichain A (or a subset$\tilde{A}$ of A) from $\mathcal{U}((V,\leq))$ by considering not all upsets $\mathcal{U}((V,\leq)) = \{\uparrow B \mid B \subseteq V\}$ but only the family $\mathcal{S}' = \{\uparrow B \mid B \subseteq V\backslash A\}$ of all upsets that are generated by $V\backslash A$ (or $V\backslash\tilde{A}$). This family is generally not a closure system, anymore, but one can simply take not the family $\mathcal{S}$ but the closure system[29] $\tilde{\mathcal{S}}$ that is generated by $\mathcal{S}'$ via $\tilde{\mathcal{S}} := \mathrm{cl}(\mathcal{S}') = \bigcap\{\mathcal{F} \mid \mathcal{F}$ closure system on $V\backslash A,\ \mathcal{F} \supseteq \mathcal{S}'\}$.

For taming the Vapnik-Chervonenkis formal context dimension of a given formal context $\mathbb{K}$ one can similarly look for objects involved in a shatterable set of maximal cardinality and then take the closure system of the concept extents of the formal concept lattice generated by the modified context where the objects involved in a shatterable set of maximal cardinality are removed. Generally, two issues arise here:

Firstly, for a closure system $\mathcal{S}$ of V.C.-dimension VC one usually has more than one shatterable set of size VC. To effectively tame the closure system one therefore has to remove the first found shatterable set of size VC and then one has to look at further shatterable sets of size VC and remove them, too. In this situation, it could be the case that the result of the taming procedure depends upon which shatterable set of maximal cardinality was removed first. To avoid this problem, one can alternatively look jointly at all shatterable sets of maximal cardinality and remove them all. However, this could have the effect that in one step a huge number of sets is removed such that the V.C.-dimension becomes too small already in one step. Furthermore, if one decides for removing only subsets of shatterable sets, then it is not straightforward, which subsets exactly to remove and also here, the choice of the removed subsets could possibly have an impact on which set would be a shatterable set of maximal cardinality in the next step. Since the ability of removing not only whole shatterable sets but also subsets would be very helpful for taming in a very flexible way, this could be seen as a problem.

Secondly, the taming of the closure system is only a statistical "regularization procedure" that only cares for the purely statistical aspects. Thus, it is desirable to analyze the taming also with respect to its "conceptual behavior" in the sense that one should care for how flexible the tamed closure system is w.r.t. which sets are in the closure system and how fine-grained the tamed closure system thus is w.r.t a purely descriptive/conceptual point of view. This is clearly a matter of the concrete application. For the closure systems of upsets and the closure system of concept extents we will now give concrete proposals for taming that are in our view also acceptable from a conceptual point of view in the situations of the application examples given later in Sections 6.1 and 6.3.

[28] As shown in Section 5.2.1, for the closure system of all upsets of a poset $(V,\leq)$, the shatterable sets are exactly the antichains of $(V,\leq)$.

[29] For the computation of the test statistic on the tamed closure system $\tilde{\mathcal{S}}$ one does not need to compute $\tilde{\mathcal{S}}$ explicitly, see Section 5.3.3.

5.3.1 Taming upsets in the context of inequality analysis

The closure system of upsets played a crucial role in the context of stochastic dominance. One field of application of stochastic dominance is multivariate poverty or inequality analysis. In this context one does not start with a poset V, instead one has some (often totally ordered) "dimensions" of poverty/inequality. In our example of application given in Section 6.1., we have basically the 3 dimensions *Income*, *Education* and *Health*. The poset $(V, \leq)$ is then given by the three dimensional attributes of all persons in the survey equipped with the coordinate wise order (i.e. person x is poorer than or as equally poor as person y iff she is poorer than or as equally poor as y w.r.t. every dimension). Then, the concept of an upset codifies a "multivariate poverty line" in the sense that an upset M would be a reasonable concretion of the term *non-poor* by saying that every person in the set M could be termed *non-poor* and every person in the complement of M could be termed *poor*. The statement of stochastic dominance $X \leq_{SD} Y$ where X describes one subpopulation and Y another subpopulation would then mean $\forall M \in \mathcal{U}((V, \leq));\ P(X \in M) \leq P(Y \in M)$ which can be simply translated to the statement: "However the term *poor* is actually reasonably concretized, in every case the proportion of the *non-poor* persons in subpopulation corresponding to X is always lower than or equal to the proportion in the subpopulation related to Y." Now, how can we reasonable tame the closure system of upsets in this context? Since the closure system of upsets is getting very big already for small posets V, a taming by explicitly removing upsets seems hopeless, but one can use the fact that every upset M is generated by its minimal elements via $M = \uparrow \min(M)$ and look at antichains instead of upsets. One way to tame the closure system of all upsets, i.e., the closure system of all reasonable concretions of the term *non-poor* in a conceptually reasonable way could be to exclude some very "skew" concretions of the term *poor*: One can try to remove upsets generated by antichains consisting of very unbalanced elements, i.e. attributes that are very low in one dimension and at the same time very high in another dimension. To do so, one has to concretize here, what low and high means. One possibility would be to firstly standardize[30] every dimension to be $U[0, 1]$ distributed. Concretely, if $X \in \mathbb{R}^{n \times p}$ is the matrix containing the n attributes of dimension p, define for $j = 1, \ldots, p$ the univariate empirical distribution[31] function F^j according to the distribution of the j-th dimension in the sample and define $Z \in \mathbb{R}^{n \times p}$ via $Z_{ij} = F^j(X_{ij})$. Then Z is a transformation of X where every dimension $Z_{\bullet j}$ has values ranging from $\frac{1}{n}$ to 1 allowing for some kind of relative comparability of the transformed attributes with the simple interpretation that if $Z_{ij} = \frac{l}{n}$ the person i is the

[30] If one has any external substance matter insight into how some decrease in one dimension can be reasonably be compensated for by an increase in another dimension, one should try to reflect this substance matter insight into the taming procedure. Of course, the herein proposed taming procedure has to be understood as a general purpose procedure that could be substantially improved by modifications based on substance matter considerations.

[31] One can use here the complementary distribution function $F^j(x) = \frac{|\{i | X_{ij} \geq x\}|}{n}$ or the usual distribution function $F^j(x) = \frac{|\{i | X_{ij} \leq x\}|}{n}$, which would lead to identical results. We use here the complementary distribution function because it fits more to the notion of upsets.

$n-l+1$-th poorest person in the sample w.r.t. dimension i. To concretize the notion of an "imbalanced" multivariate poverty line one can firstly define a transformed multivariate attribute $Z_{i\cdot} = (Z_{i1}, \ldots, Z_{ip})$ as balanced if $\max\{Z_{i1}, \ldots, Z_{ip}\} - \min\{Z_{i1}, \ldots, Z_{ip}\} \leq \delta$ for a fixed threshold δ. Then, one can define a poverty line as balanced if it is generated by an antichain that consists only of balanced attributes. If one sets the threshold δ globally to one fixed value, then w.r.t. V.C.-dimension it can happen that the V.C.-dimension can vary drastically from region to region in the sense that e.g. for regions of medium transformed Z values there are big sets of balanced elements building an antichain whereas for extreme Z values there are only small sized antichains of balanced elements. For the statistical side of the taming procedure this could lead to a very brute taming of regions of extreme Z values without globally reducing the V.C.-dimension very much. Of course, in the proof of the Vapnik-Chervonenkis inequality one essentially deals with the cardinality of the closure system and this is actually sized down by the procedure, so the statistical taming would actually still be achieved, but only if one is taming very strongly which means that one would reduce far more sets in regions of extreme Z-values/low width than in regions of medium Z-values/high width where the density of upsets is already very high. (Note that every antichain of size k induces 2^k upsets). One can avoid this seemingly bad effect with the following localization method[32]:

First, fix some envisaged V.C.-dimension h_0. For given $\alpha \in [0,1]$ and for arbitrary $\varepsilon > 0$ define an ε-stratum around the center α as the set $M_\varepsilon(\alpha) = \{v_i \mid \forall j \in \{1, \ldots, p\} : |Z_{ij} - \alpha| \leq \varepsilon\}$. Then, for fixed α choose $\varepsilon(\alpha)$ such that the V.C.-dimension of $M_{\varepsilon(\alpha)}(\alpha)$ is lower than or equal to h_0 and such that $\varepsilon(\alpha)$ is maximal w.r.t. this property. Then collect in a set $T(h_0) := \bigcup\{M_{\varepsilon(\alpha)}(\alpha) \mid \alpha \in [0,1]\}$ all strata $M_{\varepsilon(\alpha)}(\alpha)$. The closure system $\mathcal{S}_{h_0} = \mathrm{cl}(\mathcal{F}_{h_0})$ generated by the family of sets $\mathcal{F}_{h_0} = \{\uparrow B \mid B \subseteq T(h_0)\}$ can then serve as a tamed subsystem of $\mathcal{S}$. Note that the V.C.-dimension of $\mathcal{S}_{h_0}$ needs not to be h_0, it can be higher, because elements of different strata can build an antichain of size bigger than h_0. A further important point is that with this taming procedure we have introduced some asymmetry: In the case of the full closure system $\mathcal{S}$ of upsets it played no role that we looked at upsets and not at downsets: If we would have dealt with downsets to model the *poor* persons instead of modeling the *non-poor* persons via upsets, we would still have got the same results. The reason for this is simply that the complements of upsets are downsets and vice versa. In contrast to this, the complement of special selected upsets generated by antichains of some subset $T(h_0)$ of V are not necessarily downsets generated by the antichains of $T(h_0)$. Thus, for practical applications, one should analyze the results of both the tamed upset and the downset approach, which we will do in the example of application given in Section 6.1.

[32]If a taming with a global threshold δ appears more reasonable from a conceptual point of view in a concrete situation of application then a global taming may still be a better choice. However, in the example of application given in Section 6.1 we see no direct conceptual advantages of a global taming.

5.3.2 Taming formal contexts in the context of cognitive diagnosis models and knowledge space theory

For taming the closure system of all concept extents of a given formal context $\mathbb{K} = (G, M, I)$ with V.C.-dimension VC, in principal one can search for all shatterable objects sets of size VC (either step by step or in one whole step, see the remarks above) and exclude the objects of these sets from the context $\mathbb{K}$ to obtain a reduced context $\tilde{\mathbb{K}} = (\tilde{G}, M, I \cap \tilde{G} \times M)$ which then has a V.C.-dimension lower than VC (If this reduced V.C.-dimension is still too high, one can repeat the taming process until the resulting V.C.-dimension is low enough). For the actual data analysis one can then firstly take the closure system $\mathfrak{B}_2(\tilde{\mathbb{K}})$ of the intents of the reduced context $\tilde{\mathbb{K}}$ and secondly define the reduced closure system $\tilde{\mathcal{S}} := \{\{g \in G \mid \forall m \in B : gIm\} \mid B \in \mathfrak{B}_2(\tilde{\mathbb{K}})\}$ generated by all intents of the reduced context $\tilde{K}$ but w.r.t. the objects of the full original context $\mathbb{K}$. In Section 5.3.3 we will show how to do this in computational terms. Another possibility would be to not remove objects but attributes. In practical applications, often objects represent data points and the attributes represent the "multidimensional" values of the data points, so in classical situations one usually has much more objects than attributes. In these situations it appears more natural to remove objects, because if one would remove attributes, then one would remove these attributes for the whole big set of all objects. Compared to this, if one removes objects, then one removes only the specific concept intents generated by these objects (and also intents that are jointly generated by removed objects and non-removed objects). If one removes objects in the above described way, then one reduces the V.C.-dimension of the closure system under which the final analysis will be done. However, from a conceptual/descriptive/substance matter point of view, one does not know if one had removed sets that are actually interesting/important or that one did not remove uninteresting/unimportant sets. In some situations one can tame a context in a more guided manner:

One interesting example where one has some kind of substance matter guidance for taming is the case of **cognitive diagnosis models** (CDM), which are some kind of non-parametric item response models. Note that cognitive diagnosis models are very closely related to the theory of knowledge spaces ([Doignon and Falmagne, 2012], see [Heller et al., 2015]) which is itself closely related to formal concept analysis (see [Rusch and Wille, 1996]). In cognitive diagnosis models one has a set G of persons which respond to a set $M = \{1, \ldots, |M|\}$ of cognitive tasks, for example math tasks like fraction addition or fraction subtraction (for one well known fraction-subtraction data set see [Tatsuoka, 1984]). In contrast to more classical item response theory (IRT), in cognitive diagnosis modeling one is not mainly interested in measuring the abilities of persons and the difficulties of items, instead one is interested in the cognitive processes that generated the observed response patterns. Here, one demand is to give persons not only one or more numbers that measure their ability but to give a more qualitative feedback about which concrete skills the persons possess and which skills they do not possess. To do so, one develops (either theory driven or data driven or, in the best case, driven by a theory that

was rigorously empirically tested and persisted the tests) a so called Q matrix that specifies for every item, what kind of skills are in principle necessary to solve this item. Concretely, for a set of K relevant skills, a Q-matrix is a $|M| \times K$ matrix of zeros and ones where an entry $Q_{ij} = 1$ means that the skill j is needed to solve item i. In the simplest case one assumes that a person is expected to solve item i if she possesses all skills that are needed to solve item i, so a lack of one skill cannot be compensated by other skills. (This is the DINA model, cf. [Haertel, 1989, Junker and Sijtsma, 2001], but there are also other compensatory variants like the DINO model, cf. [Junker and Sijtsma, 2001].) Furthermore, one assumes the possibility of slipping an item one is principally prepared to solve and of luckily guessing the right answer of an item one is not prepared to solve. If for the moment we ignore the possibility of slipping and guessing, then the Q- matrix induces some structure of the idealized item response patterns that are possible if the probabilities of guessing and slipping are zero. For example if for solving one item i one needs all skills that one also needs for solving item i' plus some more, then response patterns of the form

$$(\ldots \underbrace{1}_{\text{i-th entry}} \ldots \underbrace{0}_{\text{i'-th entry}} \ldots)$$

are only possible due to a lucky guess of item i or a slipping of item i'. This fact can be expressed by saying that the formal implication $\{i\} \longrightarrow \{i'\}$ is valid in the closure system[33] $\mathcal{S}_Q := \mathfrak{B}_2(\{1,\ldots,K\},\{1,\ldots,|M|\},1-Q^T)$ of all possible idealized response patterns. To see that the closure system $\mathfrak{B}_2(\{1,\ldots,K\},\{1,\ldots,|M|\},1-Q^T)$ of the intents of the context $\mathbb{K}_Q := (\{1,\ldots,K\},\{1,\ldots,|M|\},1-Q^T)$ is exactly the space of all possible idealized response patterns, note that the intents are generated as $\{A' \mid A \subseteq \{1,\ldots,K\}\}$ where a set A can be understood as the set of skills an imaginary person does **not** posses. Then A' is the set of all items i with $\forall j \in A : (1-Q^T)_{ij} = 1$, i.e. the set of all items i where all skills the person does not possess are actually not needed to solve the item i. Thus, the intent A' is in fact the set of all items a person not possessing exactly all skills of A would actually be able to solve and all intents are exactly all observable idealized response patterns. A valid formal implication $Y \longrightarrow Z$ of $\mathbb{K}_Q$ could be interpreted in this situation as "All skills that are not necessary for solving any item from Y are also not necessary for solving items from Z" or alternatively as "every imaginary person who possess all skills for solving all items from Y also possesses all necessary skills for solving all items from Z".

Now, one can incorporate some or all valid implications of the idealized response pattern space $\mathcal{S}_Q$ to reduce the original closure system by looking only at concept intents of the original context $\mathbb{K} = (G,M,I)$ (where $gIm \iff$ person g has solved item m) that respect all or some of the valid implications of the formal context $\mathbb{K}_Q = (\{1,\ldots,K\},\{1,\ldots,|M|\},1-Q^T)$ representing the idealized response pattern space. If one enforces that all valid implications of the idealized response pattern space should also valid in the tamed closure system for the final analysis, then the V.C.-dimension would

[33] By abuse of notation, we identify the matrix $1-Q^T$ with the relation $\{(i,j) \mid (1-Q^T)_{i,j} = 1\}$.

decrease, but maybe unnecessarily too much. To enforce only a subset of implications one has to reasonably decide, which implications to include and which implications to not include. This can be made based on theoretical substance matter considerations about which implications are expected to be more clearly valid from a cognition theoretic perspective and which implications are maybe more questionable because they are due to a less rigorous but a more schematic specification of involved skills. To do so, one can substantially make use of the technique of **attribute exploration** (see [Ganter and Wille, 2012, p.85]) known from formal concept analysis: Given the formal context $\mathbb{K}_Q$ an algorithm like the next closure algorithm (see [Ganter and Wille, 2012, p.66-68]) can compute all formal concepts and also the so called stem base (see [Ganter and Wille, 2012, p.83]) of all valid implications of this context. In attribute exploration, at every step of the computation of a new implication, the user is asked in an interactive way, if the currently computed implication is actually true. Then the user can say that the implication is actually true or provide the algorithm with an object with specific attributes that are actually contradicting the formal implication. Then the algorithm would include this counterexample into the context and proceed, but not by computing all implications from the modified context anew, but by knowing that all implications computed before the counterexample was given are still valid in the modified context.

Another possibility of selecting implications to include for taming is to do it data driven. One can look for example at all valid implications $Y \longrightarrow Z$ of $\mathcal{S}_Q$ that are respected by at least a certain proportion C of objects from the original context $\mathbb{K}$ in the sense that at least a proportion C of persons, who solved all items of Y did also solve all items from Z. Formally, this can be described as enforcing all rules $Y \longrightarrow Z$ that have a so called confidence[34] $\mathrm{conf}(Y \longrightarrow Z)$ of at least C, where $\mathrm{conf}(Y \longrightarrow Z) := \frac{\mathrm{supp}(Y \cup Z)}{\mathrm{supp}(Y)}$ and $\mathrm{supp}(A) := |A'|$ and the operator $'$ is meant w.r.t. the original context $\mathbb{K}$. Here, the issue arises that if for example the rules $Y \longrightarrow Z_1$ and $Y \longrightarrow Z_2$ have a confidence above the threshold C then they would be included and furthermore the rule $Y \longrightarrow Z_1 \cup Z_2$ would implicitly be also valid in the tamed closure system, but this rule does not necessarily have a confidence of C. One can deal with this issue in different ways. One way of taming would be to enforce a set $\mathfrak{I}$ of implications that is deductively closed (this means that if an implication follows from some implications from $\mathfrak{I}$ then it should be already in the set $\mathfrak{I}$) and that only consists of implications with confidence above C and that is furthermore maximal w.r.t. these properties. Such maximal sets are generally not unique. Furthermore, if one has the idea that response patterns that violate implications of the idealized response pattern space are due to a random guessing or slipping, then if the slipping/guessing for different items is independent, for valid idealized implications $Y \longrightarrow Z_1$ and $Y \longrightarrow Z_2$ with confidence C_1 and C_2 one would expect a confidence of the implication $Y \longrightarrow Z_1 \cup Z_2$ that is generally lower than $\min\{C_1, C_2\}$. Thus, choosing the same threshold for implications with differently sized consequents seems to be not natural. Another way to proceed is to

[34]This term is used in the field of association rule mining, cf., e.g., [Agrawal et al., 1993, Piatetsky-Shapiro, 1991] which is also related to formal concept analysis, cf., e.g., [Lakhal and Stumme, 2005].

simply take the set of all implications with support above a threshold C and accept that with this set also implications from its deductive closure that may have a confidence smaller than C are also implicitly included for taming the closure system. If one takes this route, one is faced with computing all implications with confidence above C. For an implication $Y \longrightarrow \{z_1, z_2\}$ with confidence above C it is necessary that also the implications $Y \longrightarrow \{z_1\}$ and $Y \longrightarrow \{z_2\}$ have confidence above C, so it suffices to look only at implications with a singleton consequence. (The implication $Y \longrightarrow \{z_1, z_2\}$ is included for taming if and only if both $Y \longrightarrow \{z_1\}$ and $Y \longrightarrow \{z_2\}$ have confidence above C because this is necessary and if both $Y \longrightarrow \{z_1\}$ and $Y \longrightarrow \{z_2\}$ have confidence above C they are both included and thus also $Y \longrightarrow \{z_1, z_2\}$ has to be included since it follows from the included implications $Y \longrightarrow \{z_1\}$ and $Y \longrightarrow \{z_2\}$.) Instead of computing all implications with a singleton consequence one can also compute in a first step only that implications that have a minimal antecedent. Then one could exclude such implications $Y \longrightarrow \{z\}$ with confidence lower than C and recompute the generic base. To do so, one can directly work with the formal context $\mathbb{K}_Q := (\{1, \ldots, K\}, \{1, \ldots, |M|\}, 1 - Q^T)$. One can compute the generic basis and split every implication $Y \longrightarrow \{z_1, \ldots, z_l\}$ into implications $Y \longrightarrow \{z_1\}, \ldots, Y \longrightarrow \{z_l\}$ with singleton consequents. Then, for every such implication $Y \longrightarrow \{z\}$ with confidence lower than C one can exclude it by adding a the counterexample $Y'' \backslash \{z\}$ as a further item pattern to the context $\mathbb{K}_Q$. (Here the operation $''$ is meant w.r.t. the context $\mathbb{K}_Q$.) Then one can compute again and again the generic base of the enlarged context until no rule has confidence lower than C anymore. A computationally more elegant way would be to not to recompute the whole rule base of the enlarged context anew. In the spirit of attribute exploration (see [Ganter and Wille, 2012, p.85]), one can smartly exclude implications with confidence lower than C directly during the generation of the rules. However, since one does not work with the generic base, but with the stem base, the result would then be different and would furthermore dependent one the concrete order in which the computed implications were presented to the user.

5.3.3 How to compute the test statistics for the tamed closures systems

In this section we shortly indicate, how one can compute the test statistic for a tamed closure system. We start with the example of the closure system of upsets. In Section 5.3.1 we came up with the tamed family of sets $\mathcal{F}_{h_0} = \{\uparrow B \mid B \subseteq T(h_0)\}$ that generates the closure system $\mathcal{S}_{h_0} = \mathrm{cl}(\mathcal{F}_{h_0})$. Of course, it would be intractable to explicitly compute the closure system $\mathcal{S}_{h_0}$ generated by $\mathcal{F}_{h_0}$ because it is simply too big. Fortunately, the explicit computation of $\mathcal{S}_{h_0}$ is not needed: The closure system $\mathcal{S}_{h_0}$ simply consists of all possible intersections of sets of the generating family of sets $\mathcal{F}_{h_0}$. For ease of presentation, assume that $(V, \leq)$ itself is already a complete lattice (otherwise, simply take its Dedekind-MacNeille completion, cf., e.g., [Ganter and Wille, 2012, p.48]). The closure system $\mathcal{S}_{h_0}$ is simply the set of all possible intersections of sets of the family $\mathcal{F}_{h_0}$. For a finite[35] family $(\uparrow A_i)_{i \in \{1,\ldots,n\}}$ of upsets from $\mathcal{F}_{h_0}$, the intersection $\uparrow A_1 \cap \ldots \cap \uparrow A_n$

[35]The finiteness is actually not needed, it only makes the presentation more simple, here.

can be written as $\uparrow A_1 \cap \ldots \cap \uparrow A_n = \bigcup\{\uparrow a_1 \cap \ldots \cap \uparrow a_n \mid \forall i \in \{1,\ldots,n\} : a_i \in A_i\} = \bigcup\{\uparrow \bigvee\{a_1,\ldots a_n\} \mid \forall i \in \{1,\ldots,n\} : a_i \in A_i\}$. Thus, $\mathcal{S}_{h_0}$ can be written as $\mathcal{S}_{h_0} = \{\uparrow B \mid B \subseteq \bar{T}(h_0)\}$ where $\bar{T}(h_0) = \{\bigvee A \mid A \subseteq T(h_0)\}$. Since $\bigcup_{i\in I} \uparrow B_i = \uparrow \bigvee_{i\in I} B_i$ for arbitrary families $(B_i)_{i\in I}$, the closure system $\mathcal{S}_{h_0}$ is closed under arbitrary unions and thus, because of Birkhoffs theorem, the valid implications of $\mathcal{S}_{h_0}$ are simple implications. Thus, we can firstly calculate all simple implications or a basis thereof and implement them in a linear program: For example one can compute for every $x \in V$ the set $\downarrow x \cap T(h_0)$ of all elements of $T(h_0)$ that are below x. Then, one can take from the set M of all upper bounds of $\downarrow x \cap T(h_0)$ the minimal elements $\min M$. Finally, for every $y \in \min M$ one simply has to implement the associated implication $\{x\} \longrightarrow \{y\}$ as an inequality constrain in the linear program.

For the case of non-guided taming of a closure system that is given by a generating formal context, remember that the taming was simply done by removing objects from the context (but only for the generation of the closure system of the intents, and not for the whole analysis). Let I denote the set of indices of the objects that were excluded for the generation of the closure system. To compute the statistic for the tamed context, one only has to modify the program (24) to the following program:

$$\begin{aligned} \langle (w_1^{ext},\ldots,w_m^{ext},w_1^{int},\ldots,w_n^{int}),(z_1,\ldots,z_m,z_{m+1},\ldots,z_{m+n})\rangle &\longrightarrow \max \\ w.r.t. & \\ \forall (i,j) \text{ s.t. } A_{ij}=0 : z_i &\leq 1 - z_{j+m} \\ \forall i \in \{1,\ldots,m\} : \sum_{k:A_{ik}=0} z_{k+m} &\geq 1 - z_i \\ \forall j \in \{1,\ldots,n\} : \sum_{k\notin I:A_{kj}=0} z_k &\geq 1 - z_{j+m} \end{aligned} \tag{39}$$

Here, the only difference is that in the last set of inequalities, one does not sum over every object index k but only over that indices, that were not excluded for the generation of the closure system. To see the validity of this modification, simply note that the three verbalizations directly above the linear program (24) are still exactly characterizing the situation with the only modification of point 3, which has to be modified to

"Dually, if attribute m_j does not belong to the intent, then there exists at least one object g_k that was not excluded for the generation of the closure system of intents, and that belongs to the extent, but does not have attribute m_j."

For the guided taming the computation of the test statistic is straightforward. Since the formal implications one additionally imposes are computed explicitly, one can modify the binary program described in Section 4.2 by additionally implementing the further imposed implications as inequality constraints like described in Section 4.1.

6 Examples of application

In this section we apply the developed methods to different data sets. The applications should on the one hand be not taken at face value as serious substance matter applications. On the other hand, they should also be not misunderstood as pure toy examples. The aim of the following examples of application is to show that the developed methods are in fact applicable to "real-world" data sets and that these methods are in principle very flexible and can also deal with different kinds of data deficiency. The big part that is missing to make the examples serious substance matter studies is the fact that at much stages of the analysis, some substance matter considerations have to be made or could maybe be made to make the analysis more decisive. However, since the authors are clearly no experts in the substance matter fields the applications are related to, they would like to refrain from making such substance matter decisions, if possible, or to make the actually needed substance matter decisions only for purposes of illustration. In the following examples, especially in our main example of Section 6.1, we analyze the data sets by a more generic way of proceeding and, if appropriate, shortly indicate, at which steps and in which way one could make a more refined data analysis, that of course would be dependent on some substance matter decisions.

6.1 Upsets: Relational inequality analysis

We start with our main example of multivariate inequality analysis using data from the German General Social Survey (ALLBUS) of the year 2014 (GESIS - Leibniz - Institut fur Sozialwissenschaften [2015]). In this survey, altogether 3471 persons participated. Here, we analyze systematic multivariate differences between the group of male and female participants w.r.t. the variables *Income*, *Education* and *Health*. The question about *Health* was asked in a split ballot design to test for a possible impact of different response scales on the result. The participants were asked both in split A and split B about how they would describe their health status in general. The participants of split A got the 5 different answer categories "Sehr gut" (very good), "Gut" (good), "Zufriedenstellend" (satisfactory) , "Weniger gut" (suboptimal) and "Schlecht" (bad) whereas the participants of split B got the additional category "Ausgezeichnet" (excellent). (The english categories in brackets are our own english translation.) For reasons of simplicity, we used here only the participants of split B and did a complete case analysis.

Of course, one could also use both splits for the analysis: If one has some reason to assume that both response scales adequately operationalize the same construct, one can do a joint analysis of both splits by matching the two scales to each other based on their respective empirical distribution functions. This is actually possible because the splitting was random and thus the measured construct has the same distribution in every split.[36]

[36]Note that due to measurement error, which can be different within the two splits, the actual measurements can differ in their distribution. But if the measurement errors are independent of the measured construct and from each other, this will only produce some "smearing" of the measurements, which can

For the joint analysis of the three variables *Income*, *Education* and *Health*, the complete case analysis consisted of altogether 1515 participants (706 female and 809 male) corresponding to a non-response rate of 12.2%. The variable *Income* contributed most to the non-response-rate (the non-response rate for *Income* was 11.8%.) Here, income was asked for in a two step procedure: First with an open question and then, for participants who refused to answer the open question, a categorized question with 23 answer-categories ranging from "no income" to "more than 7500 Euro" was added. This two-step procedure was done to reduce the non-response rate. Here, for simplicity we use the combined answers to the open and the list query, where for participants who answered only the list query simply the mid-points of the interval representing the categorized answer were used as a surrogate for the true income[37]. Note that for our analysis we only need the ordinal structure of the variable *Income* and furthermore we can actually deal also with a partially ordered structure of the dimension income. Thus, here one can also use more cautious approaches where one says for example that an income that is actually only categorically observed as $[a, b)$ is only lower than or equal to another observed income (no matter if precisely observed as $[c, c]$ or imprecisely observed as $[c, d)$) of $[c, d)$ iff $b \leq c$. Another possibility would be to say that categorically observed incomes $[a, b)$ are comparable to itself (i.e. $[a, b) \leq [a, b)$), but not to a precisely observed value $c \in [a, b)$). The stochastic dominance approach is thus very flexible to deal with certain kinds of non-response/interval-valued observations. Here, we do simply work with the combined values where interval-valued observed incomes are replaced by the corresponding interval mid-points.

The variable *Education* is the classification of the level of education according to the International Standard Classification of Education (ISCED) 2011 (see [UNESCO Institute for Statistics (UIS), 2012]) implemented for Germany. On the highest stage, this classification differentiates between 9 different main levels of education:

Level 0:	Less than primary education	Level 5:	Short-cycle tertiary education
Level 1:	Primary education	Level 6:	Bachelor's or equivalent level
Level 2:	Lower secondary education	Level 7:	Master's or equivalent level
Level 3:	Upper secondary education	Level 8:	Doctoral or equivalent level
Level 4:	Post-secondary non-tertiary education		

We treat here the variable *Education* as of totally ordered scale of measurement. In the sample, only the levels from 1 to 8 were observed. Note that also for this dimension the methodology of stochastic dominance would be able to deal with an only partially ordered scale: The ISCED 2011 could also be implementation in a more cautious way: For example, instead of only comparing the highest educational achievements, one could

lead to cases where stochastic dominance w.r.t. the underlying 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 cannot happen.

[37]For the answer category "below 200 Euro" a value of 150 Euro and for the category "more than 7500 Euro " a value of 8750 Euro was assigned.

alternatively look at the whole educational paths and 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 of the dimension *Education* 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*.

We begin with a marginal analysis of all 3 variables. Figure 3 shows the lower cumulative distribution function for every variable for both the male and the female group. One can see that the female group is almost dominated by the male group for the variables *Income*, *Education* and *Health*. With regard to *Income*, the extent of dominance is the highest: 66.4% of the women earn not more than 1300 Euro, but only 31.9% of the men earn not more than 1300 Euro, which is a difference of 34.5 percentage points. Only for the very high income of 12000 Euro there is a small deviation from dominance in the sense that 99.9% of the men earn not more than 12000 Euro, where this is the case for only 99.8% of the women. For the variable *Health* there is only deviation from dominance w.r.t. the percentage of women reporting a health-status *bad*: Only 2.2% of the women report a health status *bad*, which is about 0.7 percentage points lower than the amount of 2.9% for the men. The variable *Education* shows strict dominance.

Now, let us come to the joint analysis. For the statistics

$$D^+ = \max_{M \in \mathcal{U}((V,\leq))} \langle w^x - w^y, m \rangle$$
$$D^- = \min_{M \in \mathcal{U}((V,\leq))} \langle w^x - w^y, m \rangle,$$

where X describes the subpopulation of male, and Y describes the subpopulation of female persons, we obtain

$$D^+ \approx 36.48\%$$
$$D^- \approx -1.21\%,$$

which indicates an almost strict dominance for the joint distribution of the variables *Income*, *Education* and *Health*, where the small deviation from dominance is with $D^- \approx -1.21\%$ not much higher than the largest deviation of -0.7% for the variable *Health* in the marginal analysis. The maximal value of 36.48% is about 2 percentage points higher than for the largest maximal value of 34.5% for the variable *Income* in the marginal analysis. Beyond the purely quantitative analysis one can also look, at which upsets the maximum and the minimum of the test statistic is attained. The maximum of the statistic is attained at an upset U generated by the antichain A (via $U = \uparrow A$) containing 9 elements depicted in Table 3.

The minimal test statistic is attained at an upset generated by an antichain of size 4 described in Table 4. Based on a resampling scheme with 10000 replications, the test sta-

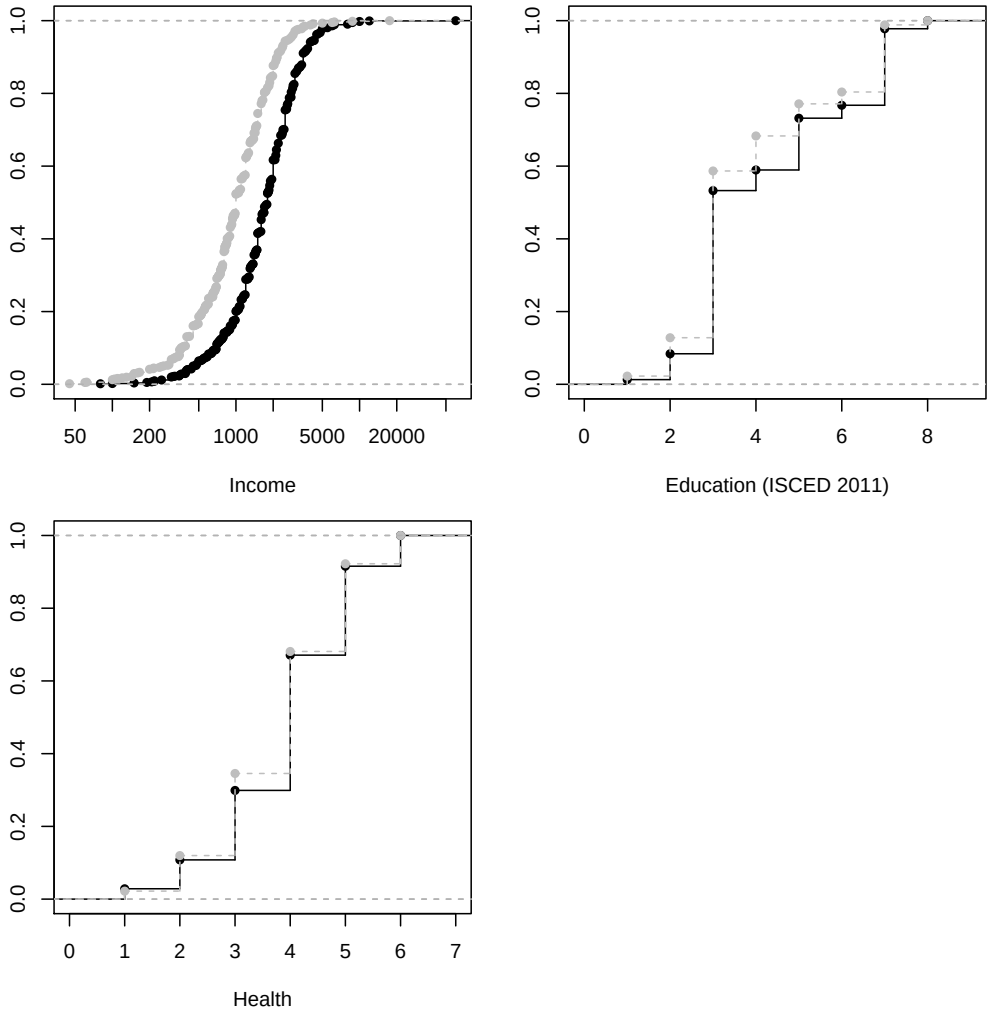

Figure 3: Empirical cumulative distribution function for all 3 considered variables for the male group (black) and the female group (grey).

tistic D^+ appears as highly significantly above 0 whereas D^- is really only non-significantly different from zero: The maximal observed value of D^+ in the resample is 17.48% and the minimal value of D^- observed in the resample is -1.63%, which is very close to -1.21% for the actually analyzed data set, actually, the value of -1.21% is closer to zero for the actual data set than the closest value of the resample. The poset generated by the actually observed data and the coordinate-wise ordering has a Vapnik-Chervonenkis-dimension (width) of 33. For such a V.C.-dimension and an n around[38] 700, the V.C.-inequality

[38]The data set included 706 female and 809 male participants. Note that actually the sampling weights for east and west have to be also taken into account, here, however, we only want to get a rough idea about how sharp the V.C.-inequality is in our situation.

	Income (Euro)	Education (ISCED 2011)	Health (self-reported)	difference	above
1	400 (0.93)	Upper secondary education (0.9)	excellent (0.08)	0.02	0.06
2	650 (0.84)	Lower secondary education (0.99)	excellent (0.08)	0.02	0.06
3	1080 (0.64)	Master's or equivalent level (0.22)	very good (0.32)	0.02	0.07
4	1100 (0.64)	Master's or equivalent level (0.22)	good (0.68)	0.06	0.14
5	1260 (0.55)	Master's or equivalent level (0.22)	satisfactory (0.89)	0.08	0.17
6	1300 (0.55)	Upper secondary education (0.9)	satisfactory (0.89)	0.3	0.49
7	1400 (0.51)	Primary education (1)	good (0.68)	0.25	0.37
8	1400 (0.51)	Upper secondary education (0.9)	bad (1)	0.33	0.49
9	1450 (0.48)	Lower secondary education (0.99)	satisfactory (0.89)	0.32	0.45

Table 3: The antichain $A = \{A_1, \ldots, A_9\}$ that generates that upset $U = \uparrow A$ where the maximum of the test statistic is attained. In brackets the marginal upper quantiles that correspond to the values are given, e.g. the 0.93 behind the 400 in the first row of the first column means that ca. 93% of the persons in the population earn at least 400 Euro. The column *difference* displays for every row i the difference between the proportion of male and the proportion of female persons that are above element A_i. The column *above* shows the proportion of all persons that are above A_i.

	Income (Euro)	Education (ISCED 2011)	Health (self-reported)	difference	above
1	100 (1)	Master's or equivalent level (0.22)	excellent (0.08)	-0.0019	0.02
2	130 (0.99)	Upper secondary education (0.9)	suboptimal (0.97)	0.0359	0.87
3	600 (0.86)	Lower secondary education (0.99)	very good (0.32)	0.0759	0.27
4	2900 (0.12)	Master's or equivalent level (0.22)	bad (1)	0.0797	0.07

Table 4: The antichain $A = \{A_1, \ldots, A_4\}$ that generates that upset $U = \uparrow A$ where the minimum of the test statistic is attained. In brackets the marginal upper quantiles that correspond to the values are given, e.g. the 1 behind the 100 in the first row of the first column means that ca. 100% of the persons in the population earn at least 100 Euro. The column *difference* displays for every row i the difference between the proportion of male and the proportion of female persons that are above element A_i. The column *above* shows the proportion of all persons that are above A_i.

(26) is too loose. For a value of the test statistic of about 36% one would have to have chosen a V.C.-dimension of about 8 to make the conservative V.C.-inequality leading to a significant result. Since we were able to compute a large enough resample, we actually do not need to rely on the V.C.-inequality. However, for the purpose of illustration, we can tame the closure system of upsets to get an insight into how this affects the behavior of the test statistic for the actually observed data and the distribution of the test statistic under H_0. Figure 4 shows a the value of the test statistic D^+ for the actually observed data, as well as the distribution[39] of D^+ under H_0 for different V.C.-dimensions ranging from 4 to 39. Note that the original V.C.-dimension was 33, which is maybe surprising, but the V.C.-dimension of 39 for the biggest tamed closure system is due to the fact that by

[39]Here, we computed a resample of size 1000 to get a rough insight into the distribution of D^+ under H_0.

taming the closure system one gets in a first step only a family of sets that is generally no closure system and one has to enlarge this family in a second step to be a closure system to make the analysis computationally feasible. One can see that, as expected, with increasing V.C.-dimension, both the value of the test statistic for the actually observed data, as well as the expectation of the test statistic under H_0 increases. The standard deviation of the test statistic has also an increasing trend for increasing V.C.-dimensions. If one standardizes the test statistic D^+ by subtracting its mean and dividing the result by its standard deviation, one sees that the shape of the distribution of D^+ is approximately independent of the V.C.-dimension. The fact that the shape of the test statistic is approximately independent of the V.C.-dimension could possibly be used to get rules of thumb for situations where the computation of large resamples is computationally intractable. However, the approximate independence of the shape of the distribution of D^+ from the V.C.-dimension may be only present in our special situation and thus may be misleading for getting a rule of thumb for the general case.

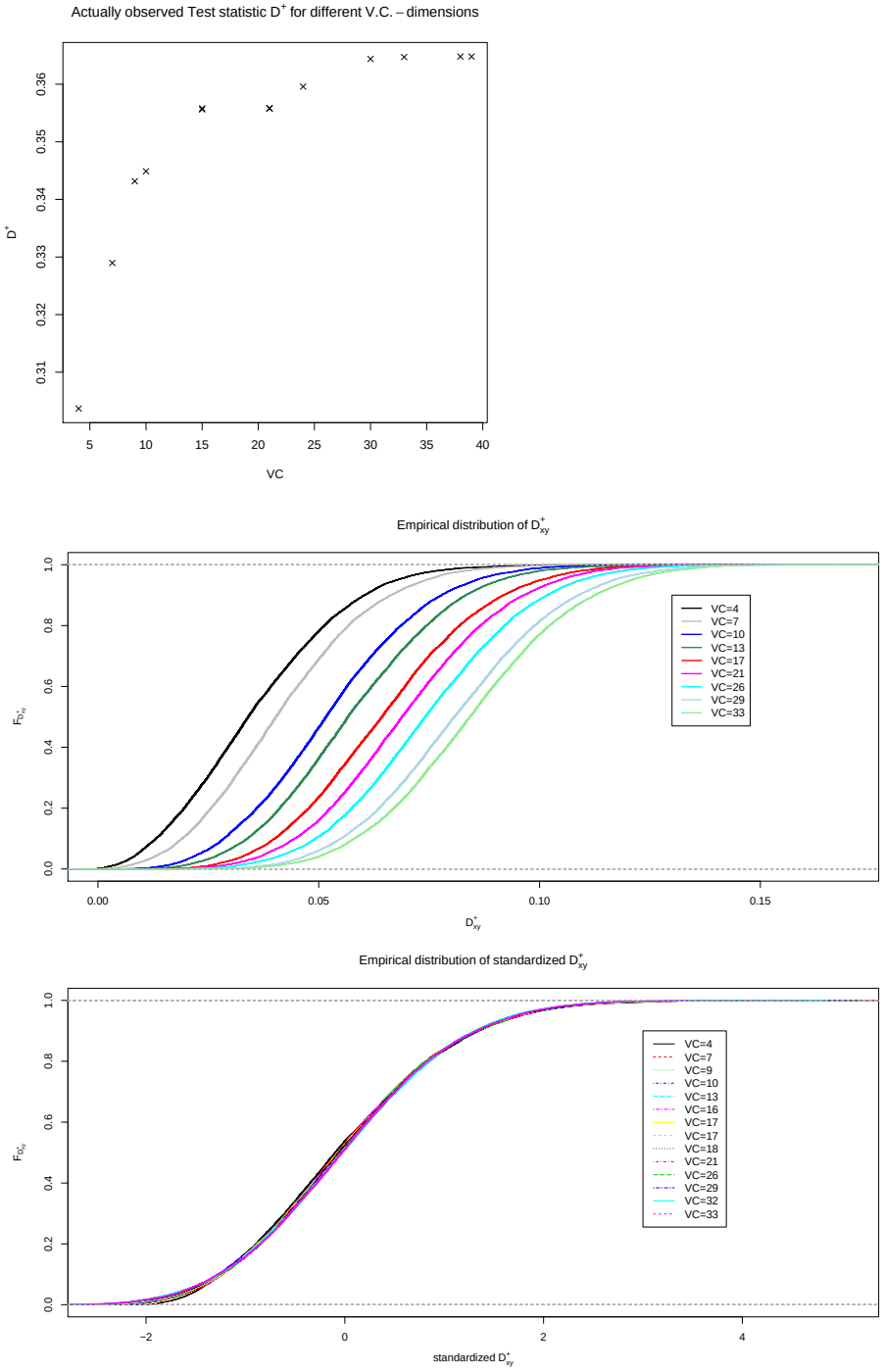

Figure 4: The value of the test statistic D^+ for the actually observed data, as well as the distributions of the test statistic D^+ and the standardized test statistic $\frac{D^+ - \overline{D^+}}{sd(D^+)}$ under H_0 for different V.C.-dimensions. One can see that the shape of the distribution D^+ is nearly independent of the V.C-dimension.

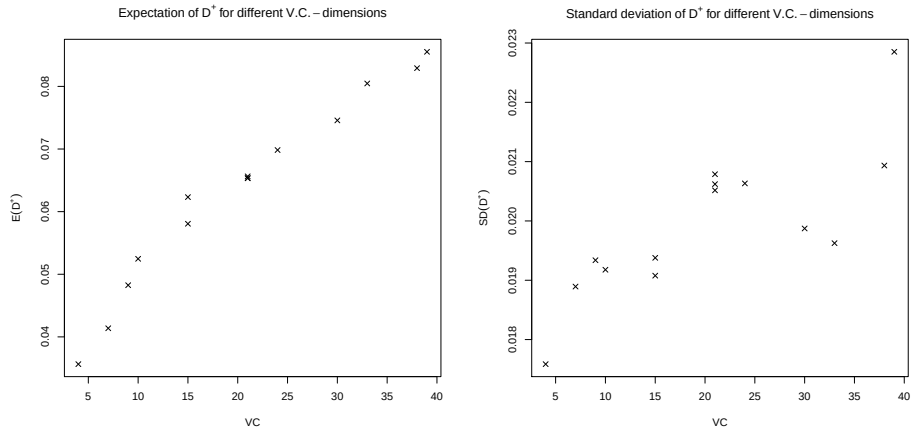

Figure 5: The expectation and the standard deviation of D^+ under H_0 for different V.C.-dimensions.

Now, we would like to illustrate a little bit, how the taming behaves w.r.t. conceptual terms. Therefore, we analyze for a tamed closure system of V.C.-dimension 7 the tamed upsets and downsets, where the maximum D^+ and the minimum D^- is attained. For the upset-approach, the maximal statistic for the tamed closure system is attained at an antichain of size 7 summarized in Table 5.

	Income (Euro)	Education (ISCED 2011)	Health (self-reported)	difference	above
1	1020 (0.65)	Short-cycle tertiary education (0.37)	excellent (0.08)	0.01	0.02
2	1050 (0.65)	Short-cycle tertiary education (0.37)	very good (0.32)	0.04	0.12
3	1050 (0.65)	Master's or equivalent level (0.22)	good (0.68)	0.06	0.14
4	1063 (0.65)	Short-cycle tertiary education (0.37)	good (0.68)	0.11	0.23
5	1200 (0.6)	Upper secondary education (0.9)	good (0.68)	0.23	0.42
6	1248 (0.56)	Upper secondary education (0.9)	satisfactory (0.89)	0.3	0.5
7	1300 (0.55)	Upper secondary education (0.9)	suboptimal (0.97)	0.32	0.52

Table 5: The antichain $A = \{A_1, ..., A_7\}$ that generates that upset $U =\uparrow A$ where the maximum of the test statistic is attained for the tamed closure system with a V.C. dimension of 7.

One can see that the maximal difference between the transformed Z-values in brackets is $0.65 - 0.08 = 0.57$ attained for the first element A_1, where the Z-value of 0.65 for an income of 1020 Euro is the largest, and a Z-value of 0.08 for a health status *excellent* is the smallest value. Compared to this, in the non-tamed case, the skewest element of the antichain generating the upset where the maximal value of the test statistic is attained is the element A_2 with a maximal Z-value of 0.99 for an education status *Lower secondary education* and a minimal Z-value of 0.08 for a health status *excellent*. The difference $0.99 - 0.08 = 0.91$ is clearly greater than for the tamed situation showing that we actually managed to reduce the skewness of elements generating the closure system for the tamed analysis.

The value of the tamed test statistic is with 32.90% not much smaller than the initial value of 36.48%, still significantly different from zero. The minimal value is -0.045% attained at the antichain consisting of only one element depicted in Table 6

	Income (Euro)	Education (ISCED 2011)	Health (self-reported)	difference	above
1	3500 (0.08)	Master's or equivalent level (0.22)	excellent (0.08)	-5e-04	0.003

Table 6: The antichain $A = \{A_1\}$ that generates that upset $U =\uparrow A$ where the minimum of the test statistic is attained for the tamed closure system with a V.C. dimension of 7.

The minimal value is not significantly different from zero. Table 7 and Table 8 finally show the results one would obtain if one would do the tamed analysis by looking at downsets instead of upsets:

Obviously, the role of the maximal and the minimal value of the test statistic will interchange: The maximal value of the test statistic of 0.15% means that the difference between the proportion of the *poor* male and the *poor* female persons is maximally 0.15% attained if one concretizes the term *poor* with the downset $D =\downarrow A$ generated by the antichain given in Table 7. The minimal value of the test statistic is -28.82% attained for the downset generated by the antichain given in Table 8. Note that for the downset-analysis we used for the construction of the Z-values not the complementary distribution function, but the usual distribution function, because this fits better to the notion of a downset. This only has an impact on the interpretation of the numbers given in brackets. For example the 0.06 beyond the income value of 360 Euro in Table 7 means now, that 6% of the population have income *below* 360 Euro. Additionally, the last column, denoted *below*, now gives the proportion of persons *below* the corresponding element of the antichain and the column *difference* gives the difference of the proportions below the corresponding element.

	Income (Euro)	Education (ISCED 2011)	Health (self-reported)	difference	below
1	360 (0.06)	Primary education (0.01)	bad (0.03)	0.0015	0.0008

Table 7: The antichain $A = \{A_1\}$ that generates that downset $D =\downarrow A$ where the maximum 0.0015 of the test statistic is attained for the tamed closure system with a V.C. dimension of 7.

6.2 Concept extents: Gender differences and differential item functioning in an item response dataset

In this section, we shortly analyze an IRT-dataset w.r.t. gender differences and **Differential item functioning** (DIF, [Osterlind and Everson, 2009]). The data set is a subsample from the general knowledge quiz *Studentenpisa* conducted online by the German weekly

	Income (Euro)	Education (ISCED 2011)	Health (self-reported)	difference	below
1	1400 (0.51)	Bachelor's or equivalent level (0.78)	good (0.68)	-0.21	0.33
2	1450 (0.52)	Short-cycle tertiary education (0.75)	very good (0.92)	-0.28	0.41
3	1460 (0.52)	Post secondary non-tertiary education (0.63)	good (0.68)	-0.20	0.3
4	1474 (0.53)	Upper secondary education (0.55)	good (0.68)	-0.16	0.28

Table 8: The antichain $A = \{A_1, ..., A_4\}$ that generates that downset $D =\downarrow A$ where the minimum -0.288 of the test statistic is attained for the tamed closure system with a V.C. dimension of 7.

news magazine SPIEGEL ([SPIEGEL Online, 2009], see also Trepte and Verbeet [2010] for a broad analysis and discussion of the original data set.) The data contain the answers of 1075 university students from Bavaria to 45 multiple choice items concerning the 5 different topics *politics*, *history*, *economy*, *culture* and *natural sciences*. For every topic, 9 questions were posed, for example question 1 of the politics topic was: "Who determines the rules of action in German politics according to the constitution?". The data set was analyzed in a number of papers, for example in Strobl et al. [2015], Tutz and Schauberger [2015], Tutz and Berger [2016], mostly from an IRT point of view. All mentioned papers identified systematic differences between the subgroups of male and female students in the sense of the presence of differential item functioning. Differential item functioning is present if the distribution of the item response patterns in two subgroups with identical latent abilities are different. Here, one cannot assume that the subgroups of male and female students that actually participated in the online quiz have the same latent abilities, because for example self selection processes can be present. To analyze the presence of differential item functioning one has to firstly somehow match persons of the two subgroups with similar abilities. One classical non-parametric procedure is the test of Mantel Haenszel (see [Holland et al., 1988].), where one takes the item scores (i.e., the number of solved items) as a matching criterion[40]. One stratifies the populations into parts with the same item score and then compares the subpopulations in every stratum. The final test statistic is then a χ^2-type statistic cumulating over all strata. The Mantel Haenszel procedure is an item-wise test, one tests for every item separately, if DIF is present for this item. For the construction of the matching score one usually does not take the whole set of items, instead one ignores items that showed DIF in a first preliminary analysis that was based on the whole set of items[41]. This process is called purification and there are different variants of purification, see, e.g., [Osterlind and Everson, 2009, p.16]. We can use the linear programming approach on formal contexts to develop a joint DIF test based on the item scores as a matching criterion. Firstly, we have to care for the different distributions of the abilities in the different subgroups. Here, we do not make a conditional analysis since conditioning would make all classes with the same item score relatively small such that a 45-dimensional multivariate

[40]Note that this will only work if the score values are a sufficient statistic for the abilities, which is for example the case for the Rasch model. For a discussion of deviations from this assumption in the context of the classical Mantel Haenszel procedure, see, e.g., [Zwick, 1990]

[41]The actually tested item should always be included for matching to make the Mantel Haenszel procedure valid under the null hypothesis of no differential item functioning, see [Holland et al., 1988, p.16].

analysis in every stratum would expectedly have very low power. Instead, we re-weight both subgroups such that the ability distributions in the male and the female group are approximately the same and then we analyze the joint distribution of item patterns and abilities (measured via the item scores). Concretely, we do the following:

1. Let $\mathbb{K}_0 = (G, M, I)$ be the formal context where $G = \{g_1, \ldots, g_{1075}\}$ is the set of persons, $M = \{m_1, \ldots, m_{45}\}$ is the set of items and gIm iff person g solved item m.

2. Separately for the male and the female group we estimate the density of the distribution of the item scores s, denoted with $\hat{f}_{male}$ and $\hat{f}_{female}$, respectively. The estimation is done here with a kernel density estimator.

3. Then we inversely re-weight the sample by giving a weight

$$W_i := \begin{cases} \hat{f}_{male}(s_i) & \text{if the } i\text{th person is female} \\ \hat{f}_{female}(s_i) & \text{if the } i\text{th person is male.} \end{cases}$$

   After this, the re-weighted distribution of the scores in the male and female group are approximately the same.

4. Then we analyze the joint distribution of response patterns and the score values in both subgroups. To do this, we use the flexibility of formal context analysis and simply conceptually scale the score values with an interordinal scale. Concretely, for every score value s we add an attribute "$\leq s$" and an attribute "$\geq s$" to the original context $\mathbb{K}_0$ with the interpretation person g has attribute "$\leq s$" if g has a score value lower than or equal to s and person g has attribute "$\geq s$" if g has a score value greater than or equal to s. Afterwards, we analyze the enlarged context $\mathbb{K}_1$ by looking at the closure system $\mathfrak{B}_1(\mathbb{K}_1)$ and computing

$$\max/\min_{M \in \mathfrak{B}_1(\mathbb{K}_1)} \langle (w^x - w^y) \cdot W, \mathbb{1}_M \rangle,$$

   where w^x are the original weights for the male and w^y are the weights for the female persons. Concretely, in the sample there were 658 male and 417 female persons, thus

$$w_i^x = \begin{cases} \frac{1}{658} & \text{if the ith person is male} \\ 0 & \text{if the ith person is female} \end{cases}$$

   and

$$w_i^y = \begin{cases} 0 & \text{if the ith person is male} \\ \frac{1}{417} & \text{if the ith person is female} \end{cases}.$$

5. In a last step we apply a purification procedure by basing the item scores for matching only on items that are not in the concept intents for which the maximal and minimal test statistic was obtained in a first run. We repeat the purification procedure until not further items are excluded.

Before showing the actual results, we firstly compute the test statistics D^+ and D^- for the context $\mathbb{K}_0$ without re-weighting the data. The context has a V.C.-dimension of 22 and has about 8.900.000.000 formal concepts (this is an estimate based on random sampling of arbitrary item sets and checking if they are a concept intent) and is thus very hardly describable explicitly and we will use the binary program described in Section 4.2 to compute the test statistics [42]. The maximal value of the test statistic is 0.335 attained at a formal concept containing the questions

F6: "Who is this? - (Picture of Horst Seehofer.)"

F26: "Which internet company took over the media group Time Warner? - AOL."

This means that the difference in the proportions of male and female persons who answered at least questions $F6$ and $F26$ correctly is the greatest observed difference between proportions of male and female persons that answered at least all items of some set of items correctly. Concretely, 53.6% of the male and 20.1% of the female persons answered these both questions rightly. The minimal value of the test statistic is -0.169 attained at a formal concept containing the questions

F40: "What is also termed Trisomy 21? - Down syndrome."

F43: "Which kind of bird is this? - Blackbird."

Here, 59.6% of the male and 76.5% of the female persons got both questions right. Both differences are significant, for a resample of size 1000 the value $\max\{D^+, -D^-\}$ had a range from 0.05 to 0.14 and a standard deviation of 0.014.). Figure 6 gives a rough idea about how a (non-guided) taming of the closure system by removing big shatterable sets of objects from the context affects the distribution of the test statistic D^+ under H_0. (Note, that this is only a very small simulation where we resampled only 100 times for every value of the V.C.-dimension.) The initial context has a V.C.-dimension of 22. One can see, that by reducing the V.C.-dimension, the mean and the standard deviation of the test statistic does not change very much for small reductions of the V.C.-dimension. Only a very strong taming to a V.C.-dimension below 8 seems to have an effect in reducing the mean of D^+ under H_0.

Now we come to the actual DIF-analysis: Figure 7 shows the distribution of the item scores for the male and female persons. The distributions are very different and thus we have to correct for this difference by re-weighting the data. However, generally, every attempt to account for such a kind of difference should be taken with some grain of salt, because initially we would like to account for differences in the abilities, but the abilities are only latent traits that cannot be observed and thus have to be estimated, in our situation

[42] To get a rough idea of computational complexity: The MIP solver `Gurobi` (see [Gu et al., 2012]) needed ca. 100 seconds to compute the statistic using one core on a 2.60 Ghz CPU (Intel(R) Xenon(R) CPU E5-2650 v2 @2.60 Ghz, 64GB RAM).

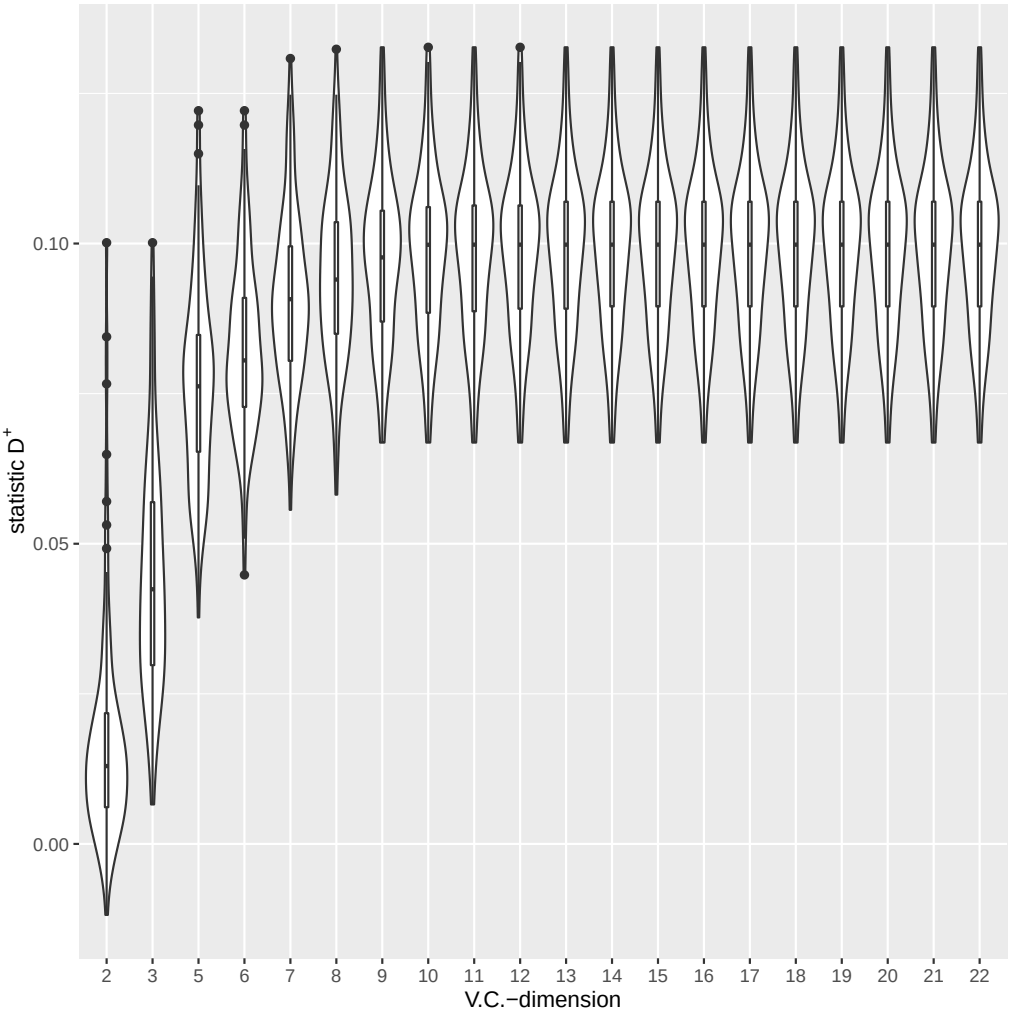

Figure 6: Distribution of the tamed statistic D^+ for different V.C.-dimensions.

from the item scores. In the unlucky case, an attempt for accounting for differences in abilities can make the analysis still more misleading if the items that suffer from DIF cannot be detected accurately enough and thus the item scores are invalidated as a surrogate for the abilities. The joint analysis of the re-weighted sample leads in the first step to a maximal value of the statistic of 0.234 attained at the intent of persons who answered the question F26: "Which internet company took over the media group Time Warner? - AOL." correctly and had a score value between 17 and 37. The minimum of the test statistic in the first step was -0.333 attained at an intent containing the 5 questions

F12: "Which form of government is associated with the French King Louis XIV? - Absolutism."

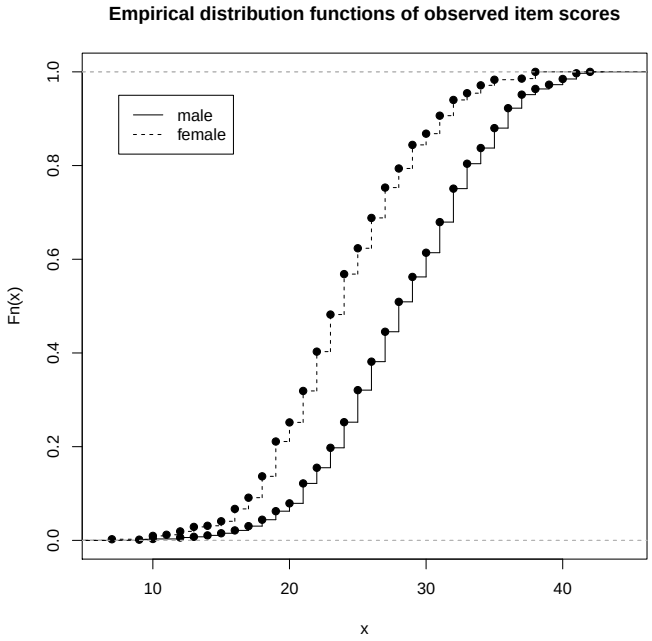

Figure 7: Empirical distribution of the item scores for the female and the male group in the subsample of the general knowledge quiz "Studentenpisa" ([SPIEGEL Online, 2009]).

F33: "What is the name of the bestselling novel by Daniel Kehlmann? - Die Vermessung der Welt (Measuring The World)."

F35: "In which city is this building located? - Paris."

F40: "What is also termed Trisomy 21? - Down syndrome."

F43: "Which kind of bird is this? - Blackbird."

and score values between 16 and 35. After excluding questions F26, F12, F33 F35, F40 and F43 for matching, in a second step, additionally the two questions

F34: "Which city is the setting for the novel 'Buddenbrooks'? - Lübeck." and

F36: "Which one of the following operas is not by Mozart? - Aida."

were excluded for matching. In the third step, the procedure stopped with a maximal final statistic D^+ of 0.241 attained for the intent containing $F26$ and (modified) score values between 16 and 29. The minimal value D^- was -0.290 attained for the intent containing questions $F12$, $F33$, $F35$, $F40$ and $F43$ and (modified) score values between 12 and 30. Thus, altogether, questions F12, F26 , F33, F34, F35, F36, F40 and F43 showed DIF. (The result was statistically significant in the sense that a bootstrap sample

of 10 samples yielded a distribution of the absolute value of the test statistic with mean 0.21 and standard deviation 0.01.)

6.3 Guided taming of concept extents in cognitive diagnosis models

In this section, we would like to illustrate a little bit, how one can tame a formal context in a more guided way in the context of cognitive diagnosis models. For illustration, we use a subsample of the *Trends in International Mathematics and Science Study* (TIMSS) of the year 2007. This study is an international assessment of the mathematics and science knowledge of students, that was firstly conducted in 1995 and has been administered every four years thereafter by the International Association for the Evaluation of Educational Achievement (IEA). It analyses math- and science knowledge of 4th and 8th grade students. We use here a subsample provided in the R-package `CDM`[43], consisting of 698 Austrian students (4th grade) answering a set of 25 math questions (dataset `data.timss07.G4.lee`). Since not all students answered all 25 questions, we restrict here the analysis to that 344 students that answered all questions. The 25 questions were the same as that used in Lee et al. [2011]. The package also provides the Q-matrix and the description of the skills used in Lee et al. [2011]. We will use this small subsample to illustrate the guided taming procedure by comparing it to the non-guided taming procedure described in section 5.3.2. The formal context $\mathbb{K}_0 = (\{g_1, \ldots, g_{344}\}, \{m_1, \ldots, m_{25}\}, I)$ has a Vapnik-Chervonenkis dimension of 14 and consists of 255712 formal concepts. Because of the small cardinality of the concept lattice, we can explicitly compute the closure system $\mathfrak{B}_1(\mathbb{K}_0)$ of all extents and thus we will analyze the taming process not w.r.t. the V.C.-dimension, but w.r.t. the cardinalities of the tamed closure systems $\mathfrak{B}_1(\tilde{\mathbb{K}})$. The data set contains also information about gender, so we will analyze differences w.r.t. gender. Figure 8 shows the value of the test statistic for the actually observed data in dependence on the cardinality of the tamed closure system for both the guided taming and the non-guided taming. One can see that, as expected, the statistic increases with increasing cardinality of the closure system. The general pictures for the non-guided and the guided taming are very similar. For the guided taming, the smallest closure system that is obtained by enforcing all valid implications of the idealized response pattern space, has a size of 127, which is much higher than the smallest possible closure system of size 2, obtainable by the strongest possible non-guided taming. Figure 9 shows the p-value one would obtain if one would do a statistical test. (Here, we did resampling with 1000 resamples to compute the p-values.) One can see, that for comparable sizes of the closure system, the guided taming procedure generally has lower p-values. One could speculate here, that the guided taming tends to exclude mainly sets that are statistically not so important in the sense that they play no crucial role w.r.t. differences between male and female participants. If one assumes that in the actually observed data set there are clear differences between male and female participants,

[43] See Robitzsch et al. [2016] for an introduction to the package `CDM`.

then it seemingly appears here, that the guided taming leads to a smart reduction of the size of the closure system that actually reduces the variability of the statistic under H_0 without reducing the test statistic under H_1, too much.

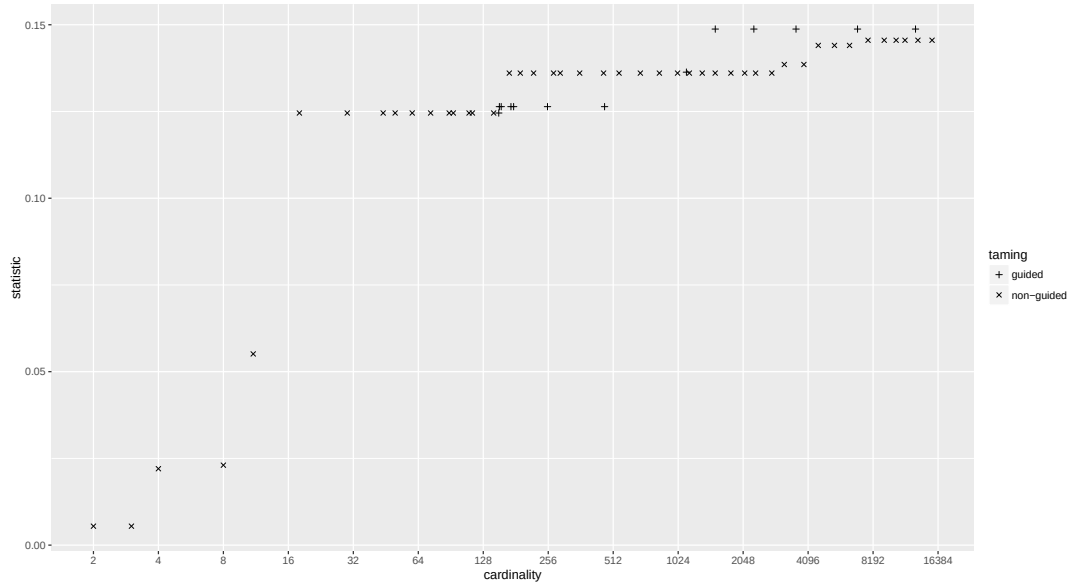

Figure 8: Value of the tamed test statistic for different cardinalities of the tamed closure system, both for the non-guided and the guided taming.

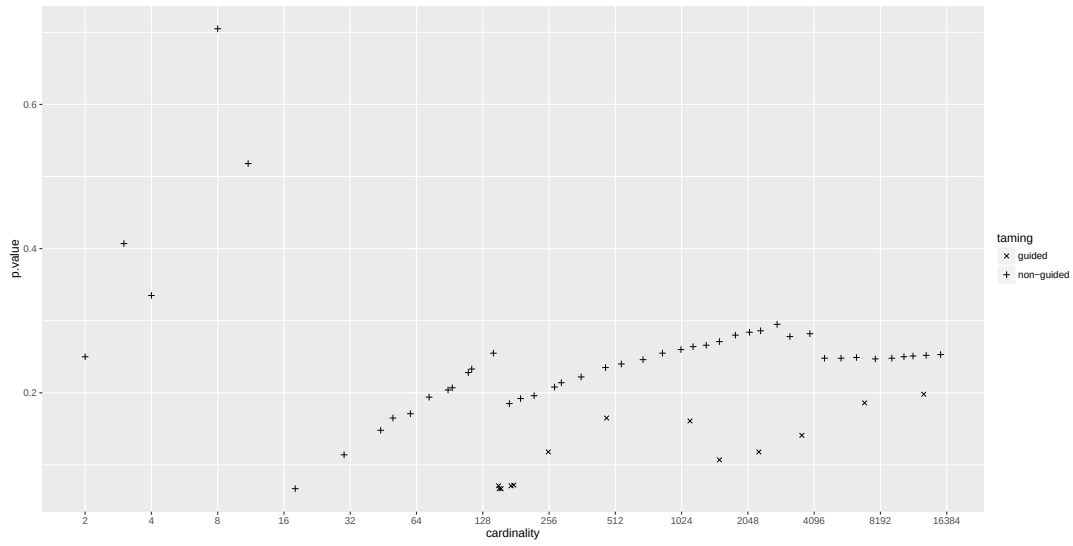

Figure 9: Obtained p-values of the test statistic for the actually observed data for different sizes of the tamed closure system, both for the non-guided and the guided procedure.

6.4 Convex sets: A geometrical generalization of the Kolmogorov-Smirnov test

Finally, we want to illustrate that for small data sizes the generalization of the Kolmogorov-Smirnov test for analyzing spatial differences between subpopulations in spatial statistics as indicated in Section 4 is also practically applicable. We use here the data set `quercusvm` which is a subsample of a larger data set analyzed in Laskurain [2008] and available in the R Package *ecespa* ([de la Cruz Rot, 2008]). This data set consists of 100 data points representing the locations of alive and dead oak trees (Quercus robur) in a secondary wood in Urkiola Natural Park (Basque country, north of Spain). The data are depicted in Figure 10. We can now compute that convex sets where the maximal and the minimal

Figure 10: Locations of altogether 100 alive and dead oak tress (Quercus robur) in a secondary wood in Urkiola Natural Park (Basque country, north of Spain).

differences in proportion of alive and dead oak trees is attained. Figure 11 shows the

results. The blue convex set is the set, where the difference is maximal (37%): In the

Figure 11: Difference in proportions of alive and dead oak trees. Blue: maximal difference of 39% (more alive than dead trees). Red: minimal difference of −37% (more dead than alive trees).

blue convex area we have a proportion of 61% alive, but only a proportion of 24% dead trees. The red convex set is the set, where the difference is minimal (−39%): In the red convex area there are 15% alive and 54% dead trees. Based on 1000 resamples, one gets an approximate p-value of 0.83, so the differences or not statistically significant. Note that also the Cramér von Mises type test proposed by Syrjala [1996] and also a classical generalization of the Kolmogorov-Smirnov test, where one only looks at rectangular areas are both non-significant. (The p-values of the tests, computed with the function `syrjala` from the R Package `ecespa` are approximately 0.84 and 0.67, respectively.) Compared to Syrjalas test, the test based on convex sets has the advantage that it is somehow better

interpretable because one can actually see, in which areas the differences in proportion are maximal or minimal.

A further modification of the test is also possible: Since the convex sets are described by formal implications and one explicitly models these implications by imposing corresponding inequality constraints in the binary program, one has the flexibility to impose not all, but only some implications. One natural way to select implications to include would be to include only implications where the data points of the premise and the conclusion are not too far away from each other. This would lead to some kind of a localization method and the associated closure system would get larger, which means more flexibility in detecting non-convex distributional features but a generally higher V.C.-dimension. We shortly illustrate this modification by imposing only implications where the distances between the points of the premises and the conclusions is not greater than $40m$. Figure 12 shows the non-convex sets where the maximal (blue) and the minimal (red) differences in the proportions of alive and dead oak trees is attained.

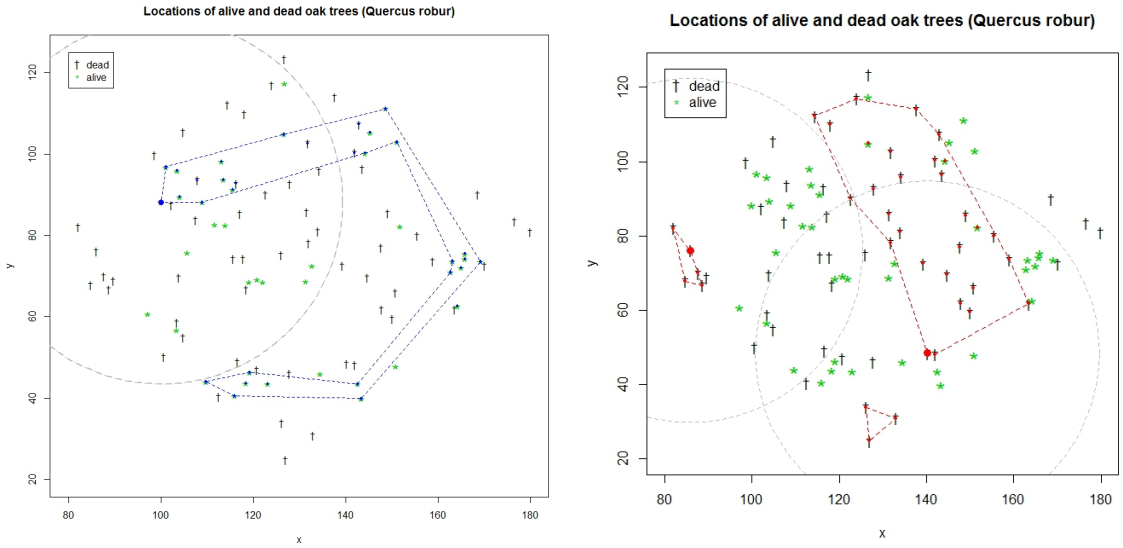

Figure 12: Non-convex sets where the maximal (blue) and the minimal (red) difference in proportions between alive and dead oak trees is attained for the modified method where only implications, where the distance between the points of the premises and the conclusions in not greater than $40m$, are used. The dashed circles around the highlighted blue and red points have a radius of $40m$ to get an impression about which implications were actually included.

For the modified version we got a maximal value of 58% and a minimal value of 50% for the difference of the proportions with an approximate p-value of 0.17.

While the computations for this data set could be done quickly enough, for larger data sets, the method quickly becomes intractable. The data set analyzed in Syrjala [1996] containing 327 spatial measuring points was already very hard to analyze, to compute the test statistic, the mixed integer solver Gurobi (Gu, Rothberg, and Bixby [2012]) took a few days to solve the binary program. Similarly to the analysis in Syrjala [1996], the result w.r.t. differences between male and female cods was not statistically significant. To assess the statistical significance of our test statistic, we did not need to do resampling, which would actually be very time demanding. Instead we could rely on the fact that for a value of 0.06 that we observed for our test statistic, still the more classical Kolmogorov-Smirnov type test statistic that only looks at rectangular areas would not be statistically significant.

However, for dealing with the computational issue, one can use the technique of attribute exploration for formal contexts: One can firstly look at the formal context $\mathbb{K} := (G, M, I)$ where G is the set of all rectangular areas, M is the set of all spatial measuring points and gIm iff measuring point m lies in the rectangular area g. The resulting closure system of all concept intents is then the set of all sets of measuring points lying in some rectangular area, which is a smaller closure system than the system of all convex sets of measuring points and in which thus more formal implications are valid. Note that despite this, a base of all implications of this smaller closure system can be given as

$$\{\{p, q\} \mapsto [p, q] \mid p, q \in M, [p, q] \supsetneq \{p, q\}\},$$

where $[p, q] := \{r \in M \mid r_1 \in [\min\{p_1, q_1\}, \max\{p_1, q_1\}] \;\&\; r_2 \in [\min\{p_2, q_2\}, \max\{p_2, q_2\}]\}$. Compared to the base for general convex sets, this base has only $\mathcal{O}(n^2)$ implications and is thus far more easy to handle.

Now, during the computation of all valid implications of $\mathbb{K}$, in the spirit of attribute exploration, one can check for every currently generated implication, if it is also approximately true with some confidence c in the context $\tilde{\mathbb{K}} = (\tilde{G}, M, \tilde{I})$, where $\tilde{G}$ is the set of all half-spaces generated by two points of M and $g\tilde{I}m$ means that measuring point m lies in the half-space g. The intents of this context are exactly all convex areas of measuring points. If the currently generated implication is also true in the larger closure system of all convex areas, then one would treat it as valid, otherwise one would provide a convex half-space $g \in \tilde{G}$ as a counterexample.

With this procedure one would generate a closure system that is larger than the system of all rectangular areas and smaller than the closure system of all convex areas and the confidence level c regulates the size of the resulting closure system and the size of the implication base.

Thus, with this modification, we have some "scalable" method for spatial statistics. (Of course, with the drawback that now the result of the method is dependent on the choice of the coordinate system.)

7 Conclusion

In this paper we analyzed the problem of detecting stochastic dominance as a prototypical example of optimizing a linear function on a closure system. Compared to the general case, for stochastic dominance, the integrality constraints of the underlying binary program could be dropped which helped in making the problem more tractable. For general closure systems the binary programs are more difficult to solve, but we managed to solve them in our concrete cases of application. Note that we did not explicitly incorporate knowledge about the underlying closure system into the mixed integer solver we used. It seems that one can make the computations far more efficient by using for example knowledge about valid formal implications of the underlying closure system. If one knows that certain formal implications are valid, then one can possibly use this knowledge to explicitly prune the search space in the branch and cut algorithm of the mixed integer solver.

The solved binary programs and the associated test statistics treated in this paper could be understood as some Kolmogorov-Smirnov type generalizations. This motivates the question if also other generalizations like weighted Kolmogorov-Smirnov type or Anderson-Darling type tests are computational tractable. Actually, it seems to be not too difficult to compute such variants of a test statistic: Firstly, one can impose one additional constraint into the underlying program that demands that the sets one is optimizing over contain at least (or at most, or exactly) an amount c of overall probability mass. Secondly, one can do the constrained optimization for every possible amount c and can then aggregate the optimal values for different c for example to

$$\sup_{c} \sup_{\substack{m: \\ \langle w^x + w^y, m\rangle \geq c}} \langle w^x - w^y, m\rangle \cdot \psi(c),$$

where ψ is some appropriately chosen weighting function.

All in all, it seems that the optimization of linear functions on closure systems has a broad range of possible applications and thus deserves further research.

References

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, J. Roche, and M. Santos. *Multidimensional Poverty Measurement and Analysis*. Oxford University Press, 2015.

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.

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, M. A. Hussain, V. Salvucci, F. Tarp, and L. P. Østerdal. Advancing small area estimation. Technical report, WIDER Working Paper, 2013. URL http://hdl.handle.net/10419/80908. accessed 28.08.2017.

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

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.

G. F. Barrett and S. G. Donald. Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104, 2003.

Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, and L. Lakhal. Mining minimal non-redundant association rules using frequent closed itemsets. In J. Lloyd, V. Dahl, U. Fuhrbach, M. Kerber, K.-K. Lau, C. Palamidessi, and L. M. Pereira, editors, *Computational Logic-CL 2000*, pages 972–986. Springer, 2000.

G. Birkhoff. Rings of sets. *Duke Mathematical Journal*, 3(3):443–454, 1937. doi: 10.1215/S0012-7094-37-00334-X.

L. Bottou. In hindsight: Doklady akademii nauk SSSR, 181(4), 1968. In B. Schölkopf, Z. Luo, and V. Vovk, editors, *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pages 3–5. Springer, 2013.

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

B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2002.

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.

M. de la Cruz Rot. Metodos para analizar datos puntuales. In F. T. Maestre, A. Escudero, and A. Bonet, editors, *Introduccion al Analisis Espacial de Datos en Ecologia y Ciencias Ambientales: Metodos y Aplicaciones.*, chapter 3, pages 76–127. Asociacion Espanola de Ecologia Terrestre, Universidad Rey Juan Carlos and Caja de Ahorros del Mediterraneo, 2008. URL `https://cran.r-project.org/web/packages/ecespa/index.html`.

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

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

B. Ganter and R. Wille. *Formal concept analysis: mathematical foundations*. Springer, 2012.

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.

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

E. H. Haertel. Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4):301–321, 1989.

G. Hansel and J. Troallic. Mesures marginales et théorème de Ford-Fulkerson. *Probability Theory and Related Fields*, 43(3):245–251, 1978.

J. Heller, L. Stefanutti, P. Anselmi, and E. Robusto. On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 80(4):995–1019, Dec 2015.

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

J. E. Hopcroft and R. M. Karp. A n5/2 algorithm for maximum matchings in bipartite. In *Switching and Automata Theory, 1971., 12th Annual Symposium on*, pages 122–125. IEEE, 1971.

B. W. Junker and K. Sijtsma. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272, 2001.

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

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

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 (Bizkaia)*. PhD thesis, Universidad del País Vasco/Euskal Herriko Unibertsitatea, 2008.

Y.-S. Lee, Y. S. Park, and D. Taylan. A cognitive diagnostic modeling of attribute mastery in massachusetts, minnesota, and the us national sample using the timss 2007. *International Journal of Testing*, 11(2):144–177, 2011.

E. Lehmann. Ordered families of distributions. *Ann Math Stat*, 26:399–419, 1955.

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: 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.

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.

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.

A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley, 2002.

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

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

J. W. Pratt and J. D. Gibbons. *Concepts of Nonparametric Theory*. Springer, 2012.

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

C. Preston. A generalization of the FKG inequalities. *Communications in Mathematical Physics*, 36(3):233–241, 1974.

T. M. Range and L. P. Østerdal. Checking bivariate first order dominance. Technical report, Discussion Papers on Business and Economics, 2013.

A. Robitzsch, T. Kiefer, A. C. George, and A. Uenlue. *CDM: Cognitive Diagnosis Modeling*, 2016. URL `http://CRAN.R-project.org/package=CDM`. R package version 4.8-0.

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.

N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972.

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

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

S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, 1972.

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.

V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

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.

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

F. Tarp and L. P. Østerdal. Multivariate discrete first order stochastic dominance. Discussion Papers 07-23, University of Copenhagen. Department of Economics, 2007. URL `http://EconPapers.repec.org/RePEc:kud:kuiedp:0723`.

K. K. Tatsuoka. *Analysis of errors in fraction addition and subtraction problems*. National Institute of Education, Washington, D.C., 1984.

S. Trepte and M. Verbeet. *Allgemeinbildung in Deutschland: Erkenntnisse aus dem SPIEGEL-Studentenpisa-Test*. VS Verlag für Sozialwissenschaften, 2010.

W. T. Trotter. *Combinatorics and Partially Ordered Sets: Dimension Theory*, volume 6. JHU Press, 2001.

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

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.

V. N. Vapnik and S. Kotz. *Estimation of Dependences Based on Empirical Data*. Springer, 1982.

N. Vayatis and R. Azencott. Distribution-dependent vapnik-chervonenkis bounds. In P. Fischer and H. U. Simon, editors, *European Conference on Computational Learning Theory*, pages 230–240. Springer, 1999.

R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In *Ordered Sets*, pages 445–470. Springer, 1982.

R. Zwick. When do item response function and Mantel-Haenszel definitions of differential item functioning coincide? *Journal of Educational Statistics*, 15(3):185–197, 1990.

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.

München, 27.04.2018
Ort, Datum

(Christoph Jansen)