Decision Making under Complex Uncertainty: Evaluation and Implementation of Different Criteria

Bachelorthesis of the Faculty for Mathematics, Computer Science and Statistics

Marc Johler

Ludwig-Maximilians-Universität München
Institute for Statistics
Chair of Method(olog)ische Grundlagen der Statistik und ihre Anwendungen

supervised by
Prof. Dr. Thomas Augustin

November 12th 2020

Abstract

The aim of this work is the presentation and implementation of a selection of principles and criteria of decision theory in the statistical programming language *R*. The selection contains the most common criteria as well as some criteria introduced by the chair for "Foundations of Statistics and Their Applications". Theoretical aspects and background information on the criteria is included, as well as explanations for each of the their implementations.

Most of the algorithms used did already exist, still this thesis contains suggestions of new approaches to compute some of the criteria. The chosen algorithms will be critically evaluated and compared to possible alternatives in terms of running time. Moreover the functions and their input structure have been defined in a way which should enable the decision maker to insert her information about the acts, states and the used utility or loss function in a straight-forward and convenient way.

Many approaches in this thesis are based on the idea of *linear programming*. The algorithms for solving these kind of problems are not part of this thesis, since one can rely on already existing linear-programming-packages like *rcdd* and *lpSolve*. For the manipulation of the data and stylistical aspects of the code the author made use of the functionality of the *tidyverse* package. The checks of input validity have been partially done with the package *checkmate*.

Contents

Introduction

Decision theory is the scientific theory of rational decision making under uncertainty. Some people see it as a part of statistics, while contrarily other opinions state that statistics is in fact part of decision theory. Independently from the specific point of view, decision theory offers the possibility to model an agents preferences about a specific decision making problem in a very simple and yet also very effective way. In theory there exists a lot of concepts of different principles and criteria which could be applied at decision making problems in order to detect one or more optimal decisions with respect to the chosen criteria/principles. However, there is no widely used package in the statistical programming-language R for decision making under uncertainty. That's why this work aims to provide a tool-kit of solutions for different kinds of these problems.

In chapter 1 there will be a short introduction into *finite decision theory* and its basic underlying assumptions. In chapter 2 there will be a brief explanation of the concept of *imprecise probabilities* and how it can be utilized for decision theory. Chapter 3 is dedicated to present the so-called *Dominance principle* and the algorithms which were chosen for its implementation in this work. Chapter 4 introduces the other different decision criteria and the corresponding algorithms which have been implemented as R functions. Chapter 5 closes this thesis, by summarizing the outcome of the work and also giving some perspective for further improvements of the implementation.

1 Finite decision theory

This section aims to provide a short introduction to *finite decision theory.* The classical model is to express the decision makers preferences with a cardinal utility function, like explained in 1.1. A more flexible concept for modelling, which doesn't require a cardinal utility, is the *preference system* in 1.2.

1.1 The basic model

A problem of finite decision theory is characterized by a finite set of possible acts $\mathbb{A} = \{a_1, \ldots, a_n\}$ out of which one could be chosen by a decision maker and a finite set of possible states of nature $\Theta = \{\theta_1, \ldots, \theta_m\}$ out of which one will actually be the true state of nature. It is assumed that the utility of every pair $(a, \theta) \in \mathbb{A} \times \Theta \to \mathbb{R}$ can be described by a known real-valued cardinal utility function $u : \mathbb{A} \times \Theta \to \mathbb{R}$. Instead of using an utility function one could also use a loss function, then the notation changes from u to l. For easier readability of this thesis, loss functions are not considered, since a loss function can also be described as a negative utility function anyway.

	θ_1	$\cdots$	θ_m
a_1	$u(a_1, \theta_1)$	$\cdots$	$u(a_1, \theta_m)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_n	$u(a_n, \theta_1)$	$\cdots$	$u(a_n, \theta_m)$

Tab. 1: The classic structure of an utility table

In some situations it may make sense to consider *randomized acts.* These acts assign a certain probability λ_i to each pure act a_i, whereas it must hold that $\sum_{i=1}^{n} \lambda i = 1$. The decision for a randomized act can be interpreted as leaving the final decision to random experiment with certain probabilities for each pure act[10, p. 2]. The resulting utility of randomized act $a_r = (\lambda_1, \ldots, \lambda_n)$ is basically a linear combination of the utilities of the pure acts for which a_r has λ-values greater than 0:

$$u(a_r, \theta_j) = \sum_{i=1}^{n} \lambda_i^r \cdot u(a_i, \theta_j) \, \forall j \in \{1, \ldots, m\}$$

Pure acts can also be seen as a special case of randomized acts, where for one $\widetilde{i}$ it holds that $\lambda_{\widetilde{i}} = 1$, while $\lambda_i = 0 \; \forall i \neq \widetilde{i}$.

In order to find out which act should be prefered the decison maker can make use of decision principles like the Dominance principle (see chapter 3) and/or decision criteria like mentioned in chapter 4.

1.2 Preference Systems

In specific situations the decision maker might not be able to express her "strength of preference" in a cardinal utility or loss function, but could still be able to put the consequences of each action-state-interaction into a complete total order. In other cases there might be a group of decision makers with different cardinal utility values but on the other hand they agree on the order of these values. For both cases *preference systems* offer a suitable solution.

1.2.1 Preference Systems - Defintion

For a non-empty set of acts A and a non-empty set of states Θ one can express ordinal relationships between the act-state-interactions $u(a_i, \theta_a), i \in \{1, \ldots, n\}, a \in \{1, \ldots, m\}$ through some

relations R_1 and R_2 that are reflexive and transitive. Denote $u(a_i, \theta_a)$ as $u_{i,a}$ $\forall i \in \{1, \ldots, n\}, a \in \{1, \ldots, m\}$, then it holds that $(u_{i,a}, u_{j,b}) \in R_1$ if and only if $u_{i,a}$ is preferable compared to $u_{j,b}$.

If it holds that $(u_{i,a}, u_{j,b}) \in R_1$ and at the same time $(u_{j,b}, u_{i,a}) \in R_1$ then $u_{i,a}$ and $u_{j,b}$ are said to be indifferent with respect to R_1. One can split R_1 into its strict part and its indifferent part, which are denoted by P_{R_1} for the strict part, respectively by I_{R_1} for its indifferent part. If neither $(u_{i,a}, u_{j,b}) \in R_1$ nor $(u_{j,b}, u_{i,a}) \in R_1$ then $u_{i,a}$ and $u_{j,b}$ are said to be incomparable with respect to R_1.

If one wants to provide further information about the strength of certain preferences (elements in R_1), one can make use of the relation R_2. It holds that $((u_{i,a}, u_{j,b}), (u_{k,c}, u_{l,d})) \in R_2$ if and only if exchanging $u_{j,b}$ through $u_{i,a}$ is preferable to exchanging $u_{l,d}$ through $u_{k,c}$. For this relation there can be a seperation of its strict part and indifferent part too, analogous to the definitions for R_1. The resulting triple $\mathcal{P} = [\mathbb{A} \times \Theta, R_1, R_2]$ is called a *preference system* on $\mathbb{A} \times \Theta$.

1.2.2 (Delta-)Consistency

Let $\mathcal{P} = [\mathbb{A} \times \Theta, R_1, R_2]$ be a preference system on $\mathbb{A} \times \Theta$ and let $a, b, c, d \in \mathbb{A} \times \Theta$. Note that compared to sec. 1.2 there is a slight change in notation, since u is used for the actual (unknown) utility function in this section. Then $\mathcal{P}$ is said to be *consistent* if there exists a function $u : \mathbb{A} \times \Theta \to [0, 1]$ such that for all a, b, c, d the following two properties hold[9, p. 7]:

- If $(a, b) \in R_1$, then $u(a) \geq u(b)$ (with $=$ instead of $\geq$ if $(a, b) \in I_{R_1}$)
- If $((a, b), (c, d)) \in R_2$, then $u(a) - u(b) \geq u(c) - u(d)$ (with $=$ instead of $\geq$ if $((a, b), (c, d)) \in I_{R_2}$)

Every such function u is then said to *represent* the preference system $\mathcal{P}$. The set of all representations u of $\mathcal{P}$ is denoted by $U_{\mathcal{P}}$. To set the scale of the used utility function to $[0, 1]$, one additionally requires that $\inf_{a \in \mathbb{A} \times \Theta} u(a) = 0$ and $\sup_{a \in \mathbb{A} \times \Theta} u(a) = 1$. The set of all u also satisfying this is denoted by $\mathcal{N}_{\mathcal{P}}$.

In order to check whether a preference system $\mathcal{P}$ is consistent or not one can use the linear optimization problem in [9, p. 8]. If and only if the optimal outcome is strictly greater than 0, then $\mathcal{P}$ is consistent. In addition to that for each value $0 \leq \delta \leq \varepsilon$ (with ε denoting the optimal outcome) $\mathcal{P}$ is also called δ-consistent.

2 Imprecise probabilities in the context of decision theory

The most common understanding of probabilities is the classical probability, where for each event A there will be single probability $P(A)$ assigned. However for many applications having a distribution which assigns a precise but also reliable probability to each possible event is kind of unrealistic. Another problem is that classical probability doesn't allow to distinguish between known symmetry of events and unknown asymmetry, means that whether one knows about each state of nature to be equally likely or one just doesn't have any evidence for one state to be more likely than another will both lead to the usage of an uniform distribution over the set of possible states. A motivation from decision theory itself is the Ellsberg-Paradoxon, which points out that with restriction to precise probabilistic information it is quite easy to construct a situation where the decision-maker is very likely to decide in an intuitive but irrational way[6, sec. 2.1].

For these reasons a more flexible theory of uncertainty has been developed - the so-called *imprecise probabilities*. A mathematically very precise explanation from the basic underlying concepts of *Desirability* and *Lower previsions* until the application in decision theory can be found in [4]. For a more philosophical point of view one can take a look at [6]. For this thesis it is totally sufficient to give a more intuitive but less general understanding of Imprecise probabilities. The main focus of the application lies in the usage of *interval probabilities* as proposed from Weichselberger[20].

2.1 Different types of probabilites

Interval probabilities are defined by assigning a lower probability $L(.) : \mathcal{A} \to [0;1], A \to L(A)$ and upper probabilitiy $U(.) : \mathcal{A} \to [0;1], A \to U(A)$ to each event $A \in \mathcal{A}$, where $\mathcal{A}$ is a sigma algebra over a sampling space Ω with $|\Omega| \geq 2$. This results in an interval of possible probabilities $[L(A); U(A)] \subseteq [0;1]$. In that sense a classic probability is a special case for which it holds that $L(A) = U(A)$. On the other hand if one has no information about any of the events' probabilities, to all of them the probability interval $[0;1]$ will be assigned. Weichselberger axiomatizizes this concept by some requirements on the set $\mathcal{M}$[2, p. 19]:

$$\mathcal{M} = \{p(.) \mid L(A) \leq p(A) \leq U(A), \forall A \in \mathcal{A}\}$$

This set of probabilities can have different grades of quality:

1. If $\mathcal{M}$ is empty $L(A)$ and $U(A)$ contradict each other, which means that this may not be interpreted and used as a generalized probability.

2. If $\mathcal{M}$ is not empty, one speaks of interval probability in the narrow sense - $\mathcal{M}$ is then called the *structure* of the interval probability. A more general term for $\mathcal{M}$, which is not only used for interval probabilities but all types of probability sets is *credal set*[15]. One can distinguish between two types of relation between the boundaries of the interval and its structure:

   (a) if $L(.)$and $U(.)$ are allowed to be "too wide" - meaning that for all events the actual structure lies definitely in the resulting interval but it could be too conservative for some of the events. This case is also called *R-probability*.

   (b) if $L(.)$and $U(.)$ are exactly the actual boundaries of the structure, this is the ideal case of a generalized probability, which is also called *F-probability*.

K-probability

Let $(\Omega, \mathcal{A})$ be a measure space. A function $p(.)$ over $\mathcal{A}$ is called *K-probability* if it holds that:

- $p(.) : \mathcal{A} \to [0;1], A \to p(A)$

- $p(\Omega) = 1$
- For all sequences $(A_i)_{i\in\mathbb{N}}$ of pairwise disjoint $A_i \in \mathcal{A}$ it holds that: $p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i)$ (σ- additivity)

...or namely the axioms of Kolmogorow. These axioms may be one of the most important contributions to statistics and probability theory - especially because they offer a definition for a classical probability without any attempt in giving an idiological interpretation of its meaning.

Note that K-probabilities will often be called "classical" probabilities and that Weichselberger also refers to them as *K-function.*

R-probability

Let $(\Omega, \mathcal{A})$ be a measure space and let $\mathcal{Z}_0 := \{[L;U] \mid 0 \leq L \leq U \leq 1\}$ be the set of all closed intervals within $[0;1]$. A function $P(.)$ over $\mathcal{A}$ is called *R-probability* if it holds that:

- $P(.) : \mathcal{A} \to \mathcal{Z}_0, A \to P(A) = [L(A); U(A)]$
- The set $\mathcal{M} := \{p(.) \mid \forall A \in \mathcal{A} : L(A) \leq p(A) \leq U(A)\}$ is not empty.

in that case $\mathcal{R} = (\Omega; \mathcal{A}; L(.), U(.))$ is called *R-probability-field* and the set $\mathcal{M}$ the structure of $\mathcal{R}$.

Weichselberger additionally distinguishes between partially determined and totally determined R-probabilities[20, p. 13]. The difference is that for totally determined R-probabilities the lower and upper boundary for every $A \in \mathcal{A}$ are specified, while for partially determined R-probabilities there is a subset $\mathcal{A}_L \subseteq \mathcal{A}$ for which $L(A)$ is defined for every $A \in \mathcal{A}_L$ and another subset $\mathcal{A}_U \subseteq \mathcal{A}$ for which $U(A)$ is defined for every $A \in \mathcal{A}_U$.

F-probability

A totally determined R-probability $P(.)$ over $\mathcal{A}$ with structure $\mathcal{M}$ is called totally determined *F-probability*, if the following conditions are fulfilled:

- $\inf_{p(.)\in\mathcal{M}} p(A) = L(A)\ \forall A \in \mathcal{A}$
- $\sup_{p(.)\in\mathcal{M}} p(A) = U(A)\, \forall A \in \mathcal{A}$

The triple $\mathcal{F} = (\Omega; \mathcal{A}; L(.))$ is called *F-probabiliy-field.* Note that it is not necessary to include $U(.)$ in $\mathcal{F}$ since it must hold that:

$$U(A) = 1 - L(\neg A)\, \forall A \in \mathcal{A}$$

Also for F-probabilities Weichselberger made a more detailed distinction between the partially determined and totally determined type. Instead of fulfilling the just mentioned conditions a partially determined R-probability over $\mathcal{A}$ is called partially determined F-probability if it holds that[20, p. 354]:

$$1 - L(\neg A) = U(A)\, \forall A \in \mathcal{A}_U$$

A totally determined R-probability, which is not a totally determined F-probability is also called redundant R-probability[20, p. 147]. In that case one can generate the corresponding F-probability-field by applying the linear programming problem of [2, p. 42]. To deal with R-probabilities this way is seen as a *rigorous point of view* by Weichselberger[20, p. 329]. The counterpart is the *conservative point of view* as proposed in [20, sec 2.8]. If one takes this perspective, the R-probability-intervals will be extended instead of shortened to generate a F-probability.

C-probability

Let $(\Omega, \mathcal{A})$ be a measure space and $P(.)$ be a F-probability: $P : \mathcal{A} \to \mathcal{Z}_0, A \to [L(A); U(A)]$.

- If $L(.)$ is *two-monotonous,* which means:
  $L(A \cup B) + L(A \cap B) \geq L(A) + L(B) \, \forall A, B \in \mathcal{A}$,
  then $P(.)$ is also called *C-probability.*
- If $L(.)$ is *total-monotonous, which means*
  $L\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} L\left(\bigcap_{i \in I} A_i\right)$,
  then $P(.)$is called *total-monotonous C-probability.*

The corresponding F-probability-field $\mathcal{C} = (\Omega; \mathcal{A}; L(.))$ is then called *C-probabiliy-field*, respectively total-monotonous C-probability-field.

In terms of decison theory, C-probability is always present when there are only lower and upper boundaries for the probability of each state of nature defined (assuming the resulting set $\mathcal{M}$ is not empty). As soon as one adds ordinal relationships between the states probabilities like $p(A) \geq p(B)$ or requirements concerning the union of two or more states like $p(A) \cup p(B) \leq c \mid c \in \mathbb{R}$ one can easily construct situations where the necessary conditions for C-probabilities are violated (see for example [2, p. 55,56]).

Relationships between types of probabilities

The relationships between the different types of probabilities is given by the following implication chain:

K-probability $\Rightarrow$ total-monotonous C-probability $\Rightarrow$ C-probability $\Rightarrow$F-probability $\Rightarrow$R-probability

However note that this implication does not hold for the opposite direction. The only exception is the possibility to generate a F-probability-field from a R-probability-field like previously mentioned.

2.2 Representation of the probability structure as a set of extreme points

This section will build the bridge between the underlying probability type and computational aspects of a decision making problem. First let $\Omega = \{\omega_{1,\ldots,}\omega_q\}$ be a finite sampling space with $2 \leq \mid \Omega \mid = q < \infty$. Let $\mathcal{A}$ be a sigma algebra over Ω. Since the structure $\mathcal{M}$ of any F-proability $P(.)$ over $\mathcal{A}$ represents a finite set of restrictions which can be also described as a finite set of linear constraints, $\mathcal{M}$ is also a convex polyhedron in $\mathbb{R}^q$. In addition to that every K-probability $\pi \in \mathcal{M}$ is a point in $\mathbb{R}^q$. Some of these points, namely the set of extreme points $\mathcal{E}(\mathcal{M})$ can represent such a convex polyhedron without any loss of information. This set can be computed by solving a specific linear programming problem according to Kofler[13] (citation adopted from [3, ch. 3, p. 349]).

The implementation has been done by utilizing the R-package *rcdd*[7] by transforming the probability boundaries and ordinal relationships of the states of nature to linear programming constraints. However the explanation and evaluation of the specific algorithm for achieving this is not part of this thesis. For the further part of this work one only has to know that the running time of that algorithm is mainly dependent on the actual number of those extreme points[20, p. 470].

3 The Dominance Principle

Probably one of the most intuitive principles in decision theory is the *Dominance Principle*. In this chapter there will be a short formal defintion of the principle and some information on its potential power as a pre-selection tool for different decision criteria. Afterwards there will be an explanation of how the principle was implemented in this work.

3.1 Dominance Principle - Definition

One compares any act $a_i \in \mathbb{A}$ to any other act $a_{j|j\neq i} \in \mathbb{A}$ in respect to every state of the initial utility table. If there is a a_j for which holds that $u(a_j, \theta_k) \geq u(a_i, \theta_k) \, \forall k$ and there is at least one $\widetilde{k}$ for which holds that $u\left(a_j, \theta_{\widetilde{k}}\right) > u\left(a_i, \theta_{\widetilde{k}}\right)$ then a_i is called strictly dominated by a_j (denoted by $a_j \succ a_i$). In other words this means that there is no state in the set of possible states Θ for which a_i leads to a better outcome than a_j, but there is at least one state for which the opposite holds. Of course this means that when comparing act a_i to a_j that it is rational to prefer a_j. If one makes a comparison between every pair of acts in $\mathbb{A}$ and drops every strictly dominated act, one will receive a new set of actions $\widetilde{\mathbb{A}}$ where no act is strictly dominated by another. This set is also called the set of *admissible* acts.

For most of the criteria in this thesis the non-admissible acts cannot be optimal, and even if they are: according to dominance principle there would still be at least one better act to choose. However one cannot simply use $\widetilde{\mathbb{A}}$ instead of $\mathbb{A}$ in any case since the exclusion of certain acts may change the criteria values of the other acts e.g. for Joint Statistical Preference (see Chapter 4.2.3). In situations like this one has to individually decide wheter it makes more sense to include or exclude the non-admissible acts. In other situations excluding every strictly dominated act will also declare in fact optimal acts as non-optimal (e.g. for the second method to compute $\mathcal{M}$- maximal acts). Therefore another version, the *strong dominance* is required. An act a_i is said to be strongly dominated by another act a_j (denoted by $a_j \succ\succ a_i$) if it holds that $u(a_j, \theta_k) > u(a_i, \theta_k) \, \forall k$.

3.2 Dominance Principle - Implementation

The most-straightforward approach is to simply compare every act in $\mathbb{A}$ to any other by successively comparing the utility value of those two acts for every state in Θ. Whenever an act is proven to be dominated, it can be deleted from the utility table (note that this is possible because both relations $\succ$ and $\succ\succ$ are transitive). However this way of comparing two acts will have the worst possible running time when either one of the acts strictly dominates the other or when both acts are equal on all possible states. Another more global problem of this method is the influence of the order of acts. Assume there is in fact an act which strictly dominates all (or at least a very high number of) acts but it is the last one of the utility table. In addition to that assume that no other act dominates any other act. Then one could save a lot of running time if the algorithm would start comparing acts from the lowest row in the utility table instead of starting at the beginning, since the other acts would get removed from the utility table before they could be compared to any other act. However this choice is not trivial, since an utility table does ussually not come in a specific order. Of course this effect will hardly ever be that strong in practice, still there is a high potential to save running time.

Algorithm 1 deals with both of the previously mentioned problems by pre-ordering the acts of $\mathbb{A}$ in an advantageous way for detecting strict dominance. One simply calculates sums for every act (assume acts to be represented by rows) of the utility table. It is trivial that no act can be strictly dominated by another act with a lower row sum in case of utility (respectively by an act with a higher row sum for loss). This means that if the row sum of a_i is greater than that one of a_j one only needs to check wheter the utility of a_j exceeds that of a_i in any state or not. However this will lead to the exclusion of a_j if a_j and a_i are equal in every possible

state. If one wants to avoid that, there must also be a check if there is a state $\theta_{k|k\in\{1,\ldots,m\}}$ where $u(a_i, \theta_k) > u(a_j, \theta_k)$. Even then the comparison of two acts will only have worst case running time if a_i strictly dominates a_j or a_i is equal to a_j.

Algorithm 1 Strict Dominance

```
UT <- ... *table with utility values for each pair
of act and state (acts as rows and states as columns)*
ns <- ... *number of possible states of nature*
na <- ... *number of possible acts*
AS <- ... *a data frame/table with the names/numbers of the acts
and the sum over their utility values for all states of nature*
AS <- ... *arrange AS in decreasing order of the acts' sums*

i <- 1
while (i < length(AS){
        j <- i + 1
        while (j <= length(AS){
        reminder_equal <- TRUE
        *only required if all duplicates shall be conatained*
        reminder <- FALSE
                for (k in 1:ns){
                        if (UT[AS[j],k]>UT[AS[i],k]){
                                reminder <- TRUE
                                break *remainder of for-loop is redundant*
                        }
                        else if (UT[AS[j],k]<UT[AS[i],k]){
                        *only required if all duplicates shall be conatained*
                                reminder_equal <- FALSE
                        }
                }
                if (reminder == FALSE){
                        if(reminder_equal == FALSE){
                        *only required if all duplicates shall be conatained*
                                AS <- AS[-j]
                                j <- j - 1
                        }
                }
                j <- j + 1
        }
        i <- i + 1
}

*The remaining acts in AS will be the optimal acts
according to strict dominance*
```

The second mentioned problem is solved by this approach too, if one believes in the following hypothesis: Let's assume the row sum of act a_i is represented by rs_i $\forall i \in \{1, 2, 3, \ldots, n\}$. If it holds that $rs_1 > rs_2 > rs_3$ then a_3 is more likely to be strictly dominated by a_1 than by a_2 .

In practical applications it could occur that an act a_i performs better than almost every other act in every state except one where it is worse than the majority. Thus there might be

Algorithm 2 Strong Dominance

```
UT <- ... *table with utility values for each pair
of act and state (acts as rows and states as columns)*
ns <- ... *number of possible states of nature*
na <- ... *number of possible acts*
AS <- ... *a data frame/table with the names/numbers of the acts
and the sum over their utility values for all states of nature*
AS <- ... *arrange AS in decreasing order of the acts' sums*

i <- 1
while (i < length(AS){
        j <- i + 1
        while (j <= length(AS){
                for (k in 1:ns){
                        reminder <- FALSE
                        if (UT[AS[j],k]>=UT[AS[i],k]){
                                reminder <- TRUE
                                break
                                *remainder of for-loop is redundant*
                        }
                }
                if (reminder == FALSE){
                        AS <- AS[-j]
                        j <- j - 1
                }
                j <- j + 1
        }
        i <- i + 1
}

*The remaining acts in AS will be the optimal acts
according to strong dominance*
```

not a single act strictly dominated by a_i even though it has the highest row sum of all acts. For many acts similar to a_i and many acts with low row sums who still dominate some other acts, the pre-ordering could even extend the running time. Though it seems like one would have to create an utility table on purpose to fit these conditions. In contrast if we consider the example from above that there is in fact an act which strictly dominates all other acts but is placed at the end of the table, it will automatically be on top of the utility table after pre-ordering and the pairs of acts to be compared decreases from $\frac{n\cdot(n-1)}{2}$ to $n-1$.

If the hypothesis from above does not hold, one could consider some further pre-ordering of the utility table. Assume there is a rather high number of acts with high row sums which for whatever reason perform relatively bad in just one state of the utility table. If one re-orders the columns (states) in such a manner, that this "worst-case-state" will be considered first when comparing such an act to another act with lower row sum, then it might occur that the number of necessary comparisons for each pair of acts will decrease from a maximum of $2\cdot(m-1)+1$ to just 1 (m being the number of states/columns) respectively from m to 1 when equal acts may be excluded. However ordering before every iteration of comparisons is also kind of expensive concerning running time. Therefore one could consider ordering the whole utility table only once by calculating some sort of score for each column.

In this work an approach has been tried out where one uses normalization on every column of the table. First one subtracts the minimum of the respective column and then divides through the maximum. This will lead to the score value for an act to be 1 if it's utility equals the columns maximum. The score value will be 0 if the utility value of the act is the same as the columns minimum. Then all acts receive a weight according to how many other acts have a smaller row sum, so the weights will range from $n - 1$ for the act with the highest row sum to 0 for the act with the lowest row sum. Then the main score will be computed by multiplying the weights with the score vectors and summing them. The order of the columns scores will define in which order the utility comparisons of the acts will be made (lower scores first for utility, higher scores first for loss). Of course the improvements strongly depends on certain requirements, like many acts with high row sums having low utility or high loss for the same states. In any other case this pre-ordering usually doesn't lead to an improvement for randomly generated utility tables.

4 Different Criteria of Decision Theory

Only excluding the non-admissible acts will leave way too many optimal acts most of the time. To decrease the size of set of optimal acts one can make use of the *optimality or decision criteria*. A decision criteria is defined as a mapping $\Phi : \mathbb{A} \to \mathbb{R}$, $a \to \Phi(a)$, for which one is looking for acts a^* for which it holds that $\Phi(a^*) \geq \Phi(a)\ \forall a \in \mathbb{A}$. The set of all a^* which fulfill this condition, is called the set of optimal acts with respect to Φ. Some of those criteria are based on the idea of precise probabilistic information (K-probabilities), e.g. Bayes-criterion and Hodges-and-Lehmann criterion. Some other enable the usage of imprecise probabilistic information, e.g. Gamma-Maxi-Min–criterion and E-admissibility. A third group (e.g. Maxi-Min, Maxi-Max and Hurwicz-criterion) doesn't require probabilistic information at all.

All of the linear programming problems have been implemented using the *lpSolve*[5] package. All of these implementations can be found on the GitHub repository mentioned in the Attachments. It is important to point out, that some of the following criteria are special cases of others, which makes their seperate implementation kind of redundant.

Also note that all of the implementation is based on the idea of making it as easy as possible for the decision maker to input her probabilistic information. Instead of specifying a F- or C-probability on her own, the probability will be generated by combining a complete set of lower und upper boundaries, an optional ordinal relationship between the states' probabilies, as well as optional further linear constraints for the states' probabilities in accord to the rigorous point of view. If both optional parts are not defined, the set of boundaries represents a totally determined R-probability - totally determined because the set of lower and upper boundaries must be complete - which results in a C-probability. However generating the corresponding F- respectively C-probability will not be done explicitly, but implicitly by the linear programming methods which are used to compute the criteria.

The following chapter was split into two subsections. Section 4.1 contains all of the criteria, for which the assumption of cardinal utility in chapter 1 is fulfilled. Section 4.2 in contrast contains those criteria which focuses on dealing with utility tables for which the cardinality is questionable or the decision maker simply just defined an ordinal utility instead. The remaining criteria based on preference systems will be presented in section 4.3.

4.1 Criteria based on cardinal utility

The following criteria will be based on the assumption that the underlying utility table does represent a cardinal order of the consequences of the acts for the different states of nature.

4.1.1 The Maxi-Min(Mini-Max)-criterion

This criterion introduced by Abraham Wald[19] is especially useful for situations where no information about the probability of the states is available and the decision maker has a very pessimistic point of view. Every act will be measured by its worst-case utility, meaning that the act with the highest minimal utility (respectively the lowest maximal loss) will be considered as optimal. If one only considers pure acts, the Maxi-Min-optimal act(s) is/are simply the act(s) which has/have the highest minimal utility in the utility table. For this criterion it makes sense to consider randomized-acts, since there will be often an improvement in the criteria-value compared to pure acts.

An algorithm to compute the randomized Maxi-Min-optimal act is mentioned in [3, ch. 4, p. 447]. It is important to point out that the solution of this linear programm can in fact be one of many optimal randomized acts and that any of the other optimal randomized acts could be better in any of the other decision criteria (see Table 2). Here the optimal solution of the algorithm will induce the randomized act $a^*=(0.25, 0, 25, 0.5)$ to be the Maxi-Min-optimal act with corresponding utility $u_{a^*}(\theta_k) = 3\ \forall k \in \{1, 2, 3\}$. In contrast if one chooses $a^{**} =$

$(2/3, 1/3, 0)$ instead the resulting utility will be $u_{a^{**}}(\theta_k) = 3\ \forall k \in \{2,3\}$ and $u_{a^{**}}(\theta_k) = 4$ if $k = 1$. Of course this means that a^{**} strictly dominates a^*, so it is reasonable to choose the first over the latter. To avoid choosing a non-admissible Maxi-Min-optimal act one could e.g. use the Hurwicz-criterion with a very low optimism parameter instead (see 4.1.7). Since this criterion is a special-case of three other criteria in this thesis, there has no separate implementation been done.

	θ_1	θ_2	θ_3
a_1	5	3	2
a_2	2	3	5
a_3	2.5	3	2.5

Tab. 2: Table to illustrate sub-optimality of randomized Maxi-Min acts for other criteria

4.1.2 The Maxi-Max(Mini-Min)-criterion

This criterion is the counterpart to the Maxi-Min-criterion. Instead of having a look at the worst-cases, the decision maker focuses on the best-case of each individual act. In contrast to the Maxi-Min-criterion there will always be an optimal pure act for the Maxi-Max-criterion. A randomized act can only be an optimal Maxi-Max act if it only gives strictly positive probabilities to acts which have the maximal utility (respectively minimal loss) value in the same state of nature. One would then be interested in all possible combinations of pure acts which lead to a Maxi-Min-optimal randomized act.

An example is shown in table 2. All of the randomized acts assigning strictly positive probabilities only to a_1, a_2 and a_3 would be Maxi-Min-optimal as well as all randomized acts assigning strictly positive probabilities only to a_3 and a_4. Assigning a strictly postive probability to a_6 would always lead to a non-optimal randomized act. The Hurwicz-criterion is a more general form of the Maxi-Max-criterion, therefore there is no separate implementation necessary.

	θ_1	θ_2	θ_3
a_1	10	0	2
a_2	10	2	0
a_3	10	1	1
a_4	0	10	2
a_5	2	10	0
a_6	5	5	5

Tab. 3: Utility table for explantion of randomized Maxi-Max acts

4.1.3 The Bayes-criterion

In contrast to the Maxi-Min and Maxi-Max-criterion the decision-maker uses information about the probability of occurence for each state of nature. This priori-distribution of states in form of a K-probability is used to calculate expectations of utility for each act. The act with the highest utility-expectation (respectively lowest for loss) is considered as Bayes-optimal. There is always at least one pure act, which is Bayes-optimal, so there is no need to consider randomized acts. This criterion is also a special-case of two other criteria, so there is no need for its implementation.

4.1.4 The Hodges and Lehmann-criterion

This criterion proposed by Joseph Hodges and Erich Lehmann[8] combines two different aspects of decision making into one. Even though one uses precise probabilistic information for this

criterion, this information is assumed to be uncertain[10, p. 3], which is represented by the *trust-parameter* and tells how strongly the decision maker "believes" in the priori-distribution of the states. For a parameter-value of 0 the criterion is equal to Mini-Max and for a parameter value of 1 it's equal to the Bayes-criterion.

The algorithm in [10, p. 4] was used for implementation. On a computational point of view this algorithm could very easily be extended to also work for imprecise probabilistic information, though there must be a good reasoning to legitimate mixing both aspects, which was neither suggested by any of the sources of this work, nor will be part of the work itself.

4.1.5 The Gamma-Maxi-Min (or Max-E-Min)-criterion

This is basically a more general version of the Maxi-Min-criterion. For this criterion one can specify imprecise probabilistic information on the states of nature, which means that the "worst-case-states" of each act can only be the true state of nature with a certain probability (which could be < 1 for specific cases), which leads to the conclusion that the criteria-value for each act will be at least as good as its value for the Maxi-Min-criterion. For the same reason as for Maxi-Min-critierion it makes sense to consider randomized acts.

If one has no non-trivial restrictions on the probabilities, which results in a C-probability where every state has a lower probability of 0 and an upper probability of 1, the criterion is equal to the Maxi-Min-criterion. If the probability restrictions only allow one probability for each state (which means there is precise and sure probabilistic information), the Gamma-Maxi-Min-criterion is equal to the Bayes-criterion.

If the probability-inducing variables are defined in a non-trivial way, but randomized acts are not considered, one can make a single optimization problem for each act to compute its highest lower boundary of utility (respectively its lowest higher boundary of loss) and compare the objective values of each pure act. For the case that there are non-trivial probability requirements as well as randomized acts to be considered one can use the algorithm suggested by Jansen[10, p. 8]. In that case one could run into the same problem as illustrated in Table 2 for Maxi-Min-optimal acts.

4.1.6 The Gamma-Maxi-Max (or Max-E-Max)-criterion

Like the Gamma-Maxi-Min for the Maxi-Min-criterion, the Gamma-Maxi-Max criterion is a more generalized version of the Maxi-Max criterion. If the probability for the union of the "best-case-states" of an act is bounded to be strictly smaller than 1, the Gamma-Maxi-Max score of the act will be lower than the Maxi-Max-score of that act. If, again, one has no non-trivial restrictions on the probabilities, the criterion is equal to the Maxi-Max-criterion. If the probability restrictions imply a single valid probability, the Gamma-Maxi-Max-criterion is equal to the Bayes-criterion.

In the case of non-trivial probabilistic information and without considering randomized acts one can compute the optimal acts analogous as for the same case with Gamma-Maxi-Min by maximizing expectation for utility and minimizing in case of loss. For the case that there are non-trivial probability requirements as well as randomized acts to be considered one needs to calculate the set of extreme points and afterwards the utility for every act at every extreme point. On the basis of that resulting utility-expectation-table one can proceed in the same way as for randomized Maxi-Max-acts.

4.1.7 The Hurwicz-criterion

This criterion by Leonid Hurwicz and Kenneth J. Arrow[1] is an approach to find a middle way between a completely pessimistic ((Gamma-)Maxi-Min) and a completely optimistic ((Gamma-)Maxi-Max) point of view. The balance between both apsects is controlled by the so-called *optimism-parameter* α. This parameter reaches from 0 to 1, whereas for 0 the criterion is equal to the Maxi-Min criterion and for 1 the criterion is equal to the Maxi-Max criterion in case of the

absence of any probabilistic information. If there is non-trivial imprecise probabilistic information the criterion equals the Gamma-Maxi-Min criterion in case of $\alpha = 0$ and the Gamma-Maxi-Max in case of $\alpha = 1$.

If α is strictly greater than 0 and strictly smaller than 1, the score values of the pure acts are weighted sums of their (Gamma-)Maxi-Min and (Gamma-)Maxi-Max-scores, whereas the (Gamma-)Maxi-Min part has a weight of $1 - \alpha$ and the (Gamma-)Maxi-Max part has a weight of α. If in addition to that randomized acts are considered, one has to use a way more complex approach[18, p. 6].

4.1.8 E-admissibility

One of the most well-known criteria in decision theory with imprecise probabilities is the E-admissibility, proposed by Levi[15]. An act a_i is called E-admissible if and only if there is at least one probability π in $\mathcal{M}$ for which holds that $\mathbb{E}_\pi(u(a_i)) \geq \mathbb{E}_\pi(u(a_j)) \forall j \in \{1, \ldots, n\}$.

The implementation has been done very similar to the algorithm in [18, p. 23], just without the auxiliary variable. In many cases running time can be saved if the acts will be checked for admissibility before, since the number of linear programming problems to be solved could be reduced drastically. One can also consider checking for $\mathcal{M}$-maximality before and then only consider these acts as possible E-admissible acts. Note that this can only save running time with the algorithm based on extreme points in 4.1.9

4.1.9 $\mathcal{M}$-maximality

This criterion is basically a weaker version of E-admissibility, meaning every act which is E-admissible is also $\mathcal{M}$-maximal but not vice versa. Formally an act a_i is called $\mathcal{M}$-maximal if for every other act $a_{j|j\in\{1,\ldots,n\}\setminus\{i\}}$ there is a $\pi_{a_j} \in \mathcal{M}$ for which holds that $\mathbb{E}_{\pi_{a_j}}(u(a_i)) \geq \mathbb{E}_{\pi_{a_j}}(u(a_j))$.

This can either be implemented by using the algorithm from [10, p. 6] or by computing the set of extreme points $\varepsilon(\mathcal{M})$, then calculating each acts expected utility for each $\pi \in \varepsilon(\mathcal{M})$. The result will be a new utility-table-like data set where one can use Algorithm 2 to exclude the strongly dominated acts in order to define the set of $\mathcal{M}$-maximal acts. Concerning running-time both methods behave very differently (see Table 4).

While the first method is quite time-expensive concerning the number of possible acts, which could in fact be improved by first checking for admissiblity and E-admissibility of acts, the second method is highly correlated with the number of extreme points (see 2.2). However the estimation of the latter number is very imprecise, while the estimation of the running time of the first method is not straight-forward. Several attempts have failed to provide a suitable solution for a precise estimation of both running times, which would enable an automatic selection of the best method.

# of acts	# of admissible acts	# of states	# of extreme points	linear programming algorithm*	extreme point algorithm*
22	22	31	3813	5,918	2,025
28	28	39	5712	20,427	4,715
34	34	18	474	49,066	0,168
8	8	27	3619	1,417	1,445
33	33	31	14068	10,559	6,512
37	33	9	16	17,892	0,010
31	31	16	3116	30,191	0,770
11	11	31	14182	9,036	6,973
8	8	22	4768	0,574	1,407
7	7	37	382042	3,748	444,254

*Running time was calculated with the *microbenchmark* package and is displayed in seconds

Tab. 4: Comparison of running time of $\mathcal{M}$-maximality algorithms

4.1.10 E_ε-admissibility

This is a mixture between E-admissibility and $\mathcal{M}$-maximality. Formally an act a_i is called E_ε-admissible (with $\varepsilon \geq 0$) if the two following conditions hold:

- $\mathbb{E}_{\pi_a}(u(a_i)) \geq \mathbb{E}_{\pi_a}(u(a)) \forall a$ and
- $\|\pi_a - \pi_b\| \leq \varepsilon \ \forall a, b \in \mathbb{A}$ where $\|.\|$ is a norm on $\mathcal{M}$

If ε equals 0, this criterion is equal to the E-admissiblity. On the other hand if ε is set to 1, there is equality to $\mathcal{M}$-maximality.

This criterion is especially useful to find acts which are not very far away from being E-admissible. If one of the E_ε-admissible (consider $\varepsilon > 0$) acts performs way better in other criteria than all of the E-admissible acts, one can argue that this act should be prefered even though it is not E-admissible.

For the implementation of this criterion based on the Manhattan norm, an algorithm similar to that one in [12, p. 8] was used. To avoid having a constraint with absolute differences of the states probabilities here addional optimization parameters, representing upper and lower limits of each states probabilities, have been used.

4.1.11 E-admissibility-extent

Assume the decision-maker wants to choose one of the E-admissible acts as their act of choice. However all of those acts are incomparable without considering another criteria which gives information about the *extent* of the E-admissibility[12, p. 10]. In this work two measures have been implemented in order to operationalize this.

Maximal extent

The first measure, the maximal extent, tries to describe the set of probability measures for which an act a maximizes the expected utility, by having a look at the two "most different" probabilities in the set according to a certain norm $\|.\|$. Formally for an E-admissible act a with a set of proabilities $\mathcal{M}_a$for which a has the maximal utility the maximal extent is defined as followed:

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

To make this solveable as a linear programming problem one uses the norm $\|.\|_\infty$. One can then use the algorithm in [12, p. 10, Prop. 3]. This will lead to the criteria value being defined by the widest range of any state in $\mathcal{M}_a$. Note that Jansen also offers a solution for using the Manhattan norm instead, however this is not possible with linear programming methods.

Uniform extent

Since the maximal extent only shows one side of the coin, there must be another perspective to measure the extent of E-admissibility. For this sake one can have a look at the largest barycentric ε-cube that can be inscribed into $\mathcal{M}_a$[12, p. 11,12]. Compared to the maximal extent this measure only reaches its maximum of 0.5, when a state a is E-admissible for every possible $\pi \in \mathcal{M}$. On the other hand if there is a state with only one possible value in $\mathcal{M}_a$ the measure will have a value of 0. The linear programming problem in [12, p. 11,12] is an option to compute the uniform extent for an E-admissible act.

4.2 Criteria based on ordinal utility

If there exists an utility table, but the cardinality of its underlying utility function is questionable, one can instead assume an ordinal utility and use the following criteria.

4.2.1 Joint-Stochastic-Dominance (Imprecise Version)

Any act a_0 for which there is a strictly increasing function $t : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}(t \circ u(a_0)) \geq \mathbb{E}(t \circ u(a)) \forall a \in \mathbb{A}$ and $\forall \pi \in \mathcal{M}$. If one computes the set of extreme points $\varepsilon(\mathcal{M})$, one can use the algorithm mentioned in [12, p. 15] to determine if a certain act is optimal with respect to Joint-Stochastic-Dominance. According to Jansen, this criterion can be viewed as global.

4.2.2 Pairwise-Stochastic-Dominance (Imprecise Version)

Jansen describes this as the local version of the Joint-Stochastic-Dominance. For an act a_0 there must be at least one $t_i : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}(t \circ u(a_0)) \geq \mathbb{E}(t \circ u(a_i)) \forall a_i \in \mathbb{A}$ and $\forall \pi \in \mathcal{M}$. Of course this condition is weaker than Joint-Stochastic-Dominance, since there could be a different function t_i for each act a_i.

The algorithm used for Joint-Stochastic-Dominance can be adapted to be used for the pairwise version. One simply solves two linear programming problems for each pair of acts (one for each ordered pair of distinct acts $(a_1, a_2) \neq (a_2, a_1)$) but only includes optimization parameters $t_{11,\ldots}, t_{2m}$ instead of $t_{11,\ldots}, t_{nm}$.

4.2.3 Joint-Statistical-Preference

Label an act a_0 as optimal, if it holds that $D_\pi(a_0) \geq D_\pi(a)\ \forall a$ where

$$D_\pi(a) := \pi\left(\left\{\theta | u(a, \theta) \geq u\left(a', \theta\right) \forall a' \in \mathbb{A}\right\}\right)$$

This means that a_0 has the highest probability to be utility dominant over all other acts and is considered as a global criterion[12, p. 14].

If $\mathcal{M}$ only consists of one point it can be used as priori distribution and one simply sums over the probabilities of the states for which a_0 has the maximal utility. In the case of imprecise probabilistic information, one must find a way to replace $D_\pi(a)$ through a legit and reasonable alternative, e.g. $\min_{\pi \in \mathcal{M}} D_\pi(a)$[12, p. 15]. This idea has been extended in this work in the spirit of the Hurwicz-criterion by including an optimism-parameter which will lead to a weighted criterion value between $\min_{\pi \in \mathcal{M}} D_\pi(a)$ and $\max_{\pi \in \mathcal{M}} D_\pi(a)$.

4.2.4 Pairwise-Statistial-Preference

This is the local version of the Joint-Statistical-Preference. Every act a_0 for which there is no other act a_1 for which holds that

$$\pi(\{\theta | u(a_1, \theta) \geq u(a_0, \theta)\}) > \pi(\{\theta | u(a_0, \theta) \geq u(a_1, \theta)\})$$

In the case of having only one $\pi \in \mathcal{M}$ the computation of the optimal acts is straight-forward like for the Joint version of the criterion. If there is imprecise probabilistic information one can consider a similar approach as proposed there, by considering a mixture of the best- and worst-case probability to dominate the other act, which can also be controlled by an optimism parameter.

4.3 Criteria based on Preference Systems

If the information on the decision makers preferences is represented by a preference system instead of an utility table, consider the following decision criteria.

4.3.1 Generalized Interval Expectation and corresponding criteria

Let $\mathcal{P} = [\mathbb{A} \times \Theta, R_1, R_2]$ be a preference on system $\mathbb{A} \times \Theta$ and let $X : \Theta \to [0,1]$ be a function representing an act $a \in \mathbb{A}$. Let $\mathcal{M}$ be the probabilistic information on Θ and $\mathcal{P}$ be δ-consistent for every $\delta | 0 \leq \delta \leq \varepsilon$ for a certain $\varepsilon > 0$. Then for a corresponding δ and $\mathcal{D}_\delta = \mathcal{N}_\mathcal{P}^\delta \times \mathcal{M}$:

$$\mathbb{E}_{\mathcal{D}_\delta}(X) := \left[\underline{\mathbb{E}}_{\mathcal{D}_\delta}(X), \overline{\mathbb{E}}_{\mathcal{D}_\delta}(X)\right] := \left[\inf_{(u,\pi)\in\mathcal{D}_\delta} \mathbb{E}_\pi(u \circ X), \sup_{(u,\pi)\in\mathcal{D}_\delta} \mathbb{E}_\pi(u \circ X)\right]$$

is called the *generalized interval expectation* of X with respect to $\mathcal{P}, \mathcal{M}$ and granularity δ[9, p. 11].

The boundaries of this interval can be computed by applying the linear programs from [9, p. 12, Prop. 3].

One can define three new decision criteria based on the generalized interval expectation. An function X like above representing an act $a \in \mathbb{A}$. Then a is called:

- $\mathcal{D}_\delta$-*maximin* if and only if $\forall Y : \Theta \to [0,1]$ representing any act $a' \in \mathbb{A} \setminus \{a\}$:

$$\underline{\mathbb{E}}_{\mathcal{D}_\delta}(X) \geq \underline{\mathbb{E}}_{\mathcal{D}_\delta}(Y)$$

- $\mathcal{D}_\delta$-*maximax* if and only if $\forall Y : \Theta \to [0,1]$ representing any act $a' \in \mathbb{A} \setminus \{a\}$:

$$\overline{\mathbb{E}}_{\mathcal{D}_\delta}(X) \geq \overline{\mathbb{E}}_{\mathcal{D}_\delta}(Y)$$

- $\mathcal{D}_\delta^\alpha$-*maximix* if and only if $\forall Y : \Theta \to [0,1]$ representing any act $a' \in \mathbb{A} \setminus \{a\}$:

$$(1-\alpha) * \underline{\mathbb{E}}_{\mathcal{D}_\delta}(X) + \alpha * \overline{\mathbb{E}}_{\mathcal{D}_\delta}(X) \geq (1-\alpha) * \underline{\mathbb{E}}_{\mathcal{D}_\delta}(Y) + \alpha * \overline{\mathbb{E}}_{\mathcal{D}_\delta}(Y)$$

  where $\alpha \in [0,1]$ is some fixed parameter.

From a computational point of view, one creates an utility table with every act $a \in \mathbb{A}$ and two columns representing the corresponding lower and upper boundary from the corresponding generalized interval expectation. Then one can simply apply the same algorithm as for the Maximin-criterion without randomized acts for detecting $\mathcal{D}_\delta$-maximin, the algorithm for non-randomized Maximax-acts for detecting $\mathcal{D}_\delta$-maximax and the algorithm for the Hurwicz-criterion without randomization for detecting the $\mathcal{D}_\delta^\alpha$-maximix optimal acts (with α as optimism parameter).

4.3.2 Criteria based on global comparisons of acts

Let $\mathcal{P}, \mathcal{M}$ and X be defined as in 4.3.1. Then act a is called:

- $\mathcal{P}|\mathcal{M}$-*admissible* if and only if $\exists u \in \mathcal{U}_\mathcal{P}, \exists \pi \in \mathcal{M} \ldots$
- $\mathcal{P}$-*admissible* if and only if $\exists u \in \mathcal{U}_\mathcal{P} \forall \pi \in \mathcal{M} \ldots$
- $\mathcal{M}$-*admissible* if and only if $\exists \pi \in \mathcal{M} \forall u \in \mathcal{U}_\mathcal{P} \ldots$
- $\mathcal{P}|\mathcal{M}$-*dominant* if and only if $\forall u \in \mathcal{U}_\mathcal{P} \forall \pi \in \mathcal{M} \ldots$

$\ldots \forall Y$ representing any act $a' \in \mathbb{A} \setminus \{a\}$:

$$\mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$$

For further explanations of these criteria and their connections to each other see [9, p. 14]. For the computation of $\mathcal{P}$-admissible acts one can use the algorithm proposed on [9, p. 15-16]. This

algorithm can easily be adapted for the detection of $\mathcal{P}|\mathcal{M}$-admissible acts. One solves instead an optimization problem for each extreme point $\pi \in \mathcal{E}(\mathcal{M})$. If and only if the optimal outcome of at least one of these problems is strictly greater than 0, the corresponding act is $\mathcal{P}|\mathcal{M}$-admissible.

One can find the set of $\mathcal{M}$-admissible acts by using similar optimization problems like in [9, p. 18, Prop. 5] (consider maximization in case of utility). For each act-representation Y and each extreme point $\pi \in \mathcal{E}(\mathcal{M})$ one checks, if she can find another act representation X for which the optimal outcome of the problem is strictly greater than 0. If so, the act represented from Y is not $\mathcal{M}$-admissible and in any other case it is.

Checking for $\mathcal{P}|\mathcal{M}$-Dominance can be done in a similar way. If there is at least one combination of an extreme point $\pi \in \mathcal{E}(\mathcal{M})$ and an act representation X for which the optimal outcome is strictly greater than 0, then the act represented by Y is not $\mathcal{P}|\mathcal{M}$-dominant and in any other case it is.

5 Conclusion

The outcome of this work is a set of functions which could the starting point for a R-package for decision theory problems. Most of the implemented criteria are restricted to the usage on already defined utiliy tables. For preference systems there exists more criteria than contained in this thesis (see e.g. [9, p. 17]). According to Jansen et. al there is also some special treatment for the case of group decisions[11] required. Chapter 8 of [4] also shows some concepts of sequential decision making, a field which is not covered by this work at all. Of course a good R-package should offer the possibility to model and deal with all of these situations.

Two aspects, concerning the already implemented algorithms, are the current way of prefiltering of acts (like e.g. for E-admissiblity and E_ε-admissibility) and the choice for the algorithm used to compute $\mathcal{M}$-maximal acts. Both has to be done manually at the current state of implementation, which is suboptimal because the user needs to know when each of the options is preferable. This should definitely be automatized in further work, so the decision maker doesn't have to deal with methodical questions which are not directly related to the decision process itself.

Acknowledgements

First of all I want to thank my thesis supervisor Prof. Dr. Thomas Augustin for introducing me to the very interesting fields of *decision theory* and *imprecise probabilities* in his lecture "Entscheidungstheorie" and also for taking his time to discuss some of my thoughts concerning the *Ellsberg-Paradox,* where the idea for this Bachelors thesis came up. Furthermore I am very thankful for the possibility of doing some very exciting and enjoyable tasks and the very good support throughout the whole duration of this thesis.

Furthermore I want to thank Dr. Georg Schollmeyer for his help concerning the implementation of everything related to preference systems and ordinal utilities. Without him some essential parts couldn't have been included in this work.

I also want to thank the *R*-Core-Team for providing users all around the world the possibility to use *R*[17] for free and also the responsible teams for *rcdd*[7], *lpSolve*[5], *tidyverse*[21], *microbenchmark*[16], and *checkmate*[14] for contributing to the *R*-open-source community.

A big credit goes to Sergej Levashov for the proofreading.

Last but not least I want to thank my family for enabling me to focus completely on my studies and for always supporting me without ever expecting anything in return.

List of Algorithms

List of Tables

References

[1] KJ Arrow. Hurwicz. l.(1972), an optimality criterion for decision-making under ignorance. *Uncertainty and Expectations in Economics. Oxford*, 1954.

[2] Thomas Augustin. *Optimale Tests bei Intervallwahrscheinlichkeit*. Number 46. Vandenhoeck und Ruprecht, 1998.

[3] Thomas Augustin. Entscheidungstheorie script, Oct 29 2020. URL: https://moodle.lmu.de/course/view.php?id=7271.

[4] Thomas Augustin, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

[5] Michel Berkelaar et al. *lpSolve: Interface to Lp_solve v. 5.5 to Solve Linear/Integer Programs*, 2020. R package version 5.6.15. URL: https://CRAN.R-project.org/package=lpSolve.

[6] Seamus Bradley. Imprecise probabilities. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edition, 2019.

[7] Charles J. Geyer, Glen D. Meeden, and incorporates code from cddlib written by Komei Fukuda. *rcdd: Computational Geometry*, 2019. R package version 1.2-2. URL: https://CRAN.R-project.org/package=rcdd.

[8] Joseph L Hodges, Erich L Lehmann, et al. The use of previous experience in reaching statistical decisions. *The Annals of Mathematical Statistics*, 23(3):396–407, 1952.

[9] Christoph Jansen, Georg Schollmeyer, and Thomas Augustin. Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences. In *Proceedings of the tenth international symposium on imprecise probability: theories and applications*, pages 181–192, 2017.

[10] Christoph Jansen, Georg Schollmeyer, and Thomas Augustin. Decision theory meets linear optimization beyond computation. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 329–339. Springer, 2017.

[11] Christoph Jansen, Georg Schollmeyer, and Thomas Augustin. A probabilistic evaluation framework for preference aggregation reflecting group homogeneity. *Mathematical Social Sciences*, 96:49–62, 2018.

[12] Christoph Jansen, Georg Schollmeyer, and Thomas Augustin. Quantifying degrees of e-admissibility in decicion making with imprecise probabilities. 2018.

[13] Eduard Kofler. *Prognosen und Stabilität bei unvollständiger Information*. Campus-Verlag, 1989.

[14] Michel Lang. checkmate: Fast argument checks for defensive r programming. *The R Journal*, 9(1):437–445, 2017. URL: https://journal.r-project.org/archive/2017/RJ-2017-028/index.html.

[15] Isaac Levi. *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. MIT press, 1983.

[16] Olaf Mersmann. *microbenchmark: Accurate Timing Functions*, 2019. R package version 1.4-7. URL: https://CRAN.R-project.org/package=microbenchmark.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL: https://www.R-project.org/.

[18] Lev V Utkin and Thomas Augustin. Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In *ISIPTA*, volume 5, pages 349–358, 2005.

[19] Abraham Wald. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, pages 265–280, 1945.

[20] Kurt Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept*, volume 1. Springer-Verlag, 2013.

[21] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain Francois, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. https://doi.org/10.21105/joss.01686 doi:10.21105/joss.01686.

Attachments

All of the code produced in this work can be found on:

https://github.com/MarcJohler/decisionmakeR

in the file "all_criteria.R" (commit version **bbb93bff2a26d5e5b6e10a95aacaa03716b3c6a8**).

Statement of originality

I hereby confirm that I have written the accompanying thesis with the title

Decision Making under Complex Uncertainty: Evaluation and Implementation of Different Criteria

by myself, without contributions from any sources other than those cited in the text and acknowledgements.

This applies also to all algorithms, graphics, drawings, maps and images included in the thesis.

Marc Johler 

Munich, November 12th 2020