Decision Making under Weakly Structured Information with Applications to Robust Statistics

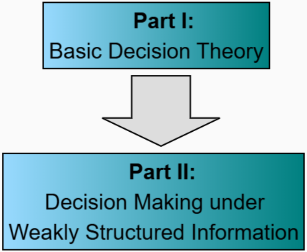
Presentation given at the **GOR AG FMI& FNAI Workshop**
Held on 25.03.2026 – 26.03.2026 at Lancaster University Leipzig

Christoph Jansen
School of Computing & Communications
Lancaster University Leipzig

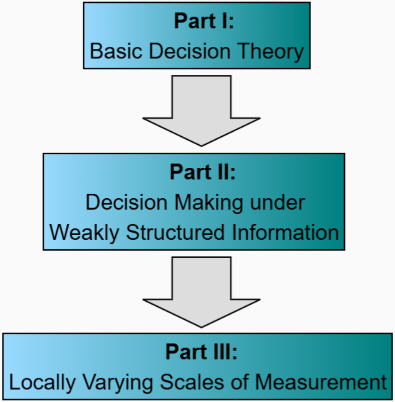
Outline

Part I:
Basic Decision Theory

Outline


Part I:
Basic Decision Theory
Part II:
Decision Making under
Weakly Structured Information

Outline


Part I:
Basic Decision Theory
Part II:
Decision Making under
Weakly Structured Information
Part III:
Locally Varying Scales of Measurement

Basic Decision Theory

Informal description of the model:

- An agent has to choose among different acts X from a set $\mathcal{G}$.
- The consequence that choosing X yields depends on which state of nature s from a set S is the true one.

Basic Decision Theory

Informal description of the model:

- An agent has to choose among different acts X from a set $\mathcal{G}$.
- The consequence that choosing X yields depends on which state of nature s from a set S is the true one.

Formal description of the model:

- Let C denote some non-empty set of consequences.
- Each act X corresponds to a mapping $X : S \to C$.
- The set $\mathcal{G}$ is a subset of $C^S = \{X : S \to C\}$.

Basic Decision Theory

Informal description of the model:

- An agent has to choose among different acts X from a set $\mathcal{G}$.
- The consequence that choosing X yields depends on which state of nature s from a set S is the true one.

Formal description of the model:

- Let C denote some non-empty set of consequences.
- Each act X corresponds to a mapping $X : S \to C$.
- The set $\mathcal{G}$ is a subset of $C^S = \{X : S \to C\}$.

Goal: Determining a 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.

Interpreting Choice Functions

Depending on the quality of the underlying information, the choice sets $ch(\mathcal{D})$ can be given two different interpretations:

Interpreting Choice Functions

Depending on the quality of the underlying information, the choice sets $ch(\mathcal{D})$ can be given two different interpretations:

Strong interpretation:

- $ch(\mathcal{D})$ is the set of equally optimal acts from $\mathcal{D}$.
- The agent is indifferent between these acts.

Interpreting Choice Functions

Depending on the quality of the underlying information, the choice sets $ch(\mathcal{D})$ can be given two different interpretations:

Strong interpretation:

- $ch(\mathcal{D})$ is the set of equally optimal acts from $\mathcal{D}$.
- The agent is indifferent between these acts.

Weak interpretation:

- $ch(\mathcal{D})$ is the set of all non-neglectable acts from $\mathcal{D}$.
- These acts are incomparable for the agent.

Interpreting Choice Functions

Depending on the quality of the underlying information, the choice sets $ch(\mathcal{D})$ can be given two different interpretations:

Strong interpretation:

- $ch(\mathcal{D})$ is the set of equally optimal acts from $\mathcal{D}$.
- The agent is indifferent between these acts.

Weak interpretation:

- $ch(\mathcal{D})$ is the set of all non-neglectable acts from $\mathcal{D}$.
- These acts are incomparable for the agent.

Obvious comment: If only weakly structured information is available, we often have to work with weakly interpretable choice functions.

Two Standard Choice Functions

Expected utility:

If a probability π on S and a cardinal scale $u : C \to [0, 1]$ are available, set

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

and choose that acts from $\mathcal{D}$ that maximize expected utility.

Two Standard Choice Functions

Expected utility:

If a probability π on S and a cardinal scale $u : C \to [0,1]$ are available, set

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

and choose that acts from $\mathcal{D}$ that maximize expected utility.

Stochastic Dominance:

If a probability π on S and a preorder $\succsim$ on C are available, set

$$ch_{\succsim,\pi}(\mathcal{D}) = \left\{ Y : \nexists X \text{ s.t. } \begin{array}{ll} \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y) & \text{for all } \quad u \in \mathcal{U}_\succsim \\ \mathbb{E}_\pi(u \circ X) > \mathbb{E}_\pi(u \circ Y) & \text{for some } u \in \mathcal{U}_\succsim \end{array} \right\}$$

where $\mathcal{U}_\succsim$ is the set of all $\succsim$-isotone $u : C \to [0,1]$. Choose acts that are not excluded by every compatible EU-maximizer.

Example: Expected Utility

Expected utility:

If a probability π on S and a cardinal scale $u : C \to [0,1]$ are available, set

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

and choose that acts from $\mathcal{D}$ that maximize expected utility.

Example: Expected Utility

Consider the following decision problem with $\mathcal{G} = \{X_1, X_2, X_3\}$: $\mathcal{G} = \{X_1, X_2, X_3\}$, $S = \{s_1, s_2, s_3\}$, $C = \{c_1, \ldots, c_9\}$:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9

Example: Expected Utility

Consider the following decision problem with $\mathcal{G} = \{X_1, X_2, X_3\}$: $\mathcal{G} = \{X_1, X_2, X_3\}$, $S = \{s_1, s_2, s_3\}$, $C = \{c_1, \ldots, c_9\}$:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9

Assume we know

- a probability measure $\pi : 2^S \to [0, 1]$ on $S = \{s_1, s_2, s_3\}$,
- a utility function $u : C \to [0, 1]$ on $C = \{c_1, \ldots, c_9\}$.

Example: Expected Utility

Consider the following decision problem with $\mathcal{G} = \{X_1, X_2, X_3\}$: $\mathcal{G} = \{X_1, X_2, X_3\}$, $S = \{s_1, s_2, s_3\}$, $C = \{c_1, \ldots, c_9\}$:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9

Assume we know

- a probability measure $\pi : 2^S \to [0,1]$ on $S = \{s_1, s_2, s_3\}$,
- a utility function $u : C \to [0,1]$ on $C = \{c_1, \ldots, c_9\}$.

	$\pi(s_1) = \frac{1}{3}$	$\pi(s_2) = \frac{1}{3}$	$\pi(s_3) = \frac{1}{3}$
$u \circ X_1$	6000	3000	-6000
$u \circ X_2$	8000	1000	-3000
$u \circ X_3$	5000	4000	0

Example: Expected Utility

Consider the following decision problem with $\mathcal{G} = \{X_1, X_2, X_3\}$: $\mathcal{G} = \{X_1, X_2, X_3\}$, $S = \{s_1, s_2, s_3\}$, $C = \{c_1, \ldots, c_9\}$:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9

Assume we know

- a probability measure $\pi : 2^S \to [0,1]$ on $S = \{s_1, s_2, s_3\}$,
- a utility function $u : C \to [0,1]$ on $C = \{c_1, \ldots, c_9\}$.

	$\pi(s_1) = \frac{1}{3}$	$\pi(s_2) = \frac{1}{3}$	$\pi(s_3) = \frac{1}{3}$	$\mathbb{E}_\pi(u \circ X_i)$
$u \circ X_1$	6000	3000	-6000	1000
$u \circ X_2$	8000	1000	-3000	2000
$u \circ X_3$	5000	4000	0	3000

Example: Expected Utility

Consider the following decision problem with $\mathcal{G} = \{X_1, X_2, X_3\}$: $\mathcal{G} = \{X_1, X_2, X_3\}$, $S = \{s_1, s_2, s_3\}$, $C = \{c_1, \ldots, c_9\}$:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9

Assume we know

- a probability measure $\pi : 2^S \to [0, 1]$ on $S = \{s_1, s_2, s_3\}$,
- a utility function $u : C \to [0, 1]$ on $C = \{c_1, \ldots, c_9\}$.

	$\pi(s_1) = \frac{1}{3}$	$\pi(s_2) = \frac{1}{3}$	$\pi(s_3) = \frac{1}{3}$	$\mathbb{E}_\pi(u \circ X_i)$
$u \circ X_1$	6000	3000	-6000	1000
$u \circ X_2$	8000	1000	-3000	2000
$u \circ X_3$	5000	4000	0	3000

Expected utility maximization: $\mathrm{ch}_{u,\pi}(\mathcal{G}) = \mathrm{argmax}_{X \in \mathcal{G}} \mathbb{E}_\pi(u \circ X) = \{X_3\}$

Example: Stochastic Dominance

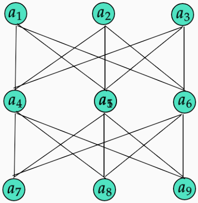
Now assume, the preferences allow only for a preorder $\succsim$ on C.

Example: Stochastic Dominance

Now assume, the preferences allow only for a preorder $\succsim$ on C.

Then, the situation looks for instance like this:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9



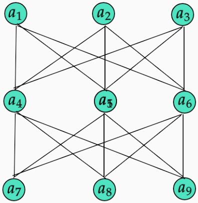
Every $\succsim$-isotone function $u : C \to [0, 1]$ is a compatible scale.

Example: Stochastic Dominance

Now assume, the preferences allow only for a preorder $\succsim$ on C.

Then, the situation looks for instance like this:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9



Every $\succsim$-isotone function $u : C \to [0, 1]$ is a compatible scale.

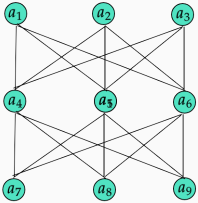
$\Rightarrow$ Maximizing expected utility is not-well-defined!

Example: Stochastic Dominance

Now assume, the preferences allow only for a preorder $\succsim$ on C.

Then, the situation looks for instance like this:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9



Every $\succsim$-isotone function $u : C \to [0, 1]$ is a compatible scale.

$\Rightarrow$ Maximizing expected utility is not-well-defined!

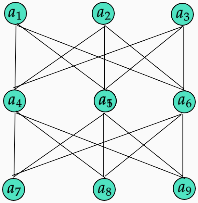
$\Rightarrow$ We can still apply FSD, but now $\mathrm{ch}_{\succsim,\pi}(\mathcal{G}) = \mathcal{G}$, since for every act there exists a compatible scale making it the unique EU-maximizer.

Example: Stochastic Dominance

Now assume, the preferences allow only for a preorder $\succsim$ on C.

Then, the situation looks for instance like this:

	s_1	s_2	s_3
X_1	c_1	c_4	c_7
X_2	c_2	c_5	c_8
X_3	c_3	c_6	c_9



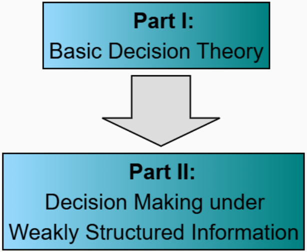
Every $\succsim$-isotone function $u : C \to [0, 1]$ is a compatible scale.

$\Rightarrow$ Maximizing expected utility is not-well-defined!

$\Rightarrow$ We can still apply FSD, but now $\mathrm{ch}_{\succsim,\pi}(\mathcal{G}) = \mathcal{G}$, since for every act there exists a compatible scale making it the unique EU-maximizer.

$\Rightarrow$ None of the acts can be excluded. (Weak interpretation!)

Outline

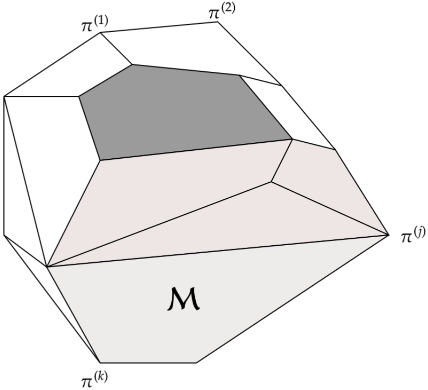


Part I:
Basic Decision Theory
Part II:
Decision Making under
Weakly Structured Information

Weakly Structured Information (WSI)

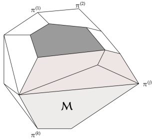
Imprecise probabilistic models: If not possible to specify **one** probability on S, work with the **credal set** $\mathcal{M}$ of all **compatible** probabilities.

Weakly Structured Information (WSI)



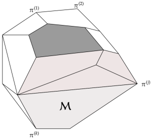
Weakly Structured Information (WSI)

Imprecise probabilistic models: If not possible to specify **one** probability on S, work with the **credal set** $\mathcal{M}$ of all **compatible** probabilities.



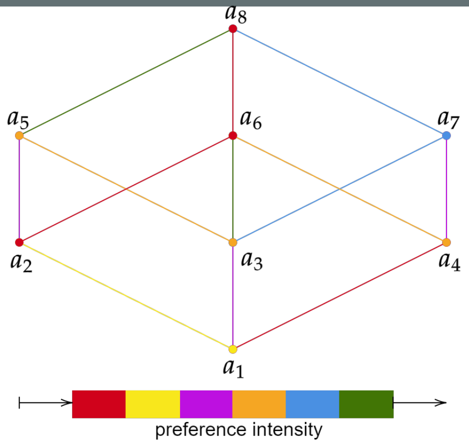
Weakly Structured Information (WSI)

Imprecise probabilistic models: If not possible to specify **one** probability on S, work with the **credal set** $\mathcal{M}$ of all **compatible** probabilities.



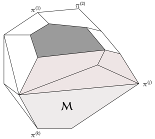
Complexly ordered consequences: If not possible to specify **one** utility function on C, work with the **set** $\mathcal{U}$ of all **compatible** utilities.

Weakly Structured Information (WSI)

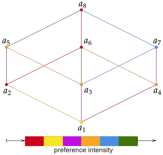


Weakly Structured Information (WSI)

Imprecise probabilistic models: If not possible to specify **one** probability on S, work with the **credal set** $\mathcal{M}$ of all **compatible** probabilities.



Complexly ordered consequences: If not possible to specify **one** utility function on C, work with the **set** $\mathcal{U}$ of all **compatible** utilities.



Modelling $\mathcal{U}$: Preference Systems

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

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

The triplet $\mathcal{A} = [C, R_1, R_2]$ is called a **preference system** on C.

We call $\mathcal{A}$ **consistent** if there is $u : C \to [0, 1]$ with for all $a, b, c, d \in C$:

$$(a, b) \in R_1 \Rightarrow u(a) \geq u(b) \qquad (\text{with} = \textit{iff} \in I_{R_1}).$$

$$((a, b), (c, d)) \in R_2 \Rightarrow u(a) - u(b) \geq u(c) - u(d) \qquad (\text{with} = \textit{iff} \in I_{R_2}).$$

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

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 $\mathcal{U}$: Preference Systems

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

$R_1 \subseteq C \times C$ be a binary relation on C

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

The triplet $\mathcal{A} = [C, R_1, R_2]$ is called a **preference system** on C.

We call $\mathcal{A}$ **consistent** if there is $u : C \to [0, 1]$ with for all $a, b, c, d \in C$:

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

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

Example: Multi-target decision making

Let $\mathcal{G}$ be a decision problem with C as its consequence set.

Example: Multi-target decision making

Let $\mathcal{G}$ be a decision problem with C as its consequence set.

A target evaluation is a function $\phi : C \to [0, 1]$.

Example: Multi-target decision making

Let $\mathcal{G}$ be a decision problem with C as its consequence set.

A target evaluation is a function $\phi : C \to [0, 1]$.

Let $(\phi_j)_{j=1,\ldots,r}$ be target evaluations. Assume – w.l.o.g. – that

- $\phi_1, \ldots, \phi_z$ are of cardinal scale, while
- $\phi_{z+1}, \ldots, \phi_r$ are purely ordinal.

Example: Multi-target decision making

Let $\mathcal{G}$ be a decision problem with C as its consequence set.

A target evaluation is a function $\phi : C \to [0, 1]$.

Let $(\phi_j)_{j=1,\dots,r}$ be target evaluations. Assume – w.l.o.g. – that

- $\phi_1, \dots, \phi_z$ are of cardinal scale, while
- $\phi_{z+1}, \dots, \phi_r$ are purely ordinal.

Define a preference system

$$\text{pref}(C) = [C, R_1^*, R_2^*]$$

$$R_1^* = \Big\{(c, c') : \phi_j(c) \geq \phi_j(c') \ \forall j \leq r\Big\}$$

$$R_2^* = \left\{((c, d), (c', d')) : \begin{array}{l} \phi_j(c) - \phi_j(d) \geq \phi_j(c') - \phi_j(d') \ \ \forall j \leq z \\ \phi_j(c) \geq \phi_j(c') \geq \phi_j(d') \geq \phi_j(d) \ \forall j > z \end{array} \right\}.$$

Modelling $\mathcal{M}$: Credal sets

Credal set

The uncertainty among the elements of S is described by a polyhedral *credal set* of probability measures of the form

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

where Δ_S is the set of all probability measures on $(S, \sigma(S))$ 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.

Modelling $\mathcal{M}$: Credal sets

Credal set

The uncertainty among the elements of S is described by a polyhedral *credal set* of probability measures of the form

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

where Δ_S is the set of all probability measures on $(S, \sigma(S))$ 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.

If S is finite, such $\mathcal{M}$ is a convex polyhedron with extreme points

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

Modelling $\mathcal{M}$: Credal sets

Credal set

The uncertainty among the elements of S is described by a polyhedral *credal set* of probability measures of the form

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

where Δ_S is the set of all probability measures on $(S, \sigma(S))$ 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.

If S is finite, such $\mathcal{M}$ is a convex polyhedron with extreme points

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

Special cases: *Classical probability – Interval probability – Lower previsions – Linear partial information – Neighbourhood models*

Example: Contamination models

Let π be a probability measure on $(S, \sigma(S))$, the central distribution.

Example: Contamination models

Let π be a probability measure on $(S, \sigma(S))$, the central distribution.

For every $X \in \mathcal{G}$ and $\varepsilon \in [0, 1]$ fixed, define the credal set

$$\mathcal{M}_{X,\varepsilon} = \left\{(1 - \varepsilon)\pi_X + \varepsilon\mu : \mu \in \Delta_C\right\}$$

where π_X is the **push-forward measure** of π under X.

Example: Contamination models

Let π be a probability measure on $(S, \sigma(S))$, the central distribution.

For every $X \in \mathcal{G}$ and $\varepsilon \in [0, 1]$ fixed, define the credal set

$$\mathcal{M}_{X,\varepsilon} = \Big\{(1 - \varepsilon)\pi_X + \varepsilon\mu : \mu \in \Delta_C\Big\}$$

where π_X is the **push-forward measure** of π under X.

We can then define a credal set of the form before by setting

$$\mathcal{M}_{\mathcal{G},\varepsilon} = \Big\{\pi' \in \Delta_S : \pi'_X \in \mathcal{M}_{X,\varepsilon} \text{ for all } X \in \mathcal{G}\Big\}$$

Call $\mathcal{M}_{\mathcal{G},\varepsilon}$ the ε-contamination model for $\mathcal{G}$.

Example: Contamination models

Let π be a probability measure on $(S, \sigma(S))$, the central distribution.

For every $X \in \mathcal{G}$ and $\varepsilon \in [0, 1]$ fixed, define the credal set

$$\mathcal{M}_{X,\varepsilon} = \left\{ (1 - \varepsilon)\pi_X + \varepsilon\mu : \mu \in \Delta_C \right\}$$

where π_X is the **push-forward measure** of π under X.

We can then define a credal set of the form before by setting

$$\mathcal{M}_{\mathcal{G},\varepsilon} = \left\{ \pi' \in \Delta_S : \pi'_X \in \mathcal{M}_{X,\varepsilon} \text{ for all } X \in \mathcal{G} \right\}$$

Call $\mathcal{M}_{\mathcal{G},\varepsilon}$ the ε-contamination model for $\mathcal{G}$.

For each $\mathcal{M}_{X,\varepsilon}$, the extreme points are given as:

$$\mathcal{E}(\mathcal{M}_{X,\varepsilon}) = \left\{ (1 - \varepsilon)\pi_X + \varepsilon\delta_c : c \in C \right\}$$

Decision Making under Weakly Structured Information

Theory for optimal decision making based on the sets $\mathcal{U}$ 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

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/mss

A probabilistic evaluation framework for preference aggregation reflecting group homogeneity

C. Jansen, G. Schollmeyer, T. Augustin

Home > Symbolic and Quantitative Approaches to Reasoning with Uncertainty > Conference paper

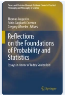
Decision Theory Meets Linear Optimization Beyond Computation

Conference paper | First Online: 15 June 2017

pp 329–339 | Cite this conference paper

Home > Reflections on the Foundations of Probability and Statistics > Chapter

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



A Choice Function under WSI

A choice function between the extremes can be obtained as follows. Let

- $\mathcal{M}$ be a (well-behaving) credal set on S, and
- $\mathcal{U}$ be a set of (well-behaving) utility functions $u : C \to [0, 1]$.

Generalized Stochastic Dominance (GSD)

Define the following preorder among ps-valued variables $X, Y : S \to C$:

$$X \succsim_{\text{GSD}} Y \quad :\Leftrightarrow \quad \forall u \in \mathcal{U} \; \forall \pi \in \mathcal{M} : \; \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$$

The relation $\succsim_{\text{GSD}}$ is called **generalized stochastic dominance (GSD)**.

A Choice Function under WSI

A choice function between under WSI can be obtained as follows. Let

- $\mathcal{M}$ be a credal set of the form before, and
- $\mathcal{U}$ be a set of utility functions $u : C \to [0, 1]$.

Generalized Stochastic Dominance (GSD)

Define the following preorder among ps-valued variables $X, Y : S \to C$:

$$X \succsim_{\text{GSD}} Y \quad :\Leftrightarrow \quad \forall u \in \mathcal{U}\ \forall \pi \in \mathcal{M} : \ \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$$

The relation $\succsim_{\text{GSD}}$ is called **generalized stochastic dominance (GSD)**.

The GSD relation now directly induces the GSD choice function by setting

$$\text{GSD}(\mathcal{U}, \mathcal{M}, \mathcal{D}) := \left\{ X \in \mathcal{D} : \nexists Y \in \mathcal{D} \text{ such that } Y \succ_{\text{GSD}} X \right\}$$

[Jansen et al., 2018, Jansen et al., 2022]

Checking for GSD for finite C

Fix $X_i, X_j \in \mathcal{G}$, $\pi^{(t)} \in \mathcal{E}(\mathcal{M}) = \{\pi^{(1)}, \dots \pi^{(K)}\}$, and $C = \{c_1, \dots, c_n\}$.

Consider the LP:

$$\sum_{\ell=1}^{n} v_\ell \cdot [\pi_{X_i}^{(t)}(c_\ell) - \pi_{X_j}^{(t)}(c_\ell)] \longrightarrow \min_{(v_1,\dots,v_n)\in[0,1]^n}$$

under the constraints $(v_1, \dots, v_n) \in [0,1]^n$ and

- $v_i = v_j$ for every pair $(c_i, c_j) \in I_{R_1}$,
- $v_i - v_j \geq 0$ for every pair $(c_i, c_j) \in P_{R_1}$,
- $v_k - v_l = v_p - v_q$ for every pair of pairs $((c_k, c_l), (c_p, c_q)) \in I_{R_2}$ and
- $v_k - v_l - v_p + v_q \geq 0$ for every pair of pairs $((c_k, c_l), (c_p, c_q)) \in P_{R_2}$.

Denote by $opt_{ij}(t)$ its optimal value. It then holds:

$$X_i \succsim_{GSD} X_j \quad \Leftrightarrow \quad \min\{opt_{ij}(t) : t = 1, \dots, K\} \geq 0.$$

Example: Contamination models again

In this setting, assume

- $\mathcal{M} := \mathcal{M}_{\mathcal{G},\varepsilon}$, and
- $\exists c_{k_1}, c_{k_2} \in C$ such that c_{k_1} is R_1-maximal and c_{k_2} is R_1-minimal.

Set $f(i,j,\ell) := \pi_{X_i}(c_\ell) - \pi_{X_j}(c_\ell)$ and consider the LP:

$$v_{k_1} \cdot [f(i,j,k_1) - \varepsilon] + v_{k_2} \cdot [f(i,j,k_2) + \varepsilon] + \sum_{\ell \notin \{k_1,k_2\}} v_\ell \cdot f(i,j,\ell) \longrightarrow \min_{(v_1,\ldots,v_n)\in[0,1]^n}$$

- $v_i = v_j$ for every pair $(c_i, c_j) \in I_{R_1}$,
- $v_i - v_j \geq 0$ for every pair $(c_i, c_j) \in P_{R_1}$,
- $v_k - v_l = v_p - v_q$ for every pair of pairs $((c_k, c_l), (c_p, c_q)) \in I_{R_2}$ and
- $v_k - v_l - v_p + v_q \geq 0$ for every pair of pairs $((c_k, c_l), (c_p, c_q)) \in P_{R_2}$.

Denote by opt_{ij} its optimal value. It then holds: $X_i \succsim_{GSD} X_j \quad \Leftrightarrow \quad opt_{ij} \geq 0$.

Toy Example: Algorithm Comparison under Contamination

We assume that six different algorithms $A_1, \dots, A_6$ are to be compared with respect to three different targets, more precisely:

- ϕ_1 **Running Time** (cardinal target, higher is better)
- ϕ_2 **Performance** (cardinal target, higher is better)
- ϕ_3 **Explainability** (Ordinal target with labels $\{0, 0.1, \dots, 0.9, 1\}$).

Toy Example: Algorithm Comparison under Contamination

We assume that six different algorithms $A_1, \dots, A_6$ are to be compared with respect to three different targets, more precisely:

- ϕ_1 **Running Time** (cardinal target, higher is better)
- ϕ_2 **Performance** (cardinal target, higher is better)
- ϕ_3 **Explainability** (Ordinal target with labels $\{0, 0.1, \dots, 0.9, 1\}$).

Performance wrt to (ϕ_1, ϕ_2, ϕ_3) is affected by scenarios $S = \{s_1, s_2, s_3, s_4, s_5\}$:

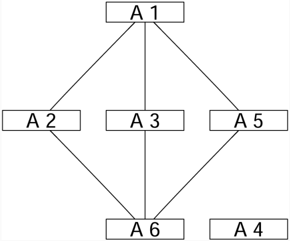
ϕ_1	s_1	s_2	s_3	s_4	s_5
A_1	0.81	0.86	0.68	0.72	0.56
A_2	0.64	0.82	0.62	0.93	0.68
A_3	0.60	0.66	0.62	0.73	0.58
A_4	0.75	0.66	0.97	0.83	0.64
A_5	0.33	0.30	0.53	0.38	0.44
A_6	0.00	0.21	0.56	0.12	0.72

ϕ_2	s_1	s_2	s_3	s_4	s_5
A_1	0.71	0.88	0.82	0.90	0.91
A_2	0.52	0.67	0.68	0.72	0.88
A_3	0.56	0.45	0.81	0.83	0.47
A_4	0.36	0.12	0.54	0.60	0.17
A_5	0.79	0.30	0.47	0.68	0.46
A_6	0.14	0.58	0.30	0.66	0.29

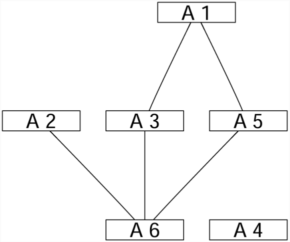
ϕ_3	s_1	s_2	s_3	s_4	s_5
A_1	0.70	1.00	0.80	0.60	0.90
A_2	0.50	0.80	0.60	0.50	0.80
A_3	0.50	0.40	0.60	0.40	0.70
A_4	0.70	0.40	0.70	0.70	0.30
A_5	0.60	0.20	0.20	0.30	0.30
A_6	0.10	0.20	0.40	0.30	0.20

Results

Solving the linear program for all pairs $(A_i, A_j), i \neq j$ gives Hasse diagrams:

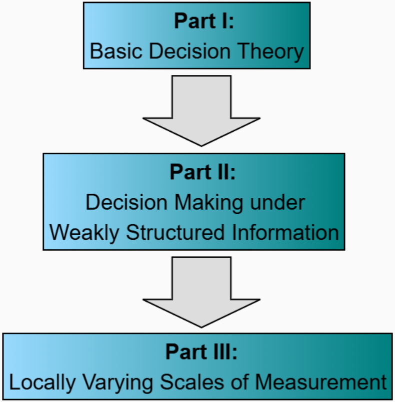


$\mathsf{GSD}(\mathcal{A}, \mathcal{M}, \mathcal{G}), \varepsilon = 0$
$\{A_1, A_4\}$



$\mathsf{GSD}(\mathcal{A}, \mathcal{M}, \mathcal{G}), \varepsilon = 0.05$
$\{A_1, A_2, A_4\}$

Outlook



Part I:
Basic Decision Theory
Part II:
Decision Making under
Weakly Structured Information
Part III:
Locally Varying Scales of Measurement

Standard Scales of Measurement

Consider some random variable $X : \Omega \to A$ mapping to some set A.

Standard Scales of Measurement

Consider some random variable $X : \Omega \to A$ mapping to some set A.

Ordinal Data:

If A is structured only by a preorder $\succsim$, we call X of ordinal scale.

Then: Any analysis of the variable X has to be invariant under $\succsim$-monotone scale transformations $u : A \to [0, 1]$, as these encode the same information.

Standard Scales of Measurement

Consider some random variable $X : \Omega \to A$ mapping to some set A.

Ordinal Data:

If A is structured only by a preorder $\succsim$, we call X of ordinal scale.

Then: Any analysis of the variable X has to be invariant under $\succsim$-monotone scale transformations $u : A \to [0, 1]$, as these encode the same information.

Cardinal Data:

If there exists a natural scale $u^* : A \to [0, 1]$, we call X of cardinal scale.

Then: Any analysis of the variable X can be based on u^* alone.

Preference Systems in Statistics

Question: What if X does not belong to either extreme pole?

In other words: What if the structuredness of A varies along its subsets?

Preference Systems in Statistics

Question: What if X does not belong to either extreme pole?

In other words: What if the structuredness of A varies along its subsets?

A **preference system** $\mathcal{A} = [A, R_1, R_2]$ helps to formalize:

- R_1 formalizes the available ordinal information.
- R_2 describes the available cardinal information.
- A is locally almost cardinal on subsets where R_1 and R_2 are dense.
- A is locally at most ordinal on subsets where R_2 is sparse.

Comparing PS-Valued Variables

Idea: Weaken the expectation order to a preorder by demanding expectation dominance for all scales u compatible with the preference system $\mathcal{A}$.

Comparing PS-Valued Variables

Idea: Weaken the expectation order to a preorder by demanding expectation dominance for all scales u compatible with the preference system $\mathcal{A}$.

This idea leads to the following precise version of GSD.

Generalized Stochastic Dominance (GSD)

Define the following preorder among ps-valued variables $X, Y : \Omega \to A$:

$$X \succsim_{\text{GSD}} Y \quad :\Leftrightarrow \quad \forall u \in \mathcal{U}_{\mathcal{A}} : \ \mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$$

The relation $\succsim_{\text{GSD}}$ is called **generalized stochastic dominance (GSD)**.

Addressing Sampling Uncertainty (JMLR 2023)

Practical Problem: Usually, we do not know π but only *i.i.d.* samples

$$\mathsf{X} = (X_1, \ldots, X_n) \quad \text{and} \quad \mathsf{Y} = (Y_1, \ldots, Y_m)$$

of X and Y are available.

Addressing Sampling Uncertainty (JMLR 2023)

Practical Problem: Usually, we do not know π but only *i.i.d.* samples

$$\mathsf{X} = (X_1, \ldots, X_n) \quad \text{and} \quad \mathsf{Y} = (Y_1, \ldots, Y_m)$$

of X and Y are available.

Approach: Perform a statistical test for GSD.

Addressing Sampling Uncertainty (JMLR 2023)

Practical Problem: Usually, we do not know π but only *i.i.d.* samples

$$\mathbf{X} = (X_1, \ldots, X_n) \quad \text{and} \quad \mathbf{Y} = (Y_1, \ldots, Y_m)$$

of X and Y are available.

Approach: Perform a statistical test for GSD.

Hypotheses:

$$H_0 : Y \succsim_{\text{GSD}} X \quad \textbf{vs.} \quad H_1 : \neg Y \succsim_{\text{GSD}} X$$

Addressing Sampling Uncertainty (JMLR 2023)

Practical Problem: Usually, we do not know π but only *i.i.d.* samples

$$\mathbf{X} = (X_1, \ldots, X_n) \quad \text{and} \quad \mathbf{Y} = (Y_1, \ldots, Y_m)$$

of X and Y are available.

Approach: Perform a statistical test for GSD.

Hypotheses:

$$H_0 : Y \succsim_{\text{GSD}} X \quad \textbf{vs.} \quad H_1 : \neg Y \succsim_{\text{GSD}} X$$

Good News: As the worst case under the null hypothesis H_0 is

$$\pi_X = \pi_Y$$

performing a permutation test gives a valid level α test.

The Choice of the Test Statistic

Observation: It holds $X \succsim_{\text{GSD}} Y$ if and only if

$$D(X, Y) := \inf_{u \in \mathcal{U}} \left(\mathbb{E}_\pi(u \circ X) - \mathbb{E}_\pi(u \circ Y)\right) \geq 0.$$

The Choice of the Test Statistic

Observation: It holds $X \succsim_{\mathrm{GSD}} Y$ if and only if

$$D(X,Y) := \inf_{u \in \mathcal{U}} \left(\mathbb{E}_\pi(u \circ X) - \mathbb{E}_\pi(u \circ Y) \right) \geq 0.$$

Consequence: A natural test statistic is the empirical version of $D(X,Y)$, i.e.,

$$d_{\mathrm{X,Y}} : \Omega \to \mathbb{R}$$

$$\omega \mapsto \inf_{u \in \mathcal{A}_\omega} \sum_{z \in (\mathrm{XY})_\omega} u(z) \cdot \left(\hat{\pi}_X^\omega(\{z\}) - \hat{\pi}_Y^\omega(\{z\}) \right)$$

where $\hat{\pi}_X^\omega$, $\hat{\pi}_Y^\omega$, $(\mathrm{XY})_\omega$, and $\mathcal{A}_\omega$ are the corresponding empirical analoga.

The Choice of the Test Statistic

Observation: It holds $X \succsim_{\text{GSD}} Y$ if and only if

$$D(X,Y) := \inf_{u \in \mathcal{U}} \left(\mathbb{E}_\pi(u \circ X) - \mathbb{E}_\pi(u \circ Y)\right) \geq 0.$$

Consequence: A natural test statistic is the empirical version of $D(X,Y)$, i.e.,

$$d_{\mathrm{X,Y}} : \Omega \to \mathbb{R}$$

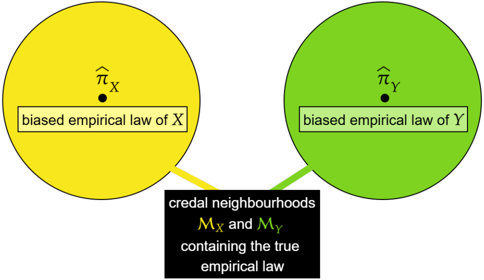
$$\omega \mapsto \inf_{u \in \mathcal{A}_\omega} \sum_{z \in (\mathrm{XY})_\omega} u(z) \cdot \left(\hat{\pi}_X^\omega(\{z\}) - \hat{\pi}_Y^\omega(\{z\})\right)$$

where $\hat{\pi}_X^\omega$, $\hat{\pi}_Y^\omega$, $(\mathrm{XY})_\omega$, and $\mathcal{A}_\omega$ are the corresponding empirical analoga.

Good news: The test statistic $d_{\mathrm{X,Y}}$ can be computed by solving the very same linear programming problem we seen before.

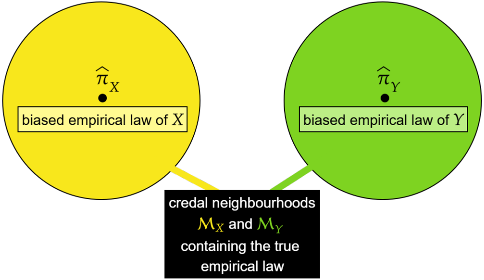
Credal Sets For Robustified Testing (UAI 2023)

Idea: Use credal sets to robustify the permutation test to small deviations from the *i.i.d.* assumption.



Credal Sets For Robustified Testing (UAI 2023)

Idea: Use credal sets to robustify the permutation test to small deviations from the *i.i.d.* assumption.



Adapted Resampling Scheme: Replace

- $d_{\mathrm{X,Y}}(\omega)$ by $\inf_{(\pi_1,\pi_2)\in\mathcal{M}_X\times\mathcal{M}_Y} \tilde{d}_{\mathrm{X,Y}}(\pi_1,\pi_2)$
- d_{re} by $\sup_{(\pi_1,\pi_2)\in\mathcal{M}_X\times\mathcal{M}_Y} \tilde{d}_{re}(\pi_1,\pi_2)$

Results in: Valid (yet conservative) statistical test!

Example: ε-contamination models (again)

Let both $\mathcal{M}_X$ and $\mathcal{M}_Y$ be contamination models, i.e., assume

$$\mathcal{M}_Z = \left\{(1-\varepsilon)\hat{\pi}_Z + \varepsilon\mu : \mu \in \Delta_A\right\}$$

for $Z \in \{X, Y\}$ and $\varepsilon \in [0, 1]$ fixed.

Example: ε-contamination models (again)

Let both $\mathcal{M}_X$ and $\mathcal{M}_Y$ be contamination models, i.e., assume

$$\mathcal{M}_Z = \left\{(1-\varepsilon)\hat{\pi}_Z + \varepsilon\mu : \mu \in \Delta_A\right\}$$

for $Z \in \{X, Y\}$ and $\varepsilon \in [0,1]$ fixed.

Then, the observed p-values of the robustified test for (H_0, H_1) can then be computed as a function of the contamination size ε:

$$f(\varepsilon) := 1 - \frac{1}{N} \cdot \sum_{I \in \mathcal{I}_N} \mathbb{1}_{\left\{d_{\mathsf{X},\mathsf{Y}}(\omega) - d_{re} > \frac{2\varepsilon}{(1-\varepsilon)}\right\}}$$

Example: Inferential Benchmarking (NeurIPS 2024)

We use 80 binary classification datasets from OpenML ([Van Rijn et al., 2013]).

Example: Inferential Benchmarking (NeurIPS 2024)

We use 80 binary classification datasets from **OpenML** ([Van Rijn et al., 2013]).

We compare the performance of **Support Vector Machine (SVM)** with

- **Random Forest (RF)**,
- **Decision Tree (CART)**,
- **Logistic Regression (LR)**,
- **Elastic net (GLMNet)**,
- **Extreme Gradient Boosting (xGBoost)**, and
- **k-Nearest Neighbors (kNN)**.

Example: Inferential Benchmarking (NeurIPS 2024)

We use 80 binary classification datasets from **OpenML** ([Van Rijn et al., 2013]).

We compare the performance of **Support Vector Machine (SVM)** with

- **Random Forest (RF)**,
- **Decision Tree (CART)**,
- **Logistic Regression (LR)**,
- **Elastic net (GLMNet)**,
- **Extreme Gradient Boosting (xGBoost)**, and
- **k-Nearest Neighbors (kNN)**.

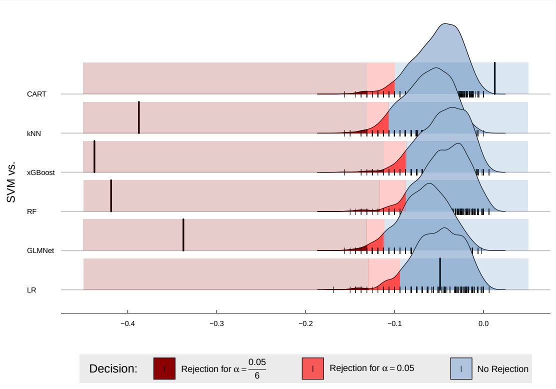
Comparison is based on the multivariate metric Φ composed of

- Accuracy (cardinal)
- Test Computation Time (treated as ordinal)
- Training Computation Time (treated as ordinal)

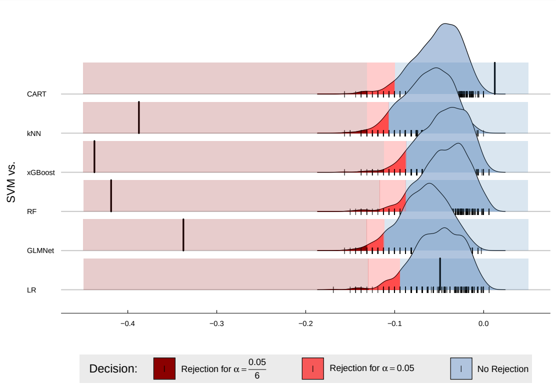
Structure of the Data

data sets / classifier	D_1	$\dots$	D_{80}
RF	$\begin{pmatrix} \phi_1(\text{RF}, D_1) \\ \phi_2(\text{RF}, D_1) \\ \phi_3(\text{RF}, D_1) \end{pmatrix}$	$\dots$	$\begin{pmatrix} \phi_1(\text{RF}, D_{80}) \\ \phi_2(\text{RF}, D_{80}) \\ \phi_3(\text{RF}, D_{80}) \end{pmatrix}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
kNN	$\begin{pmatrix} \phi_1(\text{kNN}, D_1) \\ \phi_2(\text{kNN}, D_1) \\ \phi_3(\text{kNN}, D_1) \end{pmatrix}$	$\dots$	$\begin{pmatrix} \phi_1(\text{kNN}, D_{80}) \\ \phi_2(\text{kNN}, D_{80}) \\ \phi_3(\text{kNN}, D_{80}) \end{pmatrix}$

Example: Inferential Benchmarking



Example: Inferential Benchmarking

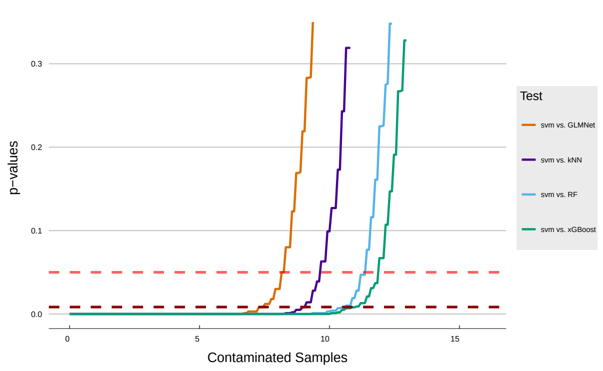


Result: On a 5% level, **SVM** significantly lies in

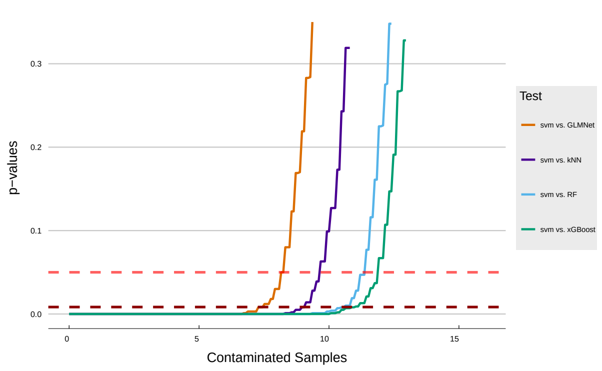
$$\mathrm{GSD}(\mathcal{U}, \mathcal{M})(\{\mathbf{SVM}\ , \mathbf{kNN}\ , \mathbf{xGBoost}\ , \mathbf{RF}\ , \mathbf{GLMNet}\})$$

if quality is measured by trading-off time and accuracy.

Benchmarking Experiments: Robustness



Benchmarking Experiments: Robustness



Result: Our test decision remains valid if up to

$$7 = \min\{7, 8, 11\}$$

data sets from our benchmark suite come from arbitrary distributions.

Thank you for your attention!

I am happy to receive your comments and questions.

[Jansen et al., 2024] [Jansen et al., 2023b] [Jansen et al., 2023a]

References i

Jansen, C., Blocher, H., Augustin, T., and Schollmeyer, G. (2022).
Information efficient learning of complexly structured preferences: Elicitation procedures and their application to decision making under uncertainty.
International Journal of Approximate Reasoning, 144:69–91.

Jansen, C., Nalenz, M., Schollmeyer, G., and Augustin, T. (2023a).
Statistical comparisons of classifiers by generalized stochastic dominance.
Journal of Machine Learning Research, 24(231):1–37.

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

Jansen, C., Schollmeyer, G., Blocher, H., Rodemann, J., and Augustin, T. (2023b).
Robust statistical comparison of random variables with locally varying scale of measurement.
In *Uncertainty in Artificial Intelligence (UAI)*. PMLR.
To appear.

Jansen, C., Schollmeyer, G., Rodemann, J., Blocher, H., and Augustin, T. (2024).
Statistical multicriteria benchmarking via the GSD-front.
In *NeurIPS 2024*, pages 98143–98179.

References iii

Van Rijn, J. N., Bischl, B., Torgo, L., Gao, B., Umaashankar, V., Fischer, S., Winter, P., Wiswedel, B., Berthold, M. R., and Vanschoren, J. (2013).
OpenML: A collaborative science platform.
In *ECML PKDD 2013*, pages 645–649.