Decision making with state-dependent preference systems

Christoph Jansen & Thomas Augustin
Department of Statistics, LMU Munich
IPMU 2022, Milano

Decision Theory: Basic model & classical solution

We consider the basic model of finite **Decision Theory**:

- $A = \{a_1, \ldots, a_n\}$ set of consequences
- $S = \{s_1, \ldots, s_m\}$ set of states
- $\mathcal{G} \subseteq A^S = \{X : S \to A\}$ set of acts

Decision Theory: Basic model & classical solution

We consider the basic model of finite **Decision Theory**:

- $A = \{a_1, \dots, a_n\}$ set of consequences
- $S = \{s_1, \dots, s_m\}$ set of states
- $\mathcal{G} \subseteq A^S = \{X : S \to A\}$ set of acts

Goal: Find optimal acts via some *choice function*

$$ch : 2^{\mathcal{G}} \to 2^{\mathcal{G}} \text{ with } ch(\mathcal{D}) \subseteq \mathcal{D} \text{ for all } \mathcal{D} \in 2^{\mathcal{G}}$$

that best possibly utilizes the available information.

Decision Theory: Basic model & classical solution

We consider the basic model of finite **Decision Theory**:

- $A = \{a_1, \ldots, a_n\}$ set of consequences
- $S = \{s_1, \ldots, s_m\}$ set of states
- $\mathcal{G} \subseteq A^S = \{X : S \to A\}$ set of acts

Goal: Find optimal acts via some *choice function*

$$ch : 2^{\mathcal{G}} \to 2^{\mathcal{G}} \text{ with } ch(\mathcal{D}) \subseteq \mathcal{D} \text{ for all } \mathcal{D} \in 2^{\mathcal{G}}$$

that best possibly utilizes the available information.

Classical approach: *If both*

I) preferences on A are characterized by a cardinal utility $u : A \to \mathbb{R}$ *and*

II) beliefs on S are characterized by a classical probability π,

then one commonly maximizes expected utility, i.e. defines

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

Problems with the classical solution

Obviously: If I) and/or II) are not satisfied, then $\text{ch}_{u,\pi}(\mathcal{D})$ is not well-defined.

Problem: In practice, this will often be the case.

(I) and II) require strong axiomatic assumptions, e.g. the axioms of Savage)

Problems with the classical solution

Obviously: If I) and/or II) are not satisfied, then $\mathrm{ch}_{u,\pi}(\mathcal{D})$ is not well-defined.

Problem: In practice, this will often be the case.

(I) and II) require strong axiomatic assumptions, e.g. the axioms of Savage)

Idea: Replace

- u by a set $\mathcal{U}$ of compatible utility functions on A and
- π by a set $\mathcal{M}$ of compatible probability measures on S

and generalize

- $\mathrm{ch}_{u,\pi}$ to a choice function $\mathrm{ch}_{\mathcal{U},\mathcal{M}}$

utilizing exactly the information that is encoded in the two sets $\mathcal{U}$ and $\mathcal{M}$ (and nothing more than that).

Modelling the set $\mathcal{U}$

Notation: Binary relation R has *strict part* P_R and *indifference part* I_R.

Preference system & Consistency

Let A denote a set of consequences. Let further

- $R_1 \subseteq A \times A$ be a binary relation on A
- $R_2 \subseteq R_1 \times R_1$ be a binary relation on R_1

The triplet $\mathcal{A} = [A, R_1, R_2]$ is called a **preference system** on A. We call $\mathcal{A}$ **consistent** if there exists $u : A \to [0,1]$ such that for all $a, b, c, d \in A$:

- $(a, b) \in R_1 \Rightarrow u(a) \geq u(b) \quad$ (with $=$ *iff* $\in I_{R_1}$).
- $((a, b), (c, d)) \in R_2 \Rightarrow u(a) - u(b) \geq u(c) - u(d) \quad$ (with $=$ *iff* $\in I_{R_2}$).

The set of all representations u of $\mathcal{A}$ is denoted by $\mathcal{U}_{\mathcal{A}}$.

Modelling the set $\mathcal{U}$

Notation: Binary relation R has *strict part* P_R and *indifference part* I_R.

Preference system & Consistency

Let A denote a set of consequences. Let further

- $R_1 \subseteq A \times A$ be a binary relation on A
- $R_2 \subseteq R_1 \times R_1$ be a binary relation on R_1

The triplet $\mathcal{A} = [A, R_1, R_2]$ is called a **preference system** on A. We call $\mathcal{A}$ **consistent** if there exists $u : A \to [0,1]$ such that for all $a, b, c, d \in A$:

- $(a, b) \in R_1 \Rightarrow u(a) \geq u(b) \quad$ (with $=$ *iff* $\in I_{R_1}$).
- $((a, b), (c, d)) \in R_2 \Rightarrow u(a) - u(b) \geq u(c) - u(d) \quad$ (with $=$ *iff* $\in I_{R_2}$).

The set of all representations u of $\mathcal{A}$ is denoted by $\mathcal{U}_{\mathcal{A}}$.

Interpretation of the components of $\mathcal{A}$:

- $(a, b) \in R_1$: *“a is at least as desirable as b”*
- $((a, b), (c, d)) \in R_2$: *“exchanging b by a is at least as desirable as d by c”*

Modelling the set $\mathcal{M}$

The uncertainty about S is characterized by a credal set of probabilities:

$$\mathcal{M} = \left\{\pi \in \mathcal{P} : \underline{b}_\ell \leq \mathbb{E}_\pi(f_\ell) \leq \overline{b}_\ell \text{ for } \ell = 1, \dots, r\right\}$$

where $\mathcal{P}$ denotes the set of all probability measures on $(S, 2^S)$ and

- $f_1, \dots, f_r : S \to \mathbb{R}$ are real-valued mappings and
- $\underline{b}_\ell \leq \overline{b}_\ell$, $\ell = 1, \dots, r$, are lower and upper expectation bounds.

Such $\mathcal{M}$ is a convex and finitely generated polyhedron with extreme points

$$\mathcal{E}(\mathcal{M}) = \{\pi^{(1)}, \dots \pi^{(K)}\}$$

Modelling the set $\mathcal{M}$

The uncertainty about S is characterized by a credal set of probabilities:

$$\mathcal{M} = \left\{ \pi \in \mathcal{P} : \underline{b}_\ell \leq \mathbb{E}_\pi(f_\ell) \leq \overline{b}_\ell \text{ for } \ell = 1, \ldots, r \right\}$$

where $\mathcal{P}$ denotes the set of all probability measures on $(S, 2^S)$ and

- $f_1, \ldots, f_r : S \to \mathbb{R}$ are real-valued mappings and
- $\underline{b}_\ell \leq \overline{b}_\ell$, $\ell = 1, \ldots, r$, are lower and upper expectation bounds.

Such $\mathcal{M}$ is a convex and finitely generated polyhedron with extreme points

$$\mathcal{E}(\mathcal{M}) = \{\pi^{(1)}, \ldots \pi^{(K)}\}$$

$\to$ Very general uncertainty model capturing special cases such as:

Classical probability – Interval probability – Lower previsions – Linear partial information – Neighbourhood models

Decision making based on $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{M}$

Theory for optimal decision making based on the sets $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{M}$ as well as efficient computation algorithms have been developed in:


Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Concepts for decision making under severe uncertainty with partial ordinal and partial cardinal preferences ☆,☆☆

C. Jansen *, G. Schollmeyer, T. Augustin

Methods for efficient elicitation of the underlying preference system and their theoretical properties have been investigated in:


Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty

C. Jansen *, H. Blocher, T. Augustin, G. Schollmeyer

Department of Statistics, LMU Munich, Ludwigsstr. 33, 80539 Munich, Germany

Problem: All these models only work for state-independent preferences!

Today: State-dependent preference systems

In many applications, the agent's preferences in a decision problem under uncertainty can not be modeled independently of the true state of nature.

Prominent examples:

- Insurance science: Often, a policyholder's preferences are modelled to be dependent on her health status.
- Portfolio selection: The agent's attitude towards risky choices (and therefore indirectly the underlying preferences) are seen as depending on some exogenous environment.

Basic Definitions I

We start by giving the fundamental definition of our basic model.

State-dependent decision system

Let

- $\mathcal{A}_s = [A, R_1^s, R_2^s]$ be a preference system for every state $s \in S$, and
- $\mathcal{G} \subseteq A^S := \{f : S \to A\}$ non-empty.

We call the pair

$$\mathcal{D} = \left[\mathcal{G}, (\mathcal{A}_s)_{s \in S}\right]$$

a **decision system**. We call $\mathcal{D}$

- **state-independent** if $\mathcal{A}_s = \mathcal{A}_{s'}$ for all $s, s' \in S$ and
- **state-dependent** otherwise.

Basic Definitions II

Especially in the case of a state-dependent decision system, it is useful to consider only utility functions that measure the utility on the same scale.

Commonly scalable, consistent

$\mathcal{D} = \left[\mathcal{G}, (\mathcal{A}_s)_{s\in S}\right]$ is called **commonly scalable** if there exist with $a_*, a^* \in A$

$$(a^*, a) \in R_1^s \ \wedge (a, a_*) \in R_1^s$$

for all $a \in A$ and $s \in S$.

Further, $\mathcal{D}$ is called **consistent** if

$$\mathcal{N}_{\mathcal{A}_s} := \left\{u \in \mathcal{U}_{\mathcal{A}_s} : u(a_*) = 0 \ \wedge \ u(a^*) = 1\right\} \neq \emptyset$$

for all states $s \in S$.

Structural assumption (wlog)

Dealing with the state-independent parts:

We assume, *without restricting generality of what follows*, that for some $\ell \in \{1, \ldots m\}$ there is a partition $\mathbb{S} := \{S_1, \ldots, S_\ell\}$ of S satisfying:

i) For all $d \in \{1, \ldots, \ell\}$ and all $s_{i_1}, s_{i_2} \in S_d$ it holds $\mathcal{A}_{s_{i_1}} = \mathcal{A}_{s_{i_2}}$.

ii) For all $c \neq d \in \{1, \ldots, \ell\}$ and all $s_{i_1} \in S_c$ and $s_{i_2} \in S_d$ it holds $\mathcal{A}_{s_{i_1}} \neq \mathcal{A}_{s_{i_2}}$.

iii) For $c < d \in \{1, \ldots, \ell\}$, if $s_{i_1} \in S_c$ and $s_{i_2} \in S_d$, then $i_1 < i_2$.

We then denote by $\mathcal{A}_{S_d}$ the preference system $\mathcal{A}_s$ for arbitrary $s \in S_d$.

The criterion of $(\mathcal{D}, \mathcal{M})$-dominance

Preparation: Let

- $\mathcal{D}$ be commonly scalable and consistent and
- π be a probability measure on $(S, 2^S)$ and
- $u := (u_d)_{d=1,\dots,\ell}$ be such that $u_d \in \mathcal{N}_{\mathcal{A}_{S_d}}$ for each $d = 1, \dots, \ell$.

The (π, u)-expectation of an act $X \in \mathcal{G}$ is the expression:

$$E_{(\pi,u)}(X) = \sum_{d=1}^{\ell} \left(\sum_{s \in S_d} u_d(X(s)) \cdot \pi(\{s\}) \right)$$

$(\mathcal{D}, \mathcal{M})$-dominance

Let $\mathcal{M}$ be a convex and finitely generated credal set.

For $X, Y \in \mathcal{G}$, say that Y is **$(\mathcal{D}, \mathcal{M})$-dominated** by X if

$$E_{(\pi,u)}(X) \geq E_{(\pi,u)}(Y)$$

for every $u := (u_d)_{d=1,\dots,\ell}$ with $u_d \in \mathcal{N}_{\mathcal{A}_{S_d}}$ and every $\pi \in \mathcal{M}$.

Remarks and special cases

If we have a **state-independent DS**...

- ... with $\mathcal{M} = \{\pi\}$ and $R_2 = \emptyset$

  $\rightarrow$ criterion reduces to (first-order) **stochastic dominance**

- ... with $\mathcal{M} = \{\pi\}$ and R_1 and R_2 guaranteeing utility unique up to plts

  $\rightarrow$ criterion reduces to comparing **expected utility**

Checking for $(\mathcal{D}, \mathcal{M})$-dominance: Preparation

Now, let

- $\mathcal{A} = [A, R_1, R_2]$ be a consistent decision system and
- $a_{k_1}, a_{k_2} \in A$ such that $(a_{k_1}, a) \in R_1$ and $(a, a_{k_2}) \in R_1$ for all $a \in A$.

A vector $(v_1, \ldots, v_n)$ containing exactly the images of a utility function $u \in \mathcal{N}_{\mathcal{A}}$ is then describable by the system of linear (in-)equalities given through

- $v_{k_1} = 1$ and $v_{k_2} = 0$,
- $v_i = v_j$ for every pair $(a_i, a_j) \in I_{R_1}$,
- $v_i - v_j \geq 0$ for every pair $(a_i, a_j) \in P_{R_1}$,
- $v_k - v_l = v_p - v_q$ for every pair of pairs $((a_k, a_l), (a_p, a_q)) \in I_{R_2}$ and
- $v_k - v_l - v_p + v_q \geq 0$ for every pair of pairs $((a_k, a_l), (a_p, a_q)) \in P_{R_2}$.

Denote by $\Delta_{\mathcal{A}}$ the set of all $(v_1, \ldots, v_n) \in [0, 1]^n$ satisfying these (in)equalities.

Checking for $(\mathcal{D},\mathcal{M})$-dominance: Linear Program

Checking for $(\mathcal{D},\mathcal{M})$-dominance

Let $\mathcal{D}$ be consistent and commonly scalable.

For $X, Y \in \mathcal{G}$, denote by x_j, y_j the unique i_X, i_Y with $X(s_j) = a_{i_X}$ and $Y(s_j) = a_{i_Y}$.

For every fixed $t \in \{1,\dots,K\}$, consider the linear optimization problem

$$\sum_{d=0}^{\ell-1}\left(\sum_{j=c_d+1}^{c_{d+1}}(v_{x_j}^d - v_{y_j}^d)\cdot\pi^{(t)}(\{s_j\})\right)\longrightarrow \min_{(v_1^1,\dots,v_n^1,\dots,v_1^\ell,\dots v_n^\ell)}$$

with constraints

- $(v_1^d,\dots,v_n^d) \in \Delta_{\mathcal{A}_{S_d}}$ for every $d \in \{1,\dots,\ell\}$

and the conventions $S_0 = \emptyset$ and $c_d = |\cup_{j=0}^d S_j|$.

Denote by $opt(t)$ the optimal value for t fixed. It then holds:

$$X \geq_{(\mathcal{D},\mathcal{M})} Y \quad\Leftrightarrow\quad \min\Big\{opt(t) : t \in \{1,\dots,K\}\Big\} \geq 0$$

Approximating the linear program

Challenge: The LPs have separate variables and constraints for each $\mathcal{A}_{S_d}$ under each $S_d \in \mathbb{S}$. This may produce high computational costs.

Idea: Approximate the LPs by grouping the preference systems under (in a certain sense) similar states of nature.

How exactly? Find partitions $\mathbb{V}$ of S of which the partition $\mathbb{S}$ is a refinement: For every element $S_d \in \mathbb{S}$ there exists an element $V \in \mathbb{V}$ such that $S_d \subseteq V$.

Then replace the LPs from before by

$$\sum_{d=0}^{\ell-1} \left(\sum_{j=p_d+1}^{p_{d+1}} (v_{x_j}^d - v_{y_j}^d) \cdot \pi^{(t)}(\{s_j\}) \right) \longrightarrow \min_{(v_1^1,\dots,v_n^1,\dots,v_1^r,\dots v_n^r)}$$

with constraints

- $(v_1^d, \dots, v_n^d) \in \Delta_{\mathcal{A}^{\mathbb{V}}_{V_d}}$ for every $d \in \{1, \dots, r\}$

and, again, $V_0 = \emptyset$ and $p_d = |\cup_{j=0}^d V_j|$.

Different choices for the partition

Pattern clustering: Partition the state space by grouping preference systems containing a predefined preference pattern.

Distance-based clustering: Partition the state space to groups of states $s \in S$ with 'similar' R_1^s, where similarity is defined by some distance between preorders and a threshold $\xi \in (0, 1)$ bounding it from above.

A small example

Let $A = \{a_*, b, c, d, a^*\}$ and consider the decision system

	s_1	s_2	s_3
X_1	d	c	b
X_2	a^*	d	a_*

where

- $R_1^S = R_1^{S_1} = R_1^{S_2} = R_1^{S_3}$ are all induced by $a^* P_{R_1^S} d P_{R_1^S} c P_{R_1^S} b P_{R_1^S} a_*$,
- $R_2^{S_1}$ is induced by $e_{ba_*} I_{R_2^{S_1}} e_{cb} I_{R_2^{S_1}} e_{dc} I_{R_2^{S_1}} e_{a^* d}$,
- $R_2^{S_2}$ is induced by $e_{ba_*} P_{R_2^{S_2}} e_{cb} P_{R_2^{S_2}} e_{a^* d} P_{R_2^{S_2}} e_{dc}$,
- $R_2^{S_3}$ is induced by $e_{ba_*} P_{R_2^{S_3}} e_{a^* d} P_{R_2^{S_3}} e_{cb} P_{R_2^{S_3}} e_{dc}$.

Assume the uncertainty about S is described by the credal set

$$\mathcal{M} = \{\pi : \pi(\{s_1\}) \leq 0.2 \ \wedge \ \pi(\{s_2\}) \leq 0.2\}.$$

A small example, continued

Three observations:

(1) $\mathcal{A}_{s_1}$ uniquely specifies a $u_{s_1} \in \mathcal{N}_{\mathcal{A}_{s_1}}$ given by

$$(u_{s_1}(a_*), u_{s_1}(b), u_{s_1}(c), u_{s_1}(d), u_{s_1}(a^*)) = (0, 0.25, 0.5, 0.75, 1).$$

(2) $\mathcal{A}_{s_2}$ restricts all $u_{s_2} \in \mathcal{N}_{\mathcal{A}_{s_2}}$ to satisfy $u_{s_2}(d) - u_{s_2}(c) \leq 0.25$.

(3) $\mathcal{A}_{s_3}$ restricts all $u_{s_3} \in \mathcal{N}_{\mathcal{A}_{s_3}}$ to satisfy $u_{s_3}(b) - u_{s_3}(a_*) \geq 0.25$.

Thus: For any $\pi \in \mathcal{M}$, $u_{s_1} \in \mathcal{N}_{\mathcal{A}_{s_1}}$, $u_{s_2} \in \mathcal{N}_{\mathcal{A}_{s_2}}$ and $u_{s_3} \in \mathcal{N}_{\mathcal{A}_{s_3}}$ the expression

$$E_{(\pi,u)}(X_1) - E_{(\pi,u)}(X_2)$$

can be computed by

$$-\underbrace{\pi_1(u_{s_1}(a^*) - u_{s_1}(d))}_{\leq 0.2 \cdot 0.25} - \underbrace{\pi_2(u_{s_2}(d) - u_{s_2}(c))}_{\leq 0.2 \cdot 0.25} + \underbrace{\pi_3(u_{s_3}(b) - u_{s_3}(a_*))}_{\geq 0.6 \cdot 0.25} > 0.$$

This gives $X_1 \geq_{(\mathcal{D},\mathcal{M})} X_2$.

An approximation under distance-based clustering yields the same.

Directions for future research

Some directions for future research are:

- **Comparison of cluster techniques:** Investigate which technique to use in what type of concrete application example.
- **Other approximation approaches:** Utilize existing approximations for the special case of two-monotone lower probabilities.
- **Adapt other decision criteria:** An adaptation of other criteria to the state-dependent setting would certainly deserve further research.
- **Real world application:** Test the model and its approximations in real world decision making problems.