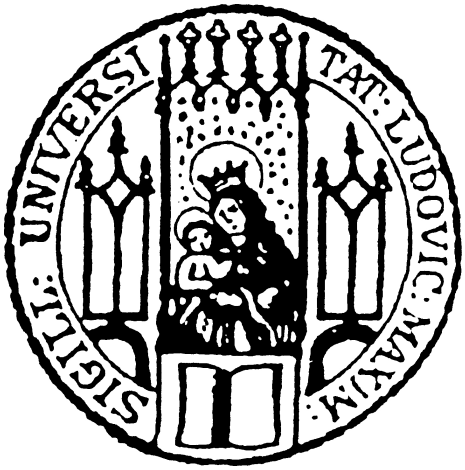
Ludwig-Maximilians-Universität München
Fakultät für Mathematik, Informatik und Statistik
Institut für Statistik



Decision making under partial information using precise and imprecise probabilistic models

Author: Christoph Jansen
Supervisor: Prof. Dr. Thomas Augustin
Date: February 17, 2015

Decision making under partial information using precise and imprecise probabilistic models

Master's Thesis

Ludwig-Maximilians-Universität München
Fakultät für Mathematik, Informatik und Statistik
Institut für Statistik

Author:
Christoph Jansen

Supervisor:
Prof. Dr. Thomas Augustin

February 17, 2015

Abstract

In this work we recall and discuss optimality criteria for decision making under uncertainty with respect to different assumptions concerning the structure of the information available. First, an overview of the basic concepts of classical decision theory is given. Here, particular emphasis is placed on explaining classical decision criteria (i.e. criteria for the case that the uncertainty can be characterized by using classical probabilistic models) from literature and discussing the assumptions underlying them (specifically Bernoulli-, Maximin and Hodges & Lehmann-criterion are discussed). Afterwards, in order to establish the mathematical basis necessary, a brief introduction to the theory of linear optimization is provided. Thereby, particular interest lies in recalling theoretical results concerning the resolvability of linear optimization problems. Subsequently, we demonstrate how linear optimization theory can be used to construct algorithms for determining optimal decisions (with respect to the classical criteria) in finite decision problems. Examples of concrete decision problems are given. Next, we give some (theoretical and practical) examples that support the idea of introducing imprecise probabilistic models to decision theory. Accordingly, two common generalizations of classical probability theory are explained: Credal sets and interval probability. Connections of the two concepts are shown up. Finally, we recall criteria for optimal decision making, if the uncertainty is characterized by a credal set or an interval probability respectively (specifically interval dominance, E-admissibility, maximality, Γ-Maximin-criterion, Γ-Maximax-criterion and a criterion combining the two latter). Again, for all criteria discussed, we illustrate and explain algorithms for determining optimal decisions.

Contents

0 Introduction

The present work deals with *decision making under partial information* and, therefore, contributes to the scientific field of *Decision Theory*. Accordingly, a straightforward way of introducing to this thesis is to briefly summarize (our standpoint on) the basic characteristics of Decision Theory. Particularly, this helps to clarify where this work is located within the broad field of this scientific discipline.

So, what is Decision Theory? A commonly accepted (and very general) definition is the following: Decision Theory is the science of *rational* decision making in situations under *uncertainty* (see e.g. [2, p.3] or [33, § 1.1]). However, due to its immense generality, this definition immediately gives rise to (at least) three (highly related) questions:

1. What do we mean by *rational* decision making?
2. How can we formalize a vague concept like *uncertainty* adequately?
3. Where is the connection between rational decision making and the (yet formalized) uncertainty underlying a situation?

Throughout the present work, the focus will mainly lie on dealing with question two and three posed above. In contrast, the first question will be *assumed* to be answered already (with the exception of the short discussion in Remark 3 on Definition 1).

Clearly, this seems to be a rather comfortable point of view: Defining and formalizing rationality is a highly *non-trivial* task and constitutes a fundamentally important part of many different scientific disciplines, such as *economics*, *(subjective) probability theory*, *sociology* and *philosophy*. So, how can we justify this incompleteness?

Like many others before, we do so by appealing to a well-established scientific principle (or excuse?): '*To divide the difficulties, i.e. to concentrate on one (the subject proper of the investigation in hand), and to reduce all others as far as reasonable possible, by simplifying and schematizing assumptions*' ([31, p. 16, l. 7-10]).

Nevertheless, let us (*very* briefly) clarify our sight on what we understand by *rationality* and *rational decision making* in the present work (following for example [6, Ch. 2.2] or [31, Ch. 1.3]):

Consider a decision situation under uncertainty and suppose the set of available decisions (or *actions*, see Definition 1) is given by $\mathbb{A}$. Furthermore, suppose the uncertainty underlying the situation is about which of the elements of a set Θ (the *set of states*, see Definition 1) corresponds to the true description of reality.

Clearly, every pair $(a, \theta) \in \mathbb{A} \times \Theta$ induces a *consequence* $c(a, \theta)$, namely the consequence of choosing decision a under the condition that θ corresponds to the true description of reality. Accordingly, the set of all possible consequences is given by

$$\mathcal{C} := \Big\{ c(a, \theta) : (a, \theta) \in \mathbb{A} \times \Theta \Big\} \tag{1}$$

Note that, in general, the set $\mathcal{C}$ won't consist of numbers. Instead, $\mathcal{C}$ might for example be a listing of statements describing the consequences of the pairs by words. Particularly, there does not exist a meaningful way of measuring distances between the elements of $\mathcal{C}$ in general.

Now, for any *actor* (here the term *actor* labels the person having to choose between the actions), we assume that he can order the elements of the set $\mathcal{C}$ (that is the possible consequences) by *individual preference* in a certain way (if this assumption is violated decision making is not possible at all, since there is complete indifference between the possible consequences). For example, an actor might prefer consequence $c(a_1, \theta_1)$ before consequence $c(a_2, \theta_2)$ (or vice versa). Note that, at this point, this ordering doesn't have to be meaningful at all.

Next, we define an *actor* to be *rational* if, and only if, his individual *ordering of preferences* on the set of consequences $\mathcal{C}$ is in accordance with the axioms of *von Neumann and Morgenstern* (see e.g. [31, § 3.6] or [33, p. 43]). This set of axioms of rationality for example ensures the preference order to be *transitive* and *total* (see Definition 5 for further details).

This is where the famous *Theorem of Morgenstern and von Neumann* (see [31, p. 617-632]) comes into game. It states the following: For any rational actor A there exists a mapping $u_A : \mathbb{A} \times \Theta \to \mathbb{R}$ such that for all $a_1, a_2 \in \mathbb{A}$ and $\theta_1, \theta_2 \in \Theta$ we have

$$A \text{ prefers } c(a_1, \theta_1) \text{ before } c(a_2, \theta_2) \quad \Leftrightarrow \quad u_A(a_1, \theta_1) \geqslant u_A(a_2, \theta_2) \tag{2}$$

Hence, the theorem guarantees the existence of a real-valued *utility function* perfectly specifying a (rational) actor's ordering of preferences. Note that the map u_A is not unique. Instead, any linear transformation of u_A characterizes the same preference order (see [31, p. 617-632]).

The Theorem of Mogenstern and von Neumann justifies the following approach: The problem of a rational actor A having to decide in a situation under uncertainty can be formalized by a triplet

$$(\mathbb{A}, \Theta, u_A(\cdot)) \tag{3}$$

where u_A denotes a utility function for actor A. Note that this is in accordance with Definition 1 and, therefore, justifies the approach chosen in the present work.

But when does a rational actor A make *rational decisions*? Clearly, he should try to achieve the highest possible value of the utility function u_A (remember that u_A perfectly characterizes his individual preferences between the possible consequences). This can be used as a definition: A rational actor makes *rational decisions* if, and only if, by his decisions, he always *tries* to achieve the best possible consequence and, therefore, *tries* to maximize his utility function.

This gives us definitions of rationality and rational decision making. However, remember that an actor's influence is restricted on choosing between the elements of the set $\mathbb{A}$. That is, in general, the actor doesn't know which element of Θ corresponds to the true description of reality and, therefore, is *uncertain* about the consequence of his decision.

For this reason, the above definition of a rational decision strongly depends on what is known about the set Θ. For example, consider the extreme case where Θ consists of one single element. Here, a rational decision rule is not too hard to make out: Choose the action that leads to the best consequence under the only possible description of reality. But what if there are concurring descriptions of reality? How can we use information about the set Θ best possible in order to determine rational decisions? How can we best decide if there is no information available at all?

This allows us to refine the very general definition from the beginning for our purposes: Decision Theory is the science of rational actors making decisions that use the available information in the best possible manner in order to receive the best expectable consequence (according to their individual preference ordering).

This is where our story begins.

In the present work, we recall and discuss criteria for rational (or *optimal*) decision making with respect to different assumptions on the structure of the available information. Furthermore, we recall (and extend) algorithms that allow to determine such optimal decisions computationally. The thesis is structured as follows:

- **Chapter 1:** In the first Chapter, the basic definitions and concepts from *Classical Decision Theory* are recalled. Here, the term 'classical' particularly relates to the assumptions made on the structure of the available information: Chapter 1 treats the case that the uncertainty between the different states of reality (that is the elements of the set Θ) is either adequately characterizable by a classical probability measure (see Paragraph 1.2.3) or can be compared to

a game against an omniscient and all mighty enemy (see Paragraph 1.2.4). For both cases, the standard criteria of optimality are recalled and discussed. Additionally, we recall an approach for decision making under mixed uncertainty types (see Paragraph 1.2.5).

- **Chapter 2:** Often, the task of determining optimal decisions can be reformulated as the task of optimizing a real-valued linear function under linear constraints (remember, any rational preference ordering is representable by a real-valued function). This task coincides with the basic problem setting of the well-investigated theory of linear optimization (see Definition 11). Therefore, Chapter 2 deals with some concepts and results from this mathematical discipline that are relevant for our purposes. Within the considerations in the Chapters 3 and 5, the theory of linear optimization will turn out to be a very help- and powerful tool for both computing optimal decisions and proving their existence.

- **Chapter 3:** After having given an overview on the required mathematical background in Chapter 2, the third Chapter explicitly treats how optimal decisions can be determined by solving suitable linear optimization problems. Specifically, suitable optimization problems will be recalled and explained for every optimality criterion discussed in the first Chapter. Additionally, concrete examples illustrating how optimal actions can be determined by the usage of standard statistical software (for example using R) will be given.

- **Chapter 4:** In many applications, the uncertainty underlying a decision situation turns out to be not characterizable within the framework of classical probability theory. The axioms of classical probability theory demand a degree of precision that often can't be justified by the available information. In such cases, more general descriptions of uncertainty seem to be more adequate. Therefore, the fourth Chapter is divided into two parts: In Paragraph 4.1, we give some examples from different scientific disciplines that support the idea of a generalized theory of uncertainty. Consequently, in Paragraph 4.2, we explain two common generalizations of the classical theory, namely the concepts of credal sets (see Paragraph 4.2.1, Definition 16) and interval probability (see Paragraph 4.2.2, Definitions 18 and 19).

- **Chapter 5:** In the fifth Chapter of the present work, we recall criteria for optimal decision making if the uncertainty is described by a convex credal set (see Paragraph 5.1) or an interval probability field respectively (see Paragraphs 5.2-

5.6). It turns out that, applying the generalized theory, the choice of a suitable criterion is strongly dependent of the actor's attitude towards risk. Therefore, many different concurring criteria exist. Additionally, we recall (and extend) algorithms to determine optimal decisions with respect to these criteria using linear optimization theory. Again, we illustrate the computational treatment of these algorithms by many examples. Finally, we list a couple of results on the connection of the different criteria.

Throughout the thesis, some basic knowledge in the mathematical disciplines *Probability Theory*, *Calculus*, *Measure Theory* and *Linear Algebra* is required. Here, we refer to the standard textbooks [26], [16], [18] and [11] respectively. However, advanced application of the corresponding discipline are recalled at the appropriate places in the text.

For the sake of readability, the following symbols are used to indicate the end of a specific passage throughout the whole thesis:

$\square$: Indicates the end of a proof.

$\bigtriangledown$: Indicates the end of a definition.

$\star$: Indicates the end of an example.

$\circ$: Indicates the end of a remark.

$\bullet$: Indicates the end of an excursus.

Furthermore, the following letters have a fixed meaning:

n: Number of actions in a finite decision problem.

m: Number of states of nature in a finite decision problem.

k: Number of possible observations when considering finite data-based extensions (that is $k := |\mathcal{X}|$, where $\mathcal{X}$ denotes the space of observations).

π: Probability measure on the set of states Θ of a decision problem (together with some suitable σ-field).

p: Randomized action for some basic decision problem.

1 Fundamental principles of Classical Decision Theory

In the first chapter of the present work a (not too) short compilation of the basic definitions of *Classical Decision Theory* is given. In the first paragraph the classical decision problem and two of its extensions are defined. Subsequently, we recall and discuss some common criteria for the optimality of decisions in a given decision problem. Finally, we recall the *Fundamental Theorem of Baysian Decision Theory* that makes up a connection between classical decision theory and *Baysin Statistics*.

The chosen presentation of Decision Theory doesn't claim completeness, but focusses on concepts needed here. Particularly, many results that could be shown in a greater generality are only shown for the *finite* case. We mainly follow [40], [2], [3], [6], [33], [35], [13], [29] and [41]. More precise references are given at the appropriate places in the text.

1.1 The classical decision problem and its extensions

Informally, the classical decision problem in a *situation under uncertainty* can be described as follows: An actor has to choose an *action* from an (often finite) set of alternatives. However, the *utility* of the chosen action depends on the true *state of nature*, which, in general, is unknown to the actor. How should he decide?

The following definition, which is used e.g. in [41, § 2], formalizes this situation.

Definition 1. A *classical (no-data) decision problem (CDP)* in utility form is a triplet

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

consisting of

- an arbitrary non-empty set $\mathbb{A}$,
- an arbitrary non-empty set Θ and
- a map $u : (\mathbb{A} \times \Theta) \to \mathbb{R}\ ,\ (a, \theta) \mapsto u(a, \theta)$.

The sets $\mathbb{A}$ and Θ are called *set of actions* and *set of states*, their elements are referred to as *actions* and *states of nature*. The map u is called *utility function*. The decision problem $\mathfrak{A}$ is *finite*, if $|\mathbb{A}| < \infty$ and $|\Theta| < \infty$. ▽

Remark. 1.) By analogy, the CDP could be defined in *loss form*. If this is the case, the map contained in the triplet is no longer denoted by u (like *utility*), but by l (like *loss*). This alternative way of defining a CDP is especially used in (the literature

on) *Statistical Decision Theory* (see e.g. [6]), where l could for example measure the *expected quadratic loss* of an estimator (see [12] for an exact definition).

From a mathematical point of view, there is no difference in the two definitions, since there are no formal restrictions concerning the utility/loss function. Semantically, it is of course essential to know whether the values should be interpreted as loss or as utility.

2.) If $\mathfrak{A}$ is finite, now and and from now on *always*

$$\mathbb{A} := \{a_1, \ldots, a_n\} \quad \text{and} \quad \Theta := \{\theta_1, \ldots, \theta_m\}$$

the following compact presentation of a CDP can be used:

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

Here, the entry (i, j) of the table equals the utility of the pair (a_i, θ_j). From now on this presentation is used without further commentary . The letters n and m always denote the cardinalities of the sets $\mathbb{A}$ and Θ respectively.

3.) The above definition implicitly makes a couple of idealizing assumptions. For example, it is assumed that both, the set of possible states and the set of possible actions, are completely known to the actor (this is sometimes referred to as *closed-world-assumption*, see for example [2, p. 17]).

Furthermore, it is assumed that the utility of each pair of action and state (a, θ) can be perfectly specified by one real number $u(a, \theta)$. More precisely, we assume that any *preference ordering* on the set of *consequences* (see introduction) can be adequately characterized by a real-valued function. This assumption is justified by the famous and well-known *Theorem of von Neumann and Morgenstern* (see e.g. in [31, p.617-632]). Here, the authors prove that any preference order satisfying a set of four (not too far taken) rationality axioms can be represented by such a function (a more detailed discussion of this was given in the introduction).

Finally, the term utility shouldn't be interpreted in a too restrictive way (for example only *monetary*: A classical example for purely monetary utility functions contradicting rationality is given by the famous *Saint-Petersburg Paradox*, see for example [34,

§ 1.3]). Instead, this *abstract* utility is assumed to be a *perfect weighing* of all forms of utility relevant to the actor.

Of course, as for any other scientific model, these assumption should be discussed critically in each concrete decision problem. In the following we want to assume them to be true. A far more in-depth discussion of the topic can be found in [36, Ch.1] and/or in [31, Ch.1, Part 3]. ∘

For each CDP $\mathfrak{A}$ (satisfying certain conditions of measurability) one can define its *mixed extension* $G(\mathfrak{A})$. The idea is the following: Instead of having to choose an action from the set $\mathbb{A}$ (the actions contained in $\mathbb{A}$ are said to be *pure actions* in this context), it is now possible to the actor to choose a *randomized action*. Formally, each randomized action equals a probability measure $p : \sigma(\mathbb{A}) \to [0,1]$ on the measurable space $(\mathbb{A}, \sigma(\mathbb{A}))$, where $\sigma(\mathbb{A})$ is σ-field including all the singletons. This motivates the following definition, which is taken from [2, p.53] (with adapted notation).

Definition 2. Let $\mathfrak{A} := (\mathbb{A}, \Theta, u(\cdot))$ be a CDP and $\sigma(\mathbb{A})$ a σ-field on $\mathbb{A}$ such that $\{a\} \in \sigma(\mathbb{A})$ for all $a \in \mathbb{A}$. Define

$$G(\mathbb{A}, \sigma(\mathbb{A})) := \Big\{ p(\cdot) : p \text{ is a probability measure on } (\mathbb{A}, \sigma(\mathbb{A})) \Big\}$$

Let, for each fixed $\theta \in \Theta$, the map

$$u_\theta(\cdot) := u(\cdot, \theta) : \mathbb{A} \to \mathbb{R} \ ; \ a \mapsto u(a, \theta)$$

be p-integrable for all $p \in G(\mathbb{A})$. Define the map

$$G(u) : (G(\mathbb{A}, \sigma(\mathbb{A})) \times \Theta) \to \mathbb{R} \ , \ (p, \theta) \mapsto \mathbb{E}_p(u_\theta) := \int_{\mathbb{A}} u_\theta \, dp \tag{4}$$

Then the triplet

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

is called the *mixed extension* of the CDP $\mathfrak{A}$. The set $G(\mathbb{A}, \sigma(\mathbb{A}))$ is referred to as *set of randomized actions*. If no confusions can appear one writes $G(\mathbb{A})$ instead of $G(\mathbb{A}, \sigma(\mathbb{A}))$. This particularly is the case if $\mathfrak{A}$ is finite, because then always $\sigma(\mathbb{A}) := \mathcal{P}(\mathbb{A})$ is used. ▽

Remark. 1.) The mixed extension of a CDP is again a CDP. If the underlying CDP is finite, this is not the case for its mixed extension: Even on a finite set there exist uncountably many different probability measures. However, note that

the set of states of a CDP $\mathfrak{A}$ and the set of states of its mixed extension $G(\mathfrak{A})$ coincide by definition (Θ remains unchanged under the transition to the mixed extension). Particularly, this implies that the set of states of the mixed extension of a finite decision problem still admits only finitely many states of nature, that is $|\Theta| < \infty$. This will turn out to be very important, since in the following many results (theorems and algorithms) are shown for CDPs that satisfy this condition. Specifically, all theorems that only demand the condition $|\Theta| < \infty$ as a prerequisite can be applied to either a finite CDP or its mixed extension.

2.) Every pure action $a \in \mathbb{A}$ can uniquely be identified with the randomized action $\delta_a \in G(\mathbb{A})$, where

$$\delta_a : \sigma(\mathbb{A}) \to [0,1] \ , \ A \mapsto \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{else} \end{cases}$$

δ_a is called the *Dirac-measure* in the action a (see for example [26, p.12]). Then, for all $(a, \theta) \in (\mathbb{A} \times \Theta)$, the equation

$$u(a, \theta) = G(u)(\delta_a, \theta) \tag{5}$$

holds. Using this identification the set $\mathbb{A}$ can be understood as a subset of $G(\mathbb{A})$.

3.) If $\mathbb{A}$ is finite and $\sigma(\mathbb{A}) := \mathcal{P}(\mathbb{A})$, then all conditions of the above definition are satisfied and the equation

$$G(u)(p, \theta) = \sum_{i=1}^{n} u(a_i, \theta) \cdot p(\{a_i\}) \tag{6}$$

holds for all pairs $(p, \theta) \in (G(\mathbb{A}) \times \Theta)$. This clarifies the fact that the term $G(u)(p, \theta)$ equals the expectation of the randomized action p, if θ is the true state of nature.

4.) In the case of a finite CDP one can interpret the randomized action p as the following rule: Choose an action a_i with probability $p(\{a_i\})$, i.e. simulate a random number Z from the set $\{1, \ldots, n\}$ with probabilities $(p(\{a_1\}), \ldots, p(\{a_n\}))$ and choose the action a_z (where z is the realisation of the simulation).

5.) The transition from a CDP to its mixed extension may appear counter-intuitive at first sight: Why the risk of a bad choice shouldn't be avoided completely? As we will see subsequently, the question of the sense or non-sense of randomization strongly depends on the chosen criterion of optimality. The criterion itself depends on the *type* of uncertainty underlying the decision problem. It turns out that randomization can lead to a superior utility in situations where the process generating

the states of nature can be compared to a *game against nature*. A more in-depth discussion on the connection between randomization and the *type of uncertainty* underlying a CDP can be found in Paragraph 1.2.2. ◦

Next, we want to describe how additional information (in form of data) can be included in a CDP. Informally, the idea is the following: Instead of directly choosing one of the actions in $\mathbb{A}$, a *random experiment* is interposed. Which of the actions is chosen in the end then depends on the result of this experiment. Thus, each action now is a *decision function* from the set of all possible realisations of the experiment $\mathcal{X}$ to the set of actions $\mathbb{A}$. More precisely, the new set of actions equals the set of all possible decision functions.

In addition, a *parametric statistical model* on the space of observations $\mathcal{X}$ is assumed. The corresponding *parametric space* is given by Θ, i.e. every state of nature $\theta \in \Theta$ induces a probabilistic model p_θ for the random experiment described above. The utility of a decision function d, given the information that θ is the true state of nature, then equals the expectation of d under p_θ.

To formalize this, some preparation work has to be done: Let $\mathfrak{A}$ be a CDP and Ω and $\mathcal{X}$ be arbitrary non-empty sets. Note that the space Ω will get re-interpreted in the context of the considerations concerning the *Fundamental Theorem of Baysian Decision Theory* in Paragraph 1.3. Further, let

- $X : \Omega \to \mathcal{X}$ denote a $\sigma(\Omega)$-$\sigma(\mathcal{X})$-measurable map, where $\sigma(\Omega)$ and $\sigma(\mathcal{X})$ are σ-fields on Ω and $\mathcal{X}$, such that $\{x\} \in \sigma(\mathcal{X})$ for all $x \in \mathcal{X}$.
- $\mathcal{Q}(\Theta) := \{q_\theta(\cdot) : \theta \in \Theta\}$ denote a Θ-parametrized set of probability measures on the measurable space $(\Omega, \sigma(\Omega))$
- $\mathcal{L}(\Theta) := \{p_\theta(\cdot) := X[q_\theta](\cdot) : \theta \in \Theta\}$ denote the (again Θ-parametrized) set of image measures from $Q(\Theta)$ under X on $(\mathcal{X}, \sigma(\mathcal{X}))$
- $\sigma(\mathbb{A})$ denote a σ- field on $\mathbb{A}$ (where $\{a\} \in \sigma(\mathbb{A})$ for all $a \in \mathbb{A}$) and let, for every $\theta \in \Theta$ fixed, the map

$$u_\theta(\cdot) := u(\cdot, \theta) : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto u(a, \theta)$$

  $\sigma(\mathbb{A})$-$\mathcal{B}$-measurable (where $\mathcal{B}$ denotes the *Borel σ-field* on $\mathbb{R}$).

Then, every $\sigma(\mathcal{X})$-$\sigma(\mathbb{A})$-measurable map $d : \mathcal{X} \to \mathbb{A}$ is called a *decision function* for the decision problem $\mathfrak{A}$. The set of all decision functions is denoted by $\mathcal{D}(\mathbb{A}, \mathcal{X})$. Under the additional assumption that, for every pair $(\theta, d) \in \Theta \times \mathcal{D}(\mathbb{A}, \mathcal{X})$ fixed, the

map

$$((u_\theta \circ d) \circ X) : \Omega \to \mathbb{R}$$

is q_θ- integrable, we get a well-defined mapping

$$U : (\mathcal{D}(\mathbb{A}, \mathcal{X}) \times \Theta) \to \mathbb{R}\ ,\ (d, \theta) \mapsto \int_\Omega ((u_\theta \circ d) \circ X)\ dq_\theta = \int_\Omega u(d(X(\omega)), \theta)\ q_\theta(d\omega) \quad (7)$$

This allows us (and motivates) the following definition. It is taken from [2, p. 66] or [43, § 3.9.8] respectively (however, both authors use a notation that is slightly different of the one used here).

Definition 3. Under the conditions described above, the triplet

$$\mathcal{D}(\mathfrak{A}) := (\mathcal{D}(\mathbb{A}, \mathcal{X}), \Theta, U(\cdot))$$

is called the *data-based extension* of the CDP $\mathfrak{A}$. ▽

Remark. 1.) According to the *Theorem of Measure Transformation* (see e.g. [26, p. 92]) for all $(d, \theta) \in (\mathcal{D}(\mathbb{A}, \mathcal{X}) \times \Theta)$ the identity

$$U(d, \theta) := \int_\Omega ((u_\theta \circ d) \circ X)\ dq_\theta = \int_{\mathcal{X}} (u_\theta \circ d)\ dp_\theta := \int_{\mathcal{X}} u(d(x), \theta)\ p_\theta(dx) \quad (8)$$

holds. This justifies the following approach: In statistical modelling the set Ω, in general, is unknown. So, due to the above equation, one can build the model directly on the space $\mathcal{X}$ of possible realisations of the corresponding random experiment. This is backed-up by intuition: In general, it will be much easier to build a probabilistic model on $\mathcal{X}$ than on Ω, because $\mathcal{X}$ is the space of *observations* and, therefore, known to the actor.

2.) If both $\mathfrak{A}$ and $\mathcal{X}$ are finite, then (together with Remark 1.) the equation

$$U(d, \theta) = \sum_{x \in \mathcal{X}} u(d(x), \theta) \cdot p_\theta(\{x\}) \quad (9)$$

holds for all pairs $(d, \theta) \in (\mathcal{D}(\mathbb{A}, \mathcal{X}) \times \Theta)$.

Additionally, if $\mathcal{X} := \{x_1, \ldots, x_k\}$, the set $\mathcal{D}(\mathbb{A}, \mathcal{X})$ contains exactly $K := k^n$ elements. Thus, $\mathcal{D}(\mathbb{A}, \mathcal{X})$ can be written in the form

$$\mathcal{D}(\mathbb{A}, \mathcal{X}) := \{d_1, \ldots, d_K\}$$

and the following presentation for the data-based extension can be used:

$U(d_i, \theta_j)$	θ_1	$\dots$	θ_m
d_1	$U(d_1, \theta_1)$	$\dots$	$U(d_1, \theta_m)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
d_K	$U(d_K, \theta_1)$	$\dots$	$U(d_n, \theta_m)$

The structure of the corresponding random experiment (sometimes referred to as *information structure*, see e.g. [2, p.67]) then is of the form:

$p_{\theta_j}(\{x_i\})$	θ_1	$\dots$	θ_m
x_1	$p_{\theta_1}(\{x_1\})$	$\dots$	$p_{\theta_m}(\{x_1\})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$p_{\theta_1}(\{x_k\})$	$\dots$	$p_{\theta_m}(\{x_k\})$

Again, the above representations for data-based extensions of finite decision problems in the following will be used without further commentary.

3.) Formally, the data-based extension of a (no-data) CDP is again a (no-data) CDP (all required formal conditions are satisfied by the data-based extension as well). If the underlying (or *basic*) CDP is finite, the same holds for its data-based extension. Particularly, the set of states remains unchanged under the transitions to a data-based extension. Thus, all theorems and algorithms that demand a CDP such that $|\Theta| < \infty$ is satisfied can also be applied to data-based extensions of finite decision problems.

However, the cardinality of the set of all decision functions increases exponentially in the number of actions in $\mathbb{A}$: For example, if $|\mathbb{A}| = |\mathcal{X}| = 10$, then the set $\mathcal{D}(\mathbb{A}, X)$ has exactly $10^{10} = 10000000000$ elements.

But the situation can get even more absurd: Suppose our random experiment not only consists in observing the realisation of one single random variable X, but in observing the realisations of an i.i.d. sample of such variables. More precisely, let $X_1, \dots, X_s : \Omega \to \mathcal{V}$ denote i.i.d. random variables on Ω, where $|\mathcal{V}| = w \in \mathbb{N}$. Set $\mathcal{X} := \times_{i=1}^{s} \mathcal{V}$ and $X := (X_1, \dots, X_s) : \Omega \to \mathcal{X}$. Then, observing a sample $(X_1(\omega), \dots, X_s(\omega)) \in \mathcal{X}$ again fits in the above framework. However, as $\mathcal{X}$ now contains exactly w^s elements, the set $\mathcal{D}(\mathbb{A}, X)$ now contains $w^{s \cdot n}$ different decision functions. For example, consider an i.i.d. sample consisting of 10 observation and let again $|\mathbb{A}| = |\mathcal{X}| = 10$. Then, there even exist 10^{100} different decision functions,

which is an absurdly high number for a very small decision problem. This *combinatorial explosion* presents a vast computational challenge. However, the calculation of an optimal decision function can be avoided for some purposes, as we will see in Theorem 3 and the corresponding example (see Paragraph 1.3).

Another way of addressing this challenge, is to restrict the set $\mathcal{D}(\mathbb{A}, \mathcal{X})$ in a meaningful way. How this restriction can be done without loosing too much generality, depends on the concrete situation: Some decision function might be unreasonable in the context under considerations and, therefore, don't have to be taken into account. As a concrete example, consider the task of estimating the parameter of a parametric statistical model. Here, both the set of states of nature Θ and the set of actions $\mathbb{A}$ coincide with the parametric space of the model. Additionally, the set $\mathcal{D}(\mathbb{A}, \mathcal{X})$ consists of all possible estimating functions. However, there exists a huge variety of quality criteria for such estimating functions. For example, one could restrict the set $\mathcal{D}(\mathbb{A}, \mathcal{X})$ in a meaningful way by only taking *unbiased* estimating functions (see e.g. [12] for an exact definition) into account. In this way, one receives a much smaller decision problem without loosing anything: Why should we care about biased estimating functions? ◦

1.2 Criteria for the optimality of actions

After having introduced the classical decision problem and two of its common extensions in the last paragraph, we now recall and discuss criteria for determining optimal decisions in a given CDP. All considerations focus on CDPs in *utility form*. The translation of the results to CDPs in *loss form* can be done canonically. This will be the way of proceeding throughout the whole work.

So, let $\mathfrak{A} := (\mathbb{A}, \Theta, u(\cdot))$ be a CDP in utility form. Now, the crucial question is: Which action $a \in \mathbb{A}$ should be chosen? Clearly, it is very hard to answer this without further information. So, as a first approximation to a solution of this problem, it is conducive to remove all the actions that shouldn't be chosen, independent of what the true state of nature is. But, under which conditions an action shouldn't be chosen at all?

This question motivates the following definition, which also can be found (under the name *pointwise dominance* in a slightly different notation) e.g. in [40, p.18].

Definition 4. Let $\mathfrak{A}$ be any classical decision problem. An action $a \in \mathbb{A}$ is said to be *inadmissible*, if there exists an action $a^* \in \mathbb{A}$, such that

1. $u(a^*, \theta) \geq u(a, \theta)$ for all $\theta \in \Theta$

2. $u(a^*, \theta) > u(a, \theta)$ for at least one $\theta \in \Theta$

In this case one writes $a^* \gg a$ and calls a *strictly dominated* by a^*. $\triangledown$

Why shouldn't inadmissible actions be chosen at all? Under the assumption that the utility function is a perfect representation of the preference order of the actor (see introduction for further details), he will always try to maximize his utility by his choice (otherwise he would wilfully accept a loss of utility). Thus, he will never choose a strictly dominated action, since he could decide for the strictly dominating action without any risk. This action has at least the same utility for every state of nature, but a strictly greater utility for at least one of them.

However, by removing inadmissible actions from $\mathbb{A}$, the corresponding CDP in general isn't solved: Not all actions can be compared with respect to the relation $\gg$. For example, consider a CDP with only two possible states of nature θ_1 and θ_2. Here, the two actions (identified with the corresponding utility vectors) $a_1 := (10, 20)^T$ and $a_2 := (20, 10)$ are incomparable with respect to $\gg$.

Nevertheless, one can always work with the reduced set of action

$$\mathbb{A}^{ad} := \{a \in \mathbb{A} : \nexists a^* \in \mathbb{A} \text{ s.t. } a^* \gg a\}$$

instead of $\mathbb{A}$, since for every inadmissible action there exists another action dominating its utility. Immediately, two questions come into mind: Do there always exist admissible actions? And: Can we construct criteria that make all actions comparable and being still compatible with the concept of admissibility? Before being able to answer these questions, we need to recall some basics from order and relation theory.

1.2.1 Excursus: Relations and orderings

For some aspects of decision theory, it is useful to introduce some basic termini from order and relation theory. The presentation chosen here is kept to an absolute minimum. Let V be any non-empty set (consider for example the set of actions $\mathbb{A}$ of a decision problem). Then, every non-empty subset $R \subset V \times V$ of the product space of V is called a *(binary) relation* on V. Instead of $(u, v) \in R$ one often writes uRv. This is sometimes referred to as the so called *infix notation* for relations. Since the definition of a relation nearly makes no assumptions, there exist many different types of relations.

The following two definitions introduce a couple of properties to be able to distin-

guish between these different types of relations. They are taken from [19, Ch. 2.2].

Definition 5. Let R denote a binary relation on a non-empty set V. Then, R is said to be

i) *reflexive*, if vRv holds for all $v \in V$.

ii) *symmetric*, if uRv implies vRu for all $v, u \in V$.

iii) *irreflexive*, if $(v, v) \notin R$ holds for all $v \in V$.

iv) *transitive*, if vRu and uRw imply vRw for all $v, u, w \in V$.

v) *antisymmetric*, if vRu and uRv imply $v = u$ for all $v, u \in V$.

vi) *total*, if vRu or uRv holds for all $v, u, w \in V$. ▽

Definition 6. Let R denote a binary relation on a non-empty set V. Then, R is said to be

i) an *equivalence relation*, if it is reflexive, symmetric and transitive.

ii) a *partial ordering*, if it is reflexive, antisymmetric and transitive.

iii) a *strict ordering*, if it is irreflexive, antisymmetric and transitive.

iv) a *linear ordering*, if it is reflexive, antisymmetric, transitive and total. ▽ •

After having recalled some basic definitions from order and relation theory in the previous excursus, we can now reformulate the above mentioned inability of the relation $\gg$ to make all actions comparable: The relation $\gg$ is irreflexive, antisymmetric and transitive, but it is not total. This leads us to the first of the questions posed before: When do admissible actions exist? That, at least in the case of a finite CDP, this always is the case, is the statement of the following theorem. A more general version of the result can be found in [40, p. 18].

Theorem 1. Let $\mathfrak{A}$ be a finite CDP. Then $\mathbb{A}^{ad} \neq \emptyset$.

Proof. Let $\mathfrak{A}$ be finite. Assume, for contradiction, $\mathbb{A}^{ad} = \emptyset$. Then for all $a \in \mathbb{A}$, there exists an action $a^* \in \mathbb{A}$ such that $a^* \gg a$ $(:= (*))$. Otherwise, the action a^* would be admissible, that is $a^* \in \mathbb{A}^{ad}$, which would be a contradiction to $\mathbb{A}^{ad} = \emptyset$. Since the relation $\gg$ is known to be irreflexive, this always implies $a \neq a^*$. Otherwise, $(a, a) \in \gg$ would hold, which contradicts the irreflexivity of $\gg$.
Now, consider the following construction:

1.) Choose an arbitrary action $a^{(1)} \in \mathbb{A}$.

2.) Choose an arbitrary action $a^{(2)} \in \mathbb{A} \setminus A_1$, where $A_1 := \{a^{(1)}\}$. If $a^{(2)} \gg a^{(1)}$, set $A_2 := \{a^{(2)}\}$. Otherwise, set $A_2 := \emptyset$.

3.) Choose an arbitrary action $a^{(3)} \in \mathbb{A} \setminus \{a^{(1)}, a^{(2)}\}$. If $a^{(3)} \gg a^*$ for all $a^* \in A_1 \cup A_2$, set $A_3 := \{a^{(3)}\}$. Otherwise, set $A_3 := \emptyset$.

$\vdots$

n.) Choose the remaining action $a^{(n)} \in \mathbb{A} \setminus \{a^{(1)}, \dots, a^{(n-1)}\}$. If $a^{(n)} \gg a^*$ for all $a^* \in \bigcup_{k=1}^{n-1} A_k$, set $A_n := \{a^{(n)}\}$. Otherwise, set $A_n := \emptyset$.

The constructed set $A^+ := \bigcup_{k=1}^{n} A_k$ then has the following properties:

i) There exists $a_{max} \in A^+$ such that $a_{max} \gg a$ for all $a \in A^+ \setminus \{a_{max}\}$.

ii) For all $a \in A^- := \mathbb{A} \setminus A^+$ we have $\neg(a \gg a_{max})$.

Here, the properties i) and ii) directly follow from the construction of the set A^+. Now, we apply property $(*)$: Since $a_{max} \in \mathbb{A}$ there exists an action $a^* \in \mathbb{A}$ such that $a^* \gg a_{max}$. According to property ii) we then have $a^* \notin \mathbb{A} \setminus A^-$, that is $a^* \in A^+$. Then, according to property i), we have $a_{max} \gg a^*$. Since the relation $\gg$ is known to be antisymmetric, this gives us $a^* = a_{max}$. But then we have $a^* = a_{max}$ and $a^* \gg a_{max}$. This contradicts the irreflexivity of $\gg$. Hence, $\mathbb{A}^{ad}$ is non-empty. $\square$

A big advantage of the described *principle of excluding inadmissible actions* is its not doubtable claim for rationality: If an inadmissible action appears to be a rational choice, then the underlying utility function is necessarily misspecified.

As a consequence of this big generality, the principle turns out to be not sufficient to determine optimal decisions: The relation $\gg \subset \mathbb{A} \times \mathbb{A}$ is not total and, therefore, not all actions are comparable w.r.t. $\gg$. This leads us to the second question posed before: If we want to *construct* criteria that induce a *total* ordering on $\mathbb{A}$ (that is, any two elements of $\mathbb{A}$ are comparable with respect to the constructed criteria), we will have to accept some loss of generality. However, since the exclusion of inadmissible actions is undoubtedly reasonable, any *reasonable* construction should be compatible with the concept of admissibility, in the following sense: An inadmissible action should never be labelled optimal by any constructed ordering.

This motivates the following definition. It can be found e.g. in [2, p. 99].

Definition 7. Let $\mathfrak{A}$ be a CDP. Any function

$$\Phi^{\mathfrak{A}} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \Phi^{\mathfrak{A}}(a)$$

is called *criterion*.
An action $a^* \in \mathbb{A}$ is *optimal* for the criterion $\Phi^{\mathfrak{A}}$ (short: $\Phi^{\mathfrak{A}}$-*optimal*), if

$$\Phi^{\mathfrak{A}}(a^*) \geq \Phi^{\mathfrak{A}}(a)$$

holds for all $a \in \mathbb{A}$. ▽

Remark. 1.) The *image* $\Phi^{\mathfrak{A}}(\mathbb{A}) \subset \mathbb{R}$ of every criterion $\Phi^{\mathfrak{A}}$ is a real subset. Thus, any criterion $\Phi^{\mathfrak{A}}$ induces a linear ordering $\leqslant^{\Phi^{\mathfrak{A}}}$ on the set of actions $\mathbb{A}$ via: For all $a_1, a_2 \in \mathbb{A}$

$$a_1 \leqslant^{\Phi} a_2 \quad :\Leftrightarrow \quad \Phi^{\mathfrak{A}}(a_1) \leqslant \Phi^{\mathfrak{A}}(a_2)$$

where $\leqslant$ denotes the usual *smaller-equal* relation on the real numbers $\mathbb{R}$.

Here, the following convention is used: Actually, the relation $\leqslant^{\Phi}$ is not antisymmetric, since $\Phi^{\mathfrak{A}}(a_1) \leqslant \Phi^{\mathfrak{A}}(a_2)$ and $\Phi^{\mathfrak{A}}(a_2) \leqslant \Phi^{\mathfrak{A}}(a_1)$ does not necessarily imply $a_1 = a_2$, but only $\Phi^{\mathfrak{A}}(a_1) = \Phi^{\mathfrak{A}}(a_2)$. However, after having chosen a criterion Φ, we are totally indifferent between actions with coinciding criterion values. Hence, we write $a_1 =^{\Phi} a_2$ if, and only if,the equation $\Phi^{\mathfrak{A}}(a_1) = \Phi^{\mathfrak{A}}(a_2)$ holds. Note that $=^{\Phi}$ defines an equivalence relation on $\mathbb{A}$. The relation $\leqslant^{\Phi}$ then is antisymmetric with respect to the relation $=^{\Phi}$.

2.) Semantically, the definition of a criterion doesn't make any assumptions. Thus, the linear ordering induced by a criterion doesn't need to be meaningful at all. Particularly, it is easy to construct criteria, which label inadmissible actions as optimal. Hence, the choice of a criterion in a concrete decision problem should always be based on well substantiated arguments. For a (far) more in-depth discussion of the optimality of criteria, see for example [39]. A much shorter discussion of the topic can be found in the following excursus. ◦

1.2.2 Excursus: Different types of uncertainty

The following excursus mainly refers to [2, p. 21]. Particularly, the classification of uncertainty into different types has been adopted from this source. The crucial question is: Which of the (uncountably many) criteria available should be chosen in a concrete decision problem? According to the previous considerations, the chosen criterion should at least be compatible with the concept of admissibility. This

compatibility is *independent* of the type of uncertainty underlying a CDP. Thus, we can specify our question: Which of the criteria being compatible with the concept of admissibility should be chosen in a concrete situation?

The answer to this question strongly depends on the *nature* of the process generating the states of nature. Firstly, we can observe two extremes:

- *Type I*: The process generating the states of nature can be compared to an *ideal lottery*. Every state occurs with a *fixed* and *known* (classical) probability. Thus, the realized state of nature is independent of the action chosen by the actor. The *nature* (as an abstract concept) *cannot* influence the generation of the states.

- *Type II*: The nature acts as an *antagonist* to the actor. For every action the actor chooses, the nature will pick (one of the) state(s) minimizing the actors utility (this state won't be unique in general). In this case, the nature can fully influence the process generating the states of nature and is *omniscient* concerning the actor's order of preferences. Otherwise, it wouldn't be possible for the nature to choose the actor's *least favourable* state.

Both of the types of nature described above seem to be *idealizing* and, therefore, very *restrictive*. The most (non-academic) examples for CDPs will neither fit strictly to type I nor to type II. In the case of a situation under strict type I uncertainty, especially the assumption of a known probability distribution often seems to be unrealistic. In situations, which are rather of the second type, the assumption of omniscience can't be kept up in many cases. Therefore, we want to define two less idealizing modifications of the types of uncertainty described above.

- *Type I*:* Exactly like in the type I situation, the process generating the states of nature can be compared to an ideal lottery. However, in this modification, the exact probability mass function on the set Θ is assumed to be unknown (or ill-known) to the actor. Depending on the quality of information concerning the state generating process, the actor might be able to specify a *best subjective guess* for the true probability measure. If this is the case, he can simply proceed by *acting as if* this guess was the true measure (see Remark 1 on Definition 8). If such a measure cannot be guessed in a consistent way, one has to consider more general descriptions of uncertainty (for examples of such situations see Paragraph 4.1).

- *Type II*:* Again, the nature acts as an antagonist to the actor. However, the assumption of omniscience can be weakened in different ways. For example,

consider the situation of playing chess against a friend: As both players want to win the game, the friend will always try to make a move that maximizes the actor's utility. However, as the friend is no computer, he won't always know this least favourable move.

As we will see later, even these weakened types of uncertainty are too restrictive in many situations. Instead, very often a combination of the different types seems to be suitable to characterize a situation under uncertainty adequately.

Another important point to mention in this context is the following: In reality, there might be a vast *difference* in what the actor *assumes* about the type of uncertainty underlying the CDP and the *true* type of uncertainty. Constructing criteria for the optimality of decisions based on wrong beliefs about the true type of uncertainty might have horrible consequences. This is a serious problem.

Nevertheless, in the following we *assume* that the actor's beliefs about the type of uncertainty and the true type of uncertainty coincide. •

1.2.3 Optimal criteria under type I/I* uncertainty: Bernoulli-/Bayes-Actions

Consider a CDP under strict type I uncertainty and let ξ be the probability measure on $(\Theta, \sigma(\Theta))$ characterizing the *purely stochastic* uncertainty between the different states of nature belonging to the set Θ. Here, $\sigma(\Theta)$ denotes a σ-field on Θ including all the singletons.
A very straightforward criterion for the optimality of an action under this type of uncertainty then is the following: An action $a^* \in \mathbb{A}$ is optimal if, and only if, it maximizes the expected utility under ξ, i.e. for all $a \in \mathbb{A}$ the inequality

$$\mathbb{E}_\xi(u(a^*, \theta)) \geqslant \mathbb{E}_\xi(u(a, \theta))$$

holds. This motivates the following definition. Similar definitions can be found in e.g. [43, § 3.9.4] or [2, p.184].

Definition 8. Let $\mathfrak{A}$ a CDP and ξ a probability measure on $(\Theta, \sigma(\Theta))$, where $\{\theta\} \in \sigma(\Theta)$ for all $\theta \in \Theta$. Let

$$u_a(\cdot) := u(a, \cdot) : \Theta \to \mathbb{R} \ , \ \theta \mapsto u(a, \theta)$$

be a ξ-integrable map for all $a \in \mathbb{A}$. Then the criterion

$$\Phi^{\mathfrak{A}}_{B(\xi)} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \mathbb{E}_\xi(u_a) := \int_\Theta u_a \, d\xi \tag{10}$$

is called *Bernoulli-criterion* with respect to ξ.
Every $\Phi^{\mathfrak{A}}_{B(\xi)}$-optimal action is called *Bernoulli-action* with respect to ξ. The set of all Bernoulli-actions w.r.t. ξ is denoted by $\mathbb{A}_{B(\xi)}$. $\bigtriangledown$

Remark. 1.) Under Type I* uncertainty, the exact probability measure ξ is unknown (or at least ill-known). Therefore, the Bernoulli-criterion is not applicable. In this case, one has to content oneself with a *prior-distribution* π estimating ξ. This *subjective* assignment of probability mass should take into account possible information about the state generating process in the best possible manner.

In this case, the criterion $\Phi^{\mathfrak{A}}_{B(\pi)}$ is called *Bayes-criterion* with respect to the prior π. Every $\Phi^{\mathfrak{A}}_{B(\pi)}$-optimal action then is called *Bayes-action* with respect to the prior π. For more information, see for example [3, p.6].

From a philosophical point of view, this needs some attention: Implicitly, the above approach makes the assumption that any kind of prior information can be formalized by using a classical probability distribution. And even more is assumed: *Any situation under uncertainty can be characterized by a classical probability.* This assumption is known as the *Bayesian Paradigm* or the *Bayesian Dogma of Precision* (see for example [43, § 1.1.3], or [4, p. 32]).

This assumption is closely related to the (classical) *subjective* approach to probability theory established by *Bruno de Finetti (1906-1985)* (see [14] and [15]): Here, the probability of an event is interpreted as the unique(!) *degree of belief* of a rational actor in the occurrence of the event. According to this theory, a suitable prior distribution π can be gained by simply assigning each singleton event in $\mathcal{P}(\Theta)$ the corresponding degree of belief. If the assignment is done by a rational actor, this will induce a classical probability measure.

However, in practice, this assumption often is violated: Actors that are undoubtedly rational sometimes produce degrees of belief that simply aren't embeddable into the restrictive theoretical framework of classical probability theory. Such situations will be discussed in detail within the motivating examples in Paragraph 4.1. Of particular interest in the context of de Finetti's theory is Paragraph 4.1.2.

Note that, from a mathematical point of view, there is no difference in the definition of the Bernoulli- and the Bayes-criterion. Semantically, the Bernoulli-criterion can be viewed as the special case of the Bayes-criterion where perfect information on the mechanism generating the states of nature is available.

2.) If Θ is finite and $\sigma(\Theta) := \mathcal{P}(\Theta)$ then all conditions of the above definition are

satisfied and the equation

$$\Phi^{\mathfrak{A}}_{B(\pi)}(a) = \sum_{j=1}^{m} u(a, \theta_j) \cdot \pi(\{\theta_j\}) \tag{11}$$

holds for all $a \in \mathbb{A}$.
Again, it makes no difference, if $\Phi^{\mathfrak{A}}_{B(\pi)}$ is a Bernoulli- or a Bayes-criterion. ∘

The following theorem makes a statement on the relationship between the Bernoulli- or Bayes-criterion and the mixed extension of a finite CDP: The transition from a CDP to its mixed extension does *not* improve the expected utility with respect to an arbitrary prior probability measure on the set of states.

Assuming that the Bernoulli-criterion is optimal when considering decision problems under type I uncertainty, the theorem gives certain evidence that there is no use in randomization in such situations. An alternative proof of the theorem can for example be found in [35, Theorem 2.4.2]. Another alternative proof for the theorem using duality theory (see Paragraph 2.2) will implicitly be given within the considerations in Paragraph 3.1.

Theorem 2. Let $\mathfrak{A}$ be any finite CDP and $G(\mathfrak{A})$ its mixed extension. Let further π denote a probability measure on $(\Theta, \mathcal{P}(\Theta))$. Then the following holds:

$$\exists a^* \in \mathbb{A}: \quad \Phi^{G(\mathfrak{A})}_{B(\pi)}(\delta_{a^*}) \geqslant \Phi^{G(\mathfrak{A})}_{B(\pi)}(p) \quad \forall p \in G(\mathbb{A})$$

Proof. Let $p \in G(\mathbb{A})$ be arbitrary. Then, according to the equations (11) and (6), we have

$$\Phi^{G(\mathfrak{A})}_{B(\pi)}(p) = \sum_{j=1}^{m} G(u)(p, \theta_j) \cdot \pi(\{\theta_j\}) = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} u(a_i, \theta_j) \cdot p(\{a_i\}) \right) \cdot \pi(\{\theta_j\}) \tag{12}$$

A simple computation gives us

$$\sum_{j=1}^{m} \left(\sum_{i=1}^{n} u(a_i, \theta_j) \cdot p(\{a_i\}) \right) \cdot \pi(\{\theta_j\}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} u(a_i, \theta_j) \cdot \pi(\{\theta_j\}) \right) \cdot p(\{a_i\}) \tag{13}$$

Using the the notations

- $c(a_i) := \sum_{j=1}^{m} u(a_i, \theta_j) \cdot \pi(\{\theta_j\})$ for all $i = 1, \ldots, n$ and
- $c_{max} := \max\{c(a_1), \ldots, c(a_n)\}$

we derive the following inequality

$$\Phi_{B(\pi)}^{G(\mathfrak{A})}(p) = \sum_{i=1}^{n} c(a_i) \cdot p(\{a_i\}) \leqslant c_{max} \cdot \sum_{i=1}^{n} p(\{a_i\}) = c_{max}$$

As bounded functions attain their maximum on finite sets, there exists an action $a^* \in \mathbb{A}$ such that $c_{max} = c(a^*)$. Therefore, we get

$$\Phi_{B(\pi)}^{G(\mathfrak{A})}(p) \leqslant c(a^*) = \sum_{i=1}^{n} c(a_i) \cdot \delta_{a^*}(\{a_i\}) = \Phi_{B(\pi)}(\delta_{a^*})$$

As p was chosen arbitrarily, this completes the proof. □

Remark. Implicitly, the above Theorem 2 guarantees the existence of Bayes-optimal actions in arbitrary finite CDPs. Furthermore, it proves that such actions remain optimal when considering the mixed extension of a finite CDP. For any prior distribution π, there exists a pure action admitting an expected utility that is at least as big as the expected utility of any randomized action available. Hence, the theorem as well guarantees the existence of Bayes-optimal actions in the mixed extension of finite CDPs. ∘

1.2.4 Optimal criteria under type II/II* uncertainty: Maximin-Actions

Now, consider a CDP under strict type II uncertainty. Then, the occurrance of the states of nature is assumed to be not at random, but controlled by a omniscient nature trying to minimize the actor's utility.

Taking this knowledge into account, a reasonable criterion for the optimality of a decision is the following: Choose an action if, and only if, it maximizes the utility in the *worst-case-scenario*, i.e. if it acts best under the least favourable state of nature. This decision rule is called the *Maximin-principle* (or *Wald-rule*) and originally goes back to works of the statistician *Abraham Wald (1902-1950)* (see [42, p. 184-185] for further details).

This motivates the following definition. In this form (with a slightly different notation), it can for example be found in [2, p.153].

Definition 9. Let $\mathfrak{A}$ be any CDP. Then the criterion

$$\Phi_M^{\mathfrak{A}} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \inf_{\theta \in \Theta} u(a, \theta) \tag{14}$$

is called *Maximin-criterion.*

Every $\Phi_M^{\mathfrak{A}}$-optimal action then is called *Maximin-action.* $\triangledown$

Remark. 1.) If $\mathfrak{A}$ is finite, the above infimum is attained on the set Θ, i.e. there is an action $a \in \mathbb{A}$ such that

$$\Phi_M^{\mathfrak{A}}(a) = \min_{\theta \in \Theta} u(a, \theta) \tag{15}$$

An action $a^* \in \mathbb{A}$ then is $\Phi_M^{\mathfrak{A}}$-optimal if, and only if,

$$\Phi_M^{\mathfrak{A}}(a^*) = \max_{a \in \mathbb{A}} \left(\min_{\theta \in \Theta} u(a, \theta) \right) \tag{16}$$

This clarifies the following fact: The criterion $\Phi_M^{\mathfrak{A}}$ labels all the actions as optimal whose utility values are the most resistant (or *robust*) against the whims of nature.

2.) If the underlying CDP is of type II*, using the Maximin-criterion is still the best we can do. If additional information on the *degree* of omniscience of the nature is available, a modification of the criterion might pay off. We will come back to this point later (see Paragraph 1.2.5). $\circ$

The following example shows that, assuming strict type II uncertainty, the transition to the mixed extension of a CDP might generate a strict improvement of the utility. Here, we assume the Maximin-criterion to be the optimal decision criterion under this type of uncertainty.

Example 1. Consider the following finite CDP $\mathfrak{A}$

	θ_1	θ_2	$\Phi_M^{\mathfrak{A}}(a_i)$
a_1	10	30	10
a_2	25	5	5

The unique Maximin-action in the above problem is a_1 with an utility of 10 units. Now, consider the randomized action $p \in G(\mathbb{A})$ induced by the assignment $p(\{a_1\}) = 0.6$. We compute

$$\Phi_M^{G(\mathfrak{A})}(p) = \min\{0.6 \cdot 10 + 0.4 \cdot 25, 0.6 \cdot 30 + 0.4 \cdot 5\} = 16$$

Hence, the randomized action p has a strict greater Maximin-utility than any of the pure actions available. An effective method to derive optimal randomized actions in similar situations will be discussed in Paragraph 3.4. $\star$

1.2.5 Combining the two approaches: The Hodges & Lehmann criterion

As already mentioned before, the vast majority of CDPs relevant to the praxis won't be exactly of one of the discussed uncertainty types. Instead, in many cases the true type of uncertainty underlying a given CDP will be adequately described be a combination of the two extremes.

In literature, one can find many different criteria trying to take this asymmetry into account. In the present work, we will focus on only one of them: The *criterion of Hodges & Lehmann*. It goes back two the work of the statisticians *Joseph Hodges (1922-2002)* and *Erich Lehmann (1917-2009)* and was first published in [22] in 1952.

The idea is the following: Any type of uncertainty can be viewed as a *trade-off* between type I and type II uncertainty. The optimal criterion under type I uncertainty is assumed to be the Bernoulli-criterion (see the considerations in Paragraph 1.2.3), the optimal criterion under type II uncertainty is assumed to be the Maximin-criterion (see the considerations in Paragraph 1.2.4).

The weighing of the trade-off between these two extreme criteria can be controlled by a trade-off parameter $\alpha \in [0,1]$. The closer α is to 1, the more the underlying uncertainty type tends to type I and, therefore, the more the optimal criterion tends to the Bernoulli-criterion. The closer α is to 0, the more the type of uncertainty tends to strict type II uncertainty and, therefore, the more the optimal criterion coincides with the Maximin-criterion.

Definition 10. Let $\mathfrak{A}$ and ξ be defined as in definition 8 and $\alpha \in [0,1]$. The criterion

$$\Phi^{\mathfrak{A}}_{H(\alpha,\xi)} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \alpha \cdot \Phi^{\mathfrak{A}}_{B(\xi)}(a) + (1-\alpha) \cdot \Phi^{\mathfrak{A}}_{M}(a) \tag{17}$$

is called *criterion of Hodges & Lehmann.*
Every $\Phi^{\mathfrak{A}}_{H(\alpha,\xi)}$-optimal action then is called *H(α, ξ)*-action. ▽

Remark. Applying the Hodges & Lehmann criterion goes along with some strong assumptions: Recall that one of the assumptions made in the definition of a CDP is that the underlying uncertainty type is completely known to the actor. In the context of the Hodges & Lehmann criterion this would mean that the exact value for the trade-off parameter α is known. However, the Hodges & Lehmann is constructed for situations in which the exact type of uncertainty underlying the CDP is hard to make out. This somehow seems to contradict this assumption. One way of addressing this problem is the following: As we actually *don't* know the true uncertainty type

underlying a CDP of interest, but only have vague guesses that it is neither strict type I nor strict type II uncertainty, the best we can to, is to fix α best possible and *act as if* the criterion perfectly fits the underlying uncertainty type. ◦

The next example illustrates the following: If the type of uncertainty underlying a CDP is adequately characterized by the Hodges & Lehmann criterion, i.e. if the Hodges & Lehmann criterion generates optimal decisions in a given CDP, then the transition to the mixed extension of the CDP might generate strict higher utility. Therefore, randomization is useful for combined uncertainty types (for the extreme case that α equals 1, this was already shown in Example 1).

Example 2. Consider again the CDP of the previous example

	θ_1	θ_2
a_1	10	30
a_2	25	5

Let ξ denote the probability measure on $(\Theta, \mathcal{P}(\Theta))$ induced by the assignment $\xi(\{\theta_1\}) = 0.3$ and let $\alpha = 0.5$.
Next, we compute the Hodges & Lehmann utility of the pure actions a_1 and a_2:

$$\begin{aligned}\Phi^{\mathfrak{A}}_{H(\alpha,\xi)}(a_1) &= \frac{1}{2}\cdot\Phi^{\mathfrak{A}}_{B(\xi)}(a_1) + \frac{1}{2}\cdot\Phi^{\mathfrak{A}}_{M}(a_1)\\ &= \frac{1}{2}\cdot(0.3\cdot 10 + 0.7\cdot 30)\\ &+ \frac{1}{2}\cdot\min\{10,30\}\\ &= 17\end{aligned}$$

$$\begin{aligned}\Phi^{\mathfrak{A}}_{H(\alpha,\xi)}(a_2) &= \frac{1}{2}\cdot\Phi^{\mathfrak{A}}_{B(\xi)}(a_2) + \frac{1}{2}\cdot\Phi^{\mathfrak{A}}_{M}(a_2)\\ &= \frac{1}{2}\cdot(0.3\cdot 25 + 0.7\cdot 5)\\ &+ \frac{1}{2}\cdot\min\{5,25\}\\ &= 8\end{aligned}$$

Now, consider the randomized action $p \in G(\mathbb{A})$ induced by the assignment $p(\{a_1\}) = 0.8$. Again, we compute the Hodges & Lehmann utility:

$$\begin{aligned}\Phi^{G(\mathfrak{A})}_{H(\alpha,\xi)}(p) &= \frac{1}{2}\cdot\Phi^{G(\mathfrak{A})}_{B(\xi)}(p) + \frac{1}{2}\cdot\Phi^{G(\mathfrak{A})}_{M}(p)\\ &= \frac{1}{2}\cdot(0.8\cdot(10\cdot 0.3 + 30\cdot 0.7) + 0.2\cdot(25\cdot 0.3 + 5\cdot 0.7))\end{aligned}$$

$$
\begin{aligned}
+ \quad & \frac{1}{2} \cdot \min\{0.8 \cdot 10 + 0.2 \cdot 25, 0.8 \cdot 30 + 0.2 \cdot 5\} \\
= \quad & \frac{1}{2} \cdot (21.4 + 13) \\
= \quad & 17.2
\end{aligned}
$$

Hence, we get the following inequalities

$$\Phi^{G(\mathfrak{A})}_{H(\alpha,\xi)}(p) > \Phi^{\mathfrak{A}}_{H(\alpha,\xi)}(a_1) > \Phi^{\mathfrak{A}}_{H(\alpha,\xi)}(a_2)$$

Therefore, we found a randomized action p with a strict higher utility than all of the pure actions available. Thus, in general, when applying the Hodges & Lehmann criterion it doesn't suffice to take only pure actions into consideration. Hereby, the simplicity of the example even strengthens the argument: If randomization can improve the utility even in very small decision problems, the improvement in bigger CDPs might be even more drastically. An algorithm for determining optimal randomized actions will be discussed in Paragraph 3.5. ⋆

1.3 The Fundamental Theorem of Bayesian Decision Theory

In this paragraph, we recall a connection between Bayes-optimal decision functions in data-based decision problems and *Bayesian Theory* known from classical statistics: When considering data-based extensions of classical decision problems, much computational effort can be avoided by *updating* the measure before and, afterwards, determining the Bayes-action with respect to the updated measure. Our presentation is strongly influenced by [4].

For a more in-depth discussion we need to recall some basic concepts from Baysian statistics (for a detailed presentation see for example [37, § 2.4.1] or [7, Ch. 1]). Very roughly spoken, the fundamental principle (or assumption?) in this field of statistics (or philosophy?) is the following: *For any situation under uncertainty there exists a classical probability measure characterizing it.* This (very rigorous) point of view is sometimes referred to as the *Baysian Paradigm* (see e.g. [4, p. 32]). More precisely, the idea underlying this view can be summarized as follows: There exists a set Ω which contains all the possible (abstract) states of the world. The only random that exists is between the different elements of Ω. That is, once we know which concrete $\omega \in \Omega$ has occurred, there is no random left at all. Particularly, this implies that every random phenomenon which we are uncertain about can represented as a random variable $X : \Omega \to S$ from Ω to some set S. The different elements $\omega \in \Omega$ occur at random. Thus, a more precise version of the above paradigm could be: *There exists*

a probability measure q on $(\Omega, \mathcal{A})$ adequately characterizing the uncertainty between the different elements of Ω (under the information available). Here, $\mathcal{A}$ denotes a suitable σ-field. Of course both, the measure q and the space Ω, are *unknown*.

Now, consider a finite classical decision problem $\mathfrak{A}$. We are uncertain about which concrete state of nature from the set Θ occurs. According to the above assumption, this uncertainty then necessarily has to be representable by a random variable $N : \Omega \to \Theta$. Hence, if we assume π to be the probability measure on $(\Theta, \mathcal{P}(\Theta))$ that characterizes the uncertainty between the elements of Θ adequately, the following identity necessarily has to hold:

$$\pi(A) = N[q](A) := q(N^{-1}(A)) \tag{18}$$

for all $A \in \mathcal{P}(\Theta)$. That is, under this assumption π equals the image measure of p under the random variable N.

Now, consider we know that event $B \in \mathcal{A}$, where $p(B) > 0$, has occurred. Then, the map

$$q_B : \mathcal{A} \to [0, 1] \quad , \quad A \mapsto \frac{q(A \cap B)}{q(B)} \tag{19}$$

again defines a probability measure on $(\Omega, \mathcal{A})$, the so called *conditional probability measure* w.r.t. the event B. This measure then adequately characterizes the uncertainty between the elements of Ω *given the new information.* Using equation (19) one easily verifies the following identity, which is well known under the name *Bayes' Theorem:* For all $A \in \mathcal{A}$, such that $q(A) > 0$, we have

$$q_B(A) = \frac{q_A(B) \cdot q(A)}{q(B)} \tag{20}$$

Now, let $\mathcal{D}(\mathfrak{A})$ denote a data-based extension of some finite decision problem $\mathfrak{A}$. Additionally, we assume the the space $\mathcal{X}$ to be finite, so that we can use the power set $\sigma(\mathcal{X}) := \mathcal{P}(\mathcal{X})$ as a σ-field. Thus, for every $\theta \in \Theta$, we receive a probability measure p_θ on $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$. This measure is assumed to be the true measure on $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$ given the information $N(\omega) = \theta$. However, according to the above considerations, we know that $q_{N^{-1}(\{\theta\})}$ equals the true measure on $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$ given the information $N(\omega) = \theta$ as well. This, together with (19), implies the identity

$$p_\theta(A) := X[q_\theta](A) = q_{N^{-1}(\{\theta\})}(X^{-1}(A)) = \frac{q(X^{-1}(A) \cap N^{-1}(\{\theta\}))}{q(N^{-1}(\{\theta\}))} \tag{21}$$

for all $A \in \mathcal{P}(\mathcal{X})$. Now, if π again is assumed to be the true measure on $(\Theta, \mathcal{P}(\Theta))$, this, together with (18) and (1.3), implies the following identity

$$q_{X^{-1}(A)}(N^{-1}(\{\theta\})) = \frac{p_\theta(A) \cdot q(N^{-1}(\{\theta\}))}{q(X^{-1}(A))} = \frac{p_\theta(A) \cdot \pi(\{\theta\}))}{q(X^{-1}(A))} \tag{22}$$

for all $A \in \mathcal{P}(\mathcal{X})$. If $A = \{x\}$ for some $x \in \mathcal{X}$, we receive

$$q_{X^{-1}(\{x\})}(N^{-1}(\{\theta\})) = \frac{p_\theta(\{x\}) \cdot q(N^{-1}(\{\theta\}))}{q(X^{-1}(\{x\}))} = \frac{p_\theta(\{x\}) \cdot \pi(\{\theta\}))}{q(X^{-1}(\{x\}))} \tag{23}$$

Clearly, for $x \in \mathcal{X}$ fixed, the above equation equation then induces a probability measure π_x on the measurable space $(\Theta, \mathcal{P}(\Theta))$. More precisely, we have

$$\pi_x(A) := \sum_{\theta \in A} \frac{p_\theta(\{x\}) \cdot \pi(\{\theta\}))}{q(X^{-1}(\{x\}))} \tag{24}$$

for all $A \in \mathcal{P}(\Theta)$. For an event $A \in \mathcal{P}(\Theta)$, the value $\pi_x(A)$ then can be interpreted as the probability that event A occurs *given the information* $X(\omega) = x$. Hence, we get a new description of the uncertainty between the different elements of Θ that takes the new information given by our random experiment into account.

Finally, this allows us to formulate the following theorem. It is sometimes referred to as the *Fundamental Theorem of Bayesian Decision Theory* (see e.g. [37, p. 284]).

Theorem 3. Let $\mathfrak{A}$ denote a finite decision problem and let $\mathcal{D}(\mathfrak{A})$ denote a data-based extension of $\mathfrak{A}$, where $\mathcal{X} := \{x_1, \dots, x_k\}$ is finite and $q(X^{-1}(\{x\})) > 0$ for all $x \in \mathcal{X}$. Further, let π denote the true probability measure on $(\Theta, \mathcal{P}(\Theta))$. Then, the following holds:

$$d^* \in \mathcal{D}(\mathbb{A}, \mathcal{X})_{B(\pi)} \quad \Leftrightarrow \quad \forall\, x \in \mathcal{X} : \quad d^*(x) \in \mathbb{A}_{B(\pi_x)} \tag{25}$$

Particularly, a $B(\pi)$-optimal decision function for the data-based extension $\mathcal{D}(\mathfrak{A})$ can be gained by determining $B(\pi_x)$-optimal actions for the no-data CDP $\mathfrak{A}$ with respect to all updated prior measures π_x, where $x \in \mathcal{X}$.

Proof. (similar in [2, p.267-270], more general version e.g. in [37, p. 284])

Before we can start to prove the theorem, we need to prove the following statement (S): Let $s \in \mathbb{N}$, A be any set and let $f_1, \dots, f_s : A \to \mathbb{R}$ be functions such that

$\max_{x\in A} f_i(x)$ exists for all $i = 1, \ldots, s$. Further, let $c \in (\mathbb{R}^+)^s$. Define the function

$$f : A^s \to \mathbb{R} \quad , \quad (x_1, \ldots, x_s) \mapsto \sum_{i=1}^{s} c_i \cdot f_i(x_i) \tag{26}$$

Then, we have

$$\max_{x\in A^s} f(x) = \sum_{i=1}^{s} c_i \cdot \max_{x_i \in A} f_i(x_i) \tag{27}$$

That is, $f(x_1^*, ..., x_s^*)$ is maximal if, and only if, $f_i(x_i^*) = \max_{x\in A} f_i(x)$ for all $i = 1, ..., s$.

Proof of (S): Let $x_{m(i)} \in A$ such that $f_i(x_{m(i)}) = \max_{x\in A} f_i(x)$ for all $i = 1, \ldots, s$. Clearly, we have $\sup_{x\in A^s} f(x) \geqslant f(x_{m(1)}, \ldots, x_{m(s)}) =: f(x^{max})$. Assume, for contradiction, $\sup_{x\in A^s} f(x) > f(x^{max})$. Then , there exists $x^* \in A^s$ such that

$$f(x^*) - f(x^{max}) = \sum_{i=1}^{s} c_i \cdot (f_i(x_i^*) - f(x_{m(i)})) > 0 \tag{28}$$

Since $c \in (\mathbb{R}^+)^s$, this implies the existence of $i^* \in \{1, \ldots, s\}$ such that $f_{i^*}(x_{i^*}^*) > f(x_{m(i^*)})$, which yields a contradiction to the definition of $f(x_{m(i^*)})$. Thus, we have

$$\sup_{x\in A^s} f(x) = \max_{x\in A^s} f(x) = \sum_{i=1}^{s} c_i \cdot \max_{x_i \in A} f_i(x_i) \tag{29}$$

and (S) is proven.

So, let's turn to the proof of the theorem. For $i = 1, ..., k$, we define the functions

$$f_i : \mathbb{A} \to \mathbb{R} \quad , \quad a \mapsto \Phi^{\mathfrak{A}}_{B(\pi_{x_i})}(a) \tag{30}$$

Since we know that for every probability measure on $(\Theta, \mathcal{P}(\Theta))$ there exists a Bayes-action, the expression $\max_{x\in\mathbb{A}} f_i(x)$ exists for all $i = 1, \ldots, k$. Now, for $i = 1, \ldots, k$, define $c_i := q(X^{-1}(x_i))$. Thus, $c \in (\mathbb{R}^+)^s$.

Define the function

$$f : \mathbb{A}^k \to \mathbb{R} \quad (a_1, \ldots, a_k) \mapsto \sum_{i=1}^{k} c_i \cdot f_i(a_i) \tag{31}$$

Then, according to (S), we have

$$\max_{a\in\mathbb{A}^k} f(a_1, \ldots, a_k) = \sum_{i=1}^{k} c_i \cdot \max_{a_i\in\mathbb{A}} \Phi^{\mathfrak{A}}_{B(\pi_{x_i})}(a_i) = \sum_{i=1}^{k} c_i \cdot f_i(a_i^*) \tag{32}$$

where, for all $i = 1, \ldots, k$, $a_i^* \in \mathbb{A}_{B(\pi_{x_i})}$ denotes a Bayes action w.r.t. π_{x_i}. That is, $f(a^*)$ is maximal if, and only if, the vector $a^* \in \mathbb{A}_{B(\pi_{x_1})} \times \cdots \times \mathbb{A}_{B(\pi_{x_k})}$ consists of Bayes-actions only.

Now, let $d \in \mathcal{D}(\mathbb{A}, \mathcal{X})$ arbitrary. We compute

$$\begin{aligned}
\Phi_{B(\pi)}^{\mathcal{D}(\mathfrak{A})}(d) &= \sum_{j=1}^{m} U(d, \theta_j) \cdot \pi(\{\theta_j\}) \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{k} u(d(x_i), \theta) \cdot p_{\theta_j}(\{x_i\}) \right) \cdot \pi(\{\theta_j\}) \\
&= \sum_{i=1}^{k} \left(\sum_{j=1}^{m} u(d(x_i), \theta) \cdot \pi(\{\theta_j\}) \cdot p_{\theta_j}(\{x_i\}) \right) \\
&\overset{(24)}{=} \sum_{i=1}^{k} \left(\sum_{j=1}^{m} u(d(x_i), \theta) \cdot \pi_{x_i}(\{\theta_j\}) \cdot q(X^{-1}(\{x_i\}) \right) \\
&= \sum_{i=1}^{k} \left(\sum_{j=1}^{m} u(d(x_i), \theta) \cdot \pi_{x_i}(\{\theta_j\}) \right) \cdot q(X^{-1}(\{x_i\}) \\
&= \sum_{i=1}^{k} c_i \cdot \Phi_{B(\pi_{x_i})}^{\mathfrak{A}}(d(x_i)) \\
&= f(d(x_1), \ldots, d(x_n))
\end{aligned}$$

Hence, $\Phi_{B(\pi)}^{\mathcal{D}(\mathfrak{A})}(d)$ is maximal if, and only if, $f(d(x_1), \ldots, d(x_n))$ is maximal. According to (32), this is the case if, and only if, $d(x_i) \in \mathbb{A}_{B(\pi_{x_i})}$ for all $i = 1, \ldots, k$. This completes the proof. □

As already mentioned before, Theorem 3 can help to avoid vast computational effort. The following example illustrates how this can be done in a concrete data-based decision situation.

Example 3. Consider the finite CDP $\mathfrak{A}$ given by

u_{ij}	θ_1	θ_2
a_1	25	15
a_2	30	10
a_3	20	30

where $\theta_1 = 0.6$ and $\theta_2 = 0.3$. Our random experiment is given by the variable $X : \Omega \to \mathcal{X} := \{0, 1\}$. Suppose, for $\theta = \theta_1, \theta_2$, the conditional measure p_θ on

$(\mathcal{X}, \mathcal{P}(\mathcal{X}))$ is induced by the assignment

$$p_\theta(\{x\}) := \theta^x \cdot (1-\theta)^{1-x}$$

for all $x \in \mathcal{X}$. Further, let π be the probability measure on $(\Theta, \mathcal{P}(\Theta))$ induced by the assignment $\pi(\{\theta_1\}) = 0.5$. Then, the data-based extension $\mathcal{D}(\mathfrak{A})$ is given by

U_{ij}	θ_1	θ_2	$\mathbb{E}_\pi(U_{d(i,j)})$
$d(1,1)$	25	15	20
$d(1,2)$	28	13.5	20.75
$d(1,3)$	22	19.5	20.75
$d(2,1)$	27	19.5	19.25
$d(2,2)$	30	10	20
$d(2,3)$	24	16	20
$d(3,1)$	23	24	24.5
$d(3,2)$	26	24	25
$d(3,3)$	20	30	25

where $d(i,j)$ denotes the decision function that maps 0 to action a_i and 1 to action a_j, for $i, j, \in \{1,2,3\}$. The last column contains the expected utility of the corresponding decision function under the measure π. For example, the utility $U(d(1,3), \theta_1)$ of the decision function $d(1,3)$ under the state θ_1 can be computed as follows:

$$\begin{aligned} U(d(1,3),\theta_1) &= \sum_{x\in\mathcal{X}} u(d(1,3)(x),\theta_1) \cdot p_{\theta_1}(\{x\}) \\ &= u(a_1,\theta_1)\cdot 0.6^0 \cdot 0.4^1 + u(a_3,\theta_1)\cdot 0.6^1 \cdot 0.4^0 \\ &= 25 \cdot 0.4 + 20 \cdot 0.6 = 22 \end{aligned}$$

Then, the expected utility of the decision function $d(1,3)$ w.r.t. π is

$$\mathbb{E}_\pi(U_{d(1,3)}) = U(d(1,3),\theta_1)\cdot\pi(\{\theta_1\}) + U(d(1,3),\theta_2)\cdot\pi(\{\theta_2\}) = 0.5\cdot(22+19.5) = 20.75$$

Clearly, $d(3,2)$ and $d(3,3)$ are $B(\pi)$-optimal decision functions. That is, if our random experiment ends up with 0, we have to choose action a_3 and if it ends up with 1, we can either choose action a_2 or action a_3.

Next, we want to apply Theorem 3. Therefore, for each $x \in \mathcal{X}$, we compute the updated measure π_x from equation (24). So, let $x \in \mathcal{X}$ be fixed. Then, the measure

π_x is induced by the assignment

$$\pi_x(\{\theta\}) = \frac{p_\theta(\{x\}) \cdot \pi(\{\theta\})}{q(X^{-1}(\{x\}))} = \frac{\theta^x \cdot (1-\theta)^{1-x} \cdot 0.5}{q(X^{-1}(\{x\}))}$$

for all $\theta \in \Theta$. Since we have that $q(X^{-1}(\{x\})) \geqslant 0$, an action $a^* \in \mathbb{A}$ maximizes the expected utility w.r.t. π_x if, and only if, the inequality

$$q(X^{-1}(\{x\})) \cdot \mathbb{E}_{\pi_x}(u_{a^*}) \geqslant q(X^{-1}(\{x\})) \cdot \mathbb{E}_{\pi_x}(u_a)$$

holds for all $a \in \mathbb{A}$. First, let $x = 0$. We compute

$$q(X^{-1}(\{x\})) \cdot \mathbb{E}_{\pi_0}(u_{a_i}) = \begin{cases} 10.25 & \text{if } i = 1 \\ 9.5 & \text{if } i = 2 \\ 14.5 & \text{if } i = 3 \end{cases}$$

Hence, action a_3 maximizes expected utility w.r.t. π_0. Thus, we receive the same optimal action as we get by evaluating one of optimal decision functions $d(3,2)$ and $d(3,3)$ at $x = 0$.

Finally, set $x = 1$. Again, we compute

$$q(X^{-1}(\{x\})) \cdot \mathbb{E}_{\pi_1}(u_{a_i}) = \begin{cases} 9.75 & \text{if } i = 1 \\ 10.5 & \text{if } i = 2 \\ 10.5 & \text{if } i = 3 \end{cases}$$

Both, action a_2 an action a_3 maximize expected utility w.r.t. π_1. That is, both actions correspond to an optimal decision function evaluated at $x = 1$. $\star$

Remark. 1.) The previous example demonstrated how Theorem 3 can be used to determine optimal data-based decisions without computing optimal decision functions. However, another strength of the theorem turns out to be the following: In a real decision situation with data available, we aren't interested in the Bayes-actions w.r.t. every π_x (for all $x \in \mathcal{X}$). Instead, we are only interested in a Bayes-action w.r.t. π_{x_o}, where $x_o \in \mathcal{X}$ is the data that we have actually observed. Why should we care about optimal decisions in versions of reality that aren't compatible with our information base? So, instead of computing all the decision functions and solving an optimization problem in a vast number of variables, we can choose to only solve one single decision problem (in a relatively small number of variables)!

2.) Note that, in general, the theorem no longer holds when considering criteria different from the Bernoulli/Bayes-criterion. Particularly, a version of the theorem for the Maximin-criterion is not valid. To see that, consider again the situation of Example 3: Here, action a_3 is the unique Maximin-solution of the basic problem. Furthermore, since the Maximin-criterion *completely ignores* the outcome of the random experiment, action a_3 remains Maximin-optimal after having observed any arbitrary realisation $x \in \mathcal{X}$.

However, the decision $\tilde{d}$ function defined by the assignment $\tilde{d}(x) := a_3$ for all $x \in \mathcal{X}$ (that is the decision function $d(3,3)$) is not Maximin-optimal in the data-based decision problem $\mathcal{D}(\mathfrak{A})$. An unique Maximin-optimal decision functions is given by $d(3,2)$. For further details see e.g. [4, Remark 1]. ◦

2 Fundamental principles in Linear Optimization

In this chapter some basic concepts from *Linear Optimization* (or *Linear Programming*) are recalled. As we will see in later chapters (particularly in the Chapters 3 and 5), Linear Optimization turns out to be not only a very helpful (and powerful) tool to determine optimal (randomized) actions in finite decision problems. Additionally, it can be used to simplify proofs in many cases. Particularly, the methods of Linear Optimization often help to simplify proofs for the existence of optimal actions with respect to some optimality criterion under consideration.

The chapter is structured as follows: In the first paragraph two basic types of linear programming problems are defined, namely *Standard-Minimum-Problems* and *Standard-Maximum-Problems*. Afterwards, we recall sufficient criteria for the existence of optimal solutions. It will turn out that the existence of optimal solutions is already guaranteed under relatively weak conditions.

In the second paragraph, the *dual* of a linear programming problem is defined. As we will see, the dual of a linear programming problem is again a linear programming problem. Additionally, some results on the connection between a linear programming problem and its dual are recalled: The well-known *Duality Theorem of Linear Optimization* guarantees that the optimal outcome of a linear programming problem and its dual necessarily coincide.

In the last paragraph, we briefly recall and explain how optimal solutions of a concrete linear programming problem can be determined. It turns out that, instead of looking for optimal solutions on the whole domain of the linear programming problem, it suffices to check optimality on certain *finite* subsets of the domain, namely on the set of *extreme points*. This circumstance is used, for example, by the well known *Simplex Algorithm*, which is one the most common algorithms to solve linear programs in statistical software.

Again, the presentation doesn't claim completeness but focusses on results relevant for the present work. Particularly, many (almost all) of the proofs are left out, since there exists a huge variety on excellent textbooks covering this topic and the technical details of linear programming theory are not within the scope of the present work. However, references are given at the appropriate places whenever the proof of a theorem is left out. For far more in-depth presentations of Linear Optimization see for example [5, Ch.4, Einschub], [28, Ch.2], [25, Ch.2,4 and 5] and [24, Ch.2 and 3]. These are the sources that are used in the following presentation.

2.1 Linear programming problems

The basic problem in linear optimization is the following: Find a vector x that optimizes a *linear* function f, while satisfying a finite number of *linear* constraints. Depending on the concrete problem, the task can differ in maximizing or minimizing the linear function. As every minimization problem turns out to be equivalent to a maximization problem and vice versa, it suffices in many cases to discuss only one of the two types. Most of the results then can be transmitted canonically to the other case. However, especially in duality theory, it is necessary to define both types, since there exist deep (and useful) connections between them. We start with the basic definition. It can be found (similar) in [5, p.129, Definition 4.1].

Definition 11. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *linear*, if there exists $c := (c_1, ..., c_n) \in \mathbb{R}^n$ such that for all $x := (x_1, ..., x_n) \in \mathbb{R}^n$

$$f(x) = \langle c, x \rangle := \sum_{i=1}^{n} c_i \cdot x_i$$

holds. If f is a linear function, $b \in \mathbb{R}^m$ is a real vector and $A \in \mathbb{R}^{m \times n}$ is a real matrix, then the optimization problem

$$f(x) \longrightarrow \max_{x \in \mathbb{R}^n}$$

with constraints

- $x \geqslant 0$
- $A \cdot x \leqslant b$

is called *Standard-Maximum-Problem (SMP).*

By analogy, the optimization problem

$$f(x) \longrightarrow \min_{x \in \mathbb{R}^n}$$

with constraints

- $x \geqslant 0$
- $A \cdot x \geqslant b$

is called *Standard-Minimum-Problem (SMIP).*

In both cases, the function f is said to be the *objective function.*
The sets

$$Z(A,b)^+ := \{x \in \mathbb{R}^n : x \geqslant 0 \wedge A \cdot x \leqslant b\}$$

or

$$Z(A,b)^- := \{x \in \mathbb{R}^n : x \geqslant 0 \wedge A \cdot x \geqslant b\}$$

are named the *sets of admissible solutions* of the SMP or the SMIP.
The sets

$$O(A,b,c)^+ := \{x^* \in Z(A,b) : f(x^*) \geqslant f(x) \text{ for all } x \in Z(A,b)\}$$

or

$$O(A,b,c)^- := \{x^* \in Z(A,b) : f(x^*) \leqslant f(x) \text{ for all } x \in Z(A,b)\}$$

are named the *sets of optimal solutions* of the SMP or the SMIP. $\bigtriangledown$

Remark. 1.) In the above definition, we used the following convention/notation: For two vectors

$$x := (x_1, ..., x_n), y := (y_1, ..., y_n) \in \mathbb{R}^n$$

the symbols $\leqslant$ and $\geqslant$ are to be understood *component-wise*, i.e. $x \leqslant y$ or $x \geqslant y$ holds if, and only only if,

$$x_i \leqslant y_i \text{ for all } i = 1, \ldots, n \qquad \text{or} \qquad x_i \geqslant y_i \text{ for all } i = 1, \ldots, n$$

2.) In the following, a SMP or SMIP sometimes will be written in the form

$$\max \{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\}$$

or

$$\min \{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \geqslant b\}$$

Additionally, we will use the notation

$$f(x) := (c_1, \ldots, c_n) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

for the objective function, whenever this helps to simplify (or clarify) the presentation. Particularly, this will be the case in the later Chapters 3 and 5.

3.) Every SMP can be transformed into a SMIP and vice versa, since we have

$$\max\{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\} = \min\{\langle -c, x\rangle : x \geqslant 0 \wedge -A \cdot x \geqslant b\} \tag{33}$$

As already mentioned in the introduction to the chapter, the above identity now allows us to restrict on the discussion of SMPs only for many purposes (see for example in [5, p.130, Remark 4.2] for further details). ∘

Before turning to the question how optimal solutions can be determined for a given linear optimization problem, we want to give some sufficient criteria for a linear problem to be solvable at all.

First, we observe the following: If the set $Z(A,b)^+$ is empty (that is, if there doesn't exist an admissible solution at all), then the constraints of the corresponding linear programming problem are inconsistent (for example, consider the constraints demand both $x_1 \leqslant 8$ and $x_1 \geqslant 15$). Since every optimal solution is also admissible by definition, of course, in such a case there exists no optimal solution.

So, when do optimal solutions exist? It turns out that there always is an optimal solution, if the set $Z(A,b)^+$ is non-empty and bounded (with the *standard metric* of the space $\mathbb{R}^n$). This is the statement of the following theorem, which can for example be found in [5, p.130, Proposition 4.3, Part 1].

Theorem 4. Let $\max\{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\}$ be a SMP such that the set $Z(A,b)^+$ is non-empty and bounded. Then:

$$Z(A,b) \neq \emptyset \;\Leftrightarrow\; O(A,b,c) \neq \emptyset$$

Sketch of the proof. $\Leftarrow$: Trivial: $O(A,b,c) \subset Z(A,b)$ follows directly from the definition of the set of optimal solutions.
$\Rightarrow$: The SMP can be reformulated as the task of finding a vector x maximizing the function

$$f \restriction Z(A,b)^+ : Z(A,b)^+ \to \mathbb{R}\ ,\ x \mapsto f(x)$$

Now, it can be shown that the set $Z(A,b)^+$ is closed. According to the assumption, the set $Z(A,b)^+$ is also closed. According to the theorem of *Heine-Borel* (see e.g. [17, p.32]), this is equivalent with being compact for a subset of the space $\mathbb{R}^n$. Hence, $Z(A,b)^+$ is compact.
We now observe that the function $f \restriction Z(A,b)^+$ is linear as the restriction of the function f to the set $Z(A,b)^+$. But any linear function is continuous.

Thus, $f \restriction Z(A,b)^+$ is continuous.
According to the *Theorem of the Maximum* (see for example [17, p.34]), continuous functions attain their maximum on non-empty compact sets. Thus, there exists $x^* \in Z(A,b)^+$ such that

$$f(x^*) = f \restriction Z(A,b)^+(x^*) \geqslant f \restriction Z(A,b)^+(x) = f(x)$$

for all $x \in Z(A,b)^+$. This completes the proof. □

Remark. By analogy, one could prove the theorem for the case of a SMIP: If the set of admissible solutions of a given SMIP is non-empty and bounded, then the SMIP has an optimal solution. ∘

Theorem 4 gives us a sufficient criterion for the resolvability of a linear programming problem. However, to apply the theorem, the set of optimal solutions of the corresponding linear program has to be bounded. This is a serious restriction in many cases. The next theorem avoids this assumption. It can be found for example in [27, p.56].

Theorem 5. Let $\max\{\langle c,x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\}$ be a SMP such that

$$\delta := \sup\{\langle c,x\rangle : x \in Z(A,b)^+\} < \infty$$

Then there exists $x^* \in Z(A,b)^+$ such that $\langle c,x^*\rangle = \delta$.

Proof. See e.g. [27, p.56]. □

Remark. By analogy, one can prove the theorem for the case of a SMIP:
If $\min\{\langle c,x\rangle : x \geqslant 0 \wedge A \cdot x \geqslant b\}$ is a SMIP such that

$$\delta := \inf\{\langle c,x\rangle : x \in Z(A,b)^+\} > -\infty$$

holds, then the SMIP admits an optimal solution. Clearly, this directly follows by applying Theorem 5 on the SMP equivalent to the above SMIP (see Definition 11, Remark 3). ∘

Both, Theorem 4 and Theorem 5, guarantee the existence of optimal solutions of linear programming problems under relatively weak conditions. However, both of the proofs a pure proofs of existence. Specifically, the proofs do not contain a to determine optimal solutions for concrete linear optimization problems. Suitable methods to determine optimal solutions of linear programming problems are briefly recalled in Paragraph 2.3. First, we want to recall some results from *Duality Theory*.

2.2 Duality theory

For any linear programming problem one can define its *dual* linear programming problem. In this context, the origin problem will be called the *primal* linear programming problem, whenever we need to distinguish between the two. If the primal problem is a SMP, then its dual problem will be a SMIP and vice versa. This is the statement of the following definition. It can be found (similar) e.g. in [5, p.131, Definition 4.4].

Definition 12. Let $\max\{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\}$ be a SMP. Then the SMIP

$$\min\left\{\langle b, y\rangle : y \geqslant 0 \wedge A^T \cdot y \geqslant c\right\}$$

is called the *dual* (linear programming problem) of the SMP.

By analogy, if $\min\{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \geqslant b\}$ is a SMIP, then the SMP defined by

$$\max\left\{\langle b, y\rangle : y \geqslant 0 \wedge A^T \cdot y \leqslant c\right\}$$

is the dual of the SMIP. ▽

As to expect, there exist some deep connections between a linear programming problem and its dual problem: If the primal problem admits an optimal solution, so does its dual. Even more is the case: The optimal outcomes of primal and dual programming problem (that is, the value that is returned when evaluating the objective function at the optimal solution) coincide. This is the statement of the following theorem, which is often referred to as *Duality Theorem of Linear Optimization*. In this form, it is taken from [5, p.132, Proposition 4.5, Part 1 and Part 3].

Theorem 6. Let

$$\max\{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\} \tag{34}$$

denote a Standard-Maximum-Problem and let

$$\min\left\{\langle b, y\rangle : y \geqslant 0 \wedge A^T \cdot y \geqslant c\right\} \tag{35}$$

denote the corresponding dual Standard-Minimum-Problem. The following holds:

1. The dual linear programming problem of (35) is again the linear programming problem (34).
2. There exists an optimal solution x^* for problem (34) if, and only if, there exists an optimal solution y^* for problem (35).

3. If x^* and y^* are optimal solutions for the problems (34) and (35) respectively, then we have $\langle c, x^* \rangle = \langle b, y^* \rangle$. That is, the optimal outcomes of both problems coincide.

Proof. See e.g. in [28, p.425, Theorem 14.3.1]. □

Theorem 6 will prove very important at several points of the later Chapters 3 and 5 (for example in the proof of Theorem 20). Often, if one wants to prove the existence of an optimal solution of a linear programming problem, it is easier to prove the existence of an optimal solution of its dual. According to the above Theorem 6, the two approaches are equivalent. Next, we want to face the question how optimal solutions can be determined in a concrete linear optimization problem.

2.3 Convex sets and the idea of the Simplex-Algorithm

In the first paragraph of this chapter, we gave the definition of a linear optimization problem and recalled some results that guarantee the existence of optimal solutions for such a problem under certain conditions. Next, we want to very briefly recall some results on how optimal solutions for linear optimization problems can be determined in concrete situations.

Here, the main idea is the following: The set of admissible solutions of a linear optimizations problem turns out to be a *convex polyhedron*. That is, solving a linear optimization problem coincides with optimizing a linear function over a convex polyhedron. However, due to some special properties of convex polyhedra, maxima and minima of such linear functions are always attained on special subsets of the underlying polyhedra, namely on the set of *extreme points*. Geometrically, each extreme point of a convex polyhedron corresponds to one of its *corner points* (and vice versa). As there always exist only finitely many corners (and therefore extreme points), we then only have to evaluate the objective function at finitely many points. In this way, we find an optimal solution of the optimization problem by simply taking the maximum over all this evaluations.

To put this on sound theoretical ground, we need to recall some basic concepts from the theory of convex sets. We start with the central definition. It can e.g. be found in [28, p.34, Definition 3.3.1].

Definition 13. Let $M \subset \mathbb{R}^n$. Then, M is called *convex* if, and only if, the following holds for all $\lambda \in [0, 1]$ and $x, y \in M$:

$$\lambda \cdot x + (1 - \lambda) \cdot y \in M$$

For another real subset $Q \subset \mathbb{R}^n$, we define

$$K(Q) := \{M : Q \subset M \ \wedge \ M \text{ convex } \}$$

Then, the set

$$\text{conv}(Q) := \bigcap_{M \in K(Q)} M$$

is called the *convex hull* of the set Q. ▽

As already mentioned, for our purposes another concept in the context of convex sets proves very important: The concept of an *extreme point*. Informally, an extreme point a is an element of a convex set K that cannot be written as a *convex combination* of two other points x and y from the set K (where a convex combination of x and y is a term of the form $\lambda \cdot x + (1 - \lambda) \cdot y$, where $\lambda \in [0, 1]$). This motivates the following definition. It can be found (similar) in [28, p.426, Definition 14.4.1].

Definition 14. Let $M \subset \mathbb{R}^n$ be a convex set. An element $e \in M$ is said to be an *extreme point* in M if, and only if, for all $x, y \in M$ and $\lambda \in (0, 1)$ the following implication holds:

$$\lambda \cdot x + (1 - \lambda) \cdot y = e \ \Rightarrow \ x = y = a$$

The set of all extreme points of M is denoted by $\mathcal{E}(M)$. ▽

Next, we want to identify the set of admissible solutions of a linear programming problem as a special convex set, namely a so called *convex polyhedron*. One of the special properties of such a convex polyhedron turns out to be that it possesses only finitely many extreme points. We start with the definition, which can for example be found in [38, p.17, Definition 2.7].

Definition 15. Any non-empty set $P \subset \mathbb{R}^n$ of the form

$$P = \{x \in \mathbb{R}^n : A \cdot x \leqslant b\}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, is called a *convex polyhedron*. ▽

Remark. 1.) If the set of admissible solutions of a linear optimization problem is non-empty, then it defines a convex polyhedron. This is obvious, since the condition of non-negativity of the variables can easily be written in matrix form.

2.) As the name suggests, any convex polyhedron defines a convex set in the sense of Definition 13. This is not hard to prove. A proof can for example be found in [38,

p.17, Lemma 2.8, Part 3]. ∘

Now, we recall two fundamentally important results of linear optimization theory. Together, they show up a way of resolving linear programming problems without checking the whole (infinite) set of admissible solutions. We begin with a theorem that guarantees that a convex polyhedron always possesses only finitely many extreme points.

Theorem 7. Let P be a convex polyhedron of the form defined in Definition 15. Then, we have $|\mathcal{E}(P)| < \infty$. That is, any convex polyhedron has only finitely many extreme points.

Proof. See for example [3, Proposition 1], or [46, Ch. 2]. □

Next, we recall *the* central result of linear programming theory, namely the *Fundamental Theorem of Linear Optimization*. Roughly spoken, it states the following: For any (resolvable) linear programming problem, the set of optimal solutions contains an extreme point of the set of admissible solutions. In other words: The existence of an optimal solution necessarily implies the existence of an optimal extreme point. Thus, if we are only interested in finding an arbitrary optimal solution of the optimization problem, it suffices to take only the set of extreme points into account. The following theorem can be found (in a similar form) for example in [38, p.34, Theorem 2.21].

Theorem 8. Let $\max\{\langle c, x\rangle : x \geqslant 0 \wedge A \cdot x \leqslant b\}$ be a SMP such that $O(A, b, c)^+$ is non-empty. Then, there exists $x_e \in \mathcal{E}(Z(A, b)^+)$ such that $x_e \in O(A, b, c)^+$. That is, the set of optimal solutions contains at least one extreme point.

Proof. See for example [38, p.34, Theorem 2.21]. □

Together, the Theorems 7 and 8 imply the following: According to Definition 15, Remark 2, the set of admissible solutions of any consistent linear programming problem defines a convex polyhedron. Thus, according to Theorem 7, the set of admissible solutions possesses only finitely many extreme points. However, according to Theorem 8, there exists at least one optimal extreme point. That is, at least one of the (finitely many!) extreme points of the set of admissible solutions is an optimal solution.

In other words: If we want to determine an optimal solution of a linear programming problem, it suffices to check the optimality on a finite set! More precisely, if we evaluate the objective function for all extreme points of the set of admissible

solutions and, afterwards, take the maximum over all the evaluations, the extreme points generating the maximum values will be optimal solutions for the whole optimization problem. That's an enormous reduction of complexity!

Very roughly spoken, this is the main idea of the famous *Simplex-Algorithm* (and of its modifications). This algorithm to determine optimal solutions of linear optimization problems goes back to the mathematician *George Dantzig* (1914-2005) and was developed in 1947. In its basic form it can be found in [9, Ch. 5.1, p. 94-101]. Nowadays, there exist many modifications of the origin algorithm.

The idea underlying almost all of them can, *very* briefly (clearly, the techniques underlying the concrete algorithms are highly *non-trivial* and demand a large amount of *Linear Algebra*), be summarized as follows: We start with evaluating the objective function for an arbitrary extreme point of the set of admissible solutions.

Afterwards, we keep evaluating the objective function for other extreme points in a *systematic* way. The term 'systematic' is meant in the following sense: All of the algorithms try to avoid an evaluation of the objective function for all extreme points, because the number of extreme points might, though finite, become very large.

Instead, the idea is to always choose on of the extreme points that generate a higher outcome than the one chosen in the step before. In this way, it is often possible to determine an optimal extreme point without having to compute all of them (determining extreme points has high computational costs). An in-depth discussion of the Simplex-Algorithm lies not within the scope of the present work. In detail, it is described for example in [9, Ch. 5.1, p. 94-101].

2.4 The role of Linear Optimization in this thesis

In this last paragraph of the present chapter, we want to point out the fundamentally important role of the theory of linear optimization for the present thesis. Note that this paragraph should not be understood as synonymous to the considerations in Paragraph 3.1. Instead, the consideration in this paragraph have a preparing character.

In order to do so, note the following: If we want to determine optimal actions with respect to some criterion Φ (in the sense of Definition 7), naturally two (highly related) questions matter:

- Do there exist optimal actions with respect to the criterion Φ?
- And: If the answer is yes, then how can we determine Φ-optimal actions?

In many cases, the theory linear optimization allows to answer both questions simultaneously: If we succeed to reformulate the task of optimizing the criterion Φ (remember that any criterion defines a real-valued function) as a linear optimization task (in the sense of Definition 11), we always can apply the general results on the existence of optimal solutions of linear programming problems (see Theorems 4 and 5) to Decision Theory! More precisely, we can derive results concerning the existence of optimal decision by simply checking the boundedness of the set of admissible solutions of the corresponding linear programming problem.

Furthermore, embedding our problem in the theoretical framework of linear optimization allows us to determine optimal decisions computationally: Algorithms for computing optimal solutions of linear programming problems are implemented in almost every statistical standard software (e.g. R).

Finally, for our purposes, the role of linear optimization can be summarized by the following two aspects:

- **Aspect 1:** The theory of linear optimization is a very powerful tool for generating and proving theoretical statements. Particularly, at several points throughout the present work it will be used to prove the existence of optimal actions with respect to some criterion. Additionally, the usage of duality theory often allows to make up (and prove) connections between seemingly different questions relevant to decision theory (for example, see the proof of Theorem 12: Here, a connection between the Bayes-criterion and the Hodges & Lehmann-criterion is proven by using duality theory).

- **Aspect 2:** Linear optimization offers *the* theoretical framework for an effective computational treatment of (finite) decision theory. As algorithms for solving linear programming problems are standard in almost every mathematical or statistical software, optimal decisions can be computed for example in R. How this can be done for concrete decision situation will be treated at several points in the Chapters 3 and 5.

3 Linear Optimization and Decision Theory

In this chapter we demonstrate, how the theory of linear optimization treated in Chapter 2 can be used to determine optimal (randomized) actions in *finite* decision problems. It turns out that the task of determining optimal decisions (with respect to all criteria discussed in the first chapter) can be reformulated as a task of solving suitable linear programming problems (in a more or less direct way).

This reformulation then not only proves the existence of optimal actions according to the corresponding criterion by applying the results on linear optimization recalled in Chapter 2. Moreover, it can be used to compute such optimal actions by using methods implemented in standard statistical software (e.g. R).

Finally, especially in the case of the Hodges & Lehmann-criterion (see Theorem 12), we apply the duality theory recalled in Paragraph 2.2 to learn more about the characteristics of optimal actions. We mainly follow [3], [2, § 2.4] and [41]. Particularly, the techniques used to reformulate the criteria as linear programming problems are inspired and adopted by this sources.

3.1 Decision making meets Linear Optimization: A motivation

In the whole chapter, let $\mathfrak{A}$ denote a *finite* CDP and $G(\mathfrak{A})$ its mixed extension. Again, we use the power set of $\mathbb{A}$ as σ- field. Then, according to equation (5), every action $a \in \mathbb{A}$ can be identified with the randomized action $\delta_a \in G(\mathbb{A})$. Thus, one can always work with decision criteria (in the sense of Definition 7)

$$\Phi^{G(\mathfrak{A})} : G(\mathbb{A}) \to \mathbb{R}$$

defined on the set $G(\mathbb{A})$ of all randomized actions. If a pure action $a \in \mathbb{A}$ turns out to have a higher utility than all randomized actions available, then the optimization of $\Phi^{G(\mathfrak{A})}$ on the set $G(\mathbb{A})$ will have δ_a as an optimal solution.

Now recall that, for a criterion $\Phi^{G(\mathfrak{A})}$ under consideration, a randomized action $p^* \in G(\mathbb{A})$ is optimal, if, and only if, the inequality

$$\Phi^{G(\mathfrak{A})}(p^*) \geqslant \Phi^{G(\mathfrak{A})}(p)$$

holds for all $p \in G(\mathbb{A})$. Thus, determining a $\Phi^{G(\mathfrak{A})}$-optimal action coincides with computing the expression

$$\operatorname{argmax} \left\{ \Phi^{G(\mathfrak{A})}(p) : p \in G(\mathbb{A}) \right\}$$

To make up a connection to the structure of a linear programming problem, some preparation work has to be done: Since the set of action $\mathbb{A}$ is assumed to be finite, every randomized action $p \in G(\mathbb{A})$ can uniquely be identified with a point of the $(n-1)$-dimensional *simplex*

$$\Delta_n := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1 \ \wedge \ x_i \geqslant 0 \ \forall i = 1, \ldots, n \right\}$$

This is possible, since the map

$$t : G(\mathbb{A}) \to \Delta_n \ , \ p \mapsto (p(\{a_1\}), \ldots, p(\{a_n\})) =: (p_1, \ldots, p_n)$$

is bijective (this map will repeatedly be referred to throughout the whole work).

This implies that the identity

$$\max \left\{ \Phi^{G(\mathfrak{A})}(p) : p \in G(\mathbb{A}) \right\} = \max \left\{ \Phi^{G(\mathfrak{A})}(t^{-1}(p)) : p \in \Delta_n \right\}$$

holds, whenever one of the above maxima exists.

Therefore, determining a $\Phi^{G(\mathfrak{A})}$-optimal action $p^* \in G(\mathbb{A})$ is equivalent to solving the optimization problem

$$\Phi^{G(\mathfrak{A})}(t^{-1}(p)) \longrightarrow \max_{(p_1, \ldots, p_n)} \tag{36}$$

with constraints

- $p_i \geqslant 0$ for all $i = 1, \ldots, n$
- $\sum_{i=1}^{n} p_i = 1$

Particularly, this implies that there exists an optimal randomized action with respect to the criterion $\Phi^{G(\mathfrak{A})}$ if, and only if, the above optimization problem (36) admits an optimal solution.

Now, note that the second constraint of problem (36) is equivalent to the inequality

$$\begin{pmatrix} 1 & \ldots & 1 \\ -1 & \ldots & -1 \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Hence, problem (36) gives us an optimization problem with *linear* constraints (in the sense of Definition 11). Additionally, if the function $\Phi \circ t^{-1}$ is linear in p, then

determining a $\Phi^{G(\mathfrak{A})}$-optimal action is equivalent to solving a linear programming problem where Δ_n defines the set of admissible solutions.

Since Δ_n obviously is both non-empty and bounded, Theorem 4 then implies that the above SMP (36) has an optimal solution $p^* \in \Delta_N$. Hence, we get an optimal randomized action by re-transforming the point p^* into the corresponding probability measure $t^{-1}(p^*)$.

The approach just described seems to be very restrictive, since it helps us to determine optimal randomized action only for the linear case (both the constraints and the criterion itself have to be linear in the sense of Definition 11).

However, in many cases also seemingly non-linear optimization tasks can be reformulated as suitable linear programming problems. Which linear programming problem needs to be solved clearly depends on the criterion to be optimized. In the following paragraphs, common proposals for the criteria discussed in Chapter 1 are recalled.

3.2 Bernoulli/Bayes-criterion as a linear program

In this paragraph two different approaches for determining Bayes-optimal actions using linear optimization are recalled. Finally, it will turn out that the two approaches are dual to each other (in the sense of Definition 12). Implicitly, the consideration within this paragraph will give us an alternative proof of Theorem 2, that is an alternative proof of the needlessness of randomization when defining optimality in terms of the Bayes-criterion. The paragraph is orientated on [41, § 5.1].

In the whole paragraph, let π denote a probability measure on $(\Theta, \mathcal{P}(\Theta))$, where Θ again denotes the set of states of a finite CDP. Then, according to the equations (12) and (13), the Bernoulli/Bayes-criterion

$$\Phi_{B(\pi)}^{G(\mathfrak{A})} : G(\mathbb{A}) \to \mathbb{R}$$

with respect to π is given by

$$\Phi_{B(\pi)}^{G(\mathfrak{A})}(p) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} u(a_i, \theta_j) \cdot \pi(\{\theta_j\}) \right) \cdot p(\{a_i\}) \tag{37}$$

for all $p \in G(\mathbb{A})$. Here, we used again the exchangeability of the two sums in the finite Bayes-criterion (see for example the proof of Theorem 2).

Thus, for all $p \in G(\mathbb{A})$, we can rewrite the criterion by

$$\Phi^{G(\mathfrak{A})}_{B(\pi)}(p) = (c^\pi_1, \ldots, c^\pi_n) \cdot \begin{pmatrix} p(\{a_1\}) \\ \vdots \\ p(\{a_n\}) \end{pmatrix} \tag{38}$$

where

$$c^\pi_i := \sum_{j=1}^{m} u(a_i, \theta_j) \cdot \pi(\{\theta_j\})$$

for all $i = 1, \ldots, n$.

With the function $t : G(\mathbb{A}) \to \Delta_n$ being defined just like in the last paragraph, the map $\Phi \circ t^{-1}$ clearly is linear in p (on the set Δ_n). Thus, according to the considerations from the last paragraph, determining a $B(\pi)$-optimal action coincides with solving the linear program

$$\Phi^{G(\mathfrak{A})}_{B(\pi)}(t^{-1}(p)) = (c^\pi_1, \ldots, c^\pi_n) \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_p \tag{39}$$

with constraints

- $p_i \geqslant 0$ for all $i = 1, \ldots, n$
- $\begin{pmatrix} 1 & \ldots & 1 \\ -1 & \ldots & -1 \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Thus, we succeeded in reformulating the finite Bayes-criterion as a linear optimization problem. Since the set Δ_n is non-empty and bounded, according to Theorem 4, we have proven that there exists a Bayes-action with respect to π. As π was chosen arbitrarily, this implies the existence of Bayes-actions in finite decision problems for arbitrary prior distributions.

However, the approach described above completely ignores the additional information given by Theorem 2: In the case of the finite Bayes-criterion it suffices to take only pure actions into account. Instead, the linear programming problem (39) takes all the randomized actions into consideration and, therefore, seems to perform avoidable computational efforts.

Therefore, let us now to describe a different approach taking the additional knowledge given by Theorem 2 into account. Since we know that there exists a pure

action $a^* \in \mathbb{A}$, which is $B(\pi)$-optimal for the decision problem $G(\mathfrak{A})$ (using the identification $a^* \approx \delta_{a^*}$), the optimal outcome of the optimization problem

$$\sum_{j=1}^{m} u(a, \theta_j) \cdot \pi(\{\theta_j\}) \longrightarrow \max_{a \in \mathbb{A}} \tag{40}$$

equals the optimal outcome of (39). However, problem (40) is not a linear program, since the set $\mathbb{A}$ is not convex (as a subset of the real numbers). So, how can we reformulate (40) as a linear optimization problem?

Again, we need some preparing considerations: If $Q \subset \mathbb{R}$ is a finite real subset, then the maximum of the set Q coincides with the *smallest* real number, which is greater or equal than any number $z \in Q$, i.e.

$$\max(Q) = \min\{M \in \mathbb{R} : x \geqslant z \ \ \forall\ z \in Q\}$$

Now, define

$$Q := \left\{ \sum_{j=1}^{m} u(a, \theta_j) \cdot \pi(\{\theta_j\}) : a \in \mathbb{A} \right\}$$

Then, for our problem, we arrive at

$$\max(Q) = \min \left\{ M \in \mathbb{R} : M \geqslant \sum_{j=1}^{m} u(a, \theta_j) \cdot \pi(\{\theta_j\}) \ \ \forall\ a \in \mathbb{A} \right\}$$

Thus, the optimal outcome of (40) equals the optimal outcome of the optimization problem

$$M \longrightarrow \min_{M \in \mathbb{R}} \tag{41}$$

with the linear constraint

- $$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \cdot M \geqslant \begin{pmatrix} \sum_{j=1}^{m} u(a_1, \theta_j) \cdot \pi(\{\theta_j\}) \\ \vdots \\ \sum_{j=1}^{m} u(a_n, \theta_j) \cdot \pi(\{\theta_j\}) \end{pmatrix} = \begin{pmatrix} c_1^{\pi} \\ \vdots \\ c_n^{\pi} \end{pmatrix}$$

Note that (41) still isn't a linear program in the sense of Definition 11, since the constraints do not restrict the variable M to be bounded from below by 0.

Therefore, another general consideration has to be done: Any real number $M \in \mathbb{R}$ is representable as the difference of two positive real numbers $w_1, w_2 \in \mathbb{R}_0^+$. Thus,

for any set $Q \subset \mathbb{R}$ the following holds

$$\inf Q = \inf\{w_1 - w_2 : w_1, w_2 \in \mathbb{R}_0^+ \wedge w_1 - w_2 \in Q\}$$

Hence, the optimal outcome of problems (40) (and (41)) equals the optimal outcome of the linear programming problem

$$w_1 - w_2 \longrightarrow \min_{(w_1,w_2)\in\mathbb{R}^2} \tag{42}$$

- $(w_1, w_2) \geqslant 0$
- $\begin{pmatrix} 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \geqslant \begin{pmatrix} c_1^\pi \\ \vdots \\ c_n^\pi \end{pmatrix}$

Finally, we succeeded in reformulating (40) as a linear opimization problem in the sense of Definition 11.

As already mentioned in the beginning of the paragraph, the resulting SMIP (42) turns out to be exactly the dual programming problem of the SMP (39) (in the sense of Definition 12) and, therefore, admits the same optimal outcome.

This allows two different ways of interpretation: On the on hand, the above considerations could have been shortened a lot by just applying the duality theorem (that is Theorem 6). On the other hand, the duality theorem gives us the equivalence of the linear programs (42) and (39), without using the additional knowledge we get from Theorem 2. Thus, implicitly, the above considerations give an alternative proof for Theorem 2.

3.3 Least favourable prior distributions

In the last paragraph, the connection between determining Bayes-optimal actions and linear optimization was discussed. This gives us the possibility to prove a Theorem on the existence of so called *least favourable prior distributions*. At this point, the meaning of such special prior distributions might seem rather unclear: Why should there be more than one prior distribution? And: Why should I be interested in the least favourable one among them? The importance of this Theorem then will show up when considering more general descriptions of uncertainty in Chapter 5. More precisely, using the concept of least favourable prior distributions, it is possible to make up a connection between CDPs with precise and imprecise information

available. Therefore, the following paragraph should be understood as a lookahead on Chapter 5. Hereby, we mainly refer to the considerations [2, § 2.4.5] and [35, § 2.4].

Consider the following situation: Let $\mathcal{M}$ be a *set* of prior distributions on the set of states Θ of a finite CDP $\mathfrak{A}$ (for example a so called credal set, see Definition 16). Suppose, according to your information base, all of the measures contained in $\mathcal{M}$ are equally plausible to be the *true* measure describing the uncertainty between the elements of Θ. That is, the uncertainty underlying our CDP is of type I, but our information doesn't suffice to determine a unique prior distribution.

We want to answer the following question: Does there always exist a *least favourable prior*, i.e. an element of $\mathcal{M}$ that minimizes the Bayes-utility under all priors in $\mathcal{M}$? And, if there is such an element of $\mathcal{M}$, how can it be determined?

It will turn out that, at least for a linearly defined set $\mathcal{M}$, least favourable prior distributions always exist and can be determined by solving suitable linear programming problems. As we will see later, the assumption of a *linearly* defined set is not a too serious restriction (see Theorem 15).

So, let $\mathfrak{A}$ be any finite CDP. Then, for all $i = 1, ..., r$ let

- $(\underline{b}_i, \overline{b}_i) \in \mathbb{R}^2$ such that $\underline{b}_i \leqslant \overline{b}_i$
- $f_i : \Theta \to \mathbb{R}$

Now, define the set

$$\mathcal{M} := \Big\{\pi : \pi \text{ is pm on } (\Theta, \mathcal{P}(\Theta)) \wedge \underbrace{\underline{b}_i \leqslant \mathbb{E}_\pi(f_i) \leqslant \overline{b}_i}_{=:R(i)} \ \forall i = 1, ..., r\Big\} \tag{43}$$

of all probability measures (pm) satisfying the constraints $R(i)$ for all $i = 1, ..., r$.

As shown in the last paragraph, for every measure $\pi \in \mathcal{M}$ there exists a Bayes-action $p^*_\pi \in G(\mathbb{A})$ with respect to π. Specifically, we have

$$M(\pi) := \Phi^{G(\mathfrak{A})}_{B(\pi)}(p^*_\pi) \geqslant \Phi^{G(\mathfrak{A})}_{B(\pi)}(p)$$

for all $p \in G(\mathbb{A})$. That is, the expression $M(\pi)$ equals the expected utility of an Bayes-action with respect to π. Now, we show that the lowest possible Bayes-utility on the set $\mathcal{M}$ is indeed attained.

Theorem 9. There exists $\pi^* \in \mathcal{M}$ such that

$$M(\pi^*) \leqslant M(\pi) \tag{44}$$

for all $\pi \in \mathcal{M}$. π^* is called *least favourable prior* with respect to $\mathcal{M}$.

Proof. First note that the existence of a prior $\pi^* \in \mathcal{M}$ satisfying condition (44) is equivalent for the optimization problem

$$M(\pi) \longrightarrow \min_{\pi \in \mathcal{M}} \tag{45}$$

to have an optimal solution. Now, let $\pi \in \mathcal{M}$ be arbitrary. For the sake of readability, we define the notation $a' := -a$ for any number $a \in \mathbb{R}$.

According to the considerations from the last paragraph, $M(\pi)$ equals the optimal outcome of the SMP

$$\sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{m} u(a_i, \theta_j) \cdot \pi(\theta_j) \right)}_{=:c_i^\pi} \cdot p_i \;:=\; (c_1^\pi, \ldots, c_n^\pi) \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_p \tag{46}$$

with constraints

- $p_i \geqslant 0$ for all $i = 1, \ldots, n$
- $\begin{pmatrix} 1 & \ldots & 1 \\ 1' & \ldots & 1' \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Thus, by duality, $M(\pi)$ equals the optimal outcome of the dual program

$$(1, -1) \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \longrightarrow \min_w$$

with constraints

- $w_k \geqslant 0$ for $k = 1, 2$ $(:= R_1^\pi)$
- $w_1 - w_2 \geqslant c_i^\pi$ for all $i = 1, ..., n$ $(:= R_2^\pi)$

Hence, the optimization problem (45) has an optimal solution if, and and only if, the optimization problem

$$\min \{w_1 - w_2 : R_1^\pi \wedge R_2^\pi\} \longrightarrow \min_{\pi \in \mathcal{M}} \tag{47}$$

has an optimal solution. We have the following identity

$$\begin{aligned}\min\{M(\pi) : \pi \in \mathcal{M}\} &= \min\{\min\{w_1 - w_2 : R_1^\pi \wedge R_2^\pi\} : \pi \in \mathcal{M}\} \\ &= \min\{w_1 - w_2 : R_1^\pi \wedge R_2^\pi \wedge \pi \in \mathcal{M}\}\end{aligned}$$

Thus, the optimization problem (47) has an optimal solution if, and only if, there exists an optimal solution for the following optimization problem:

$$(1, -1, \underbrace{0, ..., 0}_{m \text{ times}}) \cdot \begin{pmatrix} w_1 \\ w_2 \\ \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \longrightarrow \min_{w,\pi} \tag{48}$$

with constraints

- R_1^π, R_2^π
- $\pi \in \Pi_m$

where, again, we used the one-to-one correspondence of the set $\mathcal{M}$ and the set

$$\Pi_m := \Big\{(\pi(\{\theta_1\}), ..., \pi(\{\theta_m\})) : \pi \in \mathcal{M}\Big\}$$

induced by the bijection

$$b : \mathcal{M} \to \Pi_m \ , \ \pi \mapsto (\pi(\{\theta_1\}), ..., \pi(\{\theta_m\})) =: (\pi_1, ..., \pi_m)$$

This gives us an optimization problem with a linear objective function (clearly, the function to be optimized in problem (9) is linear in the variable $(w_1, \ldots, p_n)$ in the sense of Definition 11). But is the optimization problem (48) linear in the sense of Definition 11? In other words: Can we rewrite its constraints by a finite system of linear inequalities?

To see the linearity of the constraints of the optimization problem (48) (in the sense of Definition 11), note that they can equivalently be written as the following system of linear inequalities

- $w_1, w_2, \pi_1, ..., \pi_m \geqslant 0$

$$\bullet \underbrace{\begin{pmatrix} 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 1' & \dots & 1' \\ 0 & 0 & f_{11} & \dots & f_{1m} \\ 0 & 0 & f'_{11} & \dots & f'_{1m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & f_{r1} & \dots & f_{rm} \\ 0 & 0 & f'_{r1} & \dots & f'_{rm} \\ 1 & 1' & u'_{11} & \dots & u'_{1m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 1' & u'_{n1} & \dots & u'_{nm} \end{pmatrix}}_{=:H} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \geqslant \underbrace{\begin{pmatrix} 1 \\ 1' \\ \underline{b}_1 \\ \overline{b}'_1 \\ \vdots \\ \underline{b}_r \\ \overline{b}'_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:b}$$

where the following notations are used

- $f_{ij} := f_i(\theta_j)$ for all $i = 1, ..., r$ and $j = 1, ..., m$
- $u_{ij} := u(a_i, \theta_j)$ for all $i = 1, ..., n$ and $j = 1, ..., m$

Next, we show that the above SMIP (48) admits an optimal solution. The following holds:

$$\inf\{w_1 - w_2 : R_1^\pi \wedge R_2^\pi \wedge \pi \in \Pi_m\} \geqslant \inf\{c_i^\pi : i = 1, ..., n \wedge \pi \in \Pi_m\} \geqslant \min\{u_{ij}\} > -\infty$$

Thus, according to Theorem 5, the SMIP (48) admits an optimal solution, i.e. there exists

$$u^* := (w_1^*, w_2^*, \underbrace{\pi_1^*, \dots, \pi_m^*}_{=:\pi^* \in \Pi_m}) \in Z(H, b)^-$$

such that

$$(1, -1, \underbrace{0, ..., 0}_{m \text{ times}}) \cdot u^* \leqslant (1, -1, \underbrace{0, ..., 0}_{m \text{ times}}) \cdot u$$

for all $u \in Z(H, b)^-$. Thus, we proved the existence of a least favourable prior.

Finally, we compute

$$\begin{aligned} \min\{M(\pi) : \pi \in \mathcal{M}\} &= \min\{w_1 - w_2 : R_1^\pi \wedge R_2^\pi \wedge \pi \in \mathcal{M}\} \\ &= \min\left\{w_1 - w_2 : R_1^{\pi^*} \wedge R_2^{\pi^*}\right\} \\ &= \max\left\{\Phi^{G(\mathfrak{A})}_{B(b^{-1}(\pi^*))}(p) : p \in G(\mathbb{A})\right\} \\ &= M(b^{-1}(\pi^*)) \end{aligned}$$

Hence, a least favourable prior distribution in $\mathcal{M}$ is given by $b^{-1}(\pi^*)$. This completes the proof. □

As an immediate consequence, we derive the following

Corollary 1. There exists a pair $(a^*, \pi^*) \in \mathbb{A} \times \mathcal{M}$ such that

$$\Phi^{G(\mathfrak{A})}_{B(\pi^*)}(p) \leqslant \Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) \leqslant M(\pi)$$

for all $(\pi, p) \in \mathcal{M} \times G(\mathbb{A})$.
In this work, (a^*, π^*) will be called a *least favourable combination* from $\mathbb{A} \times \mathcal{M}$.

Proof. Let, according to Theorem 9, $\pi^* \in \mathcal{M}$ be a least favourable prior distribution. That is, we have $M(\pi^*) \leqslant M(\pi)$ for all $\pi \in \mathcal{M}$. Now, according to Theorem 2, choose an action $a^* \in \mathbb{A}$ such that

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) = \Phi^{G(\mathfrak{A})}_{B(\pi^*)}(\delta_{a^*}) \geqslant \Phi^{G(\mathfrak{A})}_{B(\pi^*)}(p)$$

for all $p \in G(\mathbb{A})$. Then, by the definition of $M(\pi^*)$, we get

$$\Phi^{G(\mathfrak{A})}_{B(\pi^*)}(p) \leqslant \Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) = \Phi^{G(\mathfrak{A})}_{B(\pi^*)}(\delta_{a^*}) \leqslant M(\pi^*) \leqslant M(\pi)$$

for all $(\pi, p) \in \mathcal{M} \times G(\mathbb{A})$. This completes the proof. □

The proof of Theorem 9 not only shows the existence of least favourable prior distributions for finite decision problems, but also describes a method to determine such a distribution for a concrete decision problem: Again, least favourable prior distributions can be gained by solving suitable linear optimization problems and, therefore, by applying statistical standard software. Let's illustrate this by an example.

Example 4. Consider the finite CDP $\mathfrak{A}$ defined by

u_{ij}	θ_1	θ_2	θ_3	θ_4
a_1	20	15	10	30
a_2	30	10	10	20
a_3	20	40	0	20
a_4	10	30	50	30
a_5	0	30	20	40

and let

$$\mathcal{M} := \{\pi : \pi_1 + \pi_3 \leqslant 0.8 \wedge \pi_4 \geqslant 0.5\}$$

be a linearly restricted set of probability measures on the set of states. The linear constraints $R(i)$, $i = 1, 2$ defining the set $\mathcal{M}$ are given by

- $f_1(\theta) := \begin{cases} 1 & \text{if } \theta \in \{\theta_1, \theta_3\} \\ 0 & \text{else} \end{cases}$
- $f_2(\theta) := \begin{cases} 1 & \text{if } \theta = \theta_4 \\ 0 & \text{else} \end{cases}$
- $(\underline{b}_i, \overline{b}_i) := \begin{cases} (0, 0.8) & \text{if } i = 1 \\ (0.5, 1) & \text{if } i = 2 \end{cases}$

Now, according to the proof of Theorem 9, a least favourable prior $\pi^* \in \mathcal{M}$ can be found by solving the following linear programming problem

$$(1, -1, 0, 0, 0, 0) \cdot \begin{pmatrix} w_1 \\ w_2 \\ \pi_1 \\ \vdots \\ \pi_4 \end{pmatrix} \longrightarrow \min_{w,\pi}$$

with constraints

- $w_1, w_2, \pi_1, ..., \pi_4 \geqslant 0$
- $$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & -1 & -20 & -15 & -10 & -30 \\ 1 & -1 & -30 & -10 & -10 & -20 \\ 1 & -1 & -20 & -40 & 0 & -20 \\ 1 & -1 & -10 & -30 & -50 & -30 \\ 1 & -1 & 0 & -30 & -20 & -40 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \pi_1 \\ \vdots \\ \pi_4 \end{pmatrix} \geqslant \begin{pmatrix} 1 \\ -1 \\ 0 \\ -0.8 \\ 0.5 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

This linear programming problem can be solved using, for example, the lpSolve package in the statical software R. As an optimal solution we get

$$(w_1^*, w_2^*, \pi_1^*, \ldots, \pi_m^*) = (24, 0, 0.4, 0, 0.1, 0.5)$$

Hence, a $\mathcal{M}$-least favourable prior distribution for the above decision problem is given by

$$\pi^*(\{\theta_i\}) = \begin{cases} 0.4 & \text{if } i = 1 \\ 0 & \text{if } i = 2 \\ 0.1 & \text{if } i = 3 \\ 0.5 & \text{if } i = 4 \end{cases}$$

Now, there are (at least) two possibilities to determine a π^*-optimal action. The first one is the following: Pick a pure action in $\{a_1, \ldots, a_5\}$, which maximizes the expected utility with respect to the measure π^*. This is possible, since we know that there exists at least one optimal pure action. We arrive at

$$\mathbb{E}_{\pi^*}(u_{a_i}) = \begin{cases} 24 & \text{if } i = 1 \\ 23 & \text{if } i = 2 \\ 18 & \text{if } i = 3 \\ 24 & \text{if } i = 4 \\ 22 & \text{if } i = 5 \end{cases}$$

We see that both, a_1 and a_4 (respectively δ_{a_1} and δ_{a_4}), are π^*-optimal actions in the decision problem $G(\mathfrak{A})$.

The second possibility to determine an π^*-optimal action is to solve the linear optimization problem (39). We get

$$(24, 23, 18, 24, 22) \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_5 \end{pmatrix} \longrightarrow \max_p$$

with constraints

- $p_i \geqslant 0$ for all $i = 1, \ldots, 5$
- $\begin{pmatrix} 1 & \ldots & 1 \\ -1 & \ldots & -1 \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ \vdots \\ p_5 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Computing in R gives the optimal solution

$$(p_1^*, ..., p_5^*) = (0, 0, 0, 1, 0)$$

Hence, we receive a_4 as an optimal solution. Finally, least favourable combinations from $\mathcal{M} \times \mathbb{A}$ are given by (π^*, a_1) and (π^*, a_4). $\star$

3.4 Maximin-criterion as a linear optimization problem

In this paragraph, we recall how Maximin-actions in finite decision problems can be determined by the usage of linear optimization theory. The presentation is strongly orientated on [2, § 2.3, especially Proposition 2.4.3].

As a first step, recall that Example 1 proved that Maximin-actions doesn't necessarily have to be pure. There might exist randomized actions having a strictly higher utility than all pure actions available. Therefore, all considerations in this paragraph are based on randomized actions. More precisely, we always work on the mixed extension $G(\mathfrak{A})$ of a finite decision problem $\mathfrak{A}$.

Now, recall that a randomized action $p^* \in G(\mathbb{A})$ is Maximin-action if, and only if,

$$\Phi_M^{G(\mathfrak{A})}(p^*) = \min_{\theta \in \Theta} G(u)(p^*, \theta) \geqslant \min_{\theta \in \Theta} G(u)(p, \theta) = \Phi_M^{G(\mathfrak{A})}(p) \tag{49}$$

holds for all $p \in G(\mathbb{A})$. Thus, determining a Maximin-action is equivalent to the optimization problem

$$\underbrace{\min_{\theta \in \Theta} G(u)(p, \theta)}_{=:\alpha(p)} \longrightarrow \max_{p \in G(\mathbb{A})} \tag{50}$$

Particularly, this implies that there exists a randomized Maximin-action for the decision problem $G(\mathfrak{A})$ if, and only if, the optimization problem (50) admits an optimal solution.

Now, note that, for any $p \in G(\mathbb{A})$ fixed, the term $\alpha(p)$ equals the greatest real number M smaller than the utility of p under all possible states of nature, i.e.

$$\alpha(p) = \max\{M : M \leqslant G(u)(p, \theta) \ \forall \theta \in \Theta\} \tag{51}$$

Hence, the optimization problem (50) possesses an optimal solution if, and only if, the optimization problem

$$(1, \underbrace{0, \ldots, 0}_{\text{n times}}) \cdot \begin{pmatrix} M \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_{(M, p_1, \ldots, p_n)} \tag{52}$$

with constraints

- $p \in \Delta_n$
- $M \leqslant G(u)(t^{-1}(p), \theta)\ \forall \theta \in \Theta$

possesses an optimal solution. In this case, the optimal outcome of the optimization problems (52) and (50) coincide. Note that we again used the one-to-one correspondence of the sets Δ_n and $G(\mathbb{A})$.

Exactly like in the case of problem (41), the constraints do not restrict the variable M to be bounded from below by 0. Again, in order to construct a linear optimization problem in the sense of Definition 11, we substitute the variable M by the difference of two non-negative variables w_1, w_2.

We arrive at the SMP

$$(1, -1, \underbrace{0, \dots, 0}_{\text{n times}}) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_{(w_1, w_2, p_1, \dots, p_n)} \tag{53}$$

with constraints

- $p, w_1, w_2 \geqslant 0$
- $p \in \Delta_n$
- $w_1 - w_2 \leqslant G(u)(t^{-1}(p), \theta)\ \forall \theta \in \Theta$

To see the linearity of the constraints, recall that

$$G(u)(t^{-1}(p), \theta) = \sum_{i=1}^{n} u(a_i, \theta) \cdot p_i$$

Hence, the the second and third constraint can equivalently be described by the inequality

$$\underbrace{\begin{pmatrix} 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 1' & \dots & 1' \\ 1 & 1' & u'_{11} & \dots & u'_{n1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1' & u'_{1m} & \dots & u'_{nm} \end{pmatrix}}_{=:G} \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant \underbrace{\begin{pmatrix} 1 \\ 1' \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:b \in \mathbb{R}^{(n+2)\times 1}}$$

Finally, we arrive at the *linear* optimization problem (53) (in the sense of Definition 11). But does the linear optimization problem (53) admit an optimal solution?

To see that this indeed is the case, note that the objective function $w_1 - w_2$ of problem (53) is bounded on the set $Z(G,b)^+$, since

$$w_1 - w_2 \leqslant \sum_{i=1}^{n} u(a_i,\theta)\cdot p_i \leqslant \max_{i,j} u(a_i,\theta_j)\cdot \sum_{i=1}^{n} p_i = \max_{i,j} u(a_i,\theta_j) < \infty$$

for all $(w_1,\ldots,p_n) \in Z(G,b)^+$. Hence, there always exists an optimal solution of problem (53) according to Theorem 5. By equivalence, this implies the existence of an optimal solution for problem (50). Thus, there always exists a randomized Maximin-action for the decision problem $G(\mathfrak{A})$.

Particularly, every optimal solution of the SMP (53) then is of the form

$$(w_1^*, w_2^*, \underbrace{p_1^*,\ldots,p_n^*}_{=:p^*\in\Delta_n}) \in Z(G,b)^+$$

Thus, solving the SMP gives us an optimal randomized action by simply re-transforming p^* into $t^{-1}(p^*)$, where t is the bijective map defined in the beginning of the chapter.

All together, we just gave a *constructive* proof for the following theorem. It also can be found in [2, Proposition 2.4.3].

Theorem 10. For any finite CDP $\mathfrak{A}$ there exists a Maximin-action $p^* \in G(\mathbb{A})$. □

Implicitly, the above considerations recalled an algorithm for determining Maximin-actions in finite CDPs with the methods of linear optimization. Again, the computation of such an action can be done within standard statistical software. A concrete application of this algorithm is demonstrated in the following example.

Example 5. Consider again the CDP from Example 4. According to the above considerations, we can determine a (randomized) Maximin-action by solving the linear optimization problem (53). Thus, we arrive at the linear programming problem

$$(1,-1,0,0,0,0,0)\cdot\begin{pmatrix} w_1\\ w_2\\ p_1\\ \vdots\\ p_5\end{pmatrix} \longrightarrow \max_{(w_1,\ldots,p_n)}$$

with constraints

- $p, w_1, w_2 \geqslant 0$
- $$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -20 & -30 & -20 & -10 & 0 \\ 1 & -1 & -15 & -10 & -40 & -30 & -30 \\ 1 & -1 & -10 & -10 & 0 & -50 & -20 \\ 1 & -1 & -30 & -20 & -20 & -30 & -40 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_5 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

As an optimal solution we receive

$$(w_1^*, w_2^*, p_1^*, \ldots, p_5^*) = (21.\bar{6}, 0, 0, 0.5, 0.1\bar{6}, 0.\bar{3}, 0)$$

Hence, a randomized Maximin-action $p^* \in G(\mathbb{A})$ is given by the probability measure

$$p(\{a_i\}) = \begin{cases} 0 & \text{if } i = 1 \\ \frac{1}{2} & \text{if } i = 2 \\ \frac{1}{6} & \text{if } i = 3 \\ \frac{1}{3} & \text{if } i = 4 \\ 0 & \text{if } i = 5 \end{cases}$$

Finally, the Maximin-utility of the randomized action p^* can be gained by evaluating the objective function for the optimal solution just computed. Thus, the Maximin-utility is given by $w_1^* - w_2^* = 21.\bar{6}$. ⋆

3.5 Hodges & Lehmann criterion as a linear optimization problem

Finally, we want to construct linear programming problems optimizing the Hodges & Lehmann criterion discussed in Chapter 1. As shown in Example 6 in Chapter 1, taking randomized actions into account sometimes might create a strict improvement of the Hodges & Lehmann utility.

Again, since all pure actions can be uniquely identified with randomized actions (namely the corresponding Dirac-measure, see Remark 2 on Definition 2), it suffices to determine optimal actions on the mixed extension of a given CDP.

Recall that, for fixed $\alpha \in [0, 1]$ and a probability measure π on $(\Theta, \mathcal{P}(\Theta))$, a randomized action $p^* \in G(\mathbb{A})$ is $H(\alpha, \pi)$-optimal if, and only if, the inequality

$$\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p^*) \geqslant \Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p)$$

holds for all $p \in G(\mathbb{A})$, where

$$\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p) = \alpha \cdot \Phi^{G(\mathfrak{A})}_{B(\pi)}(p) + (1-\alpha) \cdot \Phi^{G(\mathfrak{A})}_{M}(p) \tag{54}$$

Hence, determining a $H(\alpha,\pi)$-optimal action is equivalent to resolving the optimization problem

$$\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p) \longrightarrow \max_{p \in G(\mathbb{A})} \tag{55}$$

Particularly, this implies that there exists a randomized $H(\alpha,\pi)$-action for the decision problem $G(\mathfrak{A})$ if, and only if, the above optimization problem (55) possesses an optimal solution.

For the case of a finite CDP this optimization problem always has an optimal solution. This is the statemant of the following theorem. Additionally, the method used to prove the theorem can be used to compute $H(\alpha,\pi)$-optimal actions applying the methods of linear programming theory.

Theorem 11. Let $\mathfrak{A}$ be any finite CDP and let $\alpha \in [0,1]$. Let further π denote a probability measure on $(\Theta, \mathcal{P}(\Theta))$. Then there exists an $H(\alpha,\pi)$-optimal action $p^* \in G(\mathbb{A})$.

Proof. According to the above considerations, it suffices to show that the optimization problem (55) has an optimal solution. Using equation (54) and the definition of the Maximin- and the Bayes-criterion, we arrive at

$$\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p) = \alpha \cdot \Big(\sum_{i=1}^{n} c_i^{\pi} \cdot p(\{a_i\})\Big) + (1-\alpha) \cdot \min_{\theta \in \Theta} G(u)(p,\theta) \tag{56}$$

where c_i^{π} is defined like in the proof of Theorem 9. According to equations (50) and (51) we get

$$\min_{\theta \in \Theta} G(u)(p,\theta) = \max\{M : M \leqslant G(u)(p,\theta)\ \forall \theta \in \Theta\} \tag{57}$$

Hence, the optimal outcome of the optimization problem (55) equals the optimal outcome of the optimization problem

$$\alpha \cdot \Big(\sum_{i=1}^{n} c_i^{\pi} \cdot p_i\Big) + (1-\alpha) \cdot M = \big((1-\alpha), \alpha c_1^{\pi}, \ldots, \alpha c_n^{\pi}\big) \cdot \begin{pmatrix} M \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_{M,p} \tag{58}$$

with constraints

- $p \in \Delta_n$
- $M \leqslant G(u)(t^{-1}(p), \theta)$ for all $\theta \in \Theta$

As already seen in the context of problem (41), in order to receive an equivalent linear programming problem in the sense of Definition 11, one has to substitute the variable M by the difference of two non-negative variables w_1, w_2. We arrive at

$$\big((1-\alpha), (\alpha-1), \alpha c_1^\pi, \ldots, \alpha c_n^\pi\big) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_{w,p} \tag{59}$$

with constraints

- $w, p \geqslant 0$
- $p \in \Delta_n$
- $w_1 - w_2 \leqslant G(u)(t^{-1}(p), \theta)$ for all $\theta \in \Theta$

These are the same constraints as for the Maximin-criterion. Thus, the constraints are linear, as already shown before. Hence, the optimization problem (59) is a SMP. But are there optimal solutions? According to Theorem 2, there exists a pure action $a^* \in \mathbb{A}$ such that

$$U_1 := \Phi^{\mathfrak{A}}_{B(\pi)}(a^*) \geqslant \sum_{i=1}^{n} c_i^\pi \cdot p_i$$

for all $p \in \Delta_n$. Additionally, we have that

$$U_2 := \max_{i,j} u(a_i, \theta_j) \geqslant \min_{\theta \in \Theta} G(u)(p, \theta)$$

for all $p \in G(\mathbb{A})$. Thus, we receive the inequality

$$\alpha \cdot \Big(\sum_{i=1}^{n} c_i^\pi \cdot p_i\Big) + (1-\alpha) \cdot (w_1 - w_2) \leqslant \alpha \cdot U_1 + (1-\alpha) \cdot U_2 < \infty$$

for all $(p, w) \in Z(G, b)^+$, where G and b are defined as in the last paragraph. Hence, the objective function of the optimization problem (59) is bounded from above on the set $Z(G, b)^+$ of admissible solutions. Therefore, the SMP has an optimal solution

according to Theorem 5.
Now, let

$$(w_1^*, w_2^*, \underbrace{p_1^*, \ldots, p_n^*}_{=:p^* \in \Delta_n}) \in Z(G,b)^+$$

be an optimal solution of (59). Then, by construction, the following holds

$$\max\Big\{\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p) : p \in G(\mathbb{A})\Big\} \quad = \quad \alpha \cdot \Phi^{G(\mathfrak{A})}_{B(\pi)}(t^{-1}(p^*)) + (1-\alpha)\cdot(w_1^* - w_2^*)$$

But

$$\begin{aligned} w_1^* - w_2^* &= \max\{M \in \mathbb{R} : M \leqslant G(u)(t^{-1}(p^*), \theta) \text{ for all } \theta \in \Theta\} \\ &= \min_{\theta\in\Theta} G(u)(t^{-1}(p^*), \theta) \\ &= \Phi^{G(\mathfrak{A})}_{M}(t^{-1}(p^*)) \end{aligned}$$

Finally, we arrive at

$$\begin{aligned} \max\Big\{\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p) : p \in G(\mathbb{A})\Big\} &= \alpha \cdot \Phi^{G(\mathfrak{A})}_{B(\pi)}(t^{-1}(p^*)) + (1-\alpha)\cdot\Phi^{G(\mathfrak{A})}_{M}(t^{-1}(p^*)) \\ &= \Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(t^{-1}(p^*)) \end{aligned}$$

Hence, $t^{-1}(p^*) \in G(\mathbb{A})$ is a $H(\alpha,\pi)$-optimal action. This completes the proof. □

The proof of Theorem 11 contains a method for determining optimal randomized Hodges & Lehmann-actions in finite decision problems by using the methods of linear optimization theory. This algorithm can be app lied for arbitrary values of the optimism parameter $\alpha \in [0,1]$.

Again, this allows to determine randomized Hodges & Lehmann-actions by using standard statistical software. How this can be done for a concrete decision problem is demonstrated in the following example.

Example 6. Consider again the CDP defined in Example 4. Additionally, let π be the probability measure on $(\Theta, \mathcal{P}(\Theta))$ induced by the assignment

$$\pi(\{\theta_j\}) = \begin{cases} \frac{2}{5} & \text{if } j = 1 \\ \frac{1}{5} & \text{if } j = 2 \\ \frac{1}{10} & \text{if } j = 3 \\ \frac{3}{10} & \text{if } j = 4 \end{cases}$$

Assume the uncertainty type underlying the CDP is perfectly characterized by setting $\alpha = 0.7$. We compute

$$c_i^\pi = \mathbb{E}_\pi(u_{a_i}) = \begin{cases} 21 & \text{if } i = 1 \\ 21 & \text{if } i = 2 \\ 22 & \text{if } i = 3 \\ 24 & \text{if } i = 4 \\ 20 & \text{if } i = 5 \end{cases}$$

Then, according to Theorem 11, a $H(\alpha, \pi)$-optimal action can be determined by solving the linear programming problem (59), that is

$$(0.3, -0.3, 0.7 \cdot 21, \ldots, 0.7 \cdot 20) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_5 \end{pmatrix} \longrightarrow \max_{w,p} \tag{60}$$

with constraints

- $w, p \geqslant 0$
- $$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -20 & -30 & -20 & -10 & 0 \\ 1 & -1 & -15 & -10 & -40 & -30 & -30 \\ 1 & -1 & -10 & -10 & 0 & -50 & -20 \\ 1 & -1 & -30 & -20 & -20 & -30 & -40 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_5 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Computing this linear programming problem in R, we receive the following optimal solution:

$$s^* := (w_1^*, w_2^*, p_1^*, \ldots, p_5^*) = (21.\bar{6}, 0, 0, 0.5, 0.1\bar{6}, 0.\bar{3}, 0)$$

Like in the case of the Maximin-criterion, the probability measure $p^* \in G(\mathbb{A})$ induced

by the assignment

$$p^*(\{a_i\}) = \begin{cases} 0 & \text{if } i = 1 \\ \frac{1}{2} & \text{if } i = 2 \\ \frac{1}{6} & \text{if } i = 3 \\ \frac{1}{3} & \text{if } i = 4 \\ 0 & \text{if } i = 5 \end{cases}$$

is an optimal randomized action. Again, the Hodges & Lehmann utility can be gained by evaluating the objective function for the optimal solution. We arrive at:

$$\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p^*) = (0.3, -0.3, 0.7 \cdot 21, \ldots, 0.7 \cdot 20) \cdot \begin{pmatrix} 21.\bar{6} \\ 0 \\ 0 \\ 0.5 \\ 0.1\bar{6} \\ 0.\bar{3} \\ 0 \end{pmatrix} = 22.01\bar{6}$$

$\star$

To complete the paragraph, we want to show up a connection between the Hodges & Lehmann-criterion and the Bayes-criterion. More precisely, given the prior π and the trade-off parameter α, we describe a method to construct a probability measure $\lambda_{\pi,\alpha}$ on $(\Theta, \mathcal{P}(\Theta))$ such that

$$\Phi^{\mathfrak{A}}_{B(\lambda_{\pi,\alpha})}(a^*) = \Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p^*) \tag{61}$$

where $a^* \in \mathbb{A}$ denotes a Bayes-action w.r.t. $\lambda_{\pi,\alpha}$ and $p^* \in G(\mathbb{A})$ denotes a randomized $H(\alpha,\pi)$-action. That is, for every pair (α,π) consisting of a parameter $\alpha \in [0,1]$ and a probability measure π on $(\Theta, \mathcal{P}(\Theta))$ there exists a probability measure $\lambda_{\pi,\alpha}$ on $(\Theta, \mathcal{P}(\Theta))$ such that the Bayes-utility w.r.t. $\lambda_{\pi,\alpha}$ equals the $H(\alpha,\pi)$-utility. This is the statement of the following

Theorem 12. Let $\mathfrak{A}$ denote any finite CDP. Further, let π be a probability measure on $(\Theta, \mathcal{P}(\Theta))$ and let $\alpha \in [0,1]$. Let, according to Theorem 11, p^* denote a randomized $H(\alpha,\pi)$-action. Then, there exists a probability measure $\lambda_{\pi,\alpha}$ on $(\Theta, \mathcal{P}(\Theta))$ and a pure action $a^* \in \mathbb{A}$ such that equation (61) is satisfied.

Proof. According to the proof of Theorem 11, the expression $\Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p^*)$ equals the optimal outcome of the linear optimization problem (59). Thus, by duality, it equals

the optimal outcome of the dual programming problem of (59), namely

$$h(u_1, \ldots, \sigma_m) := (1, -1, \underbrace{0, \ldots, 0}_{\text{m-times}}) \cdot \begin{pmatrix} u_1 \\ u_2 \\ \sigma_1 \\ \vdots \\ \sigma_m \end{pmatrix} \longrightarrow \min_{(u_1, \ldots, \sigma_m)} \tag{62}$$

with constraints

- $(u_1, \ldots, \sigma_m) \geqslant 0$
- $G^T \cdot \begin{pmatrix} u_1 \\ u_2 \\ \sigma_1 \\ \vdots \\ \sigma_m \end{pmatrix} \geqslant \underbrace{\begin{pmatrix} 1-\alpha \\ \alpha - 1 \\ \alpha \cdot c_1^\pi \\ \vdots \\ \alpha \cdot c_n^\pi \end{pmatrix}}_{=:\vec{s}}$

where G is the constraint matrix from the proof of Theorem 10. A simple computation shows that the second constraint is equivalent to the (in)equalities

$$\sum_{j=1}^{m} \sigma_j = 1 - \alpha \tag{63}$$

$$u_1 - u_2 \geqslant \alpha \cdot c_i^\pi + \sum_{j=1}^{m} u_{ij} \cdot \sigma_j = \sum_{j=1}^{m} u_{ij} \cdot (\alpha \cdot \pi_j + \sigma_j) \quad \text{for all } i = 1, \ldots, n \tag{64}$$

Now, let $(u_1^*, u_2^*, \sigma_1^*, \ldots, \sigma_m^*)$ denote an optimal solution of problem (62). By duality, such an solution always exists, since (62) is the dual of (59), which always admits an optimal solution according to Theorem 11. Then, according to equation (63), we have

$$\sum_{j=1}^{m} (\alpha \cdot \pi_j + \sigma_j^*) = \alpha \cdot \sum_{j=1}^{m} \pi_j + \sum_{j=1}^{m} \sigma_j^* = \alpha + 1 - \alpha = 1 \tag{65}$$

Together with the non-negativity of $\alpha \cdot \pi_j + \sigma_j^*$ for all $j = 1, \ldots, m$, this implies that $\alpha \cdot \pi + \sigma^* \in \Delta_m$. Hence, the assignment $\lambda_{\pi,\alpha} := b^{-1}(\alpha \cdot \pi + \sigma^*)$ defines a probability measure on $(\Theta, \mathcal{P}(\Theta))$.

Next, note that for the optimal outcome of problem (62) the following identity holds:

$$u_2^* - u_2^* =$$

$$
\begin{aligned}
&= \min\Big\{u_1 - u_2 : (u_1, \ldots, \sigma_m) \in Z^-(G^T, \vec{s})\Big\} \\
&= \min\Big\{h(u_1, \ldots, \sigma_m) : (u_1, \ldots, \sigma_m) \in Z^-(G^T, \vec{s})\Big\} \\
&= \min\left\{h(u_1, u_2, \sigma_1^*, \ldots, \sigma_m^*) : u_1 - u_2 \geqslant \sum_{j=1}^{m} u_{ij} \cdot (\alpha \cdot \pi_j + \sigma_j) \text{ for all } i = 1, \ldots, n\right\} \\
&= \min\Big\{u_1 - u_2 : u_1 - u_2 \geqslant c_i^{\lambda_{\pi,\alpha}} \text{ for all } i = 1, \ldots, n\Big\}
\end{aligned}
$$

However, the minimum in the last bracket coincides with the optimal outcome of the linear programming problem (42) for determining a Bayes-action with respect to $\lambda_{\pi,\alpha}$. That is, the optimal outcome of problem (62) coincides with the Bayes-utility w.r.t. the measure $\lambda_{\pi,\alpha}$. Now, according to Theorem 2, let $a^* \in \mathbb{A}$ denote a pure Bayes-action with respect to $\lambda_{\pi,\alpha}$. Then, we arrive at the following identity:

$$\Phi^{\mathfrak{A}}_{B(\lambda_{\pi,\alpha})}(a^*) = u_1^* - u_2^* = \Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p^*) \tag{66}$$

This completes the proof. □

So, we succeeded in showing that, instead of applying the Hodges & and Lehmann criterion, one always can maximize expected utility with respect to a suitable prior distribution. Furthermore, the proof of Theorem 12 contains a method to construct such a prior distribution using the linear optimization (and duality theory in particular).

Let's apply the described method to our example.

Example 7. Consider once again the CDP from Example 4. Additionally, let π be defined like in Example 6 and let $\alpha = 0.7$. Next, we set up the optimization problem (62) adopted for our example. We arrive at

$$(1, -1, 0, 0, 0, 0) \cdot \begin{pmatrix} u_1 \\ u_2 \\ \sigma_1 \\ \vdots \\ \sigma_4 \end{pmatrix} \longrightarrow \min_{(u_1, \ldots, \sigma_4)}$$

with constraints

- $(u_1, \ldots, \sigma_4) \geqslant 0$

- $$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & -1 & -20 & -15 & -10 & -30 \\ 1 & -1 & -30 & -10 & -10 & -20 \\ 1 & -1 & -20 & -40 & 0 & -20 \\ 1 & -1 & -10 & -30 & -50 & -30 \\ 1 & -1 & 0 & -30 & -20 & -40 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ \sigma_1 \\ \vdots \\ \sigma_4 \end{pmatrix} \geqslant \begin{pmatrix} 1-\alpha \\ \alpha - 1 \\ \alpha \cdot c_1^{\pi} \\ \vdots \\ \alpha \cdot c_5^{\pi} \end{pmatrix} = \begin{pmatrix} 0.3 \\ -0.3 \\ 0.7 \cdot 21 \\ 0.7 \cdot 21 \\ 0.7 \cdot 22 \\ 0.7 \cdot 24 \\ 0.7 \cdot 20 \end{pmatrix}$$

Running the above optimization problem in R returns the optimal solution

$$(u_1^*, \ldots, \sigma_4^*) := (22.01666667\ ,\ 0\ ,\ 0.21583333\ ,\ 0.05750000\ ,\ 0.02666667\ ,\ 0)$$

Hence, the probability measure $\lambda_{\pi,0.7}$ on $(\Theta, \mathcal{P}(\Theta))$ is induced by the assignment

$$\lambda_{\pi,0.7}(\{\theta_j\}) = 0.7 \cdot \pi(\{\theta_j\}) + \sigma_j^* = \begin{cases} 0.7 \cdot \frac{2}{5} + 0.2158\bar{3} = 0.4958\bar{3} & \text{if } j = 1 \\ 0.7 \cdot \frac{1}{5} + 0.0575 = 0.1975 & \text{if } j = 2 \\ 0.7 \cdot \frac{1}{10} + 0.02\bar{6} = 0.09\bar{6} & \text{if } j = 3 \\ 0.7 \cdot \frac{3}{10} + 0 = 0.21 & \text{if } j = 4 \end{cases}$$

The expected utilities of the actions $a_1, \ldots, a_5$ with respect to $\lambda_{\pi,0.7}$ are given by

$$\mathbb{E}_{\lambda_{\pi,0.7}}(u_{a_i}) = \begin{cases} 20.14583 & \text{if } i = 1 \\ 22.01667 & \text{if } i = 2 \\ 22.01667 & \text{if } i = 3 \\ 22.01667 & \text{if } i = 4 \\ 16.25833 & \text{if } i = 5 \end{cases}$$

Hence, we have the Bayes-actions a_2, a_3 and a_4 and the following holds:

$$\Phi^{\mathfrak{A}}_{B(\lambda_{\pi,0.7})}(a_2) = \Phi^{\mathfrak{A}}_{B(\lambda_{\pi,0.7})}(a_3) = \Phi^{\mathfrak{A}}_{B(\lambda_{\pi,0.7})}(a_4) = \Phi^{G(\mathfrak{A})}_{H(\alpha,\pi)}(p^*)$$

where p^* denotes the $H(\pi, 0.7)$-optimal solution computed in Example 6. ⋆

Remark. Theorem 12 shows up a (purely?) mathematical connection between the Bayes-criterion and the Hodges & Lehmann-criterion. However, the following question comes to mind: What does the theorem teach us about the philosophical relation of the two criteria? In other words: Can we interpret the Hodges & Lehmann-criterion as a way of constructing prior distributions that take *scepticism* (as part

of the information base) into account?

Let us deepen this idea little by little: According to the considerations in Paragraph 1.2.3, applying the Bayes-criterion is a reasonable choice, whenever the uncertainty underlying the CDP is of type I^* and there is a prior measure available that consistently uses our information in the best possible manner. That is, if we know that there exists a true (but unknown) classical probability measure on the set of states and our information is sufficient to specify a best estimator for it.

However, suppose your information is sufficient to specify a prior measure, but it is not sufficient to be sure that this specification has been done best possible. That is, there exist several measures that are equally plausible candidates to be the true one. Thus, there is no unique choice possible without accepting a certain degree of arbitrariness.

There are many ways of addressing this problem, famous ones among them are the usage of *credal sets* or *interval probabilities* respectively. Such concepts, namely so called *imprecise probabilistic models*, will be recalled and applied in detail in the later Chapters 4 (in particular, Paragraphs 4.2.1 and 4.2.2) and 5 (throughout the whole chapter).

According to Theorem 12, another way of proceeding could be the following: First, choose one of the compatible measures arbitrarily. Let π denote this measure. Next, decide (depending on how many concurring measures exist) how certain you are about π being the true measure. Express this (un)certainty by a parameter $\alpha \in [0, 1]$.

More precisely, choose α to be small, if you're very uncertain and choose it to be large, if you're very certain about that description. Then, α is part of the information base and, therefore, should be taken into account when constructing a prior distribution characterizing the underlying uncertainty. Next, construct the measure $\lambda_{\pi,\alpha}$ just as described in the proof of Theorem 12. Finally, determine a (pure) Bayes-action with respect to the constructed measure. In this way, we constructed a prior distribution that also takes scepticism into account. ◦

4 Generalized definitions of probability

The *Bayesian Paradigm* claims the following: *Any* situation under uncertainty can be described by a classical probability distribution, i.e. by a set function satisfying *Kolmogorov's Axioms* (see also the discussions in Remark 1 on Definition 8 or in the beginning of Paragraph 1.3). If new data (or more general: information) is available this set function is *updated* in order to get an even better description of the situation. This is not a contradiction, since the availability of new information creates a *new* situation under uncertainty. This, very roughly, describes one of the basic procedures in the field of *Bayesian Statistics*.

However, it turns out that there are many situations in which the available information can't be adequately characterized by using classical probability theory. The axioms demand a *degree of precision* that simply can't be guaranteed by the data. Particularly, the *additivity* of disjoint events demanded by the third axiom creates a far too restrictive framework for certain situations.

So, is there a way to use such *imprecise information*? There are (at least) two different ways of addressing this problem: The first way to proceed is to simply ignore the imprecision of the information by using a classical probability which represents the available information in the *best possible* manner. But what is the best possible manner? And is this always possible without generating inconsistencies? Subsequently, we will see it is not.

The second way to proceed is to *generalize* the concept of probability: If the classical concept of probability turns out to be too restrictive to describe certain situations under uncertainty, then one has to widen the concept. Since the classical concept still is sufficient and suitable to describe situations with precise information available, the generalization should be done in a way which contains the classical concept as a special case (see [5, p. 17]). There are several approaches to do so. In this chapter we recall two of them, namely the theory of *interval probability* and the theory of *credal sets*.

The chapter is structured as follows: In the first paragraph we recall some motivating examples from the literature, i.e. we show concrete examples for prior information that is not describable by classical probability theory. Within these examples, the scope stretches from classical examples, like the *Ellsberg Paradox*, up to examples occuring in even more recent scientific work from various disciplines. Subsequently, we recall some theoretical arguments supporting the idea of a generalization of probability theory. One of them will be the inability of classical probability theory to

describe situations in which no prior information is available at all.

In the second paragraph, we briefly explain two of the most common generalization of probability theory, namely the theory of *credal sets* and the theory of *interval probability*. Note that there exists many different other approaches for generalizing descriptions of uncertainty. An overview is for example given in [44].

In a situation under uncertainty, a *credal set* (see Definition 16) is the set of all probability measures being compatible with the information available. In contrast to classical probabilistic descriptions, this set, in general, contains more than one single element. Consequently, then the whole credal set is used as a more adequate description of the uncertainty underlying the situation. Between the different measures contained in the credal set there is complete indifference.

An *interval probability* is a map (satisfying a set of axioms, see Definitions 18 and 19) that assigns each event of some σ-field an interval-valued probability component. Every such assignment is a subset of the real unit interval $[0, 1]$. Roughly spoken, such an interval then contains all the values that are, under the available information, equally plausible to be the true probability.

As we will see, both approaches turn out to be closely related to each other: Every credal set induces an interval-valued probability assignment and, vice versa, every (consistent) interval probability is compatible with a (credal) set of classical probability measures. Finally, we demonstrate how such generalized descriptions can avoid the inconsistencies arising from the classical theory. The description mainly follows [5, Ch. 1], [10], [2, § 3.3], [47, § 2], [46, § 1.4] and [48]. Detailed references are given at the appropriate places in the text.

4.1 Some motivating examples

As the scope of the present work mainly lies in determining optimal actions in decision problems under partial information, the following question comes to mind: Why do we need a generalized concept of modelling uncertainty *in decision theory*? At what point in decision theory is it necessary to build models of (possibly imprecise) prior information?

To answer this question, recall the discussion in Remark 1 on Definition 8: If the exact probability distribution on the set of states is unknown (or ill-known), one has to work with an estimated uncertainty model. Now, if the additional information suffices to estimate an prior probability distribution, one can simply proceed by applying the Bayes-criterion instead of the Bernoulli-criterion. That is, we choose a

classical probability measure that is consistent with respect to the available information and *act as if* it was the true one.

But what if this is not possible? In other words: What if any classical probability measure on the set of states contradicts the available information? Can we construct criteria that use the information available without translating it into a probability measure? As we will see in Chapter 5, this is possible in many different ways.

To show that these questions are not only interesting from a theoretical point of view, but are highly relevant for applications in various scientific disciplines, is the aim of the considerations in the following paragraphs. We give examples for situation under uncertainty in which the available information is not describable by classical probability theory in a consistent way. We start with *the* classic one.

4.1.1 Ellsberg's Paradox

The following example (or experiment) goes back to the economist *Daniel Ellsberg* (* 7. April 1931) and was published in [10] in 1961. The presentation of the example that is chosen in the present work is strongly influenced by [2, p.322-334], [48] and [47, Example 2.6.33]. It is one of the best known examples when it comes to supporting the idea of a generalized concept of describing uncertainty.

Before we can start, we need some short preparation: Assume the Bayesian Paradigm to be true. That is, for each situation under uncertainty, there exists a classical probability distribution perfectly representing it (for a more in-depth explanation, see Paragraph 1.3 or Remark 1 on Definition 8).

Now, consider a finite CDP under type I* uncertainty. Then, the mechanism generating the states of nature, together with all our information concerning it, describes a situation under uncertainty. Thus, according to the Bayesian Paradigm, there exists a probability distribution characterizing it.

Now, according to the considerations in Chapter 1, the best we can do to determine an optimal decision in this situation is to apply the Bayes-criterion with this probability distribution on the set of states, i.e. maximizing our expected utility according to the probability measure fully representing our information. Hence, the *ordering of preferences* induced by the Bayes-criterion (recall that each criterion induces a linear ordering on the set of actions) fully characterizes the actor's preferences (under the information available).

Now, consider the following situation under uncertainty of type I*: An urn contains an unknown (but *finite*) number of balls. Some of these are red, some are yellow

and some are black. Every ball in the urn is of exactly one of these three colors. Additionally, we have the information that the proportion of red balls f_r contained in the urn exactly equals $\frac{1}{3}$.

Thus, we conclude that both, the proportion of yellow balls f_y and the proportion of black balls f_b, lay in the interval $[0, \frac{2}{3}]$ and we have $f_b = \frac{2}{3} - f_y$.
Next, a ball is drawn. One may choose between the following alternatives:

- $a_1 :=$ Receive 100 \$, if the ball is red.
- $a_2 :=$ Receive 100 \$, if the ball is black.

Which one would you choose?

Ellsberg performed this as a *thought* experiment (no real balls were drawn). The experimental subjects were a group of economists and statisticians from the *Harvard University*, i.e. *experts* in the field of probability theory. The result was that a vast majority of these experts preferred alternative a_1 before alternative a_2.

So, if there *was* a probability measure π on measurable space $(\Omega, \mathcal{P}(\Omega))$, where $\Omega = \{r, y, b\}$, which fully characterizes the ordering of preferences of *rational actors* (or experts), the following would have to hold

$$\pi(\{r\}) > \pi(\{b\}) \tag{67}$$

Otherwise, the Bayes-criterion according to π clearly would prefer alternative a_2 before alternative a_1 and, thus, wouldn't describe our experts' ordering of preferences at all.

Now, another decision situation is proposed: In the same situation as described above, one may choose between the following alternatives:

- $b_1 :=$ Receive 100 \$, if the ball is red or yellow.
- $b_2 :=$ Receive 100 \$, if the ball is black or yellow.

Note that the underlying situation under uncertainty hasn't changed at all: There is still the same information concerning the states of nature available. So, if there was a distribution describing the situation, it would still have to be the same as above, namely π.

Again, Ellsberg performed this as a thought experiment with the same group of experts. But this time a vast majority preferred alternative b_2 before alternative b_1.

So, if π would characterize the experts' preference order, the following would have to hold

$$\pi(\{b, y\}) > \pi(\{r, y\})$$

using an analogous argumentation as above. Since π is assumed to be a (classical) probability and therefore *additive* for disjoint events, this is equivalent to

$$\pi(\{b\}) + \pi(\{y\}) > \pi(\{r\}) + \pi(\{y\})$$

or more simple

$$\pi(\{b\}) > \pi(\{r\}) \tag{68}$$

which, together with (67), yields a contradiction.

So, what can we learn from this experiment? At least, two different points of view are possible:

- *Classical point of view:* The majority preference order of the group of experts is irrational as it contradicts the laws of classical probability theory. It is not possible to model irrational preference orders.

- *Imprecise point of view:* The majority preference order of the group of experts is rational, but classical probability theory doesn't offer a suitable theoretical framework to formalize it. Other frameworks (more general) might be suitable to formalize such seemingly inconsistent prior information.

The strongest argument against the first (or classic) point of view is the *expert status* of the experimental subjects: It is real hard to argue that experts for probability theory act irrational when it comes to their field of expertise (of course, only under the assumption that every expert does not willingly bias the experiment).

Additionally, the first point of view is very rigid and therefore non-constructive: Is any kind of knowledge that can't be embedded in classical probability theory worthless or/and inconsistent? Surely not.

So, if one accepts the second (or imprecise) point of view, one arrives at the following negation of the Bayesian Paradigm: *There are situations under uncertainty that cannot be described using classical probability theory.* As such situations might be highly relevant, this is a strong argument for the usage of generalized descriptions of uncertainty in decision theory.

The following examples illustrate the need for a generalized probability theory in

different scientific disciplines even more. We start with an example which concerns both *linguistic* and *artificial intelligence*.

4.1.2 Dealing with linguistic uncertainty

Next, we recall the *problem of modelling linguistic uncertainty*. Much effort has been put in formalizing this kind of uncertainty in a suitable way in the last sixty years. Pioneer work in this field has been done by *Lotfi Zadeh (*1921)* in [50] and [49]. Note that also classical probabilistic approaches have been proposed (see e.g. [21]). Additionally, we refer to [45].

Roughly spoken, we have the following: Many of the expressions used in natural language tend to be *vague*, i.e. inexact concerning their information content. However, in many situations such vague information will be the best information available.

For example, consider the expression *tall*. Here, it is hard to make out an exact threshold value separating the set of all people in two classes, namely the class of people being tall and the class of people being not tall. Thus, the expression 'subject is tall' generates a situation under uncertainty, where the uncertainty is about the subject's exact body size.

However, as natural language is the main tool of exchanging (and storing) information, the question how *vague linguistic uncertainty* can be formalized (and therefore used for the statistical analysis) seems to be of great importance. Consider again our example: If I have the information of a person being tall, what can I conclude about his or her exact size, which I am still uncertain about?

Now, the question is: Is classical probability theory a suitable framework to formalize such kind of knowledge? In other words: Is there always a probability measure characterizing situations under uncertainty with given vague linguistic information adequately (clearly, according to the Bayesian Paradigm, the answer has to be yes, see the discussion in Remark 1 on Definition 8)?

In our example, there are several aspects to consider: Let $\Omega = \{50, 51, \ldots, 299, 300\}$ be the (for the sake of simplicity) finite set of possible body sizes of adult persons. Additionally, we have the information that our subject's exact body size, denoted by b, has been labelled *tall*.

Firstly, we can state the following: In this situation (or any similar one) any *objective* probabilistic description of the uncertainty about the exact body size necessarily has to be meaningless, since the body size of our subject is not a *random entity*. The subject is of unknown but *fixed* size. Thus, statements concerning the probability

of b lying in some interval always have to equal one or zero, depending on whether the corresponding interval *covers* the exact value or not.

Nevertheless, there exists uncertainty about the exact value of the body size, since *we* don't know it. One way of addressing the problem of the exact body size b not being a random entity, is interpreting uncertainty in a *subjective* way. Following de Finetti (see for example [14] or the discussions in Remark 1 on Definition 8 or Paragraph 1.3), uncertainty is no longer interpreted as a property of the phenomenon to describe, but of the subject describing it.

In other words: The probability of an event A equals the *maximum buying price* of a *rational actor* for a bet paying out 1, if A occurs and 0, if it doesn't. Note that for this interpretation of probability no randomness is necessary at all, since the probability describes the *degree of belief* of the actor in the truth of event A (again, see the discussion in Remark 1 on Definition 8).

However, in the context of linguistic uncertainty, being a rational actor necessarily goes along with using the available (vague) information in the best possible manner. Again, the question is: Is it always feasible to assume this could be done using classical probability theory?

For the moment, assume the answer is yes. Then, in our example, a classical probability measure $\mathbb{P}_{tall}$ characterizing the situation perfectly could be constructed as follows: Every event $A \subset \Omega$ is assigned our maximum buying price of the bet described above. Since the assignment is assumed to represent our vague knowledge perfectly, we know that events containing greater values receive higher assignments, for example we have

$$\mathbb{P}_{tall}(\{50, ..., 100\}) < \mathbb{P}_{tall}(\{180, ..., 230\})$$

Again, the problem comes along with the additivity of disjoint events demanded by Kolmogorov's third axiom: Given the information of the subject of interest being tall, I will be pretty sure about the truth of event $A := \{180, ..., 230\}$. Thus, my maximum rational buying price $\mathbb{P}_{tall}(A)$ will be very close to 1. However, as the information *tall* is very vague, I will be pretty unsure about the truth of every singleton event $A_i := \{i\}$, $i = 180, ..., 230$. The degree of precision of the information *tall* simply isn't sufficient to label any of the values contained in A more plausible than others. Thus, my maximal buying price for any event A_i, $i = 180, ..., 230$, will be close to 0.

But as $\mathbb{P}_{tall}$ is assumed to be a classical probability measure, Kolmogorov's third

axiom implies the following identity

$$\underbrace{\mathbb{P}_{tall}(A)}_{\approx 1} = \sum_{i=180}^{230} \underbrace{\mathbb{P}_{tall}(A_i)}_{\approx 0}$$

which yields a contradiction. Thus, we found another class of situations under uncertainty that can't be adequately described using classical probabilities.

In a summarized form, we arrive at the following: In many situations under uncertainty, the uncertainty is given by the *vagueness* of natural language. This is the case, since the exchange of information often is done orally and, thus, by the usage of natural language. This vague linguistic information often suffices to label certain subsets A of the possibility space Ω as being very *plausible*, but not to split the plausibility assignment on the singleton events contained in A. However, such kinds of assignment contradict the additivity axiom of classical probability theory. Therefore, classical probability theory, in general, isn't a suitable framework to formalize vague linguistic uncertainty.

4.1.3 Modelling uncertainty in expert systems

In this paragraph, we mainly follow [32], [5, p.12], [20] and [1]. Decision theory is applied in many different scientific disciplines. One of them is the field of *artificial intelligence*. More precisely, the question how optimal decisions can be derived from a given (possibly imprecise or vague) information base is highly relevant for the implementation of so called *expert systems*.

There exist many different definitions of what an expert system is, all of them slightly different. In [32] the authors try to make out some properties, which all of the different definitions have in common. According to them an expert system is a computer program that (besides other properties)

- is based on knowledge
- implements human experiences and skills
- explains the decision made in natural language

For our purposes, the following vague definition is sufficient: An expert system is a computer program trying to emulate the process of decision making of experts in certain scientific discipline. Theoretically, using such programs it is then possible for non-experts to make decisions just like an expert. Of course, in practice, a computer system never can perfectly emulate the expert's decision process, so that the

proposals of the expert system should only be used as a guideline.

One of the main fields of application of such expert systems is the field of *medical diagnosis*. One of the first and best known medical expert systems is the *MYCIN* system. It was developed at the *Stanford University* in the early 1970s as the doctoral thesis of *Edward Shortliffe (*1947)*. Although it performed well in experimental runs, it was never actually used in practice.

MYCIN was implemented for the following situation: The exact diagnose of certain bacterial blood infection diseases of a patient often requires to grow cultures of the infecting organism. This growing process lasts up to 48 hours. As there are very serious blood infection diseases that require a quick therapy, growing cultures in many cases takes too much time. So, in order to save the patients life, medical experts often have to make a *best guess* for a diagnosis out of the available information base using their *experience*. Of course, this guess should be done in a way that avoids too huge risks for the patient (e.g. completely wrong medication).

Now, the problem is the following: Not every doctor is an expert for guessing suitable diagnosis out of vague information bases. Particularly, doctors just having finished their studies often don't have enough experience available. Additionally, if the disease is a very rare one, almost no doctor (with the exception of a few experts) will have the necessary experience available.

This is where MYCIN comes into game: If it is possible to emulate the decision and explanation process of an expert for blood infection diseases computational, non-experienced medical staff could use this programs to support (or improve) their own guesses.

So, how does this motivate to think about generalized uncertainty measures besides classical probability theory? Imagine the following situation: A patient comes into the practice of a young general practitioner. The patient suffers of a bacterial blood infection disease, which the doctor maybe never has heard of, but certainly is no expert about. However, the patient seems to be seriously ill and needs a quick therapy. Using the patient's description of the symptoms, the doctor will be able to eliminate certain classes of diseases completely (remember, he is a doctor). Specifically, he will be able to list up a couple of diseases that might be responsible for the symptoms. However, in general, it won't be possible to the doctor to say which of the diseases is the most plausible one. More precisely, we will end up with a couple of lists containing diseases which the doctor is totally indifferent about to be the true disease.

Hence, this exemplifies another situation where the available information is not sufficient to split the probability component of a union of disjoint events (the classes of diseases) to the single ones. The doctor, in general, won't be able to describe his information by a classical probability measure on the 'space of possible diseases'. Hence, a classical Bayesian learning model seems to be unsuitable in such a situation. Instead, the MYCIN system makes use of another description of uncertainty, so called *certainty factors*. A discussion of this concept is not within the scope of the present work. It can for example be found in [20]. However, the use of such alternative concepts clarifies that expert systems, like for example MYCIN, profit a lot (and are maybe only possible with) generalized descriptions of uncertainty.

4.1.4 Some theoretical reasons

Finally, we want to complete our paragraph on motivating examples with two theoretical reasons for using imprecise uncertainty descriptions. Here, the use the term 'theoretical' can be explained in the following sense: In contrast to the examples recalled before, generalized uncertainty description are no longer motivated by modelling problems arising from concrete practical applications, but by theoretical considerations that might concern many different scientific disciplines. Hereby, we mainly follow the presentation in [5, p.9-10 and p.16-17] and [46, Ch. 1.4].

The first problem is particularly relevant for Bayesian statistics (and, therefore, for all related areas as, for example, decision theory) and can be summarized as follows: How can one specify a (classical) prior probability measure on a measurable space $(\Omega, \mathcal{A})$ under consideration, if there is no information available at all? Does such a measure exist? Or: Do we need more general uncertainty models to formalize such a *complete lack of information* ?

One common way to proceed in such a situation is applying *Bernoulli's Principle of Indifference* (also called *Principle of Insufficient Reason*) (see [8]). It goes back to the famous mathematician *Jacob Bernoulli* (1654-1705) and, roughly, states the following: If the space Ω is finite and there is no further information available, the best we can do is to choose the *uniform* probability measure as a prior. More precisely, if Ω consists of m elements, we choose the measure π induced by the assignment $\pi(\{\omega_j\}) = \frac{1}{m}$ for all $j = 1, ..., m$.

The problem here is simple to state: This is the same description of uncertainty as we would use given the *perfect information* that every state has exactly the same probability of occurring (i.e. in a situation under *perfect symmetry*). Thus, such a description is not able to distinguish between certain situations with perfect informa-

tion available and situations with a complete lack of knowledge/information. Hence, this can't be a suitable description of ignorance! This turns out to be structural inability: The framework of classical probability theory is too restrictive to model complete ignorance.

This problem is highly relevant for decision theory and, particularly, for the present work: As our aim is to determine optimal decisions with respect to a given information base, it is essential to know how to decide when there is no information available at all (this is a information base as well). More precisely, if we find ourselves in a decision situation where our only information consists of the fact that there exists a classical probability measure generating the states of nature, the classical theory is not able to formalize this knowledge different from the situation where each state occurs with the same probability. So, this inability of the classical theory is a strong argument for considering more general frameworks *in decision theory*!

The second problem is known as *the problem of partially identified probability*. Let $(\Omega, \mathcal{A})$ denote some measurable space. Suppose, there exists a probability measure π on $(\Omega, \mathcal{A})$ characterizing the uncertainty between the different elements of Ω adequately. Further, suppose this measure π is *unknown*. However, we know the probability components

$$\{\pi(A) : A \in \mathcal{A}_0\} \tag{69}$$

for some subset $\mathcal{A}_0 \subset \mathcal{A}$. That is, the true measure is only known on a subset of the set of all events under consideration.

Now, the problem is the following: In general, this partial information won't be sufficient to specify the true measure π on the whole space on $(\Omega, \mathcal{A})$. Instead, there will be a whole set of measures on on $(\Omega, \mathcal{A})$ being compatible with the available information. So, if we want to choose a prior measure, how can we decide between all these concurring measures? Again, we described a situation under uncertainty that cannot adequately be characterized by the use of classical probability theory.

However, as we will see later, the problem of partially identified probabilities is highly relevant to decision theory: Often, the information on the mechanism generating the states of nature will be only partial. Suppose, for example, a situation in which we only know that one state is surely at least as probable as another one. Such an example seems not to be too far-taken, but can't be formalized within the framework of classical probability theory. In the next paragraph, we recall two generalizations of the classical theory that allow us to formalize such situations more adequate.

4.2 Generalized models of uncertainty

As seen in the previous examples there exist situations under uncertainty that can't be characterized by using classical probability theory. One of the main reasons for this inability of the classical theory turned out to be the third of Kolmogorov's axioms, namely the additivity of disjoint events: Often the available information doesn't suffice to split the uncertainty assignment of a union of disjoint events consistently on the single events. But how can we formalize such situations more adequate? And is it possible to keep this formalization consistent with cases, where the uncertainty is perfectly describable by classical probability theory?

There are several approaches trying to do so. A good overview is given e.g. in [44]. In the present work, we focus on only two of them: The concept of *interval probability* and the concept of *credal sets*. The basic ideas of both concepts are briefly discussed in the following two paragraphs. Afterwards, we explain how the two seemingly unrelated concepts are connected to each other. Additionally, we illustrate how these generalized uncertainty concepts can help to avoid the inconsistencies arising from the classical theory.

4.2.1 Credal sets

The theory of *credal sets* that is briefly recalled in the following paragraph goes back to *Isaac Levi (*1930)* and was mainly developed in [30]. The presentation chosen here doesn't claim completeness, but focusses on results relevant for our purposes.

Consider a situation under uncertainty. What if the available information is not sufficient to specify a classical probability measure that characterizes the uncertainty adequately? More precisely: What if there are many different measures that are equally plausible candidates for being the 'true' description of the uncertainty? One natural way of addressing this problem is the following: As a description of the uncertainty underlying the situation, we use the set of *all* probability measures being compatible with the information base.

Hence, the description of the uncertainty is no longer given by a single probability measure, but by a whole set of such measures. As every such measure is compatible with our information base, all of them are assumed to be equally plausible. That is, between the measures contained in this *credal set*, we are totally indifferent.

Within this framework, the classical concept of probability is naturally contained: If we have perfect information, the set of compatible measure shrinks to only one element, i.e. one single probability measure describing the uncertainty. Additionally, it

is possible to model situations in which there is no information available at all. Since in such situations every probability measure is equally plausible, the corresponding credal set then consists of all possible probability measures on the measurable space under consideration.

Note that, in this way, we receive different characterizations for situations under complete ignorance and situations under perfect symmetry and, therefore, already managed to address one of the inconsistencies arising from the classical theory (see Paragraph 4.1.4).

When it comes to the interpretation of credal sets, there are two fundamentally different points of view to distinguish (see for example [2, p. 355-356]):

- *Epistemic point of view:* The credal set is the best we can derive about the *true* (but ill-known) probability measure characterizing the uncertainty. That is, for any situation under uncertainty there *exists* such a true measure. However, the available information is not sufficient to specify it uniquely. Hence, the credal set is interpreted as the smallest set of possible uncertainty descriptions being compatible with the information base. Here, we distinguish between the situation under uncertainty and the information available.

- *Ontological point of view:* The credal set is interpreted as an entity on its own. That is, we no longer distinguish between situations under uncertainty and the information available: The situation under uncertainty is fundamentally determined by the information available. Hence, the credal set on its own is a suitable description for a specific pair consisting of uncertainty/information.

No matter what point of view is preferred, we can give the following, very general, definition of a credal set. A similar version of this definition can for example be found in [44, § 5].

Definition 16. Let $(\Omega, \mathcal{A})$ be a measurable space. Further, let $G(\Omega, \mathcal{A})$ denote the set of all probability measures on $(\Omega, \mathcal{A})$. Then, every non-empty subset $\mathcal{C} \subset G(\Omega, \mathcal{A})$ is called a *credal set* on $(\Omega, \mathcal{A})$. $\triangledown$

The previous definition now allows us to define an extension of the concept of the expectation of a random variable to the case, where the uncertainty is described by a credal set. This generalized expectation then equals the set of expectations of the random variable under all measures contained in the credal set. As probabilities of events can be viewed as expectations of the corresponding indicator functions, this

then gives rise to a natural description of the uncertainty about an event by sets of classical probability assignments.

Definition 17. Let $\mathcal{C} \subset G(\Omega, \mathcal{A})$ is be a credal set on some measurable space $(\Omega, \mathcal{A})$. Further, let X denote a random variable, such that $\mathbb{E}_p(X) < \infty$ for all $p \in \mathcal{C}$. Then, the set

$$\mathbb{E}_{\mathcal{C}}^{C}(X) := \Big\{ \mathbb{E}_p(X) : p \in \mathcal{C} \Big\} \tag{70}$$

is called Cred($\mathcal{C}$)-*expectation* of X.
If $X = \mathbb{1}_A$ for some event $A \in \mathcal{A}$, that is $\mathbb{E}_p(X) = p(A)$ for all $p \in \mathcal{C}$, the set

$$P^{C}(A) := \mathbb{E}_{\mathcal{C}}^{C}(\mathbb{1}_A) = \Big\{ p(A) : p \in \mathcal{C} \Big\} \tag{71}$$

is called the Cred($\mathcal{C}$)-*probability* of A. ▽

Remark. The sets $\mathbb{E}_{\mathcal{C}}^{C}(X)$ and $P^{C}(A)$ can be interpreted in the following way: As the credal set contains all the probability measures that are compatible with our information base, the set $\mathbb{E}_{\mathcal{C}}^{C}(X)$ contains the expectations of the random variable X under all the compatible measures. As we are totally indifferent between the different elements of the credal set $\mathcal{C}$, each of the values contained in $\mathbb{E}_{\mathcal{C}}^{C}(X)$ is equally plausible to be the true expectation of X. By analogy, the set $P^{C}(A)$ is the set of all probability assignments for the event A that are compatible with our information base. Again, all of the values contained in A are assumed to be equally plausible to be the true assignment. ◦

The following theorem makes an statement about how Cred($\mathcal{C}$)-expectations (and therefore Cred($\mathcal{C}$)-probabilities) look like if the underlying measurable space is finite and the credal set $\mathcal{C}$ is a convex set (see Definition 13). This theorem will prove very important for the decision theoretical applications of the theory discussed in Chapter 5.

Theorem 13. Let $\Omega := \{\omega_1, \ldots, \omega_m\}$ be any finite set. Further, let $\mathcal{C} \subset G(\Omega, \mathcal{P}(\Omega))$ be a convex credal set on $(\Omega, \mathcal{P}(\Omega))$ and $X : \Omega \to \mathbb{R}$ be a random variable. Then, the following relation holds:

$$\mathbb{E}_{\mathcal{C}}^{C}(X) \supset \Big(\inf_{p \in \mathcal{C}} \mathbb{E}_p(X), \sup_{p \in \mathcal{C}} \mathbb{E}_p(X) \Big) \tag{72}$$

Here, we used the notation $(a, b) := \{x \in \mathbb{R} : a < x < b\}$ for $a, b \in \mathbb{R}$.

Proof. According to the assumption, the set $\mathcal{C}$ is non-empty.
We distinguish three cases:

Case 1: $|\mathcal{C}| = 1$. That is, there is a probability measure $p^* \in G(\Omega, \mathcal{P}(\Omega))$ such that $\mathcal{C} = \{p^*\}$. Hence, we have that

$$\inf_{p\in\mathcal{C}} \mathbb{E}_p(X) = \sup_{p\in\mathcal{C}} \mathbb{E}_p(X) = \mathbb{E}_{p^*}(X)$$

Therefore, we have

$$E := \Big(\inf_{p\in\mathcal{C}} \mathbb{E}_p(X), \sup_{p\in\mathcal{C}} \mathbb{E}_p(X) \Big) = \emptyset$$

which proves the statement, since $\emptyset \subset \mathbb{E}_{\mathcal{C}}(X)$.

Case 2: $|\mathcal{C}| > 1$ and $E = \emptyset$. Then, $E \subset \mathbb{E}^C_{\mathcal{C}}(X)$ and the statement is proven.

Case 3: $|\mathcal{C}| > 1$ and $E \neq \emptyset$. Then, let $a \in E$. Assume, for contradiction, that $a \notin \mathbb{E}^C_{\mathcal{C}}(X)$. Then, there exist $p_1 \neq p_2 \in \mathcal{C}$ such that

$$\mathbb{E}_{p_1}(X) < a < \mathbb{E}_{p_2}(X)$$

Otherwise, $a \leqslant \mathbb{E}_p(X)$ or $a \geqslant \mathbb{E}_p(X)$ for all $p \in \mathcal{C}$ would have to hold, which contradicts $a \in E$ in either case.
Since $\mathcal{C}$ is a convex set, we have that $p_\gamma := \gamma \cdot p_1 + (1-\gamma) \cdot p_2 \in \mathcal{C}$ for all $\gamma \in [0,1]$. We compute

$$\begin{aligned}
\mathbb{E}_{p_\gamma}(X) &= \sum_{j=1}^{m} X(\omega_j) \cdot p_\gamma(\{\omega_j\}) \\
&= \sum_{j=1}^{m} X(\omega_j) \cdot (\gamma \cdot p_1(\{\omega_j\}) + (1-\gamma) \cdot p_2(\{\omega_j\})) \\
&= \gamma \cdot \sum_{j=1}^{m} X(\omega_j) \cdot p_1(\{\omega_j\}) + (1-\gamma) \cdot \sum_{j=1}^{m} X(\omega_j) \cdot p_2(\{\omega_j\}) \\
&= \gamma \cdot \mathbb{E}_{p_1}(X) + (1-\gamma) \cdot \mathbb{E}_{p_2}(X)
\end{aligned}$$

But, since intervals are convex as well, every point lying in the interval

$$\Big[\mathbb{E}_{p_1}(X), \mathbb{E}_{p_2}(X) \Big]$$

can be written int the form

$$\gamma \cdot \mathbb{E}_{p_1}(X) + (1-\gamma) \cdot \mathbb{E}_{p_2}(X)$$

for some $\gamma \in [0,1]$. Hence, there exist $\gamma^* \in [0,1]$ such that $a = \mathbb{E}_{p_{\gamma^*}}(X)$. This yields a contradiction, since $p_{\gamma^*} \in \mathcal{C}$, but $a \notin \mathbb{E}_{\mathcal{C}}(X)$. □

As an immediate consequence, we can derive the following

Corollary 2. Consider the situation of Theorem 13. If the expressions

$$\min_{p\in\mathcal{C}} \mathbb{E}_p(X) \qquad \text{and} \qquad \max_{p\in\mathcal{C}} \mathbb{E}_p(X)$$

exist, the following identity holds

$$\mathbb{E}_{\mathcal{C}}^{C}(X) = \Big[\min_{p\in\mathcal{C}} \mathbb{E}_p(X), \max_{p\in\mathcal{C}} \mathbb{E}_p(X)\Big]$$

□

Remark. Corollary 2 allows the following identification of credal sets and real intervals: If we have a convex credal set on a finite measurable space such that maximum and minimum expectation of arbitrary random variables are attained on the credal set, we can identify credal expectations with closed real valued intervals. Particularly, this implies that we can identify credal probabilities with closed interval-subsets of the unit interval. This identification will prove very important for the considerations of Chapter 5. ∘

The following example illustrates, how uncertainty can be described by using credal set in situations where only imperfect information on the states of nature is available.

Example 8. (Similar in [2, p. 391-396]). Consider once again the situation of the Ellsberg Paradox. The two concurring decision problems are summarized in the following tables.

u_{ij}	red	yellow	black
a_1	100	0	0
a_2	0	0	100

u_{ij}	red	yellow	black
b_1	100	100	0
b_2	0	100	100

Here, $\Theta = \{r, y, b\}$ is finite. Thus, we can work on the measureable space $(\Theta, \mathcal{P}(\Theta))$. Again, let f_r, f_y and f_b denote the proportion of red, yellow and black balls contained in the urn. Then, the *true* probability measure p_t characterizing the situation with perfect information available is induced by the assignment $p_t(\{i\}) = f_i$ for $i = r, y, b$.

However, the information available is only partial and, therefore, imperfect. More precisely, we have the following three properties:

(P1) $f_r = \frac{1}{3}$

(P2) $0 \leqslant f_i \leqslant \frac{2}{3}$ for $i = y, b$

(P3) $f_b + f_y = \frac{2}{3}$

According to the information available, every probability assignment satisfying these three properties is equally plausible to be the true one (under perfect information). Hence, we can describe the information by the (non-empty) credal set $\mathcal{M}_0$ of all classical probability measures on $(\Theta, \mathcal{P}(\Theta))$ that are compatible with the properties $(P1) - (P3)$. That is,

$$\mathcal{M}_0 := \Big\{p : p(\{r\}) = \tfrac{1}{3} \wedge 0 \leqslant p(\{i\}) \leqslant \tfrac{2}{3} \text{ for } i = y, b\Big\} = \Big\{p : p(\{r\}) = \tfrac{1}{3}\Big\}$$

Since all conditions of Corollary 2 are satisfied in this example, the $\mathrm{Cred}(\mathcal{M}_0)$-probability P^C induced by our credal set is of the form

$$P^C : \mathcal{P}(\Theta) \to \mathcal{Z}([0,1]) \quad , \quad A \mapsto \left[\inf_{p \in \mathcal{M}_0} p(A), \sup_{p \in \mathcal{M}_0} p(A)\right]$$

where $\mathcal{Z}([0,1])$ denotes the set of all closed intervals being a subset of $[0,1]$. In this way, we receive a natural description of the uncertainty underlying the situation by using the concept of credal sets. ⋆

4.2.2 Interval probability

In this paragraph we briefly recall the axiomatic approach of defining interval probability established by *Kurt Weichselberger (*1929)*. We mainly refer to [5, Ch. 1], [47, § 2.2] and [46, Ch. 1,2].

Roughly spoken, the idea is the following: Instead of a single number $p \in [0,1]$, every event $A \subset \Omega$ in some possibility space Ω is assigned a real interval $[a,b] \subset [0,1]$, where $a \leqslant b$, as a measure of its probability.

However, similar as for the interpretation of a credal set, there exist two fundamentally different points of view on how such an interval is connected to classical probability theory (see for example [2, p. 355-356]):

- *Epistemic point of view:* For every event A there exists a *true* classical probability assignment $p(A) \in [0,1]$. However, this true probability assignment is

unknown (or ill-known). The interval probability $[a, b]$ of the event A then can be interpreted as the *best* we can derive about the true probability from the available information. For this, the following is important to note: No $x \in [a, b]$ is more *plausible* to be the true assignment. That is, there is complete indifference between the values lying in $[a, b]$.

- *Ontological point of view:* Probability is understood as an *interval-valued entity*. Thus, the concept of interval probability is understood to be a theory on its own and *not* a theory explaining how true probabilities can be optimally covered by intervals when only partial information is available. According to this point of view, there is no true classical probability underlying the situation. The interval probability is assumed to be the *true* probability.

For our purposes, the two concurring points of view generate coinciding mathematical theories. Philosophically, which view to prefer of course is a highly relevant question. In the present work, we follow the epistemic way of viewing things, whenever this is necessary. Doing so, the following interpretation of the probability interval $[a, b]$ is possible: Depending on the width of the interval, i.e. depending on how big the number $b - a$ is, one can state the following:

- The closer $b-a$ is to 1, the more uncertain one is concerning the *true* probability of A. In the extreme case of $b - a = 1$, that is $b = 1$ and $a = 0$, there is no information on the probability of the event A available. The interval $[a, b]$ coincides with the interval $[0, 1]$. Thus all values in the interval $[0, 1]$ are equally plausible to be the true probability of the event A.

- The closer $b-a$ is to 0, the more certain one is concerning the *true* probability of A. In the extreme case of $b - a = 0$, that is $b = a$, there is perfect information on the probability of A available. The interval probability assignment $[a, b]$ shrinks to a point. Thus, the true probability of the event A is known.

In this context, the uncertainty between the different values contained in the interval $[a, b]$ is called *ambiguity* or *non-stochastic uncertainty* (see for example [2, p. 333]). According to the above considerations, the width of the interval can be used as a measure for the strength of ambiguity underlying an assignment.

How can this idea be put on sound theoretical ground? We begin with the basic definition, which is strongly orientated on the *axiomatic* approach to classical probability theory (in the sense that it is *independent of interpretation*). It is taken from [46, Definition 2.1.4].

Definition 18. Let $\mathcal{Z}([0,1])$ be the set of all closed intervals being a subset of $[0,1]$, that is

$$\mathcal{Z}([0,1]) := \Big\{ [a,b] : a \leqslant b \wedge [a,b] \subset [0,1] \Big\} \tag{73}$$

Now, let $(\Omega, \mathcal{A})$ be a measurable space, where $\mathcal{A}$ denotes a σ-field on Ω. Define two set functions

$$L : \mathcal{A} \to [0,1] \quad \text{and} \quad U : \mathcal{A} \to [0,1] \tag{74}$$

such that

$$L(A) \leqslant U(A) \quad \forall A \in \mathcal{A} \tag{75}$$

Thus, we get a well-defined map

$$P : \mathcal{A} \to \mathcal{Z}([0,1]) \quad , \quad A \mapsto [L(A), U(A)] \tag{76}$$

Then, the map P is called *R-probability* on $(\Omega, \mathcal{A})$ if, and only if, the set

$$\mathcal{M} := \Big\{ p : p \text{ is pm on } (\Omega, \mathcal{A}) \wedge L(A) \leqslant p(A) \leqslant U(A) \ \forall A \in \mathcal{A} \Big\} \tag{77}$$

is non-empty.
In this case, the set $\mathcal{M}$ is said to be the *structure* of the R-probability P. $\triangledown$

For our purposes, a special class of R-probability turns out to be of particular interest, namely *F-probability*. As we will see later, there is a strong connection between the concepts of F-probability and the concept of a credal set. This motivates the following definition. It is taken from [46, Definition 2.1.5].

Definition 19. Let P be an R-probability on the measurable space $(\Omega, \mathcal{A})$ and let $\mathcal{M}$ denote the structure of P. Then, P is called an *F-probability* on $(\Omega, \mathcal{A})$, if the following holds for all $A \in \mathcal{A}$:

$$L(A) = \inf_{p \in \mathcal{M}} p(A) \quad \text{and} \quad U(A) = \sup_{p \in \mathcal{M}} p(A) \tag{78}$$

That is, if the set functions L and U coincide with the *lower* and the *upper envelope* of the structure $\mathcal{M}$. $\triangledown$

Finally, the concept of F-probability naturally gives rise to a generalization of the concept of the expectation of a random variable. This will turn out to be very important when considering criteria for optimal decision making under complex uncertainty in Chapter 5. It can also be found in [5, Definition 2.2].

Definition 20. Let P be an F-probability on a measurable space $(\Omega, \mathcal{A})$ with structure $\mathcal{M}$. Let further $X : \Omega \to \mathbb{R}$ denote a real-valued random variable. Then, if it exists, the expression

$$\mathbb{E}_{\mathcal{M}}(X) := \left[\underline{\mathbb{E}}_{\mathcal{M}}(X), \overline{\mathbb{E}}_{\mathcal{M}}(X)\right] := \left[\inf_{p\in\mathcal{M}} \mathbb{E}_p(X), \sup_{p\in\mathcal{M}} \mathbb{E}_p(X)\right] \tag{79}$$

is called *interval expectation* of X w.r.t. P. ▽

In the following example, we demonstrate how the concept of F-probability can be used to generate probabilistic models of situations under uncertainty with only imperfect information available. Additionally, a first connection to credal sets is made up.

Example 9. (Similar in [2, p. 391-396]). Consider once again the situation of the Ellsberg Paradox. We want to model the uncertainty about the states of nature by using the concept of interval probability. The R-probability $P : \mathcal{P}(\Theta) \to \mathcal{Z}([0,1])$ that uses the available information in the best possible manner is given by

$$P(A) = [L(A), U(A)] = \begin{cases} [\frac{1}{3}] & \text{if } A = \{r\} \\ [0, \frac{2}{3}] & \text{if } A \in \{\{y\}, \{b\}, \{y, b\}\} \\ [\frac{1}{3}, 1] & \text{if } A \in \{\{b, r\}, \{y, r\}\} \\ [1] & \text{if } A = \Theta \\ [0] & \text{if } A = \emptyset \end{cases}$$

Since, for all $A \in \mathcal{P}(\Theta)$, the inequality $L(A) \leqslant \pi(A) \leqslant U(A)$ holds for all measures π on $(\Theta, \mathcal{P}(\Theta))$ with $\pi(\{r\}) = \frac{1}{3}$, the structure of P is simply given by

$$\mathcal{M} = \left\{p : p(\{r\}) = \tfrac{1}{3}\right\}$$

Furthermore, one easily verifies

$$P(A) = \left[\min_{\pi\in\mathcal{M}} \pi(A), \max_{\pi\in\mathcal{M}} \pi(A)\right]$$

for all $A \in \mathcal{P}(\Theta)$. Therefore, P is an F-probability. Hence, we receive a natural description of the uncertainty underlying the situation by using the concept of F-probability. Since the structure $\mathcal{M}$ of P coincides with the credal set $\mathcal{M}_0$ of the previous example, it makes no computational difference whether the uncertainty

is described by the credal set $\mathcal{M}_0$ or the F-probability P *in this example.* More precisely, we have

$$P(A) = P^C(A) \qquad \text{and} \qquad \mathbb{E}^C_{\mathcal{M}_0}(X) = \mathbb{E}_{\mathcal{M}}(X)$$

for all $A \in \mathcal{P}(\Theta)$ and $X : \Theta \to \mathbb{R}$. However, in more general examples, both descriptions no longer have to coincide. The connection between the two concepts is briefly explained in the following remark. ⋆

Remark. How are the concepts of credal sets and F-probability related to each other? Let $(\Omega, \mathcal{A})$ be a measurable space. If $\mathcal{M}$ is the credal set of all probability measures on $(\Omega, \mathcal{A})$ being compatible with our information, then the map P_1 defined by

$$P_1 : \mathcal{A} \to \mathcal{Z}([0,1]) \quad , \quad A \mapsto \Big[\inf_{\pi \in \mathcal{M}} \pi(A), \sup_{\pi \in \mathcal{M}} \pi(A) \Big] \tag{80}$$

defines an F-probability on $(\Omega, \mathcal{A})$. Vice versa, if P_2 is an F-probability on $(\Omega, \mathcal{A})$, then the structure of P_2 defines a credal set on $(\Omega, \mathcal{A})$.

However, note the following: In general, the structure $\mathcal{M}_{P_1}$ of the F-probability P_1 from (80) doesn't necessarily coincide with the set $\mathcal{M}$! More precisely, it is possible that $\mathcal{M} \subsetneq \mathcal{M}_{P_1}$. That is, the set $\mathcal{M}_P$ might contain elements that aren't contained in $\mathcal{M}$ and, therefore, might be incompatible with the given information base. For a more in-depth discussion of the topic see Paragraph 5.1. ∘

The previous examples showed how situation under uncertainty with only imperfect prior information available can be characterized by the use of credal sets and interval probabilities respectively. However, we still do not know, whether such an description avoids the inconsistencies arising by the use of classical probability. As a first step on the way to the answer of this question, another aspect seems to be of great importance: How can one derive preference orders from interval-valued probability assignments? Is there an pendant to the principle of maximizing expected utility that is applied in the precise case?

Definitions 17 and 20 give us well-defined extensions of the concept of the expectation of a random variable when the underlying uncertainty is described by a credal set or an interval probability respectively. However, the definition gives rise to new difficulties: In contrast to the precise case, we don't know how to *compare* interval-valued expectations: Given two intervals of the real line, how can we decide which is the 'greater' one? But the comparability of expectations is crucial for the concept of maximizing expected utility. There exist several approaches to generate comparabil-

ity between interval-valued expectations. A brief overview is given in the following excursus.

4.2.3 Excursus: Interval orders

The following excursus refers to [46, § 2.6, especially p. 226-246] and [2, p. 384-388]. Note that there exist many way of constructing orders on sets of real intervals. therefore, our presentation doesn't claim completeness, but focusses on interval orderings relevant in the context of a generalized concept of maximizing expected utility.

Similar as done in equation (73), one can define the set $\mathcal{Z}(\mathbb{R})$ of all closed interval being a subset of $\mathbb{R}$, that is

$$\mathcal{Z}(\mathbb{R}) := \Big\{ [a,b] : a \leqslant b \wedge a, b \in \mathbb{R} \Big\} \tag{81}$$

How can we define an ordering on the set $\mathcal{Z}(\mathbb{R})$? As a first idea, one could try the following: Given two interval $[a,b], [c,d] \in \mathcal{Z}(\mathbb{R})$ define

$$[a,b] <^I [c,d] \quad :\Leftrightarrow \quad b < c$$

That is, an interval is defined to be less than another one, if its upper bound is strictly small than the lower bound of the other one w.r.t. the ordinary $<$ ordering on the real numbers. Undoubtedly, in the context of interval expectation, this is a reasonable definition: If the interval expectation of an action lies completely below that of another action, it is not reasonable to choose this action.

However, the relation defined by $<^I$ is not total (see Definition 5): For example, the intervals $[1,3]$ and $[2,4]$ are not comparable. Clearly, there are several ways of construction total orderings on the set $\mathcal{Z}(\mathbb{R})$ (uncountably many, to be precise). Which ones are suitable?

In the context of a generalized concept of maximizing expected utility, any suitable construction of an ordering should be compatible with the ordering $<^I$ in the following sense: If $\sqsubseteq$ is a linear ordering on $\mathcal{Z}(\mathbb{R})$ that is potentially reasonable in the context of a generalized concept of maximizing expected utility, then it has at least to satisfy the implication

$$[a,b] <^I [c,d] \;\Rightarrow\; [a,b] \sqsubseteq [c,d] \wedge \neg([c,d] \sqsubseteq [a,b]) \tag{82}$$

for all $[a,b], [c,d] \in \mathcal{Z}(\mathbb{R})$. That is, if an interval is labelled strictly smaller as another on w.r.t. $<^I$, then it is necessarily labelled strictly smaller w.r.t. $\sqsubseteq$.

In the following, we recall three common ways of constructing linear interval orders satisfying condition (82). The idea underlying all of these orderings is a *representation of the interval expectation by one real number.*

As we will see in Chapter 5, which one to apply in a concrete decision problem depends on the *attitude* of the decision maker *towards ambiguity* (that is, optimistic, pessimistic or optimistic with degree α).

1. *The pessimistic interval order, short $\leqslant^p$:* Only the lower bounds of the intervals are compared. More precisely, we have:

$$[a,b] \leqslant^p [c,d] \quad :\Leftrightarrow \quad a \leqslant c \tag{83}$$

   for all $[a,b],[c,d] \in \mathcal{Z}(\mathbb{R})$. Clearly, the total ordering defined by (83) saisfies the condition (82): The inequaliy $b < c$ implies $a \leqslant c \wedge \neg(c \leqslant a)$, since $a \leqslant b$. Therefore, the pessimistic interval order is compatible with the ordering $<^I$.

2. *The optimistic interval order, short $\leqslant^o$:* Only the upper bounds of the intervals are compared. More precisely, we have:

$$[a,b] \leqslant^o [c,d] \quad :\Leftrightarrow \quad b \leqslant d \tag{84}$$

   for all $[a,b],[c,d] \in \mathcal{Z}(\mathbb{R})$. Again, condition (82) is satisfied: The inequaliy $b < c$ implies $b \leqslant d \wedge \neg(d \leqslant b)$, since $c \leqslant d$. Therefore, the optimistic interval order is compatible with the ordering $<^I$.

3. *The interval order with optimism degree $\alpha \in [0,1]$, short $\leqslant^\alpha$:* Intervals are compared with respect to a convex combination of their lower and their upper bound. More precisely, we have

$$[a,b] \leqslant^\alpha [c,d] \quad :\Leftrightarrow \quad \alpha \cdot a + (1-\alpha) \cdot b \leqslant \alpha \cdot c + (1-\alpha) \cdot d \tag{85}$$

   for all $[a,b],[c,d] \in \mathcal{Z}(\mathbb{R})$. Again, condition (82) is satisfied: Let $b < c$ and $\alpha \in [0,1]$. Since $a \leqslant b$, we have $\alpha \cdot a + (1-\alpha) \cdot b \leqslant b$ and since $c \leqslant d$, we have $\alpha \cdot c + (1-\alpha) \cdot d \geqslant c$. Hence, we arrive at $\alpha \cdot a + (1-\alpha) \cdot b < \alpha \cdot c + (1-\alpha) \cdot d$. Therefore, $\leqslant^\alpha$ is compatible with the ordering $<^I$. Note that the orders $\leqslant^p$ and $\leqslant^o$ are special cases of the ordering $\leqslant^\alpha$, namely $\leqslant^p = \leqslant^1$ and $\leqslant^o = \leqslant^0$. •

The previous excursus discussed three possible ways to extend the strict ordering $<^I$ on the set $\mathcal{Z}(\mathbb{A})$ to a linear ordering. Now, any of these linear orderings can be used to define a generalized concept of the principle of maximizing the expected

utility of an action. An in-depth discussion of what ordering is reasonable in which situation under uncertainty will be one of the topics of the next chapter. As a first intuition, the next example illustrates how the inconsistencies arising from the classical formalisation of Ellsberg's paradox can be avoided by the usage of these generalized concepts.

Example 10. (Continuation of Example 9, similar in [46, Example 2.6.33] and [5, p. 391-396]). Consider again the situation of Example 67. Now, note that any of the actions a_1, a_2, b_1 and b_2 induces a random variable on the set $\Theta = \{r, y, b\}$ via

$$u_a(\cdot) := u(a, \cdot) : \Theta \to \mathbb{R} \quad , \quad \theta \mapsto u(a, \theta)$$

for $a \in \{a_1, a_2, b_1, b_2\}$. We compute the interval expectation of the four random variables:

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{a_1}) := \inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{a_1}) = \inf_{\pi \in \mathcal{M}} (100 \cdot \pi(\{r\})) = \frac{1}{3} \cdot 100 = 33.\bar{3}$$

$$\overline{\mathbb{E}}_{\mathcal{M}}(u_{a_1}) := \sup_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{a_1}) = \sup_{\pi \in \mathcal{M}} (100 \cdot \pi(\{r\})) = \frac{1}{3} \cdot 100 = 33.\bar{3}$$

$$\Rightarrow \mathbb{E}_{\mathcal{M}}(u_{a_1}) = [33.\bar{3}]$$

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{a_2}) := \inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{a_2}) = \inf_{\pi \in \mathcal{M}} (100 \cdot \pi(\{b\})) = 100 \cdot 0 = 0$$

$$\overline{\mathbb{E}}_{\mathcal{M}}(u_{a_2}) := \sup_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{a_2}) = \sup_{\pi \in \mathcal{M}} (100 \cdot \pi(\{b\})) = 100 \cdot \frac{2}{3} = 66.\bar{6}$$

$$\Rightarrow \mathbb{E}_{\mathcal{M}}(u_{a_2}) = [0, 66.\bar{6}]$$

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{b_1}) := \inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{b_1}) = \inf_{\pi \in \mathcal{M}} (100 \cdot (\pi(\{r\})) + \pi(\{y\}))) = 100 \cdot \frac{1}{3} = 33.\bar{3}$$

$$\overline{\mathbb{E}}_{\mathcal{M}}(u_{b_1}) := \sup_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{b_1}) = \sup_{\pi \in \mathcal{M}} (100 \cdot (\pi(\{r\}) + \pi(\{y\}))) = 100 \cdot 1 = 100$$

$$\Rightarrow \mathbb{E}_{\mathcal{M}}(u_{b_1}) = [33.\bar{3}, 100]$$

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{b_2}) := \inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{b_2}) = \inf_{\pi \in \mathcal{M}} (100 \cdot (\pi(\{y\})) + \pi(\{b\}))) = 100 \cdot \frac{2}{3} = 66.\bar{6}$$

$$\overline{\mathbb{E}}_{\mathcal{M}}(u_{b_2}) := \sup_{\pi\in\mathcal{M}} \mathbb{E}_\pi(u_{b_2}) = \sup_{\pi\in\mathcal{M}} (100\cdot(\pi(\{y\})+\pi(\{b\}))) = 100\cdot\frac{2}{3} = 66.\bar{6}$$

$$\Rightarrow \mathbb{E}_{\mathcal{M}}(u_{b_2}) = [66.\bar{6}]$$

Next, we determine the preference orders for the two decision problems with respect to any of the three interval orderings discussed.

- *Preference order induced by $\leqslant^p$:* For the two decision problems we have

  $$\mathbb{E}_{\mathcal{M}}(u_{a_2}) \leqslant^p \mathbb{E}_{\mathcal{M}}(u_{a_1}) \text{ and } \mathbb{E}_{\mathcal{M}}(u_{b_1}) \leqslant^p \mathbb{E}_{\mathcal{M}}(u_{b_2})$$

  Hence, according to the pessimistic ordering, we prefer action a_1 before action a_2 and action b_2 before action b_1. This preference order coincides with the majority preference order of the experts in Ellsberg's experiment. Hence, this description of the uncertainty underlying the situation avoids the inconsistencies arising from the classical description.

- *Preference order induced by $\leqslant^o$:* For the two decision problems we have

  $$\mathbb{E}_{\mathcal{M}}(u_{a_1}) \leqslant^o \mathbb{E}_{\mathcal{M}}(u_{a_2}) \text{ and } \mathbb{E}_{\mathcal{M}}(u_{b_2}) \leqslant^o \mathbb{E}_{\mathcal{M}}(u_{b_1})$$

  Hence, according to the opimistic ordering, we prefer action a_2 before action a_1 and action b_1 before action b_2. The preference order induced by $\leqslant^o$ doesn't coincide wih the majority preference order of the experts.

- *Preference order induced by $\leqslant^\alpha$:* As we know that the previous ordering are special cases of the ordering $\leqslant^\alpha$, we can already state the following: What preference order is induced by $\leqslant^\alpha$ strongly depends on the value of $\alpha \in [0,1]$. We compute

  $$\begin{aligned} & \alpha\cdot\underline{\mathbb{E}}_{\mathcal{M}}(u_{a_2}) + (1-\alpha)\cdot\overline{\mathbb{E}}_{\mathcal{M}}(u_{a_2}) \leqslant \alpha\cdot\underline{\mathbb{E}}_{\mathcal{M}}(u_{a_1}) + (1-\alpha)\cdot\overline{\mathbb{E}}_{\mathcal{M}}(u_{a_1}) \\ \Leftrightarrow \quad & (1-\alpha)\cdot 66.\bar{6} \leqslant 33.\bar{3} \\ \Leftrightarrow \quad & \frac{1}{2} \leqslant \alpha \end{aligned}$$

  and

  $$\begin{aligned} & \alpha\cdot\underline{\mathbb{E}}_{\mathcal{M}}(u_{b_1}) + (1-\alpha)\cdot\overline{\mathbb{E}}_{\mathcal{M}}(u_{b_1}) \leqslant \alpha\cdot\underline{\mathbb{E}}_{\mathcal{M}}(u_{b_2}) + (1-\alpha)\cdot\overline{\mathbb{E}}_{\mathcal{M}}(u_{b_2}) \\ \Leftrightarrow \quad & \alpha\cdot 33.\bar{3} + (1-\alpha)\cdot 100 \leqslant 66.\bar{6} \\ \Leftrightarrow \quad & \frac{1}{2} \leqslant \alpha \end{aligned}$$

Hence, the ordering $\leqslant^{\alpha}$ is compatible with the preferences of the experts whenever $a \geqslant \frac{1}{2}$. ⋆

The previous example showed that a generalized concept of the principle of maximizing expected utility can help to avoid inconsistencies when modelling preference orders of experts. However, it also demonstrated that not every interval ordering induces a suitable preference order for every situation. The ordering that is applied has to reflect the decision maker's attitude towards ambiguity suitably.

5 Decision making under complex uncertainty

Roughly spoken, the following chapter can be viewed as a generalization of the previous Chapters 1 and 3 to the case, where only imperfect information about the mechanism generating the states of nature is available. More precisely, this is meant in the following sense:

In Chapter 1 we recalled criteria for optimal decision making, if the uncertainty type underlying the CDP is describable by using classical probability theory (Bernoulli-criterion, see Paragraphs 1.2.3 and 3.2) or can be compared to a game against nature (Maximin-criterion, see Paragraphs 1.2.4 and 3.4).

In both cases, the uncertainty underlying the process generating the states of nature is in some form predictable (and therefore *precise*): Under strict type I uncertainty (see Paragraph 1.2.2) the exact expected utility of every action is *known* to the actor. Hence, there is no uncertainty about this expectation, but only about which state occurs in a concrete situation: The uncertainty is purely *stochastic* and there is no *ambiguity* (see Paragraph 4.2.2) underlying the situation.

Under strict type II uncertainty (see Paragraph 1.2.2) actually there exists no uncertainty at all (thus, maybe the term type II *uncertainty* is a bit misleading in this context): For every action chosen, the actor *knows* the state of nature being realized by the antagonist. The antagonist will always pick one of the states minimizing the actor's utility under the chosen action.

In Chapter 3 we discussed how such optimal decisions can be determined using the methods of linear optimization. It turned out that the optimization of all of the three criteria discussed can be reformulated as a linear programming problem and, therefore, be resolved computationally by applying standard statistical software.

However, as shown in Chapter 4, there exist situations under uncertainty that are *not* characterizable by a classical probability measure (see the motivating examples of Paragraph 4.1 in particular). That is, there are situations in which the available information is neither sufficient to specify a prior probability measure on the set of states nor to be sure that the the situation can be compared to a game against an omniscient enemy. As seen before, in such situations we need a more general way of measuring uncertainty, namely the concepts of interval probability and credal sets.

This leads us to the main questions of the present chapter: If we find ourselves in a decision situation with only imprecise information available, are there still criteria to determine optimal decisions? How can the principle of maximizing expected

utility suitably be generalized to the case where the expectation of the random variables associated with the actions are interval quantities? And: Is it still possible to determine optimal actions with the methods of linear optimization? Subsequently, we will see it is. However, it also turns out that the choice of the 'right' decision criterion is by far less obvious as in the precise case, since it is strongly connected to the attitude of the decision maker towards ambiguity (see also [3, p. 13]).

The chapter is structured as follows: First, we explain the concepts of *Interval dominance* and *E-Admissibility*. Both criteria can be viewed as a generalization of the concept of admissibility, in the sense that they are *independent* of which element contained in the credal set is the true measure (and, therefore, independent of the decision maker's attitude towards ambiguity). Nevertheless, E-admissibility turns out to induce a pretty strong ordering.

Afterwards, we recall two ways of cautiously generalizing the concept of maximizing the expected utility, namely *Maximality* and the Γ-*Maximin-criterion*. More precisely, the generalization is done in a way that takes only the lower bound of the interval expectation into account and, therefore, is based on the pessimistic interval ordering $\leqslant^p$ (see Paragraph 4.2.3). Both criteria are equivalent for situations with precise (or perfect) information available (that is, if the corresponding credal set consists of one single element). However, they induce different orders when considering situations under imperfect information.

In the last paragraph, we first explain the Γ-*Maximax-criterion*. It can be viewed as the *optimistic counterpart* of the Γ-Maximin-criterion: Instead of considering only the lower expectations, here only the upper expectations are taken into account. Accordingly, the criterion is based on the optimistic interval ordering $\leqslant^o$ (see Paragraph 4.2.3). Subsequently, we recall a decision criterion that allows the decision maker to flexibly model his individual attitude towards ambiguity. Specifically, the *degree* of optimism can be modelled by a real valued parameter $\eta \in [0, 1]$. The corresponding criterion then is based on the ordering $\leqslant^\eta$ (see Paragraph 4.2.3).

For all criteria under consideration, we recall algorithms for determining optimal decisions. All of these algorithms are based on methods of linear optimization (see Chapter 2). Additionally, we apply duality theory (see Paragraph 2.2) to learn more about the characteristics of optimal solutions and to make up connections between the different criteria.

We mainly refer to the works of [23], [40] and [43, § 3.9]. Detailed references are given at the appropriate places in the text.

5.1 Assumptions on the type of uncertainty and the structure of the available information

Before we can start, we need to make a couple of assumptions: In the whole chapter, we assume decision problems $\mathfrak{A} = (\mathbb{A}, \Theta, u(\cdot))$, where $|\Theta| < \infty$, to be a CDP under *theoretical* type I uncertainty (or type I* uncertainty, see Paragraph 1.2.2). That is, given perfect information, there exists a probability measure perfectly characterizing the uncertainty between the different elements of Θ.

However, we assume our information to be *imperfect*. More precisely, we assume our information concerning the states of nature can be expressed by a set $\mathcal{M}$ of probability measures on the measurable space $(\Theta, \mathcal{P}(\Theta))$ of the following form:

$$\mathcal{M} := \Big\{\pi : \pi \text{ is pm on } (\Theta, \mathcal{P}(\Theta)) \wedge \underline{b}_i \leqslant \mathbb{E}_\pi(f_i) \leqslant \overline{b}_i \ \forall i = 1, ..., r\Big\} \tag{86}$$

where, for all $i = 1, ..., r$,

- $(\underline{b}_i, \overline{b}_i) \in \mathbb{R}^2$ such that $\underline{b}_i \leqslant \overline{b}_i$
- $f_i : \Theta \to \mathbb{R}$

That is, all the information available on the true measure generating the states of nature can be described by *lower* and *upper bounds* of expectations of real-valued random variables on the set Θ (the presentability of the available information as a set of the above form is a common assumption, see for example [41, § 2]).

Then, the following holds for the set $\mathcal{M}$.

Theorem 14. Let $\mathcal{M}$ be the set defined in equation (86). Then, $\mathcal{M}$ is a convex set and there exist $\pi_1, \pi_2 \in \mathcal{M}$ such that

$$\mathbb{E}_{\pi_1}(X) = \min_{\pi \in \mathcal{M}} \mathbb{E}_{\pi_1}(X) \quad \text{and} \quad \mathbb{E}_{\pi_2}(X) = \max_{\pi \in \mathcal{M}} \mathbb{E}_{\pi_1}(X)$$

for all random variables $X : \Theta \to \mathbb{R}$.

Proof. Convexity: As $\mathcal{M}$ is a credal set, it is non-empty by assumption. So, let $p_1, p_2 \in \mathcal{M}$ and $\alpha \in [0, 1]$. Clearly, $p_\alpha := \alpha \cdot p_1 + (1 - \alpha) \cdot p_2$ defines a probability measure on $(\Theta, \mathcal{P}(\Theta))$.
Now, let $i \in \{1, \dots, r\}$ be arbitrary. Then, just like in the proof of Theorem 13, we can derive the following identity

$$\mathbb{E}_{p_\alpha}(f_i) = \alpha \cdot \mathbb{E}_{p_1}(f_i) + (1 - \alpha) \cdot \mathbb{E}_{p_2}(f_i)$$

Since $\underline{b}_i \leqslant \mathbb{E}_{p_k}(f_i) \leqslant \overline{b}_i$ for $k = 1, 2$ by assumption, this implies that

$$\underline{b}_i \leqslant \mathbb{E}_{p_\alpha}(f_i) \leqslant \overline{b}_i$$

Hence, $p_\alpha \in \mathcal{M}$. Since i and α were chosen arbitrarily, this implies that $\mathcal{M}$ is convex.

It remains to show that, for any $X : \Theta \to \mathbb{R}$, maximum and minimum expectation are attained on the set $\mathcal{M}$. This is equivalent to the optimization problems

$$\mathbb{E}_\pi(X) = \sum_{j=1}^{m} X(\theta_j) \cdot \pi(\{\theta_j\}) \longrightarrow \min_{\pi \in \mathcal{M}} \quad \text{and} \quad \mathbb{E}_\pi(X) = \sum_{j=1}^{m} X(\theta_j) \cdot \pi(\{\theta_j\}) \longrightarrow \max_{\pi \in \mathcal{M}}$$

having optimal solutions. According to the considerations in Paragraph 3.3, this is equivalent to the optimization problems

$$(X(\theta_1), \ldots, X(\theta_m)) \cdot \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \longrightarrow \min_{\pi \in \mathbb{R}^m} \ / \ \max_{\pi \in \mathbb{R}^m} \tag{87}$$

with constraints

- $\pi \in \Pi_m$

having optimal solutions. To see the linearity of the constraints, note their equivalence to

- $\pi \geqslant 0$
- $K \cdot \pi \geqslant l \quad / \quad -K \cdot \pi \leqslant -l$

where

$$K := \begin{pmatrix} 1 & \ldots & 1 \\ 1' & \ldots & 1' \\ f_{11} & \ldots & f_{1m} \\ f'_{11} & \ldots & f'_{1m} \\ \vdots & \ldots & \vdots \\ f_{r1} & \ldots & f_{rm} \\ f'_{r1} & \ldots & f'_{rm} \end{pmatrix} \qquad l := \begin{pmatrix} 1 \\ 1' \\ \underline{b}_1 \\ \overline{b}'_1 \\ \vdots \\ \underline{b}_r \\ \overline{b}'_r \end{pmatrix}$$

Hence, the optimization problems from (87) define a SMP/SMIP. Since the set $\mathcal{M}$ (and therefore Π_m) is non-empty by assumption and the objective function is bounded on Π_m, both problems from (87) have optimal solutions according to Theorem 5. This completes the proof. □

As an immediate consequence of the above Theorem 14 and Corollary 2 from Paragraph 4.2.1 we receive the following statement. It shows up a deep connection between the concept of Cred($\mathcal{M}$)-expectation and the concept of interval expectation with respect to $\mathcal{M}$.

Corollary 3. Let $X : \Theta \to \mathbb{R}$ be any random variable. Then, the following identity holds for the $\mathcal{M}$-expectation of X:

$$\mathbb{E}^C_{\mathcal{M}}(X) = \Big[\min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(X), \max_{\pi \in \mathcal{M}} \mathbb{E}_\pi(X) \Big] =: \mathbb{E}_{\mathcal{M}}(X) \tag{88}$$

Proof. According to Theorem 14 the set $\mathcal{M}$ is convex and arbitrary random variables attain their maximum and minimum expectations on elements of $\mathcal{M}$. Thus, the result directly follows from Corollary 2. □

Finally, the following theorem proves that the structure of an F-probability on a finite space can always be written as a set of the form defined in (86).

Theorem 15. Let $P(\cdot) = [L(\cdot), U(\cdot)]$ be an F-probability on the measurable space $(\Theta, \mathcal{P}(\Theta))$. Then, the structure $\mathcal{M}_0$ of P is of the form defined in (86).

Proof. Since $|\Theta| = m$, we have that $|\mathcal{P}(\Theta)| = 2^m =: K$. So, let

$$\mathcal{P}(\Theta) = \{A_1, \ldots, A_K\}$$

be a listing of all elements contained in $\mathcal{P}(\Theta)$. For all $i = 1, \ldots, K$, define

$$(\underline{b}_i, \overline{b}_i) := (L(A_i), U(A_i)) \qquad \text{and} \qquad f_i := \mathbb{1}_{A_i}$$

Then, the following identity holds:

$$\mathcal{M}_0 = \Big\{ \pi : \pi \text{ is pm on } (\Theta, \mathcal{P}(\Theta)) \wedge \underline{b}_i \leqslant \mathbb{E}_\pi(f_i) \leqslant \overline{b}_i \ \forall i = 1, ..., K \Big\} =: D \tag{89}$$

Proof of (89): $\subseteq$: Let $\pi \in \mathcal{M}_0$. Then, by definition, we have

$$L(A_i) \leqslant \pi(A_i) \leqslant U(A_i)$$

for all $i = 1, \ldots, K$. Since $\pi(A_i) = \mathbb{E}_\pi(\mathbb{1}_{A_i}) =: \mathbb{E}_\pi(f_i)$ for all $i = 1, \ldots, K$, we then have that

$$\underline{b}_i := L(A_i) \leqslant \mathbb{E}_\pi(f_i) \leqslant U(A_i) =: \overline{b}_i$$

for all $i = 1, \ldots, K$. Hence, $\pi \in D$.

$\supseteq$: Let $\pi \in D$. That is,

$$\underline{b}_i := L(A_i) \leqslant \pi(A_i) = \mathbb{E}_\pi(f_i) \leqslant U(A_i) =: \overline{b}_i$$

for all $i = 1, \ldots, K$. Hence, $\pi \in \mathcal{M}_0$. Thus, the structure $\mathcal{M}_0$ is representable as a set of the form (86). □

The above considerations have shown that the information structure given by the set $\mathcal{M}$ from (86) can be viewed in two different ways:

1. The set $\mathcal{M}$ directly is interpreted as a credal set on the measurable space $(\Theta, \mathcal{P}(\Theta))$. Hence, we make the assumption that we can express our information as a convex credal set on which maximum and minimum expectations of arbitrary random variables are attained.

2. The set $\mathcal{M}$ is interpreted as the structure of some F-probability on the measurable space $(\Theta, \mathcal{P}(\Theta))$. As, according to Theorem 15, the structure of every F-probability can be represented in the form (86), there are no assumptions on the F-probability at all.

Due to the special form of the set $\mathcal{M}$, for the considerations in the following paragraphs it makes no difference whether interpretation 1. or 2. is used: According to Corollary 3, the concept of interval expectation of an F-probability with structure $\mathcal{M}$ and the concept of $\mathrm{Cred}(\mathcal{M})$ expectation coincide.

More precisely, we have the following: Let P be an F-probability with structure $\mathcal{M}$, Then, for all random variables $X : \Theta \to \mathbb{R}$, we have

$$\mathbb{E}^C_{\mathcal{M}}(X) = \mathbb{E}_{\mathcal{M}}(X)$$

Particularly, this implies $P^C(A) = P(A)$ for all $A \in \mathcal{P}(\Theta)$. Hence, when considering the generalized expectation, it makes no difference if the uncertainty underlying a CDP is characterized by a credal set of the form $\mathcal{M}$ or an F-probability with structure $\mathcal{M}$.

5.2 Interval dominance

In Paragraph 4.2.3 we saw the following: Any interval ordering that is reasonable in the context of a generalized concept of maximizing expected utility necessarily has (at least) to be compatible with the ordering $<^I$. Specifically, if the highest possible

expected utility of an action a_1 under all possible versions of reality (that is, the highest expectation of u_{a_1} under every $\pi \in \mathcal{M}$) is strictly smaller than the lowest possible expected utility of an action a_2 under all possible versions of reality (that is, the lowest expectation of the random variable u_{a_2} under every $\pi \in \mathcal{M}$), then it is unreasonable to choose this action.

This principle is independent of the type of uncertainty between the different measures contained in the set $\mathcal{M}$ (and therefore does not depend on the actor's attitude towards ambiguity): Such an action is unreasonable, no matter what the true measure is. This motivates the following Definition. It can for example be found in [23, Definition 8.5].

Definition 21. Let $\mathfrak{A}$ denote any CDP such that $|\Theta| < \infty$. Consider the uncertainty is described by an F-probability P on $(\Theta, \mathcal{P}(\Theta))$ with structure $\mathcal{M}$, or by the credal set $\mathcal{M}$. Then, an action $a_1 \in \mathbb{A}$ is said to *interval dominate* an action $a_2 \in \mathbb{A}$, if the following holds:

$$\mathbb{E}_{\mathcal{M}}(u_{a_2}) <^I \mathbb{E}_{\mathcal{M}}(u_{a_1})$$

According to the definition of the relation $<^I$, that is $\underline{\mathbb{E}}_{\mathcal{M}}(u_{a_1}) > \overline{\mathbb{E}}_{\mathcal{M}}(u_{a_2})$.
If an action $a^* \in \mathbb{A}$ is not interval dominated by any action $a \in \mathbb{A}$, that is

$$\overline{\mathbb{E}}_{\mathcal{M}}(u_{a^*}) \geqslant \underline{\mathbb{E}}_{\mathcal{M}}(u_a) \tag{90}$$

for all $a \in \mathbb{A}$, we say that a^* is $<^I$ $(\mathcal{M})$-*admissible*. The set of all $<^I$ $(\mathcal{M})$-admissible actions is denoted by $\mathbb{A}^{<^I}_{\mathcal{M}}$. ▽

The aim of the following example is to illustrate to different aspects: First, it shows that there might exist admissible actions that are not $<^I$ $(\mathcal{M})$-admissible. Second, it demonstrates the fact that an action, which is $<^I$ $(\mathcal{M})$-admissible in the basic problem, may become $<^I$ $(\mathcal{M})$-inadmissible when considering the mixed extension.

Example 11. Consider the CDP $\mathfrak{A}$ defined by the following table:

u_{ij}	θ_1	θ_2	θ_3	θ_4
a_1	20	15	10	30
a_2	25	10	10	20
a_3	20	31	0	5
a_4	10	30	50	30
a_5	0	30	20	40
a_6	26	0	0	1

Obviously, all of the actions $a_1, ..., a_6$ are admissible (in the sense of Definition 4). Additionally, let our information concerning the mechanism generating the states of nature be given by the following set of probability measure on $(\Theta, \mathcal{P}(\Theta))$:

$$\mathcal{M} := \{\pi : \pi_1 + \pi_3 \leqslant 0.8 \wedge \pi_4 \geqslant 0.4\}$$

Clearly, the above set $\mathcal{M}$ is of the form defined in equation (86). Thus, according to the consideration in Paragraph 5.1, we can interpret $\mathcal{M}$ as a convex credal set or the structure of an F-probability respectively: The Cred($\mathcal{M}$)-expectation and the interval expectation w.r.t. $\mathcal{M}$ coincide for arbitrary random variables $X : \Theta \to \mathbb{R}$.

Accordingly, we can compute the expectations (interval- or credal-expectation respectively) of every random variable u_{a_i}, $i = 1, ..., 6$:

$$\mathbb{E}_{\mathcal{M}}(u_{a_i}) = \begin{cases} [18, 30] & \text{if } i = 1 \\ [14, 23] & \text{if } i = 2 \\ [2, 20.6] & \text{if } i = 3 \\ [18, 42] & \text{if } i = 4 \\ [16, 40] & \text{if } i = 5 \\ [0.4, 16] & \text{if } i = 6 \end{cases}$$

Immediately, we see that action a_6 is interval dominated by both action a_1 and action a_4. Therefore, action a_6 is not $<^I$ ($\mathcal{M}$)-admissible even though it is admissible. The remaining actions $a_1, ..., a_5$ are $<^I$ ($\mathcal{M}$)-admissible.

Now, consider the randomized action $p^* \in G(\mathbb{A})$ induced by the assignment

$$p^*(\{a_i\}) = \begin{cases} 0.8 & \text{if } i = 1 \\ 0 & \text{if } i = 2 \\ 0 & \text{if } i = 3 \\ 0.2 & \text{if } i = 4 \\ 0 & \text{if } i = 5 \\ 0 & \text{if } i = 6 \end{cases}$$

We compute the lower bound of the interval expectation of $G(u)_{p^*}$:

$$\underline{\mathbb{E}}_{\mathcal{M}}(G(u)_{p^*}) \quad = \quad \inf_{\pi \in \mathcal{M}} \left[\sum_{j=1}^{4} G(u)(p^*, \theta_j) \cdot \pi(\{\theta_j\}) \right]$$

$$
\begin{aligned}
&= \inf_{\pi\in\mathcal{M}}\Big[\sum_{j=1}^{4}\Big(\sum_{i=1}^{6}u(a_i,\theta_j)\cdot p^*(\{a_i\})\Big)\cdot\pi(\{\theta_j\})\Big]\\
&= \inf_{\pi\in\mathcal{M}}\Big[18\cdot(\pi(\{\theta_1\})+\pi(\{\theta_2\})+\pi(\{\theta_3\}))+30\cdot\pi(\{\theta_4\})\Big]\\
&= 18\cdot 0.6+30\cdot 0.4\\
&= 22.8
\end{aligned}
$$

Hence, we have that $\underline{\mathbb{E}}_{\mathcal{M}}(G(u)_{p^*}) > \overline{\mathbb{E}}_{\mathcal{M}}(u_{a_3})$. Thus, there exists a randomized action p^* that interval dominates action a_3. Thus, an action that is $<^I$ $(\mathcal{M})$-admissible in a finite CDP $\mathfrak{A}$ can become $<^I$ $(\mathcal{M})$-inadmissible under the transition to the mixed extension $G(\mathfrak{A})$.

Implicitly, another aspect is illuminated by the previous example: Applying interval dominance, in general, doesn't bring us to a satisfying decision. In our example the actions $a_1,\dots,a_5$ remain incomparable with respect to this criterion. Hence, similar as admissibility, interval dominance actually does not define a decision criterion, but a criterion for labelling certain actions as being completely irrational choices. In order to be able to further distinguish between the incomparable actions, we need decision criteria that induce a stronger ordering. ⋆

5.3 E-Admissibility

In the previous paragraph, we recalled the *principle of excluding interval dominated actions*. More precisely, we saw that, independently of what we assume about the elements of $\mathcal{M}$, an interval dominated action can never be a reasonable choice. However, in general, the preference order induced by this principle is too weak to determine an optimal decision: To many action remain incomparable with respect to the order$<^I$.

Next, we recall a decision criterion, which also is independent of our assumptions on how the elements of $\mathcal{M}$ are chosen (and, therefore, of our attitude towards ambiguity). It is called *E-admissibility*. In contrast to interval dominance, the preference order induced by this criterion turns out to be pretty strong: As we will see in Theorem 19, every E-admissible action turns out to be also *maximal* in the sense of Definition 23. However, in contrast to E-admissibility, *maximality* is strongly dependent of he decision maker's assumptions on the set $\mathcal{M}$.

Before we can start, some preparation work has to be done: Recall that, according to Theorem 2, in every finite CDP there exists a Bayes-action w.r.t. any arbitrary probability measure on the set of states. Thus, for all $\pi\in\mathcal{M}$, there exists a Bayes-action $a^*\in\mathbb{A}$ w.r.t. π. But $\mathcal{M}$ contains all the measures being compatible with the

given information. Hence, measures that are not contained in $\mathcal{M}$ are incompatible with the given information and, therefore, incompatible with the reality to describe.

Now, to the crucial point: Can it be reasonable to choose an action that is not a Bayes-action for no matter what $\pi \in \mathcal{M}$? Remember, we know that there always exists such an action. In other words: Given the information that there always exists a better one, why should we choose an action that is not optimal in every version of reality being compatible with our information base? Of course, we shouldn't. Otherwise, we would wilfully accept a loss of utility, which contradicts the assumption of a *rational actor* (see Introduction and/or the discussion in Remark 3 on Definition 1 for further detail). This motivates the following definition. In a slightly different notation, it can for example be found in [23, Definition 8.7].

Definition 22. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$. Define the set $\mathcal{M}$ like in equation (86). Then an action $a^* \in \mathbb{A}$ is called *E-admissible* with respect to $\mathcal{M}$ (or short $E(\mathcal{M})$-*admissible*), if there exists a measure $\pi^* \in \mathcal{M}$ such that a^* is Bayes-optimal with respect to π^*. That is

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) \geqslant \Phi^{\mathfrak{A}}_{B(\pi^*)}(a) \tag{91}$$

for all $a \in \mathbb{A}$. The set

$$\mathbb{A}_{E(\mathcal{M})} := \{a \in \mathbb{A} : a \text{ is E-admissible w.r.t. } \mathcal{M}\} \tag{92}$$

then is called *set of* $E(\mathcal{M})$-*admissible actions*. Every $a \in \mathbb{A} \setminus \mathbb{A}_{E(\mathcal{M})}$ is said to be $E(\mathcal{M})$-*inadmissible*. ▽

Remark. Note that, in the absence of ambiguity (i.e. if the set $\mathcal{M}$ consists of one single element and the uncertainty is purely stochastic), E-admissibility reduces to the classical Bayes-criterion (with respect to the only element contained in $\mathcal{M}$). ∘

So, we have a criterion that labels certain actions as not being a reasonable choice. Two questions immediately come to mind: How can we decide, whether an action $a \in \mathbb{A}$ is $E(\mathcal{M})$-admissible or not? And: Can actions that are $E(\mathcal{M})$-admissible in a finite (basic) CDP $\mathfrak{A}$ become $E(\mathcal{M})$-inadmissible when considering the mixed extension $G(\mathfrak{A})$?

Let us answer the second question first: Suppose, the action $a^* \in \mathbb{A}$ is $E(\mathcal{M})$-admissible in a finite CDP $\mathfrak{A}$. Then, by definition, a^* is a Bayes-action w.r.t. some $\pi^* \in \mathcal{M}$. However, according to Theorem 2, we then have $\mathbb{E}_{\pi^*}(u_{a^*}) \geqslant \mathbb{E}_{\pi^*}(G(u)_p)$ for all $p \in G(\mathbb{A})$. That is, δ_{a^*} (and therefore a^*) is Bayes-action w.r.t. π^* in $G(\mathfrak{A})$.

Hence, a^* is $E(\mathcal{M})$-admissible in the mixed extension as well. Thus, randomization cannot improve the utility with respect to E-admissibility. Therefore, it suffices to take only pure actions into account.

So, how can we decide whether a pure action is $E(\mathcal{M})$-admissible or not? Again, it turns out that, at least for the case of a finite CDP, this can be done by solving suitable linear programming problems. Subsequently, we recall one possible algorithm to decide, whether an action under consideration is $E(\mathcal{M})$-admissible or not. It goes back to *Augustin* and *Utkin* in [41, § 5].

Let $\mathfrak{A}$ denote a finite CDP under consideration. Then, according to condition (91), an action $a_l \in \mathbb{A}$, where $l \in \{1, \dots, n\}$ is fixed, is $E(\mathcal{M})$-admissible, if there exists $\pi^* \in \mathcal{M}$ such that

$$\sum_{j=1}^{m} u(a_l, \theta_j) \cdot \pi^*(\{\theta_j\}) \geqslant \sum_{j=1}^{m} u(a_k, \theta_j) \cdot \pi^*(\{\theta_j\}) \tag{93}$$

for all $k \in \{1, \dots, n\}$. This is the case if, and only if, the set

$$\mathcal{H}_l := \Big\{ \pi \in \mathcal{M} : a_l \text{ is Bayes-action w.r.t. } \pi \Big\} \tag{94}$$

is non-empty.

Next, note that for an action $a_l \in \mathbb{A}$ the set $\mathcal{H}_l$ is non-empty if, and only if, the set

$$\mathcal{F}_l := \{ g(\pi) : \pi \in \mathcal{H}_l \} \subset \Delta_m \tag{95}$$

is non-empty, where b is the bijective transformation map defined in the proof of Theorem 9. But how can we check non-emptiness of the set $\mathcal{F}_l$ using linear programming?

For a vector $\pi \in [0,1]^m$, define the following three conditions:

(C1) $\sum_{j=1}^{m} \pi_j \leqslant 1$

(C2) $\underline{b}_i \leqslant \sum_{j=1}^{m} f_i(\theta_j) \cdot \pi_j \leqslant \overline{b}_i \ \forall i = 1, ..., r$

(C_l) $\sum_{j=1}^{m} u(a_l, \theta_j) \cdot \pi(\{\theta_j\}) \geqslant \sum_{j=1}^{m} u(a_k, \theta_j) \cdot \pi(\{\theta_j\})$ for all $k \in \{1, \dots, n\}$

Then, the set $\mathcal{F}_l$ (and therefore $\mathcal{H}_l$) is non-empty, if, and only if, the set

$$\mathcal{F}_l^- := \Big\{ \pi \in [0,1]^m : (C1) \ \wedge \ (C2) \ \wedge \ (C_l) \Big\} \tag{96}$$

contains an element π^* such that $\sum_{j=1}^{m} \pi_j^* = 1$. That is, if $\mathcal{F}_l$ contains an element,

which is maximal with respect to the $\|\cdot\|_1$ norm on $\mathcal{F}_l$ (and therefore corresponds to a probability measure).

Hence, the set $\mathcal{H}_l$ is non-empty if, and only if, the optimal outcome of the optimization problem

$$\sum_{j=1}^{m} \pi_j \longrightarrow \max_{\pi \in \mathcal{F}_l^-} \tag{97}$$

equals 1. With both the conditions (C1), (C2) and (C_l) and the objective function being linear, the optimization problem (97) is a SMP (in the sense of Definition 11). More precisely, we have that (97) is equivalent to the linear programming problem

$$\underbrace{(1, ..., 1)}_{\text{m-times}} \cdot \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \longrightarrow \max_{\pi \in \mathbb{R}^m} \tag{98}$$

with constraints

- $\pi \geqslant 0$
- $\begin{pmatrix} \mathbf{1} \\ J_1 \\ J_2 \end{pmatrix} \cdot \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \leqslant (1, \underline{b}_1', \overline{b}_1, \ldots, \underline{b}_r', \overline{b}_r, \underbrace{0, \ldots, 0}_{\text{n-times}})^T$

Here, we used the following notations (where f_{ij}, $i = 1 \ldots, r$ and $j = 1, \ldots, m$, and u_{ij}, $i = 1 \ldots, n$ and $j = 1, \ldots, m$, are defined just like in the proof of Theorem 9):

- $\mathbf{1} := \underbrace{(1, \ldots, 1)}_{\text{m-times}}$
- $J_1 := \begin{pmatrix} f_{11}' & \cdots & f_{1m}' \\ f_{11} & \cdots & f_{1m} \\ \vdots & \ldots & \vdots \\ f_{r1}' & \cdots & f_{rm}' \\ f_{r1} & \ldots & f_{rm} \end{pmatrix}$
- $J_2 := \begin{pmatrix} (u_{11} - u_{l1}) & \ldots & (u_{1m} - u_{lm}) \\ \vdots & \ldots & \vdots \\ (u_{n1} - u_{l1}) & \ldots & (u_{nm} - u_{lm}) \end{pmatrix}$

Note that, since $l \in \{1, \ldots, n\}$, one of the rows of the matrix J_2 equals zero in every entry. This is not a problem, since it leads to the restriction $0 \leqslant 0$, which is trivially satisfied in any case.

Finally, we have the following: An action $a_l \in \mathbb{A}$ from the set of actions of a finite CDP is $E(\mathcal{M})$-admissible, if the optimal outcome of the linear programming problem (98) equals 1. But does the SMP (98) always have an optimal solution?

Since the boundedness of the objective function is already guaranteed by the condition (C1), Theorem 5 guarantees the existence of an optimal solution whenever the set of admissible solution of the SMP (98) is non-empty. However, there doesn't always exist $\pi \in [0,1]^m$ satisfying the conditions (C1), (C2) and (C_l) (that is, there doesn't necessarily exist an admissible solution of the SMP (98)). If this is the case, we also can conclude the emptiness of the set $\mathcal{H}_l$, since any element of $\mathcal{H}_l$ induces an optimal and, therefore, admissible solution of (98).

Let us apply the algorithm just described to an example.

Example 12. Consider the following CDP $\mathfrak{A}$:

u_{ij}	θ_1	θ_2	θ_3	θ_4
a_1	20	15	10	30
a_2	30	10	10	20
a_3	20	40	0	20

We have the following information concerning the states of nation: We now that the event $\{\theta_1, \theta_2\}$ has a probability of at least 0.5. Additionally, we know that state θ_4 is at least as probable as state θ_3. Hence, the linear constraints defining the set $\mathcal{M}$ from equation (43) are given by:

- $f_1(\theta) := \begin{cases} 1 & \text{if } \theta \in \{\theta_1, \theta_2\} \\ 0 & \text{else} \end{cases}$
- $f_2(\theta) := \begin{cases} 1 & \text{if } \theta = \theta_3 \\ -1 & \text{if } \theta = \theta_4 \\ 0 & \text{else} \end{cases}$
- $(\underline{b}_i, \overline{b}_i) := \begin{cases} (0.5, 1) & \text{if } i = 1 \\ (-0.5, 0) & \text{if } i = 2 \end{cases}$

We want to check $E(\mathcal{M})$-admissibility of the actions a_1, a_2 and a_3. Hence, there are three different linear optimization problems to be solved. We start with action a_1. The corresponding linear programming problem (98) is of the form:

$$(1,1,1,1) \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \longrightarrow \max_{\pi}$$

with constraints

- $\pi \geqslant 0$
- $$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 10 & -5 & 0 & -10 \\ 0 & 25 & -10 & -10 \end{pmatrix} \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -0.5 \\ 1 \\ 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Running the above optimization problem in R gives the optimal output 1. Thus, according to the above considerations, action a_1 is E($\mathcal{M}$)-admissible. More precisely, a_1 is a Bayes-action with respect to $\pi^{(1)} \in \mathcal{M}$, where the probability measure $\pi^{(1)}$ is induced by the assignment

$$\pi^{(1)}(\{\theta_i\}) = \begin{cases} 0.4615385 & \text{if } i = 1 \\ 0.1538462 & \text{if } i = 2 \\ 0 & \text{if } i = 3 \\ 0.3846154 & \text{if } i = 4 \end{cases}$$

Next, we check E($\mathcal{M}$)-admissibility for action a_2. We arrive at the following linear programming problem:

$$(1,1,1,1) \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \longrightarrow \max_{\pi}$$

with constraints

- $\pi \geqslant 0$
- $$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ -10 & 5 & 0 & 10 \\ 0 & 0 & 0 & 0 \\ -10 & 30 & -10 & 0 \end{pmatrix} \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -0.5 \\ 1 \\ 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Again, we receive 1 as the optimal outcome of the optimization problem. Thus, action a_2 is as well $\mathrm{E}(\mathcal{M})$-admissible. More precisely, a_2 is a Bayes-action with respect to $\pi^{(2)} \in \mathcal{M}$, where the probability measure $\pi^{(2)}$ is induced by the assignment

$$\pi^{(2)}(\{\theta_i\}) = \begin{cases} 0.5 & \text{if } i = 1 \\ 0 & \text{if } i = 2 \\ 0.25 & \text{if } i = 3 \\ 0.25 & \text{if } i = 4 \end{cases}$$

It remains to check the $\mathrm{E}(\mathcal{M})$-admissibility for action a_3. We arrive at the following linear programming problem:

$$(1, 1, 1, 1) \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \longrightarrow \max_{\pi}$$

with constraints

- $\pi \geqslant 0$
- $$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & -25 & 10 & 10 \\ 10 & -30 & 10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -0.5 \\ 1 \\ 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Again, the optimal outcome equals 1. Thus, action a_3 is E($\mathcal{M}$)-admissible. It is an Bayes-action with respect to $\pi^{(3)} \in \mathcal{M}$, where the probability measure $\pi^{(3)}$ is induced by the assignment

$$\pi^{(3)}(\{\theta_i\}) = \begin{cases} 0.75 & \text{if } i = 1 \\ 0.25 & \text{if } i = 2 \\ 0 & \text{if } i = 3 \\ 0 & \text{if } i = 4 \end{cases}$$

Finally, the set of $E(\mathcal{M})$-admissible actions is given by $\mathbb{A}_{E(\mathcal{M})} = \mathbb{A}$. ⋆

Next, we recall two (well-known) results that make up a connection between the concept of E($\mathcal{M}$)-admissibility and the concept of admissibility from classical decision theory as defined in Chapter 1 and the concept of $<^I$ ($\mathcal{M}$)-admissibility respectively.

We start with the statement that every action being Bayes-action with respect to a strict positive measure from $\mathcal{M}$ necessarily has to be admissible as well. It can also be found in [40, § 3.4].

Theorem 16. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$ and let $\mathcal{M}$ be any non-empty set of probability measures on $(\Theta, \mathcal{P}(\Theta))$. If there exists $\pi^* \in \mathcal{H}_l$ such that $\pi^*(\{\theta\}) > 0$ for all $\theta \in \Theta$, then the action $a_l \in \mathbb{A}$ is admissible.

Proof. Let $\pi^* \in \mathcal{H}_l$ such that $\pi^*(\{\theta\}) > 0$ for all $\theta \in \Theta$. Assume, for contradiction, that there exists $a^* \in \mathbb{A}$ such that $a^* \gg a_l$. That is, the following two conditions are satisfied:

1. $u(a^*, \theta) \geqslant u(a_l, \theta)$ for all $\theta \in \Theta$
2. $u(a^*, \theta) > u(a_l, \theta)$ for at least one $\theta \in \Theta$

Since $\pi^*(\{\theta\}) > 0$ for all $\theta \in \Theta$, the above conditions imply the following inequality

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a_l) = \sum_{j=1}^{m} u(a_l, \theta_j) \cdot \pi^*(\{\theta_j\}) < \sum_{j=1}^{m} u(a^*, \theta_j) \cdot \pi^*(\{\theta_j\}) = \Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) \tag{99}$$

This yields a contradiction, since a_l was assumed to be a Bayes-action with respect to π^*. Hence, the action a_l is admissible. □

Remark. Theorem 16 is optimal in the following sense: The condition of the strict positivity of the measure π^* is crucial and cannot be dropped. Additionally, admissibility doesn't necessarily imply E($\mathcal{M}$)-admissibility. Both properties can easily be illustrated by the following two (minimal) examples. ∘

Example 13. Consider the following CDP $\mathfrak{A}$:

u_{ij}	θ_1	θ_2
a_1	20	15
a_2	30	10

Further, the set $\mathcal{M}$ consists of only one element, namely the probability measure π^* induced by the assignment $\pi^*(\{\theta_1\}) = 0.5$. Then, action a_1 is admissible. However, it is not $\mathrm{E}(\mathcal{M})$-admissible, since

$$17.5 = \Phi^{\mathfrak{A}}_{B(\pi^*)}(a_1) < \Phi^{\mathfrak{A}}_{B(\pi^*)}(a_2) = 20$$

That is, action a_1 is not a Bayes-actions with respect to any element of $\mathcal{M}$ (which is only one). ⋆

Example 14. Consider the following CDP $\mathfrak{A}$:

u_{ij}	θ_1	θ_2
a_1	20	10
a_2	30	10

Further, the set $\mathcal{M}$ consists of only one element, namely the probability measure π^* induced by the assignment $\pi^*(\{\theta_1\}) = 0$. Then, action a_1 is a Bayes-action w.r.t. the measure π^*, since

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a_1) = 10 = \Phi^{\mathfrak{A}}_{B(\pi^*)}(a_2)$$

Therefore, action a_1 is $\mathrm{E}(\mathcal{M})$-admissible. However, a_1 is inadmissible. This shows necessity of the condition of strict positivity in Theorem 16. ⋆

The next theorem states that the concept of E-admissibility is compatible with interval dominance. More precisely, it guarantees that an E-admissible action never appears to be interval dominated by any other action. Similar, it can be found in [40, A.1., Theorem 1].

Theorem 17. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$ and let $\mathcal{M}$ be any non-empty set of probability measures on $(\Theta, \mathcal{P}(\Theta))$. Then, for all actions $a \in \mathbb{A}$, the following implication holds:

$$a \in \mathbb{A}_{E(\mathcal{M})} \Rightarrow a \in \mathbb{A}^{<^I}_{\mathcal{M}} \tag{100}$$

That is, any $E(\mathcal{M})$-admissible action is automatically $<^I$ $(\mathcal{M})$-admissible.

Proof. Let $a^* \in \mathbb{A}_{E(\mathcal{M})}$. That is, there exists $\pi \in \mathcal{M}$ such that

$$\mathbb{E}_\pi(u_{a^*}) \geqslant \mathbb{E}_\pi(u_a)$$

for all $a \in \mathbb{A}$. Additionally, we have

$$\overline{\mathbb{E}}_\mathcal{M}(u_a) \geqslant \mathbb{E}_\pi(u_a) \geqslant \underline{\mathbb{E}}_\mathcal{M}(u_a)$$

for all $a \in \mathbb{A}$. Since $a^* \in \mathbb{A}$, this implies

$$\overline{\mathbb{E}}_\mathcal{M}(u_{a^*}) \geqslant \mathbb{E}_\pi(u_{a^*}) \geqslant \mathbb{E}_\pi(u_a) \geqslant \underline{\mathbb{E}}_\mathcal{M}(u_a)$$

for all $a \in \mathbb{A}$. Hence, $a \in \mathbb{A}^{\lesssim^I}_\mathcal{M}$. □

Finally, let's once again summarize the crucial aspects of this paragraph: An action is $E(\mathcal{M})$-admissible, if it is Bayes-optimal with respect to (at least) one prior contained in the credal set $\mathcal{M}$. Furthermore, when considering E-admissibility, it is sufficient to take only pure action into account: E-admissibility is robust under the transition to the mixed extension. Additionally, if the basic problem is finite and the credal set is of the form defined in equation (86), then E-admissibility can be checked by resolving suitable linear optimization problems.

5.4 Maximality

In the previous two paragraphs we recalled the decision criteria E-admissibility and interval dominance. For both criteria, there are no further assumptions concerning the actor's risk attitude necessary: An action that is interval dominated or E-inadmissible respectively, is an unreasonable choice with respect to no matter what measure contained in the set $\mathcal{M}$ corresponds to the true description of reality.

Similar as seen in the case of admissibility (see Paragraph 1.2 for further detail), this not doubtable claim for rationality has its price: Especially applying interval dominance, in general, won't help us to find a satisfying decision in a concrete problem. Often, there are simply too many actions that remain incomparable with respect to this criterion. Surprisingly, E-admissibility turns out to a relatively strong criterion (see Theorem 19 in particular).

So, the question is the following: How can we generate full comparability under the actions without increasing our risk to make unreasonable decisions too much? A common way to reach that goal is to compare the actions according to their expected utility under the least favourable measure in $\mathcal{M}$. In the following paragraphs, we

recall two criteria that make use of this idea, namely *Maximality* and the Γ-*Maximin-criterion*. As we will see, both criteria are closely related to each other.

Recall that, in the precise case, an action $a^* \in \mathbb{A}$ is $B(\pi)$-optimal for some prior measure π, if the inequality

$$\mathbb{E}_\pi(u_{a^*}) \geqslant \mathbb{E}_\pi(u_a) \tag{101}$$

holds for all $a \in \mathbb{A}$. As the precise expectation is a linear operator, this is equivalent to

$$\mathbb{E}_\pi(u_a - u_{a^*}) = \mathbb{E}_\pi(u_a) - \mathbb{E}_\pi(u_{a^*}) \leqslant 0 \tag{102}$$

for all $a \in \mathbb{A}$. Now, consider the uncertainty between the elements of Θ is no longer measured by a single classical probability measure π, but by a set $\mathcal{M}$ of the form defined in equation (86). Again, the set $\mathcal{M}$ can be interpreted as a convex credal set or the structure of an F-probability respectively. Then, if we want to compare actions according to their expected utility under the least favourable measure contained in $\mathcal{M}$, the equations (101) and (102) are no longer equivalent! More precisely, we have

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_a - u_{a^*}) \neq \underline{\mathbb{E}}_{\mathcal{M}}(u_a) - \underline{\mathbb{E}}_{\mathcal{M}}(u_{a^*}) \tag{103}$$

in general. Hence, we receive two non-equivalent ways of cautiously generalizing the principle of maximizing expected utility to the imprecise case! The generalization of equation (101) induces the Γ-Maximin criterion, whereas the generalization of equation (102) induces the criterion of Maximality. In this paragraph, we start with discussing the Maximality criterion. It originally goes back to [43, § 3.9.2]. Here, we use a slightly different definition taken from [40, p.22].

Definition 23. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$. Further, let $\mathcal{M}$ be of the form defined in (86). Then, an action $a^* \in \mathbb{A}$ is called $\mathcal{M}$-*maximal*, if

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_a - u_{a^*}) \leqslant 0 \tag{104}$$

holds for all actions $a \in \mathbb{A}$. The set of all maximal actions is denoted by $\mathbb{A}^{max}_{\mathcal{M}}$. ▽

Now, let $\mathfrak{A}$ denote a finite decision problem. Two questions immediately come to mind: How can we decide whether an action is maximal or not, without computing the lower expectations of all the random variables $u_{a_l} - u_{a_k}$, for all $l \neq k \in \{1, \dots, n\}$? And: Are pure actions that are maximal in $\mathfrak{A}$ still maximal when considering the mixed extension $G(\mathfrak{A})$? The following example gives an answer to the second question.

Example 15. Consider the following finite decision problem $\mathfrak{A}$:

u_{ij}	θ_1	θ_2
a_1	0	10
a_2	10	0
a_3	4	4

Further, let our information about the states of nature be given by the following set of probability measures on $(\Theta, \mathcal{P}(\Theta))$:

$$\mathcal{M} := \{\pi : 0.3 \leqslant \pi(\{\theta_1\}) \leqslant 0.7\}$$

First, we show that action a_3 is maximal in the basic problem $\mathfrak{A}$. We compute

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{M}}(u_{a_1} - u_{a_3}) &= \min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{a_1} - u_{a_3}) \\ &= \min_{\pi \in \mathcal{M}} \big((-4) \cdot \pi(\{\theta_1\}) + 6 \cdot \pi(\{\theta_2\})\big) \\ &= (-4) \cdot 0.7 + 6 \cdot 0.3 = -1 \leqslant 0 \end{aligned}$$

and, by analogy,

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{M}}(u_{a_2} - u_{a_3}) &= \min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_{a_2} - u_{a_3}) \\ &= \min_{\pi \in \mathcal{M}} \big(6 \cdot \pi(\{\theta_1\}) - 4 \cdot \pi(\{\theta_2\})\big) \\ &= 6 \cdot 0.3 - 4 \cdot 0.7 = -1 \leqslant 0 \end{aligned}$$

Hence, a_3 is maximal in $\mathfrak{A}$.

Next, consider the randomized action $p^* \in G(\mathbb{A})$ induced by the assignment

$$p^*(\{a_i\}) = \begin{cases} 0.5 & \text{if } i = 1, 2 \\ 0 & \text{if } i = 3 \end{cases}$$

We show that action a_3 is no longer maximal in $G(\mathfrak{A})$. We compute:

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{M}}(G(u)_{p^*} - G(u)_{\delta_{a_3}}) &= \min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(G(u)_{p^*} - G(u)_{\delta_{a_3}}) \\ &= \min_{\pi \in \mathcal{M}} \Bigg(\pi(\{\theta_1\}) \cdot \Big(\sum_{i=1}^{3} u(a_i, \theta_1) \cdot \big(p^*(\{a_i\}) - \delta_{a_3}(\{a_i\})\big) \Big) \\ &+ \qquad \pi(\{\theta_2\}) \cdot \Big(\sum_{i=1}^{3} u(a_i, \theta_2) \cdot \big(p^*(\{a_i\}) - \delta_{a_3}(\{a_i\})\big) \Big) \Bigg) \end{aligned}$$

$$
\begin{aligned}
&= \min_{\pi\in\mathcal{M}}\Big(\pi(\{\theta_1\})\cdot(10\cdot 0.5-4)+\pi(\{\theta_2\})\cdot(10\cdot 0.5-4)\Big)\\
&= \min_{\pi\in\mathcal{M}}\Big(\pi(\{\theta_1\})+\pi(\{\theta_2\})\Big)= 1>0
\end{aligned}
$$

Hence, action a_3 is not maximal when considering the mixed extension of $\mathfrak{A}$. $\star$

The previous example demonstrated the following: In general, maximality is not robust under the transition to the mixed extension. More precisely, if one is interested in whether an action is maximal or not, it does not suffice to take only pure actions into consideration. Hence, randomization might generate a strict improvement of the utility with respect to the maximality criterion. This brings us back to the first question posed before: How can we decide whether a (pure or randomized) action under consideration is maximal in the mixed extension of a finite CDP?

Again, this can be checked by solving suitable linear programming problems. Let $\mathfrak{A}$ denote any finite CDP and let $\mathcal{M}$ denote a set of the form defined in (86) (which can be interpreted as a convex credal set or the structure of an F-probability respectively). Now, recall that a randomized action $p^*\in G(\mathbb{A})$ is $\mathcal{M}$-maximal, if the following inequality holds

$$
\underline{\mathbb{E}}_{\mathcal{M}}(G(u)_p - G(u)_{p^*})\leqslant 0 \tag{105}
$$

for all $p\in G(\mathbb{A})$. Next, note the equivalence of condition (105) with the inequality

$$
\sup_{p\in G(\mathbb{A})}\Big(\underline{\mathbb{E}}_{\mathcal{M}}(G(u)_p - G(u)_{p^*})\Big)\leqslant 0 \tag{106}
$$

Thus, p^* is $\mathcal{M}$-optimal, if the optimal outcome of the optimization problem

$$
\underbrace{\underline{\mathbb{E}}_{\mathcal{M}}(G(u)_p - G(u)_{p^*})}_{=:\kappa(p)} \longrightarrow \sup_{p\in G(\mathbb{A})} \tag{107}
$$

is at most 0. Step by step, we want to reformulate (107) as a linear programming problem. For $p\in G(\mathbb{A})$ fixed, the following holds for the expression $\kappa(p)$:

$$
\begin{aligned}
\kappa(p) &= \inf_{\pi\in\mathcal{M}}\ \mathbb{E}_\pi(G(u)_p - G(u)_{p^*})\\
&= \inf_{\pi\in\mathcal{M}}\ \sum_{j=1}^{m}\Big(\underbrace{\sum_{i=1}^{n}u(a_i,\theta_j)\cdot p(\{a_i\})}_{=:e_j^p} - \underbrace{\sum_{i=1}^{n}u(a_i,\theta_j)\cdot p^*(\{a_i\})}_{e_j^{p^*}}\Big)\cdot\pi(\{\theta_j\})
\end{aligned}
$$

$$= \inf_{\pi \in \mathcal{M}} (e_1^p - e_1^{p^*}, \ldots, e_m^p - e_m^{p^*}) \cdot \begin{pmatrix} \pi(\{\theta_1\}) \\ \vdots \\ \pi(\{\theta_m\}) \end{pmatrix}$$

Thus, $\kappa(p)$ equals the optimal outcome of the optimization problem

$$(e_1^p - e_1^{p^*}, \ldots, e_m^p - e_m^{p^*}) \cdot \begin{pmatrix} \pi(\{\theta_1\}) \\ \vdots \\ \pi(\{\theta_m\}) \end{pmatrix} \longrightarrow \inf_{\pi \in \mathcal{M}} \tag{108}$$

Then, according to (the proof of) Theorem 14, the above infimum in (108) is actually attained on the set $\mathcal{M}$ and the optimal outcome equals the optimal outcome of the linear programming problem

$$(e_1^p - e_1^{p^*}, \ldots, e_m^p - e_m^{p^*}) \cdot \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \longrightarrow \min_{\pi \in \mathbb{R}^m} \tag{109}$$

with constraints

- $\pi \geqslant 0$
- $K \cdot \pi \geqslant l$

where K and l are defined just like in the proof of Theorem 14. Thus, by duality, $\kappa(p)$ equals the optimal outcome of the dual programming problem of (109), that is

$$s_p(w_1, ..., u_r) := l^T \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \end{pmatrix} \longrightarrow \max_{(w_1, \ldots, u_r)} \tag{110}$$

with constraints

- $(w_1, ..., u_r) \geqslant 0$

- $$\underbrace{\begin{pmatrix} 1 & 1' & f_{11} & f'_{11} & \dots & f_{r1} & f'_{r1} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 1' & f_{1m} & f'_{1m} & \dots & f_{rm} & f'_{rm} \end{pmatrix}}_{=K^T} \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \end{pmatrix} \leqslant \underbrace{\begin{pmatrix} e_1^p - e_1^{p^*} \\ \vdots \\ e_m^p - e_m^{p^*} \end{pmatrix}}_{=:\vec{e}}$$

Finally, since the equation

$$\sup_{p \in G(\mathbb{A})} \kappa(p) = \sup\Big\{ s_p(w_1, ..., u_r) : p \in G(\mathbb{A}) \wedge (w_1, ..., u_r) \in Z^+(K^T, \vec{e}) \Big\}$$

holds, the optimal outcome of the optimization problem (107) equals the optimal outcome of the optimization problem

$$(l^T, \underbrace{0, \dots, 0}_{\text{n-times}}) \cdot \begin{pmatrix} w_1 \\ \vdots \\ u_r \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \sup_{(w_1, ..., p_n)} \tag{111}$$

with constraints

- $(w_1, ..., p_n) \geqslant 0$
- $p \in G(\mathbb{A})$
- $(w_1, ..., u_r) \in Z^+(K^T, \vec{e})$

To see the linearity of the second and the third constraint, note that, together, they are equivalent to the inequality

$$\underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1' & \dots & 1' \\ 1 & 1' & f_{11} & f'_{11} & \dots & f_{r1} & f'_{r1} & u'_{11} & \dots & u'_{n1} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots' & \dots & \vdots \\ 1 & 1' & f_{1m} & f'_{1m} & \dots & f_{rm} & f'_{rm} & u'_{1m} & \dots & u'_{nm} \end{pmatrix}}_{=:D} \begin{pmatrix} w_1 \\ \vdots \\ u_r \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant \underbrace{\begin{pmatrix} 1 \\ -1 \\ -e_1^{p^*} \\ \vdots \\ -e_m^{p^*} \end{pmatrix}}_{=:z}$$

Hence, problem (111) defines an SMP. Additionally, since the objective function is bounded on the set $Z^+(D,z)$, there exists an optimal solution of (111) according to Theorem 5. That is, the supremum in expression (111) is actually attained on the set $Z^+(D,z)$.

Summarized, we just showed the following: An action $p^* \in G(\mathbb{A})$ in the mixed extension of a finite decision problem is maximal, if the optimal output of the corresponding linear programming problem (111) is at most 0. Thus, we can check the maximality of (randomized) actions computationally. Pure actions are naturally contained in the approach just described: If we want to check maximality for a pure action $a^* \in \mathbb{A}$, we just need to run the above algorithm with $p^* := \delta_{a^*}$. To get some practice, note the following example.

Example 16. Consider again the CDP discussed in Example 12. Further, let the information concerning the states of nature be given by the following set of probability measures on $(\Theta, \mathcal{P}(\Theta))$:

$$\mathcal{M} := \Big\{ \pi : 0.35 \leqslant \pi(\{\theta_1\}) + \pi(\{\theta_4\}) \leqslant 0.65 \Big\}$$

Hence, $\mathcal{M}$ is of the form (86) with $r = 1$, $m = 4$, $n = 3$, $(\underline{b}_1, \overline{b}_1) = (0.35, 0.65)$ and $(f_1(\theta_1), \ldots, f_1(\theta_4)) = (1, 0, 0, 1)$. We want to check $\mathcal{M}$-maximality of action a_1 in the mixed extension $G(\mathfrak{A})$. According to the before considerations, this can be done by solving the linear programming problem (111), where $p^* := \delta_{a_1}$. We arrive at the following SMP

$$(1, -1, 0.35, -0.65, 0, 0, 0) \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \longrightarrow \max_{(w_1, \ldots, p_3)}$$

with constraints

- $(w_1, \ldots, p_3) \geqslant 0$

- $$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -20 & -30 & -20 \\ 1 & -1 & 0 & 0 & -15 & -10 & -40 \\ 1 & -1 & 0 & 0 & -10 & -10 & 0 \\ 1 & -1 & 1 & -1 & -30 & -20 & -20 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ -1 \\ -20 \\ -15 \\ -10 \\ -30 \end{pmatrix}$$

Running the above linear programming problem in R gives the optimal outcome 0. Hence, action a_1 is maximal, even when considering the mixed extension. ⋆

To complete the paragraph, we now want recall some results that make up connections between the concept of maximality and other imprecise criteria for the optimality of actions. We start with the following theorem. It can also be found in [40, p. 25, Theorem 1].

Theorem 18. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$ and let $\mathcal{M}$ be any non-empty set of probability measures on $(\Theta, \mathcal{P}(\Theta))$. Then, for all actions $a \in \mathbb{A}$, the following holds:

$$a \in \mathbb{A}^{max} \Rightarrow a \in \mathbb{A}^{<^I}_{\mathcal{M}} \tag{112}$$

That is, any $\mathcal{M}$-maximal action is $<^I$ $(\mathcal{M})$-admissible.

Proof. Let $a^* \in \mathbb{A}^{max}$. Assume, for contradiction, that $a^* \notin \mathbb{A}^{<^I}_{\mathcal{M}}$, i.e. there exists an action $a^- \in \mathbb{A}$ such that

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{a^-}) > \overline{\mathbb{E}}_{\mathcal{M}}(u_{a^*})$$

Particularly, this implies

$$\mathbb{E}_{\pi}(u_{a^-}) > \mathbb{E}_{\pi}(u_{a^*})$$

for all $\pi \in \mathcal{M}$. As the precise expectation is a linear operator, this is equivalent to

$$\mathbb{E}_{\pi}(u_{a^-} - u_{a^*}) = \mathbb{E}_{\pi}(u_{a^-}) - \mathbb{E}_{\pi}(u_{a^*}) > 0$$

for all $\pi \in \mathcal{M}$. However, this implies

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{a^-} - u_{a^*}) := \inf_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u_{a^-} - u_{a^*}) > 0$$

which yields a contradiction to $a^* \in \mathbb{A}^{max}$. Thus, $a^* \in \mathbb{A}^{<^I}_{\mathcal{M}}$. □

Theorem 18 proves the compatibility of the maximality criterion and the concept

of interval dominance. Therefore, the maximality criterion evidently satisfies the minimum requirement for rational imprecise decision criteria, i.e. for generalized concepts of maximizing expected utility. But there are connections to other criteria as well: The following Theorem makes a statement on the close relation between maximal and E-admissible actions. It can be found e.g. in [43, § 3.9.5].

Theorem 19. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$ and let $\mathcal{M}$ be any non-empty set of probability measures on $(\Theta, \mathcal{P}(\Theta))$. Then, the following statements hold:

i) $\mathbb{A}_{E(\mathcal{M})} \subset \mathbb{A}_{\mathcal{M}}^{max}$

ii) $G(\mathbb{A})_{E(\mathcal{M})} = G(\mathbb{A})_{\mathcal{M}}^{max}$

Proof. i) Let $a^* \in \mathbb{A}_{E(\mathcal{M})}$. That is, there exists $\pi \in \mathcal{M}$ such that

$$\mathbb{E}_\pi(u_{a^*}) \geqslant \mathbb{E}_\pi(u_a)$$

for all $a \in \mathbb{A}$. As the precise expectation is a linear operator, this is equivalent to

$$\mathbb{E}_\pi(u_a - u_{a^*}) \leqslant 0$$

for all $a \in \mathbb{A}$. Since, $\underline{\mathbb{E}}_{\mathcal{M}}(X) \leqslant \mathbb{E}_\pi(X)$ for any random variable $X : \Theta \to \mathbb{R}$, this implies

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_a - u_{a^*}) \leqslant 0$$

for all $a \in \mathbb{A}$. Thus, $a^* \in \mathbb{A}_{\mathcal{M}}^{max}$.

ii) $\subset$: Since $G(\mathfrak{A})$ is again a CDP such that $|\Theta| < \infty$, this inclusion directly follows from i).

$\supset$: The proof of this inclusion is highly non-trivial. It can be found for example in [43, § 3.9.5]. □

Remark. Theorem 19 ii) shows that a randomized action is $\mathcal{M}$-maximal if, and only if, it is $E(\mathcal{M})$-admissible. Particularly, this implies the equivalence of the optimization problems (111) (algorithm for checking maximality) and (98) (algorithm for checking E-admissibility), since a maximal pure action that remains maximal under the transition to the mixed extension is also E-admissible. More precisely, for a fixed pure action a, the optimization problem (98) has the optimal outcome 1 if, and only if, the optimal outcome of (111), where $p^* := \delta_a$, is at most zero. Thus, basically, we have two different algorithms for checking the same thing. ∘

5.5 The Γ-Maximin Criterion

In the beginning of the previous paragraph, we saw that there exist (at least) two different ways of cautiously generalizing the principle of maximizing the expected utility (where cautiously is to be understood in terms of considering the lower expectation only). More precisely, if the uncertainty between the states is measured by a set of probability measures, the equations (101) and (102) are no longer equivalent. As seen before, the generalization of (102) induces the maximality criterion. Next, we want to generalize (101) to the imprecise case. This motivates the following definition, which is taken from e.g. [23, Definition 8.2].

Definition 24. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$. Define the set $\mathcal{M}$ like in equation (86). Again, $\mathcal{M}$ can be interpreted as a convex credal set or the structure of an F-probability respectively. Then the criterion

$$\Phi^{\mathfrak{A}}_{\Gamma(\mathcal{M})} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(u_a) =: \underline{\mathbb{E}}_{\mathcal{M}}(u_a) \tag{113}$$

is called $\Gamma(\mathcal{M})$-*Maximin-criterion*. Every $\Phi^{\mathfrak{A}}_{\Gamma(\mathcal{M})}$-optimal action is called $\Gamma(\mathcal{M})$-*Maximin-action*.
The set of all $\Gamma(\mathcal{M})$-Maximin-actions then is denoted by $\mathbb{A}_{\Gamma(\mathcal{M})}$. ▽

Remark. 1.) To see that the above criterion is well-defined, note the following: According to Theorem 14, for any map $X : \Theta \to \mathbb{R}$ the lower expectation $\underline{\mathbb{E}}_{\mathcal{M}}(X)$ is attained on the set $\mathcal{M}$. Thus, for every $a \in \mathbb{A}$ fixed, the lower expectation $\underline{\mathbb{E}}_{\mathcal{M}}(u_a)$ is attained on the set $\mathcal{M}$.

2.) Note that the above criterion corresponds to the ordering $\leqslant^p$ introduced in Excursus 4.2.3: An action $a^* \in \mathbb{A}$ is $\Gamma(\mathcal{M})$-Maximin-action if, and only if, the inequality $\underline{\mathbb{E}}_{\mathcal{M}}(u_a) \leqslant^p \underline{\mathbb{E}}_{\mathcal{M}}(u_{a^*})$ holds for all $a \in \mathbb{A}$. ∘

As for any other decision criterion, the following question immediately comes to mind: If an action is $\Gamma(\mathcal{M})$-Maximin in the basic problem, will it still be optimal compared to randomized actions? In other words: Can, in general, the $\Gamma(\mathcal{M})$-Maximin utility be strictly improved by also considering randomized actions? Indeed, this is the case.

As an extreme example, consider the situation where no additional information concerning the states of nature is available. That is, the set $\mathcal{M}$ consists of all possible probability measures on the space $(\Theta, \mathcal{P}(\Theta))$ ($\mathcal{M}$ then is called *vacuous*, see e.g. [4, p. 40]). In this case, the expression $\underline{\mathbb{E}}_{\mathcal{M}}(u_a)$ necessarily equals the utility of

action a under one least favourable state of nature, that is

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_a) = \min_{\theta \in \Theta} u(a, \theta) \tag{114}$$

To verify equation (114), two statements have to be proven: There exits a probability measure π^* on $(\Theta, \mathcal{P}(\Theta))$ such that $\mathbb{E}_{\pi^*}(u_a) = \min_{\theta \in \Theta} u(a, \theta)$. And: For all other probability measures π on $(\Theta, \mathcal{P}(\Theta))$ we have $\mathbb{E}_{\pi}(u_a) \geqslant \mathbb{E}_{\pi^*}(u_a)$. To see the validity of the first statement, let $\theta^* \in \Theta$ be such that $u(a, \theta^*) = \min_{\theta \in \Theta} u(a, \theta)$. This is possible, since the function u_a is defined on a finite set and, therefore, attains its minimum on it. Now, set $\pi^* := \delta_{\theta^*}$ and compute

$$\mathbb{E}_{\pi^*}(u_a) = \sum_{j=1}^{m} u(a, \theta_j) \cdot \delta_{\theta^*}(\{\theta_j\}) = u(a, \theta^*) = \min_{\theta \in \Theta} u(a, \theta) \tag{115}$$

This proves the first statement. To see the second statement, let π be any probability measure on $(\Theta, \mathcal{P}(\Theta))$. The following inequality holds:

$$\mathbb{E}_{\pi}(u_a) = \sum_{j=1}^{m} u(a, \theta_j) \cdot \pi(\{\theta_j\}) \geqslant \min_{\theta \in \Theta} u(a, \theta) \cdot \sum_{j=1}^{m} \pi(\{\theta_j\}) = \min_{\theta \in \Theta} u(a, \theta) \tag{116}$$

However, together with equation (115), this implies $\mathbb{E}_{\pi}(u_a) \geqslant \mathbb{E}_{\pi^*}(u_a)$. Since π was chosen arbitrarily, this shows the second statement and, therefore, finishes the proof of equation (114).

Note that equation (114) yields an interesting connection to precise decision theory: If there is no information on the states available at all, the precise Maximin-criterion and the imprecise Γ-Maximin criterion coincide! Hence, the question posed before is already answered: In general, randomization can generate a strict improvement of utility w.r.t. to the Γ-Maximin-criterion, since it improves utility with respect to the precise Maximin-criterion. This was illustrated in Example 1.

However, even in situation where there is some information on the states available, randomization might be useful. This is clarified in the following

Example 17. Consider the CDP $\mathfrak{A}$ defined by

u_{ij}	θ_1	θ_2	θ_3	θ_4	$\underline{\mathbb{E}}_{\mathcal{M}}(u_{a_i})$
a_1	20	15	10	30	18
a_2	30	10	10	20	14
a_3	20	40	0	20	8
a_4	10	30	50	30	18
a_5	0	30	20	40	16

and let

$$\mathcal{M} := \{\pi : \pi_1 + \pi_3 \leqslant 0.8 \wedge \pi_4 \geqslant 0.4\}$$

The last column of the above utility table lists the lower bounds of the interval expectation of the random variables u_{a_i} w.r.t. $\mathcal{M}$. Clearly, the actions a_1 and a_4 are $\Gamma(\mathcal{M})$-Maximin actions in the basic problem $\mathfrak{A}$ with a $\Gamma(\mathcal{M})$-Maximin utility of 18.
However, the randomized action $p^* \in G(\mathbb{A})$ induced by the assignment

$$p^*(\{a_i\}) = \begin{cases} \frac{4}{5} & \text{if } i = 1 \\ 0 & \text{if } i = 2 \\ 0 & \text{if } i = 3 \\ \frac{1}{5} & \text{if } i = 4 \\ 0 & \text{if } i = 5 \end{cases}$$

dominates all the pure actions available:

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{M}}(G(u)_{p^*}) &= \inf_{\pi \in \mathcal{M}} \Big(\pi_1 \cdot (\tfrac{4}{5} \cdot 20 + \tfrac{1}{5} \cdot 10) \\ &+ \qquad \pi_2 \cdot (\tfrac{4}{5} \cdot 15 + \tfrac{1}{5} \cdot 30) \\ &+ \qquad \pi_3 \cdot (\tfrac{4}{5} \cdot 10 + \tfrac{1}{5} \cdot 50) \\ &+ \qquad \pi_4 \cdot (\tfrac{4}{5} \cdot 30 + \tfrac{1}{5} \cdot 30) \Big) \\ &= \inf_{\pi \in \mathcal{M}} \Big(\pi_1 \cdot 18 + \pi_2 \cdot 18 + \pi_3 \cdot 18 + \pi_4 \cdot 30 \Big) \\ &= 0.6 \cdot 18 + 0.4 \cdot 30 = 22.8 > 18 \end{aligned}$$

Hence, the transition to the mixed extension generates a strict improvement of the $\Gamma(\mathcal{M})$-Maximin utility. ⋆

The previous two examples showed the following: When considering the Γ-Maximin-criterion, we can improve our utility by taking also randomized actions into account. In the case where no information is available at all, the criterion coincides with

the Maximin-criterion. In this context, another aspect is interesting to mention: Consider the situation with perfect information. That is, there is no ambiguity left and the set $\mathcal{M}$ only consists of one single probability measure. In this case, the Γ-Maximin-criterion reduces to the Bernoulli/Bayes-criterion known from precise decision theory: The infimal expectation under all measures from $\mathcal{M}$ simply reduces to the expectation under the only measure contained in $\mathcal{M}$. This seemingly trivial statement helps to clarify two things: As there always exists a *pure* Bayes-action according to Theorem 2, there are situations in which the Γ-Maximin-criterion labels pure actions as being optimal. Hence, randomization doesn't ensure a higher utility in any situation.

Additionally, depending on the quality of the information available (that is, how much ambiguity is left between the measures contained in $\mathcal{M}$), the Γ-Maximin-criterion tends more to the classical Maximin-criterion (lack of information) or to the classical Bayes-criterion (perfect information). Hence, the two classical criteria can be seen as the *extreme poles* of the Γ-Maximin-criterion with respect to the quality of the available information.

So, how can $\Gamma(\mathcal{M})$-Maximin optimal randomized actions be determined? And: Do there always exist such actions? The following Theorem guarantees the existence of randomized $\Gamma(\mathcal{M})$-Maximin-actions in finite decision problems. Furthermore, the procedure described in the proof shows up a way to determine these using linear optimization algorithms. This procedure goes back to [41, § 3.2].

Theorem 20. Let $\mathfrak{A}$ be a finite CDP and $\mathcal{M}$ be a set of probability measures of the form defined in (86). Again, the set $\mathcal{M}$ can be interpreted as a convex credal set or the structure of an F-probability respectively. Then there exists a $\Gamma(\mathcal{M})$-Maximin-action for the decision problem $G(\mathfrak{A})$. That is, there exists $p^* \in G(\mathbb{A})$ such that

$$\Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p^*) \geqslant \Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p)$$

for all $p \in G(\mathbb{A})$.

Proof. First, note that there exists an $\Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}$-optimal action if, and only if, the optimization problem

$$\underbrace{\inf_{\pi \in \mathcal{M}} \mathbb{E}_\pi(G(u)_p)}_{=:\alpha(p)} \longrightarrow \max_{p \in G(\mathbb{A})} \tag{117}$$

has an optimal solution. Now, fix $p \in G(\mathbb{A})$. As a first step, we show that the infimum in the term $\alpha(p)$ is actually attained on the set $\mathcal{M}$. This is equivalent for

the optimization problem

$$\mathbb{E}_\pi(G(u)_p) \longrightarrow \min_{\pi \in \mathcal{M}} \tag{118}$$

to have an optimal solution. Now compute:

$$\mathbb{E}_\pi(G(u)_p) = \sum_{j=1}^{m} G(u)(p, \theta_j) \cdot \pi(\{\theta_j\}) = \sum_{j=1}^{m} \underbrace{\Big(\sum_{i=1}^{n} u(a_i, \theta_j) \cdot p(\{a_i\}) \Big)}_{=: e_j^p} \cdot \pi(\{\theta_j\})$$

Hence, the optimization problem (118) is equivalent to the SMIP

$$(e_1^p, \ldots, e_m^p) \cdot \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \longrightarrow \min_{\pi \in \mathbb{R}^m} \tag{119}$$

with the constraint

- $\pi \in \Pi_m$

Here, we again used the one-to-one correspondence of the sets $\mathcal{M}$ and Π_m that was already described in previous paragraphs (see for example the proof of Theorem 9). To see the linearity of the constraints note their equivalence to

- $\pi \geqslant 0$
- $$\underbrace{\begin{pmatrix} 1 & \ldots & 1 \\ 1' & \ldots & 1' \\ f_{11} & \ldots & f_{1m} \\ f'_{11} & \ldots & f'_{1m} \\ \vdots & \ldots & \vdots \\ f_{r1} & \ldots & f_{rm} \\ f'_{r1} & \ldots & f'_{rm} \end{pmatrix}}_{=: L} \cdot \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} \geqslant \underbrace{\begin{pmatrix} 1 \\ 1' \\ \underline{b}_1 \\ \overline{b}'_1 \\ \vdots \\ \underline{b}_r \\ \overline{b}'_r \end{pmatrix}}_{=: v}$$

Since the set Π_m is non-empty and bounded, the SMIP 119 has an optimal solution according to Theorem 4. Thus, we proved the following: For every $p \in G(\mathbb{A})$ the infimum $\alpha(p)$ is attained on the set $\mathcal{M}$. That is, for $p \in G(\mathbb{A})$ fixed, the expression $\alpha(p)$ equals the optimal outcome of the SMIP (119).

Now, consider the dual linear programming problem of (119):

$$d(\underbrace{w_1,\ldots,u_r}_{=:\vec{w}}) := 1, 1', \underline{b}_1, \overline{b}'_1, \ldots, \underline{b}_r, \overline{b}'_r) \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \end{pmatrix} \longrightarrow \max_{(w_1,\ldots,u_r)} \tag{120}$$

with constraints

- $(w_1, \ldots, u_r) \geqslant 0$
- $L^T \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \end{pmatrix} \leqslant \underbrace{\begin{pmatrix} e_1^p \\ \vdots \\ e_m^p \end{pmatrix}}_{\vec{f}}$

According to duality, $\alpha(p)$ then also equals the optimal outcome of the linear programming problem (120). Hence, the optimization problem (117) is equivalent to the optimization problem

$$\max\Big\{d(\vec{w}) : \vec{w} \in Z^+(L^T, \vec{f})\Big\} \longrightarrow \max_{p \in G(\mathbb{A})} \tag{121}$$

That is, there exists a randomized $\Gamma(\mathcal{M})$-Maximin action if, and only if, the optimization problem (121) has an optimal solution. Since we have

$$\max_{p \in G(\mathbb{A})} \Big\{d(\vec{w}) : \vec{w} \in Z^+(L^T, \vec{f})\Big\} = \max\Big\{d(\vec{w}) : \vec{w} \in Z^+(L^T, \vec{f}) \wedge p \in G(\mathbb{A})\Big\} \tag{122}$$

the optimization problem (121) has an optimal solution if, and only if, the following

SMP has an optimal solution:

$$(1, 1', \underline{b}_1, \overline{b}'_1, \ldots, \underline{b}_r, \overline{b}'_r, \underbrace{0, \ldots, 0}_{\text{n times}}) \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_{(w_1,\ldots,p_n)} \tag{123}$$

with constraints

- $(w_1, \ldots, p_n) \geqslant 0$
- $p \in G(\mathbb{A})$
- $L^T \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \end{pmatrix} \leqslant \begin{pmatrix} e_1^p \\ \vdots \\ e_m^p \end{pmatrix}$

To see the linearity of the constraints, note that, together, the second and third constraint are equivalent to the inequality

$$H^T \cdot \begin{pmatrix} w_1 \\ w_2 \\ l_1 \\ u_1 \\ \vdots \\ l_r \\ u_r \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant (1, -1, \underbrace{0, ..., 0}_{m \text{ times}})^T$$

where H is the matrix defined in the proof of Theorem 9.
Finally, note that the SMP (123) is exactly the dual of the SMIP (48) defined in the proof of Theorem 9. So, according to duality and Theorem 9, the SMP (123) has an optimal solution. This completes the proof. □

As an immediate consequence of Theorem 20, Theorem 9 and Corollary 1 we can derive the following Corollary. It shows some interesting connection between the precise Bayes-criterion and the Γ-Maximin criterion.

Corollary 4. Let $\mathfrak{A}$ be a finite CDP and $\mathcal{M}$ be a set of probability measures of the form defined in (86). Furthermore, let $p^* \in G(\mathbb{A})$ be a $\Gamma(\mathcal{M})$-Maximin action for the mixed extension $G(\mathfrak{A})$. Then, there exists a pair $(a^*, \pi^*) \in \mathbb{A} \times \mathcal{M}$ such that

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) = \Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p^*) \tag{124}$$

Proof. Let, in accordance with Corollary 1, $(a^*, \pi^*) \in \mathbb{A} \times \mathcal{M}$ denote a least favourable combination. Particularly, this implies that π^* defines a least favourable prior with respect to $\mathcal{M}$. Again, for $\pi \in \mathcal{M}$, let $M(\pi)$ denote the Bayes-utility w.r.t. the prior π. Then, according to the proof of Theorem 9, $M(\pi^*)$ equals the optimal outcome of the SMIP (48). Since the linear programming problems (48) and (123) are dual to each other, $M(\pi^*)$ then also equals the optimal outcome of (123). But the optimal outcome of (123) equals $\Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p^*)$ by construction. Therefore, we get

$$M(\pi^*) = \Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p^*) \tag{125}$$

Since $(a^*, \pi^*) \in \mathbb{A} \times \mathcal{M}$ is a least favourable combination, we have

$$M(\pi^*) = \Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) \tag{126}$$

Thus, we get

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a^*) = \Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p^*)$$

This completes the proof. □

Remark. Corollary 4 shows a connection between the precise Bayes-criterion and the Γ-Maximin-criterion: There always exists a 'most pessimistic' precise description of the uncertainty (represented by a measure $\pi^* \in \mathcal{M}$), which is still compatible with the information base. Under this pessimistic description of uncertainty, applying the Bayes-criterion gives the same utility as applying the Γ-Maximin-criterion under the

imprecise description.

Example 18. In this example we want to use the algorithm explained in the proof of Theorem 20 to determine a Γ-Maximin action. Furthermore, we show that applying the Bayes-criterion w.r.t. a least favourable prior induces the same utility value.

Now, consider the following CDP $\mathfrak{A}$:

u_{ij}	θ_1	θ_2	θ_3	θ_4	θ_5
a_1	2000	15000	1000	3000	500
a_2	6000	5000	4000	3000	1400
a_3	5500	10000	2000	2000	800
a_4	8751	3000	3100	4500	1800
a_5	6200	300	20000	4000	160

Let ζ denote the *true* and *unknown* probability measure characterizing the uncertainty between the elements of Θ (that is, the measure describing the situation under perfect information). Further, let $\zeta_j := \zeta(\{\theta_j\})$ for $j = 1 \ldots, 5$. Suppose, for some reason, the following information about ζ is available:

- $2 \cdot \zeta_3 \geqslant \zeta_1 + \zeta_5$
- $0.25 \leqslant \zeta_4 + \zeta_5 \leqslant 0.77$
- $0.35 \leqslant \zeta_2 + 0.5 \cdot \zeta_3$

Then, the information is representable as a $\mathcal{M}$ set of probability measures of the form defined in (86). More precisely, we have $r = 3$ and

- $f_1(\theta) := \begin{cases} -1 & \text{if } \theta \in \{\theta_1, \theta_5\} \\ 0 & \text{if } \theta \in \{\theta_2, \theta_4\} \\ 2 & \text{if } \theta = \theta_3 \end{cases}$
- $f_2(\theta) := \begin{cases} 1 & \text{if } \theta \in \{\theta_4, \theta_5\} \\ 0 & \text{else} \end{cases}$
- $f_3(\theta) := \begin{cases} 1 & \text{if } \theta = \theta_2 \\ 0.5 & \text{if } \theta = \theta_3 \\ 0 & \text{else} \end{cases}$

- $(\underline{b}_i, \overline{b}_i) := \begin{cases} (0,2) & \text{if } i = 1 \\ (0.25, 0.77) & \text{if } i = 2 \\ (0.35, 1) & \text{if } i = 3 \end{cases}$

To compute a randomized $\Gamma(\mathcal{M})$-Maximin action, we set up the SMP (123) defined in the proof of Theorem 20. We arrive at

$$\underbrace{(1, -1, 0, -2, 0.25, -0.77, 0.35, -1, 0, 0, 0, 0, 0)}_{=:\vec{d}_1} \cdot \begin{pmatrix} w_1 \\ \vdots \\ p_5 \end{pmatrix} \longrightarrow \max_{(w_1, \ldots, p_5)}$$

with constraints

- $(w_1, \ldots, p_5) \geqslant 0$
- $C \cdot (w_1, \ldots, p_5)^T \leqslant \underbrace{(1, -1, 0, 0, 0, 0, 0)^T}_{=:\vec{d}_2}$

where

$$C := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & -2000 & -6000 & -5500 & -8751 & -6200 \\ 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & -15000 & -5000 & -10000 & -3000 & -300 \\ 1 & -1 & 2 & -2 & 0 & 0 & 0.5 & -0.5 & -1000 & -4000 & -2000 & -3100 & -20000 \\ 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & -3000 & -3000 & -2000 & -4500 & -4000 \\ 1 & -1 & -1 & 1 & 1 & -1 & 0 & 0 & -500 & -1400 & -800 & -1800 & -160 \end{pmatrix}$$

Running the above SMP in R returns the optimal randomized action $p^* \in G(\mathbb{A})$ that is induced by the assignment

$$p^*(\{a_i\}) := \begin{cases} 0.4389297 & \text{if } i = 1 \\ 0.5610703 & \text{if } i = 5 \\ 0 & \text{else} \end{cases}$$

with a $\Gamma(\mathcal{M})$-Maximin utility of 4677.989.

Next, we want to determine a pair $(a^*, \pi^*) \in \mathbb{A} \times \mathcal{M}$ satisfying condition (124). According to the proof of Corollary 4, it suffices to determine first a least favourable prior with respect to $\mathcal{M}$ and, afterwards, the corresponding pure Bayes action w.r.t. this prior. As already seen in the proofs of Theorem 9 and Theorem 20 respectively,

a least favourable prior w.r.t. $\mathcal{M}$ can be computed by solving the dual of the above SMP. Particularly, the following version of the SMIP (48) has to be solved:

$$\vec{d_2} \cdot \begin{pmatrix} g_1 \\ g_2 \\ \pi_1 \\ \vdots \\ \pi_5 \end{pmatrix} \longrightarrow \min_{(g_1,\ldots,\pi_m)}$$

with constraints

- $(g_1, \ldots, \pi_m) \geqslant 0$
- $C^T \cdot \vec{d_2} \geqslant \vec{d_1}$

Running the SMIP in R gives the optimal solution

$$(4677.989, 0, \underbrace{0, 0.2529996, 0.1940009, 0.1649978, 0.3880017}_{=:o^* \in \Pi_m})$$

Thus, a least favourable prior distribution is given by $\pi^* := b^{-1}(o^*)$. To find a Bayes action w.r.t. π^* we compute

$$\mathbb{E}_{\pi^*}(u_{a_i}) = \begin{cases} 4677.989 & \text{if } i = 1 \\ 3079.197 & \text{if } i = 2 \\ 3558.394 & \text{if } i = 3 \\ 2801.295 & \text{if } i = 4 \\ 4677.989 & \text{if } i = 5 \end{cases}$$

Hence, Bayes-actions w.r.t π^* are given by a_1 and a_5 and we have

$$\Phi^{\mathfrak{A}}_{B(\pi^*)}(a_1) = \Phi^{\mathfrak{A}}_{B(\pi^*)}(a_5) = \Phi^{G(\mathfrak{A})}_{\Gamma(\mathcal{M})}(p^*)$$

That is, the Bayes solutions a_1 and a_5 of the CDP under the prior π^* have exactly the same utility as the randomized $\Gamma(\mathcal{M})$-action p^*. ⋆

Theorem 20 allows another immediate conclusion: For every finite decision problem there exists a randomized action that minimizes the upper expectation with respect to $\mathcal{M}$. This is the statement of the following corollary.

Corollary 5. Let $\mathfrak{A}$ be a finite CDP and $\mathcal{M}$ be a set of probability measures of the form defined in (86). Then, there exists a randomized action $p^* \in G(\mathbb{A})$ such that

$$\overline{\mathbb{E}}_{\mathcal{M}}(G(u)_{p^*}) \leqslant \overline{\mathbb{E}}_{\mathcal{M}}(G(u)_p) \tag{127}$$

for all $p \in G(\mathbb{A})$. That is, p^* minimizes the upper interval expectation w.r.t. $\mathcal{M}$.

Proof. Given a finite CDP $\mathfrak{A}$, we define a new finite decision problem

$$\tilde{\mathfrak{A}} := (\mathbb{A}, \Theta, \tilde{u}(\cdot)) \tag{128}$$

where the new utility function $\tilde{u}$ is defined by

$$\tilde{u} : (\mathbb{A} \times \Theta) \to \mathbb{R} \quad , \quad (a, \theta) \mapsto \tilde{u}(a, \theta) := -u(a, \theta)$$

Then, according to Theorem 20, there exists a randomized $\Gamma(\mathcal{M})$-Maximin action $p^* \in G(\mathbb{A})$ for the decision problem $G(\tilde{\mathfrak{A}})$. That is,

$$\min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(G(\tilde{u})_{p^*}) \geqslant \min_{\pi \in \mathcal{M}} \mathbb{E}_\pi(G(\tilde{u})_p) \tag{129}$$

for all $p \in G(\mathbb{A})$. Now, note that the identity

$$\begin{aligned}
\mathbb{E}_\pi(G(\tilde{u})_{p^*}) &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \tilde{u}(a_i, \theta_j) \cdot p(\{a_i\}) \right) \cdot \pi(\{\theta_j\}) \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} (-1) \cdot u(a_i, \theta_j) \cdot p(\{a_i\}) \right) \cdot \pi(\{\theta_j\}) \\
&= (-1) \cdot \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \cdot u(a_i, \theta_j) \cdot p(\{a_i\}) \right) \cdot \pi(\{\theta_j\}) \\
&= (-1) \cdot \mathbb{E}_\pi(G(u)_{p^*})
\end{aligned}$$

holds for all $\pi \in \mathcal{M}$. That is,

$$\mathbb{E}_\pi(G(\tilde{u})_{p^*}) = (-1) \cdot \mathbb{E}_\pi(G(u)_{p^*}) \tag{130}$$

for all $\pi \in \mathcal{M}$. Next, note that, for any function $f : A \to \mathbb{R}$ (where A is an arbitrary non-empty set), the identity

$$-\min_{a \in A} f(a) = \max_{a \in A}(-f(a)) \tag{131}$$

holds. Now, for fixed $p \in G(\mathbb{A})$, define the function

$$e_p : \mathcal{M} \to \mathbb{R} \quad , \quad \pi \mapsto \mathbb{E}_\pi(G(\tilde{u})_p) \tag{132}$$

Then, for all $p \in G(\mathbb{A})$, we arrive at

$$\begin{aligned}
\max_{\pi\in\mathcal{M}} \mathbb{E}_\pi(G(u)_{p^*}) &\overset{(130),(132)}{=} \max_{\pi\in\mathcal{M}} \Big((-1)\cdot e_{p^*}(\pi)\Big) \\
&\overset{(131)}{=} -\min_{\pi\in\mathcal{M}} e_{p^*}(\pi) \\
&= -\min_{\pi\in\mathcal{M}} \mathbb{E}_\pi(G(\tilde{u})_{p^*}) \\
&\overset{(129)}{\leqslant} -\min_{\pi\in\mathcal{M}} \mathbb{E}_\pi(G(\tilde{u})_p) \\
&= \max_{\pi\in\mathcal{M}} \mathbb{E}_\pi(G(u)_p)
\end{aligned}$$

That is,

$$\max_{\pi\in\mathcal{M}} \mathbb{E}_\pi(G(u)_{p^*}) \leqslant \max_{\pi\in\mathcal{M}} \mathbb{E}_\pi(G(u)_p) \tag{133}$$

for all $p \in G(\mathbb{A})$. This completes the proof. □

The proof of Corollary 5 describes a method to determine randomized actions that minimize the upper interval expectation: These actions are exactly the Γ-Maximin-actions in the decision problem that is derived from the origin problem by multiplying the utility function with the factor -1. Let's apply this to an example.

Example 19. Consider again the setting of Example 17 (that is, the finite CDP $\mathfrak{A}$ together with the set $\mathcal{M}$). We want to determine a randomized action that minimizes the upper interval expectation with respect to $\mathcal{M}$. According to the proof of Corollary 5, this can be done by determining a randomized $\Gamma(\mathcal{M})$-Maximin action in the transformed decision problem $\tilde{\mathfrak{A}}$ given by

$\tilde{u}_{ij}$	θ_1	θ_2	θ_3	θ_4
a_1	-20	-15	-10	-30
a_2	-30	-10	-10	-20
a_3	-20	-40	0	-20
a_4	-10	-30	-50	-30
a_5	0	-30	-20	-40

According to Theorem 20, a randomized $\Gamma(\mathcal{M})$-Maximin-action for the decision problem $G(\tilde{\mathfrak{A}})$ can be gained by solving the linear optimization problem (123). Solving

the problem in R (similar as demonstrated in Example 18) gives the optimal randomized action $p^* \in G(\mathbb{A})$ induced by the assignment

$$p^*(\{a_i\}) = \begin{cases} 0 & \text{if } i = 1 \\ \frac{1}{2} & \text{if } i = 2 \\ \frac{1}{4} & \text{if } i = 3 \\ \frac{1}{4} & \text{if } i = 4 \\ 0 & \text{if } i = 5 \end{cases}$$

Thus, according to Corollary 5, the random variable $G(u)_{p^*}$ minimizes the upper interval expectation with respect $\mathcal{M}$. ⋆

To complete the paragraph, we want to recall a result that shows how the Γ-Maximin criterion is related to other decision criteria for the imprecise case. It turns out that every Γ-Maximin-optimal action is maximal as well. As, according to Theorem 18, every maximal action is not interval dominated by another action, this then implies the compatibility of Γ-Maximin and interval dominance. A more general version of the result can e.g. be found in [40, p. 25, Theorem 1].

Theorem 21. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$. Further, let $\mathcal{M}$ be of the form defined in (86). Then, the following implication holds for all $a \in \mathbb{A}$:

$$a \in \mathbb{A}_{\Gamma(\mathcal{M})} \quad \Rightarrow \quad a \in \mathbb{A}^{max}_{\mathcal{M}} \tag{134}$$

That is, any $\Gamma(\mathcal{M})$-Maximin action is $\mathcal{M}$-maximal.

Proof. (similar in [40, p.25]) Let $a^* \in \mathbb{A}_{\Gamma(\mathcal{M})}$. That is,

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_{a^*}) \geqslant \underline{\mathbb{E}}_{\mathcal{M}}(u_a) \tag{135}$$

for all $a \in \mathbb{A}$. Now, for all random variables $X, Y : \Theta \to \mathbb{R}$, the following holds (see for example [40, p.25]):

$$\underline{\mathbb{E}}_{\mathcal{M}}(X - Y) \leqslant \underline{\mathbb{E}}_{\mathcal{M}}(X) - \underline{\mathbb{E}}_{\mathcal{M}}(Y) \tag{136}$$

Thus, we have

$$\underline{\mathbb{E}}_{\mathcal{M}}(u_a - u_{a^*}) \overset{(136)}{\leqslant} \underline{\mathbb{E}}_{\mathcal{M}}(u_a) - \underline{\mathbb{E}}_{\mathcal{M}}(u_{a^*}) \overset{(135)}{\leqslant} 0 \tag{137}$$

for all $a \in \mathbb{A}$. Hence, $a \in \mathbb{A}^{max}_{\mathcal{M}}$. □

5.6 Γ-Maximax and a combined approach

In the previous two paragraphs, we saw two ways of cautiously generalizing the concept of maximizing expected utility (that is, generalizing equation (101) to the imprecise case). In both cases, the corresponding criterion takes only the lower bound of the interval expectation of an action into account. Therefore, both Γ-Maximin and maximality are examples for so called *ambiguity-averse* decision criteria: An actor applying such a criterion completely ignores the fact that all the measures contained in $\mathcal{M}$ are equally plausible to be the true one.

Instead, the actor *acts as if* he was playing a game against an omniscient antagonist, who may freely choose between the measures contained in $\mathcal{M}$. Although this approach might be adequate in certain situations, in general, it seems to be *over-pessimistic* and, therefore, fails to describe the process of decision making of rational actors (see for example [3, § 5.1] or Example 10: Here, the ordering $\leqslant^{\alpha}$ only characterizes the experts' preference order, if the degree of optimism is at least 50 %).

One question immediately comes to mind: Are there generalizations of the concept of maximizing expected utility that are less cautious (or *pessimistic*)? In other words: Can the cautious approach be modified (or generalized) for arbitrary beliefs concerning the *degree of appropriate scepticism*?

Obviously, one way of proceeding is the other extreme, namely: Choose an action if, and only if, it maximizes the expected utility with respect to the *most favourable* measure contained in the set $\mathcal{M}$. In other words: With respect to this concept, an action is optimal, if the corresponding random variable maximizes the *upper bound* of the interval expectation under $\mathcal{M}$. The criterion arising from this considerations is often referred to as Γ-*Maximax-criterion*. This motivates the following definition. A similar definition as the one given here, can for example be found in [23, p. 193, Definition 8.3].

Definition 25. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$. Define the set $\mathcal{M}$ like in equation (86). Again, $\mathcal{M}$ can be interpreted as a convex credal set or the structure of an F-probability respectively. Then the criterion

$$\Phi^{\mathfrak{A}}_{\mathcal{M}} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \max_{\pi \in \mathcal{M}} \mathbb{E}_{\pi}(u_a) =: \overline{\mathbb{E}}_{\mathcal{M}}(u_a) \tag{138}$$

is called $\Gamma(\mathcal{M})$-*Maximax-criterion*. Every $\Phi^{\mathfrak{A}}_{\mathcal{M}}$-optimal action is called $\Gamma(\mathcal{M})$-*Maximax-action*. The set of all $\Gamma(\mathcal{M})$-Maximax-actions then is denoted by $\mathbb{A}_{\overline{\mathcal{M}}}$. ▽

Remark. 1.) Again, according to Theorem 14, the above criterion is well-defined,

as the maximum is actually attained on the set $\mathcal{M}$ for all $a \in \mathbb{A}$ fixed. For a more in-depth argumentation, see the Remark on Definition 24.

2.) Note that the above criterion corresponds to the ordering $\leqslant^o$ introduced in Excursus 4.2.3: An action $a^* \in \mathbb{A}$ is $\Gamma(\mathcal{M})$-Maximax-action if, and only if, the inequality $\mathbb{E}_{\mathcal{M}}(u_a) \leqslant^o \mathbb{E}_{\mathcal{M}}(u_{a^*})$ holds for all $a \in \mathbb{A}$. ∘

Naturally, the above definition gives rise to two related questions: Under what conditions do $\Gamma(\mathcal{M})$-Maximax actions exist? And: How can they be determined in concrete decision problems?

At first sight, the task of determining a Γ-Maximax action seems to be very similar to the task of determining a Γ-Maximin action. Surprisingly, it seems not to be possible to formulate this task as a single linear programming problem. Instead, simultaneously maximizing the expression $\mathbb{E}_\pi(G(u)_p)$ in the variables π and p necessarily leads to a *multi-linear* optimization problem (see for example [3, Lemma 2] or [41, §4]).

For the moment, let us place back the question how Γ-Maximax-actions can be determined and, instead, face another problem: Obviously, applying the Γ-Maximax-criterion seems to be *over-optimistic* as it reflects the attitude of *ambiguity-seeking* actors only. Therefore, the criterion is vulnerable for a similar criticism as Γ-Maximin: Assuming total indifference between the elements of $\mathcal{M}$, why should it be reasonable to always expect the most favourable one among them? Maybe this optimism might be reasonable in certain situations. However, it does not seem to be suitable to offer a general framework for rational decision making under interval probability.

For this reason, we now want to recall a decision criterion that allows the actor to flexibly model his personal attitude towards ambiguity by a real-valued *caution parameter* $\eta \in [0, 1]$. According to this criterion, an action is optimal, if it maximizes a convex combination of the Γ-Maximin-criterion and the Γ-Maximax-criterion.

In this way, the decision maker can freely weight the influences of the two criteria: While values of η that are close to 0 indicate a cautious decision maker, values of η that are close to 1 indicate an optimistic decision maker. In general, such a criterion seems to be more suitable to model the decision processes of real actors, since almost no one acts purely optimistic or pessimistic respectively. The criterion can be found e.g. in [3, § 5.1] or [41, § 4].

Definition 26. Let $\mathfrak{A}$ be a CDP such that $|\Theta| < \infty$. Define the set $\mathcal{M}$ like in

equation (86) and let $\eta \in [0,1]$. Then the criterion

$$\Phi^{\mathfrak{A}}_{\mathcal{M},\eta} : \mathbb{A} \to \mathbb{R} \ , \ a \mapsto \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}}(u_a) + (1-\eta) \cdot \overline{\mathbb{E}}_{\mathcal{M}}(u_a) \tag{139}$$

is called $(\mathcal{M},\eta)$-*criterion*. Every $\Phi^{\mathfrak{A}}_{\mathcal{M},\eta}$-optimal action is called $(\mathcal{M},\eta)$-*action*. The set of all $(\mathcal{M},\eta)$-actions then is denoted by $\mathbb{A}_{(\mathcal{M},\eta)}$. ▽

Remark. 1.) As a convex-combination of two well-defined criteria, the above criterion is well-defined as well (of course, this can also be derived by directly applying Theorem 14). Additionally, note that both Γ-Maximin-criterion ($\eta = 1$) and Γ-Maximax-criterion ($\eta = 0$) are special cases of the above criterion. Particularly, this implies that any general method for computing $(\mathcal{M},\eta)$-actions can be used for determining Γ-Maximax-actions as well.

2.) Note that the above criterion corresponds to the ordering $\leqslant^{\alpha}$ introduced in Excursus 4.2.3: An action $a^* \in \mathbb{A}$ is $(\mathcal{M},\eta)$-optimal if, and only if, the inequality $\mathbb{E}_{\mathcal{M}}(u_a) \leqslant^{\eta} \mathbb{E}_{\mathcal{M}}(u_{a^*})$ holds for all $a \in \mathbb{A}$. ∘

Trying to formulate the task of finding an $(\mathcal{M},\eta)$-action as a linear optimization problem gives rise to the same difficulties as in the case of the Γ-Maximax-criterion: Again, simultaneously maximizing the expression $\mathbb{E}_{\pi}(G(u)_p)$ in the variables π and p necessarily leads to a *multi-linear* optimization problem. However, in [41, § 4], *Augustin* and *Utkin* show that solving this bilinear optimization problem can be avoided by solving a finite number of linear optimization problems instead. Their technique, which is mainly based on the fundamental theorem of linear optimization, is used in the proof of the following theorem.

Theorem 22. Let $\mathfrak{A}$ be a finite CDP. Further, let $\mathcal{M}$ denote a set of the form defined in equation (86) and let $\eta \in [0,1]$. Then, there exists a randomized $(\mathcal{M},\eta)$-action $p^* \in G(\mathbb{A})$.

Proof. Let $\eta \in [0,1]$. First, note that the set $b(\mathcal{M}) = \Pi_m$ (see the proof of Theorem 9) is a convex polyhedron, since it can be written in the form

$$\Pi_m = \Big\{ x \in \mathbb{R}^m : K \cdot x \leqslant l \Big\} \tag{140}$$

where K and l are defined like in the proof of Theorem 14. Hence, according to Theorem 7, the set Π_m admits only finitely many extreme points. Let

$$\mathcal{E}(\Pi_m) := \Big\{ \pi^{(1)}, \dots, \pi^{(d)} \Big\} \tag{141}$$

denote a listing of all extreme points. For $\pi^{(k)} \in \mathcal{E}(\Pi_m)$, consider the optimization problems

$$f_k(p) := \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}} G(u)_p + (1-\eta) \cdot \sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j^{(k)} \longrightarrow \max_{p \in G(\mathbb{A})} \tag{142}$$

Hence, we receive d optimization problems. Next, note that, for $p \in G(\mathbb{A})$ fixed, the following identity holds:

$$\underline{\mathbb{E}}_{\mathcal{M}} G(u)_p = \max\Big\{ w_1 - w_2 : w_1 - w_2 \leqslant \sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j \;\; \forall \pi \in \Pi_m \Big\} \tag{143}$$

However, according to Theorem 8, for every fixed action $p \in G(\mathbb{A})$ there exists a value $z(p) \in \{1, ..., d\}$ such that $\pi^{(z(p))} \in \mathcal{E}(\Pi_m)$ and

$$\sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j^{(z(p))} = \max\Big\{ \sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j : \pi \in \Pi_m \Big\} \tag{144}$$

Therefore, equation (143) reduces to the equation

$$\underline{\mathbb{E}}_{\mathcal{M}} G(u)_p = \max\Big\{ w_1 - w_2 : w_1 - w_2 \leqslant \sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j \;\; \forall \pi \in \mathcal{E}(\Pi_m) \Big\} \tag{145}$$

Now, this yields that the k-th optimization problem in (142) admits an optimal solution if, and only if, the optimization problem

$$\eta \cdot (w_1 - w_2) + (1-\eta) \cdot \sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j^{(k)} \longrightarrow \max_{(w_1,\ldots,p_n)} \tag{146}$$

with constraints

- $(w_1, \ldots, p_n) \geqslant 0$
- $p \in \Delta_n$
- $w_1 - w_2 \leqslant \sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j \;\; \forall \pi \in \mathcal{E}(\Pi_m)$

has an optimal solution. Here, we again used the one-to-one correspondence of the sets $G(\mathbb{A})$ and Δ_n (see paragraph 3.1).

To see that the problem (146) indeed defines a *linear* optimization problem, note

that it can equivalently be written in the following form:

$$\Big(\eta, -\eta, (1-\eta)\cdot c_1^{\pi^{(k)}}, \ldots, (1-\eta)\cdot c_n^{\pi^{(k)}}\Big)\cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \max_{(w_1,\ldots,p_n)} \tag{147}$$

with constraints

- $(w_1, \ldots, p_n) \geqslant 0$
- $$\begin{pmatrix} 0 & 0 & 1 & \ldots & 1 \\ 0 & 0 & 1' & \ldots & 1' \\ 1 & 1' & -c_1^{\pi^{(1)}} & \ldots & -c_n^{\pi^{(1)}} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & 1' & -c_1^{\pi^{(d)}} & \ldots & -c_n^{\pi^{(d)}} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ 1' \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where, for $\pi \in \mathcal{M}$ and $i = 1, \ldots, n$, the expression c_i^{π} is defined like in the proof of Theorem 9. This gives us a linear optimization problem.

To see that problem (147) possesses an optimal solution, note that, according to Theorem 2, its objective function is bounded from above by the expression

$$\eta \cdot m \cdot n \cdot |\max_{ij} u(a_i, \theta_j)| + (1-\eta)\cdot \mathbb{E}_{b^{-1}(\pi^{(k)})}(u_{a^*}) < \infty \tag{148}$$

where $a^* \in \mathbb{A}$ denotes a Bayes-action with respect to $b^{-1}(\pi^{(k)})$. Thus, according to Theorem 5, there exists an optimal solution of problem (147) and, therefore, of problem (142).

Hence, for every extreme point $\pi^{(k)} \in \mathcal{E}(\Pi_m)$, where $k \in \{1, \ldots, d\}$, the optimization problem (142) possesses an optimal solution. Thus, let $p^{(k)} \in G(\mathbb{A})$ denote an optimal solution of the k-th optimization problem in (142). Then, $f_k(p^{(k)})$ equals the optimal outcome of the k-th optimization problem in (142). Since $\mathcal{E}(\Pi_m)$ is a finite set, this implies the existence of $k^+ \in \{1, \ldots, d\}$ such that

$$f_{k^+}(p^{(k^+)}) \geqslant f_k(p^{(k)}) \geqslant f_k(p) \tag{149}$$

for all $k \in \{1, \ldots, d\}$ and $p \in G(\mathbb{A})$. Now, let $p \in G(\mathbb{A})$ be arbitrary. Choose $\pi^{z(p)} \in \mathcal{E}(\Pi_m)$ like in equation (144). Then, according to the equations (144) and

(149), the following holds:

$$\begin{aligned}
\Phi^{G(\mathfrak{A})}_{\mathcal{M},\eta}(p^{(k^+)}) &\geqslant f_{k^+}(p^{(k^+)}) \\
&\geqslant f_{z(p)}(p) \\
&= \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}} G(u)_p + (1-\eta) \cdot \max_{\pi \in \Pi_m} \left(\sum_{j=1}^{m} G(u)(p,\theta_j) \cdot \pi_j \right) \\
&= \eta \cdot \underline{\mathbb{E}}_{\mathcal{M}} G(u)_p + (1-\eta) \cdot \overline{\mathbb{E}}_{\mathcal{M}} G(u)_p \\
&= \Phi^{G(\mathfrak{A})}_{\mathcal{M},\eta}(p)
\end{aligned}$$

Here, the first inequality trivially holds, since

$$\overline{\mathbb{E}}_{\mathcal{M}}\left(G(u)_{p^{(k^+)}}\right) \geqslant \sum_{j=1}^{m} G(u)(p^{(k^+)},\theta_j) \cdot \pi_j^{(k^+)} \tag{150}$$

Since, p was chosen arbitrarily, this yields

$$\Phi^{G(\mathfrak{A})}_{\mathcal{M},\eta}(p^{(k^+)}) \geqslant \Phi^{G(\mathfrak{A})}_{\mathcal{M},\eta}(p) \tag{151}$$

for all $p \in G(\mathbb{A})$. That is, $p^{(k^+)} \in G(\mathbb{A})$ is a $(\mathcal{M},\eta)$-action.
This completes the proof. □

Remark. If we set $\eta = 0$, Theorem 22 implies a proof for the existence of randomized Γ-Maximax action as a special case. For $\eta = 1$, we get an alternative proof of Theorem 20. However, note the following: While the proof of Theorem 20 shows up a way to determine randomized Γ-Maximin actions by solving one single linear optimization problem, the same seems not to be possible neither for Γ-Maximax-actions nor for $(\mathcal{M},\eta)$-actions. As seen in the proof of Theorem 22, determining optimal actions with respect to these two criteria requires to solve one linear optimization problem *for each* extreme point. However, there also exists a way of determining $(\mathcal{M},\eta)$-actions without computing the extreme points of Π_m: In [41, § 4.2], the authors derive an algorithm that is not based on extreme points. Nevertheless, also this approach is based on solving *several* linear programming problems. ∘

Next, we want to apply the algorithm just described to an example. However, note that for applying the algorithm the extreme points of the set need to be determined. An effective algorithm for determining the extreme points of a set of the form Π_m (that is a set of the form $b(\mathcal{M})$, where $\mathcal{M}$ is of the form defined in equation (86)) can e.g. be found in [5, p. 34, Remark 1.15]. In the following example we assume

the extreme points to be already determined.

Example 20. Consider the finite decision problem $\mathfrak{A}$ defined by

u_{ij}	θ_1	θ_2	θ_3
a_1	20	30	40
a_2	10	10	60
a_3	30	30	25

Further, let our information be given by the following set of probability measures on $(\Theta, \mathcal{P}(\Theta))$:

$$\mathcal{M} := \Big\{ \pi : 0.3 \leqslant \pi_1 + \pi_2 \leqslant 0.8 \Big\}$$

Obviously, the set $\mathcal{M}$ is of the form defined in equation (86). The extreme points of the set $\Pi_3 := b(\mathcal{M})$ are given by

$$\pi^{(1)} = (0.3, 0, 0.7) \quad , \quad \pi^{(2)} = (0, 0.3, 0.7) \quad , \quad \pi^{(3)} = (0.8, 0, 0.2) \quad , \quad \pi^{(4)} = (0, 0.8, 0.2)$$

Let $\eta = 0.6$. We want to determine a randomized $(\mathcal{M}, \eta)$-action for the decision problem $G(\mathfrak{A})$. According to the proof of Theorem 22, such an action can be gained by solving the linear optimization problem (142) (and therefore problem (147)) for every extreme point and, afterwards, choosing the optimal solution of the problem with the highest optimal outcome. Thus, we arrive at the following four optimization problems (the constraints coincide for all four problems):

$$\Big(0.6, -0.6, 0.4 \cdot 34, 0.4 \cdot 45, 0.4 \cdot 26.5\Big) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \longrightarrow \max_{(w_1, \ldots, p_3)}$$

$$\Big(0.6, -0.6, 0.4 \cdot 37, 0.4 \cdot 45, 0.4 \cdot 26.5\Big) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \longrightarrow \max_{(w_1, \ldots, p_3)}$$

$$\Big(0.6, -0.6, 0.4 \cdot 24, 0.4 \cdot 20, 0.4 \cdot 29\Big) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \longrightarrow \max_{(w_1,\ldots,p_3)}$$

$$\Big(0.6, -0.6, 0.4 \cdot 32, 0.4 \cdot 20, 0.4 \cdot 29\Big) \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \longrightarrow \max_{(w_1,\ldots,p_3)}$$

with constraints

- $(w_1, \ldots, p_3) \geqslant 0$
- $$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & -1 & -1 \\ 1 & -1 & -34 & -45 & -26.5 \\ 1 & -1 & -37 & -45 & -26.5 \\ 1 & -1 & -24 & -20 & -29 \\ 1 & -1 & -32 & -20 & -29 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} \leqslant \begin{pmatrix} 1 \\ 1' \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Resolving the above optimization problems in R gives us the optimal solutions

$$(20, 0, 0, 1, 0) \quad , \quad (20, 0, 0, 1, 0) \quad , \quad (28.1\bar{8}, 0, 0, 0.0\bar{9}, 0.9\bar{0}) \quad , \quad (28, 0, 0.2, 0, 0.8)$$

Obviously, the third optimization problem has the highest optimal outcome, namely $28.1\bar{8}$. Hence, according to the proof of Theorem 22, a $(\mathcal{M}, 0.6)$-optimal randomized action $p^* \in G(\mathbb{A})$ is induced by the assignment

$$p^*(\{a_i\}) = \begin{cases} 0 & \text{if } i = 1 \\ 0.0\bar{9} & \text{if } i = 2 \\ 0.9\bar{0} & \text{if } i = 3 \end{cases}$$

$\star$

To complete the paragraph, we want to prove a theorem that makes up a connection between decision problems under complex uncertainty and decision problems under strict type I uncertainty. The proof of the theorem is essentially based on the dual linear programming problem of problem (147).

Theorem 23. Let $\mathfrak{A}$ be a finite CDP. Further, let $\mathcal{M}$ denote a set of the form defined in equation (86) and let $\eta \in [0,1]$. According to Theorem 22, let $p^* \in G(\mathbb{A})$ denote a $(\mathcal{M},\eta)$-optimal action for the decision problem $G(\mathfrak{A})$. Then, there exists a finite decision problem $\tilde{\mathfrak{A}} := (\tilde{\mathbb{A}}, \tilde{\Theta}, \tilde{u}(\cdot))$ and a probability measure $\tilde{\varphi}$ on $(\tilde{\Theta}, \mathcal{P}(\tilde{\Theta}))$ such that

$$\Phi^{G(\mathfrak{A})}_{(\mathcal{M},\eta)}(p^*) = \Phi^{\tilde{\mathfrak{A}}}_{B(\tilde{\varphi})}(\tilde{a}^*) \tag{152}$$

where $\tilde{a}^* \in \tilde{\mathbb{A}}$ denotes a Bayes-action w.r.t $\tilde{\varphi}$ for the decision problem $\tilde{\mathfrak{A}}$.

Proof. Let $\eta \in [0,1]$ and let $\mathcal{E}(\Pi_m)$ be defined like in the proof of Theorem 22. Then, according to Theorem 22, there exists $\pi^{(l)} \in \mathcal{E}(\Pi_m)$ such that $\Phi^{G(\mathfrak{A})}_{(\mathcal{M},\eta)}(p^*)$ equals the optimal outcome of the linear optimization problem (147), where $k := l$.

Thus, by duality, the expression $\Phi^{G(\mathfrak{A})}_{(\mathcal{M},\eta)}(p^*)$ equals the optimal outcome of the dual of problem (147), given by

$$(1, -1, \underbrace{0, \ldots, 0}_{\text{d-times}}) \cdot \begin{pmatrix} u_1 \\ u_2 \\ \kappa_1 \\ \vdots \\ \kappa_d \end{pmatrix} \longrightarrow \min_{(u_1,\ldots,\kappa_d)} \tag{153}$$

with constraints

- $(u_1, \ldots, \kappa_d) \geqslant 0$
- $$\begin{pmatrix} 0 & 0 & 1 & \ldots & 1 \\ 0 & 0 & 1' & \ldots & 1' \\ 1 & 1' & -c_1^{\pi^{(1)}} & \ldots & -c_1^{\pi^{(d)}} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & 1' & -c_n^{\pi^{(1)}} & \ldots & -c_n^{\pi^{(d)}} \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ \kappa_1 \\ \vdots \\ \kappa_d \end{pmatrix} \geqslant \begin{pmatrix} \eta \\ -\eta \\ (1-\eta) \cdot c_1^{\pi^{(l)}} \\ \vdots \\ (1-\eta) \cdot c_n^{\pi^{(l)}} \end{pmatrix}$$

Now, let $(u_1^*, u_2^*, \kappa_1^*, \ldots, \kappa_d^*)$ denote an optimal solution of problem (153). Then, due to optimality, the following two properties hold:

$$\sum_{k=1}^{d} \kappa_k^* = \eta \tag{154}$$

$$u_1^* - u_2^* = \max\left\{ \left(\sum_{k=1}^{d} \kappa_k^* \cdot c_i^{\pi^{(k)}} \right) + (1-\eta) \cdot c_i^{\pi^{(l)}} : i = 1, \ldots, n \right\} \tag{155}$$

Next, define the vector $\mu := (\mu_1, \ldots, \mu_d)$ by

$$\mu_k = \begin{cases} \kappa_k^* & \text{if } k \neq l \\ \kappa_k^* + (1-\eta) & \text{if } k = l \end{cases}$$

Clearly, we have $\sum_{k=1}^d \mu_k = 1$ and $\mu_k \geqslant 0$ for all $k = 1, \ldots, d$. Thus, the assignment $\tilde{\varphi}(\{\pi^{(k)}\}) := \mu_k$, for $k = 1, \ldots, d$, induces a probability measure on the measurable space $(\mathcal{E}(\Pi_m), \mathcal{P}(\mathcal{E}(\Pi_m)))$.

Now, define the finite decision problem $\tilde{\mathfrak{A}} := (\tilde{\mathbb{A}}, \tilde{\Theta}, \tilde{u}(\cdot))$ by

- $\tilde{\mathbb{A}} := \{\tilde{a}_1, \ldots, \tilde{a}_n\}$
- $\tilde{\Theta} := \mathcal{E}(\Pi_m)$
- $\tilde{u} : \tilde{\mathbb{A}} \times \mathcal{E}(\Pi_m) \to \mathbb{R} \quad , \quad (\tilde{a}_i, \pi^{(k)}) \mapsto c_i^{\pi^{(k)}} := \sum_{j=1}^m u(a_i, \theta_j) \cdot \pi_j^{(k)}$

Let, according to Theorem 2, $\tilde{a}^* \in \tilde{\mathbb{A}}$ denote a Bayes-action w.r.t $\tilde{\varphi}$ for the decision problem $\tilde{\mathfrak{A}}$. We compute

$$\begin{aligned} \Phi_{B(\tilde{\varphi})}^{\tilde{\mathfrak{A}}}(\tilde{a}^*) &= \max\left\{ \sum_{k=1}^d \tilde{u}(\tilde{a}_i, \pi^{(k)}) \cdot \tilde{\varphi}(\{\pi^{(k)}\}) : i = 1, \ldots, n \right\} \\ &= \max\left\{ \sum_{k=1}^d c_i^{\pi^{(k)}} \cdot \mu_k : i = 1, \ldots, n \right\} \\ &= \max\left\{ \left(\sum_{k=1}^d \kappa_k^* \cdot c_i^{\pi^{(k)}} \right) + (1-\eta) \cdot c_i^{\pi^{(l)}} : i = 1, \ldots, n \right\} \\ &\overset{(155)}{=} u_1^* - u_2^* \\ &= \Phi_{(\mathcal{M},\eta)}^{G(\mathfrak{A})}(p^*) \end{aligned}$$

This completes the proof. □

In the following example, we demonstrate how the technique applied in the proof of Theorem 23 can be used to construct the finite decision problem $\tilde{\mathfrak{A}}$ and the probability measure $\tilde{\varphi}$ given the finite decision problem $\mathfrak{A}$.

Example 21. Consider again the setting of Example 20 (that is, the decision problem $\mathfrak{A}$, the set $\mathcal{M}$ and the parameter value $\eta = 0.6$). According to the proof of Theorem 23, the decision problem $\tilde{\mathfrak{A}}$ is given by

$\tilde{u}_{ik}$	$\pi^{(1)}$	$\pi^{(2)}$	$\pi^{(3)}$	$\pi^{(4)}$
$\tilde{a}_1$	34	37	24	32
$\tilde{a}_2$	45	45	20	20
$\tilde{a}_3$	26.5	26.5	29	29

and the corresponding measure $\tilde{\varphi}$ can be gained by solving the dual of the third linear optimization problem from Example 20 (because the third problem is the one with the highest optimal outcome, as shown in Example 20). The corresponding dual optimization problem is given by

$$(1,-1,0,0,0,0)\cdot\begin{pmatrix}u_1\\u_2\\\kappa_1\\\kappa_2\\\kappa_3\\\kappa_4\end{pmatrix}\longrightarrow\min_{(u_1,\ldots,\kappa_4)}$$

with constraints

- $(u_1,\ldots,\kappa_4)\geqslant 0$
- $$\begin{pmatrix}0&0&1&1&1&1\\0&0&-1&-1&-1&-1\\1&-1&-34&-37&-24&-32\\1&-1&-45&-45&-20&-20\\1&-1&-26.5&-26.5&-29&-29\end{pmatrix}\cdot\begin{pmatrix}u_1\\u_2\\\kappa_1\\\vdots\\\kappa_4\end{pmatrix}\geqslant\begin{pmatrix}0.6\\-0.6\\0.4\cdot 24\\0.4\cdot 20\\0.4\cdot 29\end{pmatrix}$$

Resolving this problem in R returns the optimal solution

$$(28.\bar{18}\ ,\ 0\ ,\ 0.3\bar{27}\ ,\ 0\ ,\ 0.2\bar{7}\ ,\ 0)$$

Thus, according to the proof of Theorem 23, the measure $\tilde{\varphi}$ on $(\mathcal{E}(\Pi_3),\mathcal{P}(\mathcal{E}(\Pi_3)))$ is induced by the assignment

$$\tilde{\varphi}(\{\pi^{(k)}\}):=\mu_k:=\begin{cases}0.3\bar{27} & \text{if } k=1\\ 0 & \text{if } k=2\\ 0.2\bar{7}+0.4 & \text{if } k=3\\ 0 & \text{if } k=4\end{cases}$$

Now, in order to determine a Bayes-optimal action with respect to $\tilde{\varphi}$ for the decision problem $\tilde{\mathfrak{A}}$, we compute the following expression:

$$\mathbb{E}_{\tilde{\varphi}}(\tilde{u}_{\tilde{a}_i}) = \begin{cases} 27.2\bar{7} & \text{if } i = 1 \\ 28.1\bar{8} & \text{if } i = 2 \\ 28.1\bar{8} & \text{if } i = 3 \end{cases}$$

Hence, Bayes-action are given by $\tilde{a}_3$ and $\tilde{a}_3$ and we have

$$\Phi^{\tilde{\mathfrak{A}}}_{B(\tilde{\varphi})}(\tilde{a}_2) = \Phi^{\tilde{\mathfrak{A}}}_{B(\tilde{\varphi})}(\tilde{a}_3) = \Phi^{G(\mathfrak{A})}_{(\mathcal{M},0.6)}(p^*)$$

where p^* denotes the $(\mathcal{M}, 0.6)$-optimal action from Example 20. ⋆

6 Some concluding remarks

In this last chapter of the present work, we aim to fulfil two objectives: First, we want to give a brief summary of the main aspects treated throughout this thesis. Subsequently, we want to point out topics that weren't treated within this work.

6.1 Summary

Generally speaking, this thesis is about *optimal* decision making under different assumptions concerning the *mechanism generating the states of nature* in finite decision problems. Specifically, we focussed the question: How can one use the information available *best possible* in order to determine a *rational* decision?

In Chapter one, we treated the *classical* case of information bases either describable by a *classical probability measure* or comparable to a *game against an omniscient antagonist*. Furthermore, we recalled arguments that in such cases optimal criteria are given by the *Bernoulli/Bayes-criterion* or the *Maximin-criterion* respectively. Additionally, we recalled a criterion that allows to label actions as optimal if the information available is best described by a mixture of the types mentioned above: The *Hodges & Lehmann-criterion*.

At this point, the following question naturally comes up: How can we determine optimal actions with respect to the criteria discussed? Chapter two lies the theoretical foundation to answer this question, namely the theory of *Linear Optimization*. This well-investigated theory not only allows us to prove the existence of optimal solution of linear optimization problems, but also can be used to determine such solutions by using standard statistical software.

In Chapter three, we recall how the classical criteria discussed in Chapter one can be embedded into the theoretical framework of linear optimization theory. More precisely, the task of optimizing each of the three classical criteria can be reformulated as the task of solving a suitable linear optimization problem. Using this circumstance makes it possible to prove the existence of optimal actions by applying the general results from linear programming. Furthermore, in all cases the idea of the proofs is *constructive* and, therefore, can be used as a method for determining optimal actions.

Two results of this chapter seem to be of particular interest: The first one demonstrates how linear programming theory can be used to determine a *least favourable prior distribution* under all distributions contained in a linearly defined set (i.e. a set of the for defined in equation (86)). Such least favourable distributions then turn

out to have a deep connection to the Γ-Maximin-criterion that is recalled in Chapter five.

Additionally, a connections between the Bayes-criterion and the Hodges & Lehmann-criterion is shown up: For every prior measure π on the set of states and every trade-off (or *optimism*) parameter $\alpha \in [0, 1]$, there exists a measure $\lambda_{\pi,\alpha}$ such that the Bayes-utility with respect to $\lambda_{\pi,\alpha}$ equals the Hodges & Lehmann- utility with respect to π. Again, we demonstrate a method for determining such measures using linear optimization.

Afterwards, Chapter four seriously takes account of the fact that certain kinds of uncertainty cannot be captured by considering classical probabilistic models only. If the quality of the informations base is to weak, it often turns out not to be possible to give precise probabilistic descriptions of the situation. Instead one has to consider *generalized descriptions of uncertainty*. The mentioned inabilities of the classical theory are illustrated by a number of examples from various scientific disciplines. Furthermore, we recall some theoretical reasons that support the idea of a generalization of probability theory.

Accordingly, in the second part of Chapter four, we recall two common generalizations of the classical theory: *Credal sets* and *interval probability*. Both frameworks are strongly related to each other. Furthermore, they offer natural approaches for dealing with imperfect information avoiding the inconsistencies arising from the usage of classical probability theory.

In Chapter five, we recall criteria for optimal decision making if the uncertainty is described by a credal set or an interval probability respectively. However, the choice of a 'good' decision criterion turns out to be less obvious as in the precise case: What makes a good criterion strongly depends on the decision maker's attitude towards ambiguity (i.e. on the way he faces the complete indifference between the measure contained in the credal set).

However, even in the imprecise case there exist criteria that are independent of the actor's attitude towards ambiguity. We recall two of them, namely *interval dominance* and *E-admissibility*. Interval dominance excludes actions as being unreasonable choices, whenever the best possible (compatible) expectation of an action is dominated by the worst possible (compatible) expectation of another action. E-admissibility excludes action that aren't Bayes-optimal with respect to every measure contained in the credal set.

Afterwards, we explain two decision criteria that reflect the preference orders of

ambiguity-averse actors (by taking into account only the lower bound of the interval expectations), namely *Maximality* and Γ-*Maximin*. Both of them reduce to the classical Bayes-criterion, if the underlying credal set consists of one single element, that is if the uncertainty is *purely stochastic*. However, in the presence of ambiguity they no longer coincide. In the context of the Γ-Maximin-criterion, one result seems to be of particular interest: If π^* is a least favourable prior from the credal set $\mathcal{M}$, then the $\Gamma(\mathcal{M})$-Maximin-utility and the Bayes-utility with respect to π^* coincide. This builds a bridge between precise and imprecise decision theory.

Finally, we recall an imprecise decision criterion that allows us to reflect the preference ordering of decision makers with arbitrary attitudes towards ambiguity. Here, the degree of *ambiguity-averseness* is characterized by a parameter $\eta \in [0, 1]$. For $\eta = 0$ we arrive at the prototypical criterion for *ambiguity-seeking* decision makers, namely the Γ-*Maximax-criterion*. The case $\eta = 1$ coincides with the Γ-Maximin-criterion mentioned above.

Additionally, we recall and apply algorithms for determining optimal actions with respect to all of the imprecise decision criteria discussed. Again, these algorithms are based on linear programming theory and, therefore, implemented in standard mathematical software.

6.2 Outlook

So, which topics weren't treated within this work? To answer this, bring to mind again the *Fundamental Theorem of Bayesian Decision Theory* (see Theorem 3 in Paragraph 1.3):

When considering a data-based extension $\mathcal{D}(\mathfrak{A})$ of a finite decision problem $\mathfrak{A}$ (see Definition 3), we can avoid determining a $B(\pi)$-optimal decision function d^* by simply determining a $B(\pi_x)$-optimal action a_x^* in the basic problem with respect to the measure π_x updated for the *observed* data x. The $B(\pi_x)$-utility values of the actions a_x^* and $d^*(x)$ necessarily coincide. For arguments supporting the importance of the theorem for classical decision theory see the discussions in Remark 3 of Definition 3 and in the Remark on Theorem 3.

Naturally, the following question comes to mind: Are there similar results for the imprecise decision criteria discussed in Chapter five? More precisely: Can we avoid computing optimal decision function also when optimality is defined in terms of some imprecise decision criterion (as for example E-admissibility of Γ-Maximin)?

An in-depth discussion of this topic can be found for example in [4]. Here, we just

want to briefly summarize the main aspects. We start by generalizing the concept of updating to the imprecise case. Suppose

$$\mathcal{D}(\mathfrak{A}) := (\mathcal{D}(\mathbb{A}, \mathcal{X}), \Theta, U(\cdot))$$

to be a data-based extension of some finite decision problem $\mathfrak{A}$ (in the sense of Definition 3), where $\mathcal{X} := \{x_1, \ldots, x_k\}$ is a finite space of observations. Further, let $\mathcal{M}$ denote a set of probability measures on the measurable space $(\Theta, \mathcal{P}(\Theta))$ of the form defined in equation (86). For every pair $(\pi, x) \in \mathcal{M} \times \mathcal{X}$, define the updated measure π_x like in equation (24). Then, for fixed $x \in \mathcal{X}$, the set

$$\mathcal{M}_x := \left\{\pi_x : \pi \in \mathcal{M}\right\} \tag{156}$$

is called the *updated credal set* with respect to the data x (see [4, Definition 2]).

Now, we can give a precise formulation of the question posed above: Can we always determine $\mathcal{M}_x$-optimal actions instead of $\mathcal{M}$-optimal decision functions? (Here, '$\mathcal{M}$-optimal' means optimal w.r.t. some imprecise decision criterion under the information $\mathcal{M}$.)

Unfortunately, the answer is 'no' in many cases. That is, in general, we will not gain an $\mathcal{M}$-optimal decision function d^* by computing an $\mathcal{M}_x$-optimal action a_x for every $x \in \mathcal{X}$ and, afterwards, defining $d^*(x) := a_x$.

However, if optimality is defined in terms of $\mathrm{E}(\mathcal{M})$-admissibility or $\mathcal{M}$-maximality respectively, the procedure described above will lead to a $\mathcal{M}$-optimal decision function. Particularly, for these two imprecise decision criteria a generalization of the fundamental theorem holds.

In contrast, a simple counterexample for $\Gamma(\mathcal{M})$-Maximin is given in [4, p. 40]: Suppose the set $\mathcal{M}$ is vacuous (see Paragraph 5.5). Then, so is $\mathcal{M}_x$ (for no matter what $x \in \mathcal{X}$ we observe, see [43, p. 308]). Thus, according to the considerations in the beginning of Paragraph 5.5, both the $\Gamma(\mathcal{M})$-criterion for the decision problem $G(\mathcal{D}(\mathfrak{A}))$ and the $\Gamma(\mathcal{M}_x)$-criterion for the decision problem $G(\mathfrak{A})$ reduce to the corresponding classical Maximin-criterion.

However, according to Remark 2 on Example 3, there exists no pendant of the fundamental theorem for the classical Maximin-criterion. More precisely, a similar construction as described above would lead to a *constant* decision function, whereas, in general, a Maximin-optimal decision function doesn't have to be constant. This gives us a counterexample. Hence, the fundamental theorem cannot be generalized

to the case where optimality is defined in terms of the Γ-Maximin-criterion.

Since the $\Gamma(\mathcal{M})$-Maximin-criterion coincides with a special case of the $(\mathcal{M}, \eta)$-criterion (namely with the $(\mathcal{M}, 1)$-criterion, see Remark 1 on Definition 26 for further detail), this implies that a generalization of the fundamental theorem is not possible if optimality is defined in terms of the $(\mathcal{M}, \eta)$-criterion.

References

[1] A.U. *An Expert System: MYCIN*. 2011. URL: http://de.slideshare.net/vini89/mycin (visited on 01/13/2015).

[2] Augustin, Thomas. *Entscheidungstheorie*. 2014. URL: http://www.statistik.lmu.de/institut/ag/agmg/lehre/2014_SoSe/Entscheidungstheorie/ET14_Folien.pdf (visited on 12/31/2014).

[3] Augustin, Thomas. "Expected utility within a generalized concept of probability - a comprehensive framework for decision making under ambiguity". In: *Statistical Papers* 43.1 (2002), pp. 5–22.

[4] Augustin, Thomas. "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: *ISIPTA '03*. Ed. by Bernard, Jean-Marc, Seidenfeld, Teddy, and Zaffalon, Marco. Vol. 18. Proceedings in informatics. Ottawa: Carleton Scientific, 2003, pp. 31–45.

[5] Augustin, Thomas. *Optimale Tests bei Intervallwahrscheinlichkeit*. Göttingen: Vandenhoeck & Ruprecht, 1998.

[6] Berger, James O. *Statistical decision theory, foundations, concepts, and methods*. Springer series in statistics. New York: Springer-Verlag, 1980.

[7] Bernardo, José M. and Smith, Adrian F. M. *Bayesian theory*. Wiley series in probability and statistics. Chichester and New York: Wiley, 2000.

[8] Bernoulli, Jakob. *The art of conjecturing, together with Letter to a friend on sets in court tennis*. Baltimore, Md.: Johns Hopkins University Press, 2006.

[9] Dantzig, George B. *Linear programming and extensions*. 11th edition. Princeton Landmarks in Mathematics and Physics. Princeton, N.J. [u.a.]: Princeton Univ. Press, 1998.

[10] Ellsberg, Daniel. "Risk, ambiguity, and the savage axioms". In: *The quarterly journal of economics* 75.4 (1961), pp. 643–669.

[11] Elstrodt, Jürgen. *Maß- und Integrationstheorie*. 7th edition. Grundwissen Mathematik. Berlin [u.a.]: Springer-Verlag, 2011.

[12] Fahrmeir, Ludwig. *Statistik: Der Weg zur Datenanalyse*. 7th edition. Springer-Lehrbuch. Berlin and Heidelberg: Springer-Verlag, 2010.

[13] Ferschl, Franz. *Nutzen- und Entscheidungstheorie: Einführung in die Logik der Entscheidungen*. Vol. 10. Moderne Lehrtexte / Wirtschaftswissenschaften. Opladen: Westdt. Verl., 1975.

[14] Finetti, Bruno de. *Theory of probability: A critical introductory treatment. Vol. 1*. Wiley series in probability and mathematical statistics. London and New York, N.Y: Wiley, 1974.

[15] Finetti, Bruno de. *Theory of probability: A critical introductory treatment Vol. 2*. Wiley series in probability and mathematical statistics. London: Wiley, 1975.

[16] Fischer, Gerd. *Lineare Algebra: Eine Einführung für Studienanfänger*. 17th edition. Studium: Grundkurs Mathematik. Wiesbaden: Vieweg + Teubner, 2010.

[17] Forster, Otto. *Analysis 2*. 6th edition. Vieweg Studium: Grundkurs Mathematik. Braunschweig and Wiesbaden: Vieweg, 2005.

[18] Forster, Otto. *Analysis*. 7th edition. Vieweg Studium: Grundkurs Mathematik. Braunschweig and Wiesbaden: Vieweg, 2004.

[19] Ganter, Bernhard. *Diskrete Mathematik: Geordnete Mengen*. Springer-Lehrbuch. Berlin, Heidelberg: Imprint: Springer Spektrum, 2013.

[20] Heckerman, David. "The Certainty Factor Model". In: *Encyclopedia of artificial intelligence*. Ed. by Shapiro, Stuart C. New York, Chichester, and Brisbane: J. Wiley and Sons, 1992, pp. 131–138.

[21] Hisdal, Ellen. "Are grades of membership probabilities?" In: *Fuzzy Sets and Systems* 25 (1988), pp. 325–348.

[22] Hodges, Joseph L. and Lehmann, Erich L. "The use of Previous Experience in Reaching Statistical Decisions". In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 396–407.

[23] Huntley, Nathan, Hable, Robert, and Troffaes, Matthias C.M. "Decision making". In: *Introduction to imprecise probabilities*. Ed. by Augustin, Thomas, Coolen, Frank P. A., Cooman, Gert de, and Matthias C. M. Troffaes. Weinheim: Wiley, 2014, pp. 190–206.

[24] Jarre, Florian and Stoer, Josef. *Optimierung*. 1st edition. Berlin [u.a.], 2004.

[25] Jungnickel, Dieter. *Optimierungsmethoden: Eine Einführung*. 2nd edition. Springer-Lehrbuch. Berlin [u.a.]: Springer-Verlag, 2008.

[26] Klenke, Achim. *Wahrscheinlichkeitstheorie*. 3rd edition. Springer-Lehrbuch Masterclass. Berlin: Springer-Verlag, 2013.

[27] Korte, Bernhard and Vygen, Jens. *Kombinatorische Optimierung: Theorie und Algorithmen*. 1st edition. Berlin: Springer-Verlag, 2008.

[28] Kosmol, Peter. *Optimierung und Approximation*. 2nd edition. De Gruyter Studium. Berlin: De Gruyter, 2010.

[29] Laux, Helmut. *Entscheidungstheorie: Mit 12 Tabellen*. 7th edition. Springer-Lehrbuch. Berlin and Heidelberg: Springer-Verlag, 2007.

[30] Levi, Isaac. *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. 1st edition. Cambridge, Mass.: MIT Press, 1983, 1980.

[31] Neumann, John von and Morgenstern, Oskar. *Theory of games and economic behavior*. 60th anniversary edition. Princeton classic editions. Princeton: Princeton University Press, 2007.

[32] Popchev, Ivan and Zlatareva, Neli. "An expert system in reliability-structure and knowledge representation". In: *Artificial Intelligence*. Ed. by Bibel, Wolfgang and Petkoff, Boris. Amsterdam and New York: North-Holland, 1985, pp. 199–205.

[33] Pratt, John W., Raiffa, Howard, and Schlaifer, Robert. *Introduction to statistical decision theory*. Cambridge, Mass.: MIT Press, 1995.

[34] Rieger, Marc Oliver and Wang, Mei. "Cumulative prospect theory and the St. Petersburg paradox". In: *Economic Theory* 28.3 (2006), pp. 665–679.

[35] Robert, Christian P. *The Bayesian choice: A decision-theoretic motivation*. Springer texts in statistics. New York: Springer-Verlag, 1994.

[36] Rommelfanger, Heinrich and Eickemeier, Susanne. *Entscheidungstheorie: Klassische Konzepte und Fuzzy-Erweiterungen*. Springer-Lehrbuch. Berlin and Heidelberg [u.a.]: Springer-Verlag, 2002.

[37] Rüger, Bernhard. *Test- und Schätztheorie*. Lehr- und Handbücher der Statistik. München [u.a.]: Oldenbourg, 1999.

[38] Schöbel, Anita. *Einführung in die Optimierung*. 2010. URL: http://optimierung.math.uni-goettingen.de/skripte/optimierung.pdf (visited on 01/15/2015).

[39] Suppes, Patrick. "The Philosophical Relevance of Decision Theory". In: *The Journal of Philosophy* 58.21 (1961), pp. 605–614.

[40] Troffaes, Matthias C.M. "Decision making under uncertainty using imprecise probabilities". In: *International Journal of Approximate Reasoning* 45.1 (2007), pp. 17–29.

[41] Utkin, Lev V. and Augustin, Thomas. "Powerful algorithms for decision making under partial prior information and general ambiguity attitudes". In: *Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications*. Ed. by Cozman, Fabio G., Nau, Robert, and Seidenfeld, Teddy. Carnegie Mellon University, Pittsburgh, PA, USA, 2005.

[42] Wald, Abraham. "Statistical Decision Functions". In: *The Annals of Mathematical Statistics* 20.2 (1949), pp. 165–205.

[43] Walley, Peter. *Statistical reasoning with imprecise probabilities*. 1st edition. Vol. 42. Monographs on statistics and applied probability. London and New York: Chapman and Hall, 1991.

[44] Walley, Peter. "Towards a unified theory of imprecise probability". In: *International Journal of Approximate Reasoning* 24.2-3 (2000), pp. 125–148.

[45] Walley, Peter and Cooman, Gert de. "A behavioral model for linguistic uncertainty". In: *Information Sciences* 134 (2001), pp. 1–37.

[46] Weichselberger, Kurt. *Intervallwahrscheinlichkeit als umfassendes Konzept*. Vol. 1. Elementare Grundbegriffe einer allgemeinen Wahrscheinlichkeitsrechnung. Heidelberg: Physica-Verl, 2001.

[47] Weichselberger, Kurt. "The theory of interval-probability as a unifying concept for uncertainty". In: *International Journal of Approximate Reasoning* 24.2 (2000), pp. 149–170.

[48] Weichselberger, Kurt and Augustin, Thomas. "Analysing Ellsberg's Paradox by means of interval-probability". In: *Econometrics in theory and practice*. Ed. by Galata, Robert, Küchenhoff, Helmut, and Schneeweiss, Hans. Heidelberg: Physica, 1998, pp. 291–304.

[49] Zadeh, Lotfi A. "Fuzzy sets as a basis for a theory of possibility". In: *Fuzzy Sets and Systems* 1.1 (1978), pp. 3–28.

[50] Zadeh, Lotfi A. "Fuzzy sets". In: *Information and Control* 8.3 (1965), pp. 338–353.

Declaration of Authorship

Hereby, I affirm that I have written this thesis on my own. Cited sources of literature are perceptibly marked and listed at the end of this thesis. The work was not submitted previously in same or similar form to another examination committee and was not yet published.

Munich, February 17, 2015