Reasoning under severe uncertainty: lecture 1

Sébastien Destercke

CNRS researcher, Laboratoire Heudiasyc, Compiègne

AOS4 master courses

A simple example

How would you model your knowledge about my height?

A simple example

How would you model your knowledge about my height?

About whether I am taller than 1.85m? (A=[1.85,3])

A simple example

How would you model your knowledge about my height?

About whether I am taller than 1.85m? (A=[1.85,3])

How much would you give to play a game where you win 10 euros if *A* is true?

A simple example

How would you model your knowledge about my height?

About whether I am taller than 1.85m? (A=[1.85,3])

How much would you give to play a game where you win 10 euros if A is true?

How much would you give to play a game where you win 10 euros if $B = [2,3]$ is true?

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

Basic modelling

- The state X of the world
  - take values in some (finite or not) set $\mathscr{X}$ of possible situations
  - $\mathscr{X}$ assumed exhaustive and of sufficient granularity
  - is uncertainly known
- How to model our uncertainty about X?
  - by probabilities → why???

Basic definitions

Basic tool

A probability distribution $p : \mathscr{X} \to [0,1]$ such that

- $p(x) \geq 0$
- $\sum_x p(x) = 1$

from which for any subset we have

- $P(A) = \sum_{x \in A} p(x)$
- $P(A) = 1 - P(A^c)$: auto-dual

Example

Academic dice Assume a dice, we have $\mathscr{X} = \{1, 2, \ldots, 6\}$:

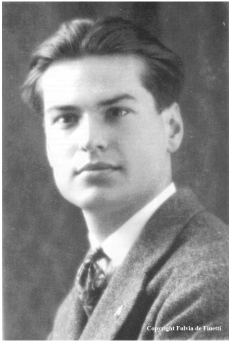
$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = {}^1\!/_6$$

$$P(\{1,3,5\}) = {}^1\!/_6 + {}^1\!/_6 + {}^1\!/_6 = {}^1\!/_2$$

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

Three important guys



L. Savage J. Von Neumann **B. De Finetti**

All justify probabilities (and expected utilities) as uncertainty models without frequencies → we will detail a bit how the second one does it

An example

A gamble/ticket f, whose reward depends on who win the most sets in next Rolland Garros




	Nadal	Ruud	Cilic	Djokovic
f=	-2	10	0	5

What price $P(f)$ do you associate to this ticket?

Acceptable transaction

The price

$$P(f)$$

is the "fair" price you associate to the ticket/gamble f:

- You would buy for any price $P(f) - \epsilon$, earning

$$f - (P(f) - \epsilon)$$

- You would sell for any price $P(f) + \epsilon$, earning

$$(P(f) + \epsilon) - f$$

→ how should a "rational" agent specify prices?

Transaction on an event

Remember the bet on $A = [1.85, 3]$?

Betting on an event A amounts to play the gamble

$$\mathbb{I}_A = \begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{else} \end{cases}$$

We can use A and $\mathbb{I}_A$ interchangeably, i.e.

$$P(\mathbb{I}_A) = P(A)$$

Avoiding the dutch book[1]

- A set of gambles $f_1, \dots, f_n$
- **You** set prices $P(f_1), \dots, P(f_n)$
- I can sell ($\lambda_i > 0$) or buy ($\lambda_i < 0$) to you any number of gambles
- **You** are **irrational** if there is a dutch book, i.e., a combination with

  $$\sup_{x \in \mathscr{X}} \sum \lambda_i \Big(f_i(x) - P(f_i) \Big) < 0,$$

  meaning that whatever happens, you lose money.
- so, a **rational** agent should avoid sure losses when setting prices $P(f_1), \dots, P(f_n)$

---

[1] History unclear

Probabilities and expectations (exercices)

Do the following:

- Prove that if you are rational, then $\inf f \leq P(f) \leq \sup f$
- Prove that if you are rational, then $P(f+g) = P(f) + P(g)$
- Deduce that $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

A little bit more:

- Show that $\sum_{x \in \mathscr{X}} P(\{x\}) = 1$
- Show that $P(f) = \sum_{x \in \mathscr{X}} f(x) P(\{x\})$

Probabilities and expectations (exercices)

Do the following:

- Prove that if you are rational, then $\inf f \leq P(f) \leq \sup f$
- Prove that if you are rational, then $P(f+g) = P(f) + P(g)$
- Deduce that $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

A little bit more:

- Show that $\sum_{x \in \mathscr{X}} P(\{x\}) = 1$
- Show that $P(f) = \sum_{x \in \mathscr{X}} f(x) P(\{x\})$

The first and second properties/axioms are enough to characterize probabilities and expectations.

Wrap-up so far

Subjective probabilities[2]:

- Betting behaviour in terms of fair price reflect (can be used to measure) your knowledge about the world
- If you are rational, those bets should conform with probabilities and expected utilities
- Those bets can be given for all kinds of events, including those that will happen only once

Yet, maybe there is a little more to the story.

---

[2] Often taken as an interpretation for Bayesian approaches

Outline

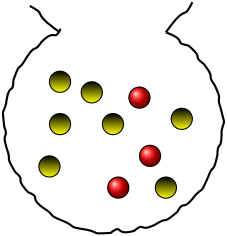
- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
  - Rationality
  - Some axiomatics
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

Experimental protocol

- Half the room goes out
- The rest pick a choice
- We exchange (inside goes outside, and vice-versa)

Urns and balls: case 1

9 balls, 3 are reds, 6 remaining are either yellow or black



What would you choose between A and B?

A

R(ed)	B(lack)	Y(ellow)
100 $	0 $	0$

B

R(ed)	B(lack)	Y(ellow)
0 $	100 $	0$

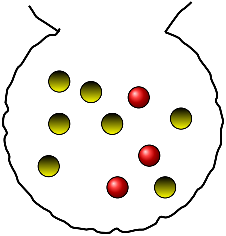
Let us bet together (buying)

- Consider the event *A*="In exactly one year from now in the same place, the outdoor temperature will be colder"
- I have a ticket that pays 100 euros if *A* happens, zero else
- How much are you willing to pay me for this ticket?

Interlude during the change

Urns and balls: case 2

9 balls, 3 are reds, 6 remaining are either yellow or black



What would you choose between C and D?

C

R(ed)	B(lack)	Y(ellow)
100 $	0 $	100$

D

R(ed)	B(lack)	Y(ellow)
0 $	100 $	100$

Let us bet together (selling)

- Consider the event *A*="In exactly one year from now in the same place, the outdoor temperature will be colder"
- I propose the following gamble:
  - I give you some money right now
  - in exchange you have to pay me 100 euros if *A* happens, zero else (you keep the money)
- How much are you willing to pay me for this ticket?

An illustration of a possible use (more latter)

Is it a lioness? a cat? a puma? a bobcat?

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
  - Rationality
  - Some axiomatics
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

Are buying and selling the same?

What if we considered that buying and selling prices for f modelling your knowledge could differ?

- For f, we now consider a maximal buying price $\underline{P}(f)$
- Meaning you would **buy** f for any price under $\underline{P}(f)$
- Any transaction $f-(\underline{P}(f)-\epsilon)$ is acceptable/desirable
- More formally:

$$\underline{P}(f)=\sup\{x\in\mathbb{R}: f-x \text{ is acceptable }\}$$

Why not caring about selling prices?

- $\overline{P}(f)$ is your minimal selling price for f:

$$\overline{P}(f) = \inf\{x \in \mathbb{R} : x - f \text{ is acceptable }\}$$

- Yet, we do have[3]:

$$\begin{aligned}\underline{P}(f) &= \sup\{x \in \mathbb{R} : f - x \text{ is acceptable }\}\\ &= -\inf\{-x \in \mathbb{R} : f - x \text{ is acceptable }\}\\ &= -\inf\{y \in \mathbb{R} : f + y \text{ is acceptable }\}\\ &= -\inf\{y \in \mathbb{R} : y - (-f) \text{ is acceptable }\}\\ &= -\overline{P}(-f)\end{aligned}$$

- By duality, we can only deal with buying prices.

---

[3] Note that it does not imply $\overline{P}(f) = \underline{P}(f)$

Being a rational agent: sure loss revisited

- A set of gambles $f_1, \dots, f_n \in \mathcal{K}$
- **You** set prices $\underline{P}(f_1), \dots, \underline{P}(f_n)$
- I can sell[4] $(\lambda_i > 0)$ to you any number of gambles for these price or lower
- **You** are **irrational** and incur sure loss if there is a combination

$$\sup_{x \in \mathcal{X}} \sum \lambda_i \Big(f_i(x) - \underline{P}(f_i) \Big) < 0, \lambda_i > 0$$

- so, a **rational** agent should avoid sure loss when setting prices $\underline{P}(f_1), \dots, \underline{P}(f_n)$
- It is strictly weaker than previously.

---

[4] But not buy anymore

Back to tennis




	Nadal (a)	Ruud (b)	Cilic (c)	Djokovic (d)
f_i	$\mathbb{I}_{\{a\}}$	$\mathbb{I}_{\{b\}}$	$\mathbb{I}_{\{c\}}$	$\mathbb{I}_{\{d\}}$
$\underline{P}(f_i) =$	0.35	0.2	0.3	0.2

Are those assessments rational? Why?

Being a reasoning agent: natural extension

- Assume prices $\underline{P}(f_i)$ avoid sure loss
- Consider a new gamble/function g
- What can I deduced about $\underline{P}(g)$ from $\underline{P}(f_i)$?
- The process of **natural extension** provides the answer:
  - Knowing that $f_i - \underline{P}(f_i)$ are acceptable
  - Find the highest price $\underline{P}'(g)$ making $g - \underline{P}'(g)$ acceptable
  - This amounts to solve

$$\underline{P}'(g) = \sup_{\alpha \in \mathbb{R}, \lambda_i \geq 0} \{\alpha : g - \alpha \geq \sum_i \lambda_i (f_i - \underline{P}(f_i))\}$$

  - We know $g - \alpha$ acceptable, because $\sum_i \lambda_i (f_i - \underline{P}'(f_i))$ acceptable
- Applying this to f_i itself, I say that prices $\underline{P}(f_i)$ are **coherent** if

$$\underline{P}'(f_i) = \underline{P}(f_i), \quad \forall f_i$$

Tennis again, rational assessments




	Nadal (a)	Ruud (b)	Cilic (c)	Djokovic (d)
f_i	$\mathbb{1}_{\{a\}}$	$\mathbb{1}_{\{b\}}$	$\mathbb{1}_{\{c\}}$	$\mathbb{1}_{\{d\}}$
$\underline{P}(f_i) =$	0.35	0.15	0.2	0.2
f_i	$\mathbb{1}_{\{b,c,d\}}$	$\mathbb{1}_{\{a,c,d\}}$	$\mathbb{1}_{\{a,b,d\}}$	$\mathbb{1}_{\{a,b,c\}}$
$\underline{P}(f_i) =$	0.5	0.7	0.6	0.6

Are those assessments coherent? Why?

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
  - Rationality
  - Some axiomatics
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

A bit of vocabulary

- $\underline{P}(f), \overline{P}(f)$ often called **lower/upper previsions**,
- A rational $\underline{P}(f)$ is said to **avoid sure loss**
- $\underline{P}(f)$ that are deductively closed (= their natural extension) are called **coherent**
- When it is the case and for reasons that will become clear,, $\underline{P}(f), \overline{P}(f)$ also called **lower/upper expectations**
- Similarly, $\underline{P}(\mathbb{I}_A) = \underline{P}(A)$ and $\overline{P}(\mathbb{I}_A) = \overline{P}(A)$ are called **lower/upper probabilities**

Coherence through betting on linear spaces

- assume space $\mathcal{K}$ of gambles is linear

$$g, f \in \mathcal{K} \implies f + g \in \mathcal{K}$$
$$g \in \mathcal{K}, \alpha g \in \mathcal{K} \text{ for } \alpha \geq 0$$

- Then $\underline{P}$ is coherent if and only if

$$\underline{P}(f) \geq \inf f \text{ (sure bet)}$$
$$\underline{P}(\lambda f) = \lambda \underline{P}(f) \text{ (positive homogeneity)}$$
$$\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g) \text{ (super-additivity)}$$

- You get back De Finetti probabilities (a.k.a. linear previsions) if super-additivity becomes additivity

Coherence through desirability

- A gamble f is desirable if $\underline{P}(f) = 0$
- A set $\mathscr{D}$ of desirable gambles is coherent if and only if

$$\text{If } \sup f \leq 0, \text{ then } f \notin \mathscr{D}, \text{ if } f > 0, \text{ then } f \in \mathscr{D}$$

$$\text{If } f, g \in \mathscr{D}, \text{ then } f + g \in \mathscr{D}$$

$$\text{If } f \in \mathscr{D}, \text{ then } \lambda f \in \mathscr{D} \text{ if } \lambda \geq 0$$

- Mathematically, a set $\mathscr{D}$ is coherent if it forms a cone.

Coherence through desirability

- A gamble f is desirable if $\underline{P}(f) = 0$
- A set $\mathscr{D}$ of desirable gambles is coherent if and only if

$$\text{If } \sup f \leq 0, \text{ then } f \notin \mathscr{D}, \text{ if } f > 0, \text{ then } f \in \mathscr{D}$$

$$\text{If } f, g \in \mathscr{D}, \text{ then } f + g \in \mathscr{D}$$

$$\text{If } f \in \mathscr{D}, \text{ then } \lambda f \in \mathscr{D} \text{ if } \lambda \geq 0$$

- Mathematically, a set $\mathscr{D}$ is coherent if it forms a cone.

Coherence through probability sets (we will stick with that)

- We can interpret $\underline{P}(f)$ as a lower bound on expectation for probabilities, i.e.,

$$\underline{P}(f) \leq P(f) = \sum_x p(x)f(x)$$

where p is a probability mass ($\sum p(x) = 1$ and $p(x) \geq 0$).

- Given $f_1, \ldots, f_n$ and $\underline{P}(f_i)$, we can define a set of dominating probabilities (a.k.a. credal sets)

$$\mathscr{M}(\underline{P}) = \{P : P(f) \geq \underline{P}(f)\}$$

- $\underline{P}$ avoids sure loss if and only if $\mathscr{M}(\underline{P}) \neq \emptyset$
- $\underline{P}$ is coherent if and only if for any f_i, we have

$$\underline{P}(f_i) = \inf_{P \in \mathscr{M}(\underline{P})} P(f_i)$$

that is if $\underline{P}$ is the lower envelope of $\mathscr{M}$

Thinking in terms of $\mathscr{M}$

If we start by specifying a set $\mathscr{M}$ of probabilities:

- $\underline{P}(f_i)$ equivalent to provide expectation (linear operator) lower bounds
- Set $\mathscr{D}$ of desirable gambles=set of random variables having positive lower expectation, i.e., $\underline{P}(f_i) = 0$

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

Probabilities

Probability mass on finite space $\mathscr{X} = \{x_1, \dots, x_n\}$ equivalent to a *n* dimensional vector

$$p := (p(x_1), \dots, p(x_n))$$

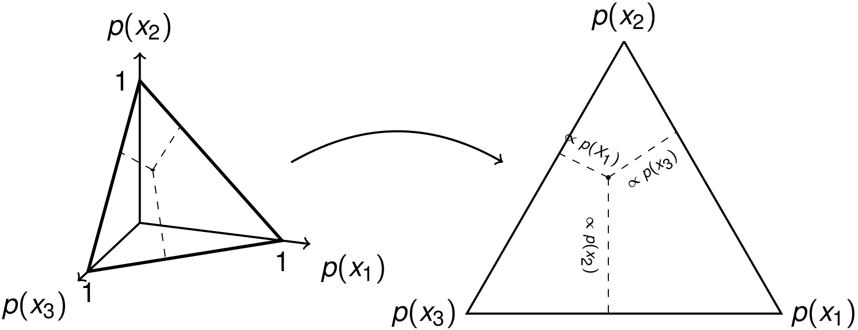
Limited to the set $\mathbb{P}_{\mathscr{X}}$ of all probabilities

$$p(x) > 0, \quad \sum_{x \in \mathscr{X}} p(x) = 1 \quad \text{and}$$

The set $\mathbb{P}_{\mathscr{X}}$ is the $(n-1)$-unit simplex.

Point in unit simplex

$$p(x_1) = 0.2,\ p(x_2) = 0.5,\ p(x_3) = 0.3$$



Imprecise probability

Set $\mathscr{M}$ defined as a set of n constraints

$$\underline{P}(f_i) \leq \sum_{x \in \mathscr{X}} f_i(x)p(x) \leq \overline{P}(f_i)$$

where $f_i : \rightarrow \mathbb{R}$ bounded functions

Example

$$p(x_2) - 2p(x_3) \geq 0$$

$$f(x_1) = 0, f(x_2) = 1, f(x_3) = -2, \underline{P}(a) = 0$$

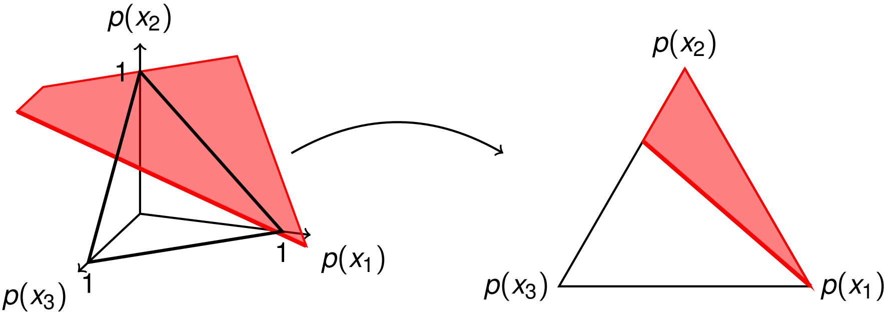
Lower/upper probabilities

Bounds $\underline{P}(A), \overline{P}(A)$ on event A equivalent to

$$\underline{P}(A) \leq \sum_{x \in A} p(x) \leq \overline{P}(A)$$

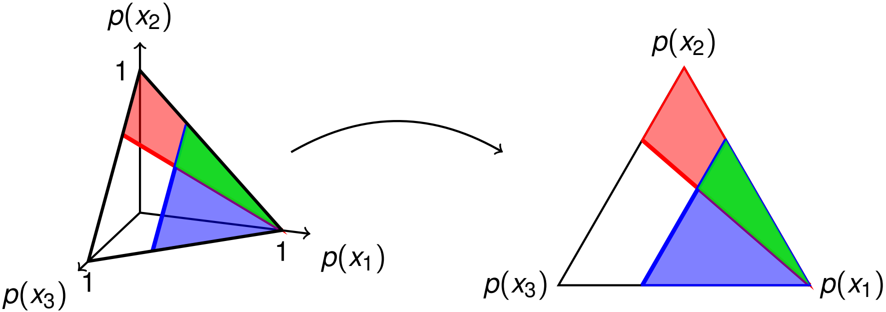
Set $\mathscr{M}$ example

$$p(x_2) \geq 2p(x_3) \Rightarrow p(x_2) - 2p(x_3) \geq 0$$



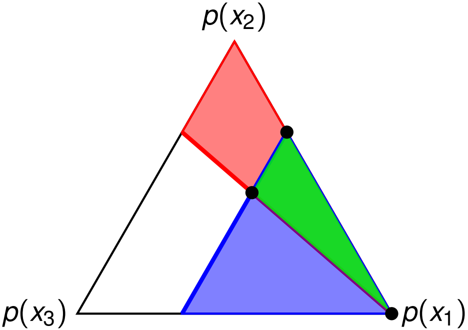
Credal set example

$$p(x_2) - 2p(x_3) \geq 0$$
$$p(x_1) \geq {}^1/_3$$
$$\mathcal{M}$$



Usual alternative presentation: extreme points

- $p(x_1) = 1, p(x_2) = 0, p(x_3) = 0$
- $p(x_1) = 1/3, p(x_2) = 2/3, p(x_3) = 0$
- $p(x_1) = 1/3, p(x_2) = 4/9, p(x_3) = 2/9$



Computing natural extension

- Given $\mathscr{M}$ and a new function g, get

$$\underline{P}(g) = \inf_{P \in \mathscr{M}} P(g) \text{ or } \overline{P}(g) = \sup_{P \in \mathscr{M}} P(g)$$

- First way: linear programming using $\underline{P}(f_i)$

$$\underline{P}(g) = \min_{p(x)} \sum_{x \in \mathscr{X}} p(x)g(x)$$

under

$$\overline{P}(f_i) \geq \sum_{x \in \mathscr{X}} p(x)f_i(x) \geq \underline{P}(f_i)$$

$$\sum_{x \in \mathscr{X}} p(x) = 1, p(x) \geq 0$$

- Second way: compute $\sum_{x \in \mathscr{X}} p(x)g(x)$ for every extreme point, take the minimum

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets

Why looking at special cases?

- Lower previsions/expectations are quite expressive uncertainty models
- Their general use, especially in large spaces, may require heavy computation (linear optimisation in the best case, often more in complex problems[5])
- Just as Gaussian makes probabilistic computations easier, so does focusing on specific lower previsions

---

[5] we will see some in the last courses

A first restriction: lower probabilities

- Lower previsions $\underline{P}(f_i)$ are defined for any function $f_i : \mathscr{X} \to \mathbb{R}$.
- Lower probabilities: focusing on events and considering $\underline{P}(A)$, i.e., restrict the space to $2^{\mathscr{X}}$.
- Upper probabilities are dual[6]:

$$\underline{P}(A) = 1 - \overline{P}(A)$$

- Already include a LOT of models used in practice

---

[6] We can focus on one of the two

A second reduction: 2-monotonicity

A lower probability $\underline{P}()$ is 2-monotone if

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B)$$

- Natural extension/lower expectation of g is given by Choquet integral

  $$\underline{P}(g) = \sum_{i=1}^{N} (g(x_{(i)}) - g(x_{(i-1)}))\underline{P}(\{x_{(i)}, \dots, x_{(N)}\})$$

  with () permutation such that $g(x_{(0)}) = 0, g(x_{(1)}) \leq \dots \leq g(x_{(N)})$
- Generating extreme points is easy. Take a permutation () of $\{1, \dots, N\}$ and compute for each i

  $$p(x_{(i)}) = \underline{P}(\{x_{(i)}, \dots, x_{(N)}\}) - \underline{P}(\{x_{(i+1)}, \dots, x_{(N)}\}),$$

  then p is an extreme point of $\mathscr{M}$

A third reduction: belief functions

A belief function is a lower probability $\underline{P}$ such that for any collection $\mathscr{A} = \{A_1, \ldots, A_K \subseteq \mathscr{X}\}$ with $K \leq 2^{\mathscr{X}}$, we do have

$$\underline{P}(\cup_{A_i \in \mathscr{A}} A_i) \geq \sum_{\mathscr{B} \subseteq \mathscr{A}} (-1)^{|\mathscr{B}|+1} \underline{P}(\cap_{A_i \in \mathscr{B}} A_i),$$

known as the property of complete (or ∞) monotonicity.

Side exercise: prove that a belief function is also 2-monotone[7]

Side bonus: everything we just said also applies to belief function

---

[7] In fact, if $\underline{P}$ is k-monotone, it is also (k-1)-monotone.

An interesting tool: Mobius inverse

The Möbius inverse[8] $m : 2^{\mathscr{X}} \to \mathbb{R}$ of a given $\underline{P}$ is

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \underline{P}(B),$$

and has some interesting properties when applied to belief functions:

- It is bijective with $\underline{P}$ (true for any $\underline{P}$), as for any B

$$\underline{P}(B) = \sum_{A \subseteq B} m(A)$$

- For a new function g, $\underline{P}(g)$ can be computed[9] as

$$\underline{P}(g) = \sum_{A \subseteq \mathscr{X}} m(A) \cdot \inf_{x \in A} g(x)$$

- m is positive (only true for belief functions) → can be seen as a random distribution over subsets → useful tool to simulate $\underline{P}$

---

[8] Apply in fact to general posets

[9] also applies as long as $\underline{P}$ is 2-monotone

Example 1: frequencies of imprecise observations

Imprecise poll: "Who will win the next Wimbledon tournament?"

○ N(adal) ○ F(ederer) ○ D(jokovic) ○ M(urray) ○ O(ther)

60 % replied $\{N,F,D\} \rightarrow m(\{N,F,D\}) = 0.6$

15 % replied "I do not know" $\{N,F,D,M,O\} \rightarrow m(\mathscr{S}) = 0.15$

10 % replied Murray $\{M\} \rightarrow m(\{M\}) = 0.1$

5 % replied others $\{O\} \rightarrow m(\{O\}) = 0.05$

...

Example 2: Imprecise Distributions [4]

A pair $[\underline{F}, \overline{F}]$ of cumulative distributions

Bounds over events $[-\infty, x]$

- Percentiles by experts;
- Kolmogorov-Smirnov bounds;

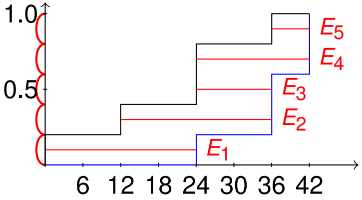
Can be extended to any pre-ordered space [2], [7] ⇒ multivariate spaces!

Expert providing percentiles

$$0 \leq P([-\infty, 12]) \leq 0.2$$

$$0.2 \leq P([-\infty, 24]) \leq 0.4$$

$$0.6 \leq P([-\infty, 36]) \leq 0.8$$



A fourth reduction: possibility measure

A possibility measure is a maxitive upper probability $\overline{P}$:

$$\overline{P}(A \cup B) = \max\{\overline{P}(A), \overline{P}(B)\}$$

This has the following consequences:

- All information is encoded in $\overline{P}(\{x\})$, as

$$\overline{P}(A) = \max_{x \in A} \overline{P}(\{x\})$$

- The associated $\underline{P}$ is a belief function
- The sets receiving positive Möbius mass are nested (form a sequence of included sets)

A simple example

A set E of most plausible values

A confidence degree $\alpha = \underline{P}(E)$

Two interesting cases:

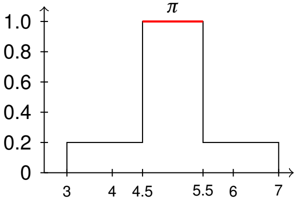
- Expert providing most plausible values E
- E set of models of a formula ϕ

Both cases extend to multiple sets $E_1, \ldots, E_p$:

- confidence degrees over nested sets [5]
- hierarchical knowledge bases [3]

pH value $\in [4.5, 5.5]$ with

$\alpha = 0.8$ ($\sim$ "quite probable")



A simple example

A set E of most plausible values

A confidence degree $\alpha = \underline{P}(E)$

Two interesting cases:

- Expert providing most plausible values E
- E set of models of a formula ϕ

Both cases extend to multiple sets $E_1, \ldots, E_p$:

- confidence degrees over nested sets [5]
- hierarchical knowledge bases [3]

variables p, q

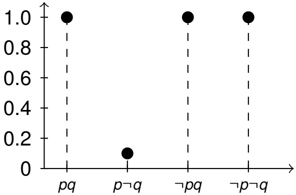
$$\Omega = \{pq, \neg pq, p\neg q, \neg p\neg q\}$$

$$\underline{P}(p \Rightarrow q) = 0.9$$

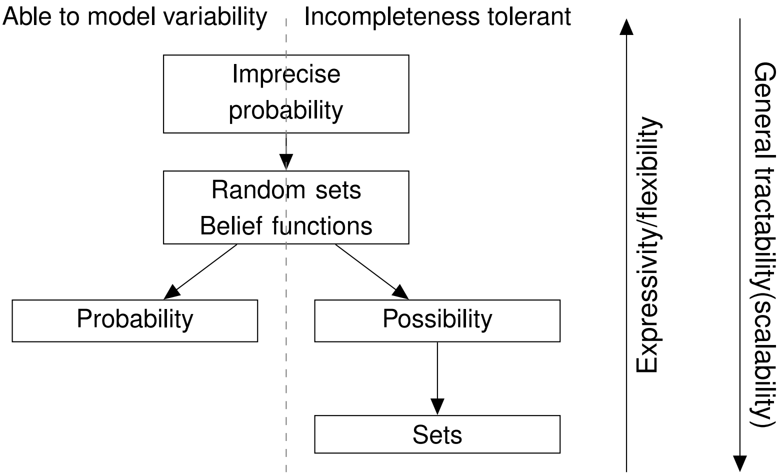
(∼ "almost certain")

$$E = \{pq, p\neg q, \neg p\neg q\}$$

- $\pi(pq) = \pi(p\neg q) = \pi(\neg p\neg q) = 1$
- $\pi(\neg pq) = 0.1$



A quick and incomplete summary



Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets
  - Example
  - Ignorance, complete order
  - Ignorance, partial orders
  - Probability sets with illustration

Decision setting

- Still a set $\mathscr{X}$ of states
- A set $\mathscr{A}$ of actions
- To each action $a : \mathscr{X} \to \mathbb{R}$ corresponds a mapping such that $a(x)$ is the reward/utility of performing a when x is true
- Possibly a set $\mathscr{M}$ modelling our knowledge about X

Decision problem (here): recommend one or multiple actions based on our knowledge about the states in $\mathscr{X}$

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets
  - Example
  - Ignorance, complete order
  - Ignorance, partial orders
  - Probability sets with illustration

An example

We want to cross a sea stretch:

- States: sea weather conditions
- Actions: type of transports

States $\mathscr{X}$

x_1 =Calm sea x_2 =Agitated sea x_3 =Stormy weather



Actions $\mathscr{A}$

a_1 =Motor boat a_2 =Catamaran a_3 =Ferry boat



The matrix $\mathscr{U}$

$$\begin{array}{c|ccc}
 & x_1 & x_2 & x_3 \\
\hline
a_1 & 12 & 0 & -10 \\
a_2 & -2 & 8 & 0 \\
a_3 & 1 & 5 & 10
\end{array}$$

Which action to choose?

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets
  - Example
  - Ignorance, complete order
  - Ignorance, partial orders
  - Probability sets with illustration

Maximin: pessimistic behaviour

- For each action a_i, compute $u_\star(a_i) = \min_j u(a_i, x_j)$
- Say that $a_k \succ_{Mm} a_\ell$ if $u_\star(a_k) > u_\star(a_\ell)$

	x_1	x_2	x_3	$u_\star(a_i)$
a_1	12	0	−10	−10
a_2	−2	8	0	−2
a_3	1	5	10	1
Max				①

- We get $a_3 \succ a_2 \succ a_1$, hence a_3 is recommended
- Pessimistic attitude: best action in the worst case

Maximax: optimistic behaviour

- For each action a_i, compute $u^\star(a_i) = \max_j u(a_i, x_j)$
- Say that $a_k \succ_{MM} a_\ell$ if $u^\star(a_k) > u^\star(a_\ell)$

	x_1	x_2	x_3	$u^\star(a_i)$
a_1	12	0	−10	12
a_2	−2	8	0	8
a_3	1	5	10	10
Max				12

- We get $a_1 \succ a_3 \succ a_2$, hence a_1 is recommended
- Optimistic attitude: best action in the best case

In-between: Hurwicz

- Pick a value $\alpha \in [0,1]$, called optimism index
- For a_i, compute

$$u_{H(\alpha)}(a_i) = \alpha u^{\star}(a_i) + (1-\alpha)u_{\star}(a_k)$$

- Say that $a_k >_{\alpha} a_{\ell}$ if $u_{H(\alpha)}(a_k) > u_{H(\alpha)}(a_{\ell})$

	x_1	x_2	x_3	$u_{\star}(a_i)$	$u^{\star}(a_i)$	$u_{H(0.5)}(a_i)$
a_1	12	0	-10	-10	12	1
a_2	-2	8	0	-2	8	3
a_3	1	5	10	1	10	5.5
Max						5.5

- We get $a_3 > a_2 > a_1$, hence a_3 is recommended
- Try to balance between optimistic and pessimistic

Savage Minimax regret

- For action a_i, compute $R(a_i,x_j) = \max_k u(a_k,x_j) - u(a_i,x_j)$ the regret of picking a_i in x_j, instead of the best possible action
- For a_i, compute $R^\star(a_i) = \max_j R(a_i,x_j)$
- Say that $a_k >_R a_\ell$ if $R^\star(a_\ell) > R^\star(a_k)$

	x_1	x_2	x_3	$R^\star(a_i)$
a_1	12	0	−10	
$R(a_1)$	0			
a_2	−2	8	0	
$R(a_2)$				
a_3	1	5	10	
$R(a_3)$				
Min				

Savage Minimax regret

- For action a_i, compute $R(a_i,x_j) = \max_k u(a_k,x_j) - u(a_i,x_j)$ the regret of picking a_i in x_j, instead of the best possible action
- For a_i, compute $R^\star(a_i) = \max_j R(a_i,x_j)$
- Say that $a_k >_R a_\ell$ if $R^\star(a_\ell) > R^\star(a_k)$

	x_1	x_2	x_3	$R^\star(a_i)$
a_1	12	0	-10	
$R(a_1)$	0	8		
a_2	-2	8	0	
$R(a_2)$				
a_3	1	5	10	
$R(a_3)$				
Min				

Savage Minimax regret

- For action a_i, compute $R(a_i,x_j) = \max_k u(a_k,x_j) - u(a_i,x_j)$ the regret of picking a_i in x_j, instead of the best possible action
- For a_i, compute $R^\star(a_i) = \max_j R(a_i,x_j)$
- Say that $a_k \succ_R a_\ell$ if $R^\star(a_\ell) > R^\star(a_k)$

	x_1	x_2	x_3	$R^\star(a_i)$
a_1	12	0	-10	
$R(a_1)$	0	8	20	
a_2	-2	8	0	
$R(a_2)$				
a_3	1	5	10	
$R(a_3)$				
Min				

Savage Minimax regret

- For action a_i, compute $R(a_i,x_j) = \max_k u(a_k,x_j) - u(a_i,x_j)$ the regret of picking a_i in x_j, instead of the best possible action
- For a_i, compute $R^\star(a_i) = \max_j R(a_i,x_j)$
- Say that $a_k >_R a_\ell$ if $R^\star(a_\ell) > R^\star(a_k)$

	x_1	x_2	x_3	$R^\star(a_i)$
a_1	12	0	-10	
$R(a_1)$	0	8	20	20
a_2	-2	8	0	
$R(a_2)$				
a_3	1	5	10	
$R(a_3)$				
Min				

Savage Minimax regret

- For action a_i, compute $R(a_i,x_j) = \max_k u(a_k,x_j) - u(a_i,x_j)$ the regret of picking a_i in x_j, instead of the best possible action
- For a_i, compute $R^\star(a_i) = \max_j R(a_i,x_j)$
- Say that $a_k \succ_R a_\ell$ if $R^\star(a_\ell) > R^\star(a_k)$

	x_1	x_2	x_3	$R^\star(a_i)$
a_1	12	0	-10	
$R(a_1)$	0	8	20	20
a_2	-2	8	0	
$R(a_2)$	14	0	10	14
a_3	1	5	10	
$R(a_3)$	11	3	0	11
Min				11

- We get $a_3 \succ a_2 \succ a_1$, hence a_3 is recommended
- Minimize regret, but sensitive to addition of non-optimal alternatives

Minimax regret vs maximin

Consider the following case:

	x_1	$\cdots$	x_{99}	x_{100}	$R^\star(a_i)$
a_1	10	$\cdots$	10	1	
$R(a_1)$					
a_2	2	$\cdots$	2	2	
$R(a_2)$					
Min					

Minimax regret and irrelevant alternatives

Before: $a_3 \succ a_2 \succ a_1$

	x_1	x_2	x_3	$R^\star(a_i)$
a_1	12	0	−10	
$R(a_1)$				
a_2	−2	8	0	
$R(a_2)$				
a_3	1	5	10	
$R(a_3)$				
a_4	−5	20	−20	
$R(a_4)$				
Min				

After a_4:

Complete ordering: summary

- Minimax=pessimistic [8]
- Maximax=optimistic
- Hurwicz=in-between [1]
- Savage=Minimizing felt regret [6]

Whatever the chosen rule, we always get one optimal action. But we need to commit to a peculiar behaviour.

What if DM does not want to commit to peculiar behaviour?

What if DM wants to only know the actions that are potentially optimal, given our uncertainty?

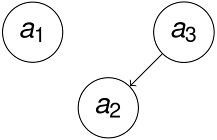
Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets
  - Example
  - Ignorance, complete order
  - Ignorance, partial orders
  - Probability sets with illustration

Lattice ordering

- Say that $a_k \geq_L a_\ell$ if $u^\star(a_k) \geq u^\star(a_\ell)$ and $u_\star(a_k) \geq u_\star(a_\ell)$

	x_1	x_2	x_3	$u_\star(a_i)$	$u^\star(a_i)$
a_1	12	0	−10	−10	12
a_2	−2	8	0	−2	8
a_3	1	5	10	1	10



- Only existing dominance is a_2 by a_3, hence only a_2 is considered non-optimal
- Can be seen as a robust Hurwicz (considering all α as possibilities)
- Note that with this criterion, we eliminate the best action in state x_2

Lattice ordering and information monotonicity

	x_1	x_2	x_3	x_4	$u_\star(a_i)$	$u^\star(a_i)$
a	10	12	14	15	10	15
b	13	11	16	14	11	16

$$b \succ a$$

All states possible

Lattice ordering and information monotonicity

	~~x_1~~	x_2	x_3	x_4	$u_\star(a_i)$	$u^\star(a_i)$
a	~~10~~	12	14	15	12	15
b	~~13~~	11	16	14	11	16

$$b \succ\prec a$$

We learn (gain info) x_1 impossible

a and b becomes incomparable.

Lattice ordering and information monotonicity

	~~x_1~~	x_2	~~x_3~~	x_4	$u_\star(a_i)$	$u^\star(a_i)$
a	~~10~~	12	~~14~~	15	12	15
b	~~13~~	11	~~16~~	14	11	14

$$b \prec a$$

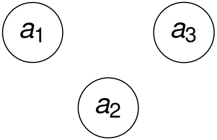
We learn (gain info) x_3 impossible

a is now preferred to b.

Interval dominance

- Say that $a_k >_{ID} a_\ell$ if $u_\star(a_k) > u^\star(a_\ell)$

	x_1	x_2	x_3	$u_\star(a_i)$	$u^\star(a_i)$
a_1	12	0	−10	−10	12
a_2	−2	8	0	−2	8
a_3	1	5	10	1	10

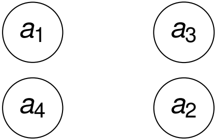


- no dominance at all
- overcautious criterion → may retain Pareto-dominated solutions

Interval dominance: drawback example

- We add a fourth possible, expensive action a_4=Helicopter

	x_1	x_2	x_3	$u_\star(a_i)$	$u^\star(a_i)$
a_1	12	0	-10	-10	12
a_2	-2	8	0	-2	8
a_3	1	5	10	1	10
a_4	8	8	4	4	8

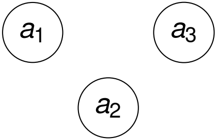


- no dominance at all, even if a_4 better (sometimes strictly) than a_2 in **every** situation!

Difference dominance

- Say that $a_k \geq_D a_\ell$ if $u(a_k, x_j) - u(a_\ell, x_j) \geq 0$ for all x_j ($\succ$ if > 0 for at least one x_j)

	x_1	x_2	x_3
a_1	12	0	-10
a_2	-2	8	0
a_3	1	5	10
$a_2 - a_1$	-14	8	10

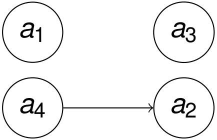


- no dominance at all, again
- do we have the same problem as with interval dominance?

Difference comparison

- We add a fourth possible, expensive action a_4=Helicopter

	x_1	x_2	x_3	$u_\star(a_i)$	$u^\star(a_i)$
a_1	12	0	−10	−10	12
a_2	−2	8	0	−2	8
a_3	1	5	10	1	10
a_4	8	8	4	4	8
$a_4 - a_2$	10	0	4		



So far. . .

Options when true state of the world completely unknown:

- Complete ordering/one top recommendation
  - Maximin: pessimistic DM
  - Maximax: optimistic DM
  - Hurwicz: attempt to in-between
- Partial ordering/multiple recommendations reflecitng lack of knowledge
  - Lattice ordering: robust hurwicz, may miss potentially optimal actions
  - Interval dominance: very conservative, may keep Pareto dominated options
  - Difference dominance: will keep every non-Pareto dominated solution

Outline

- Basics
- Probabilities as bets
- Going beyond betting probabilities: why and how?
- Probability sets, a.k.a. credal sets
- Practical models and computations
- Decision with probability sets
  - Example
  - Ignorance, complete order
  - Ignorance, partial orders
  - Probability sets with illustration

Previous decision rules adaptation

In general, replace u^* by upper expectation $\overline{P}$, u_* by lower expectation $\underline{P}$

Total order

- Maximax: $a \succeq_{MM} b$ if $\overline{P}(a) \geq \overline{P}(b)$
- Maximin: $a \succeq_{Mm} b$ if $\underline{P}(a) \geq \underline{P}(b)$
- Hurwicz: $a \succeq_{\alpha} b$ if $\alpha\overline{P}(a) + (1-\alpha)\underline{P}(a) \geq \alpha\overline{P}(b) + (1-\alpha)\underline{P}(b)$

Partial order

- Interval dominance: $a \succ_{ID} b$ if $\overline{P}(b) \leq \underline{P}(a)$
- Lattice: $a \succ_{L} b$ if $\overline{P}(b) \leq \overline{P}(a) \wedge \underline{P}(b) \leq \underline{P}(a)$
- Difference: $a \succ_{D} b$ if $\underline{P}(a-b) \geq 0$

Difference dominance

Under knowledge $\mathscr{P}$, action a_k is better than a_ℓ if

$$\underline{P}(a_k - a_\ell) = \inf_{p \in \mathscr{P}} P(a_k - a_\ell),$$

that is if in average, we gain something when exchanging a_ℓ for a_k

Special cases

- probabilities ≡ expected utility
- set ≡ difference dominance (filter out Pareto-dominated solutions)

E-admissibility

- Previous rules use orderings between alternatives
- Another way: pick potentially optimal answers
- For a given set $\mathscr{A}$ of actions and a probability p, let

$$Opt(P, \mathscr{A}) = \arg\max_{a \in \mathbb{A}} P(a)$$

- The E-admissible rule returns the set

$$Opt_E(\mathscr{M}, \mathscr{A}) = \cup_{P \in \mathscr{M}} Opt(P, \mathscr{A})$$

Links between rules

Given $\succ_i$, we denote $Opt_{\succ_i}(\mathscr{M},\mathscr{A}) := \{a \in \mathbb{A} : \nexists a' \text{ s.t. } a' \succ_i a\}$ its set of maximal elements.

We have the following relations:

- $a \succeq_{ID} b \implies a \succeq_D b \implies a \succeq_L b \implies a \succeq_\alpha b \quad \forall \alpha$
- $Opt_E(\mathscr{M},\mathscr{A}) \subseteq Opt_{\succ_D}(\mathscr{M},\mathscr{A}) \subseteq Opt_{\succ_{ID}}(\mathscr{M},\mathscr{A})$
- $Opt_{\succ_\alpha}(\mathscr{M},\mathscr{A}) \subseteq Opt_{\succ_L}(\mathscr{M},\mathscr{A}) \subseteq Opt_{\succ_D}(\mathscr{M},\mathscr{A})$

As an exercice, prove the implications of the first line, and the first inclusion of the second (other inclusions immediately follow from implications).

Back to Ellsberg

9 balls, 3 are reds, 6 remaining are either yellow or black

A

R(ed)	B(lack)	Y(ellow)
100 $	0 $	0$

B

R(ed)	B(lack)	Y(ellow)
0 $	100 $	0$

C

R(ed)	B(lack)	Y(ellow)
100 $	0 $	100$

D

R(ed)	B(lack)	Y(ellow)
0 $	100 $	100$

- What are the possible probability values? In terms of bounds over each colour?
- Compute the lower/upper expectations for each act
- What kind of comparison explain the most frequent behaviour $A \geq B$ but $D \geq C$?

Back to Ellsberg

9 balls, 3 are reds, 6 remaining are either yellow or black

A

R(ed)	B(lack)	Y(ellow)
100 $	0 $	0$

B

R(ed)	B(lack)	Y(ellow)
0 $	100 $	0$

C

R(ed)	B(lack)	Y(ellow)
100 $	0 $	100$

D

R(ed)	B(lack)	Y(ellow)
0 $	100 $	100$

Boat example

Agitated is the most likely state ($p(x_2) \geq p(x_1)$ and $p(x_2) \geq p(x_3)$ + $p(x_i) \geq 0$ + $\sum p(x) = 1$). What is the associated credal set?

Boat example

Agitated is the most likely state ($p(x_2) \geq p(x_1)$ and $p(x_2) \geq p(x_3)$ + $p(x_i) \geq 0$ + $\sum p(x) = 1$)

	x_1	x_2	x_3	$\underline{P}(a_i)$	$\overline{P}(a_i)$
a_1	12	0	−10	−5	6
a_2	−2	8	0		
a_3	1	5	10		
a_4	8	8	4		

$$\underline{P}(a_1) = 0 \cdot 12 + 0.5 \cdot 0 + 0.5 \cdot -10 = -5$$

$$\overline{P}(a_1) = 0.5 \cdot 12 + 0.5 \cdot 0 + 0 \cdot -10 = 6$$

Boat example

Agitated is the most likely state ($p(x_2) \geq p(x_1)$ and $p(x_2) \geq p(x_3)$ + $p(x_i) \geq 0 + \sum p(x) = 1$)

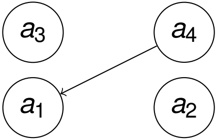
	x_1	x_2	x_3	$\underline{P}(a_i)$	$\overline{P}(a_i)$
a_1	12	0	−10	−5	6
a_2	−2	8	0	2	8
a_3	1	5	10	3	7.5
a_4	8	8	4	6	8

- Maximin: a_4
- Maximax: a_4
- Lattice ordering: $a_4 \succ \{a_2, a_3\} \succ a_1$
- Interval dominance: only $a_4 \succ a_1$ (a_2 still possibly optimal)

Example

Agitated is the most likely state ($p(x_2) \geq p(x_1)$ and $p(x_2) \geq p(x_3)$ + $p(x_i) \geq 0$ + $\sum p(x) = 1$)

	x_1	x_2	x_3
a_1	12	0	−10
a_4	8	8	4
$a_4 - a_1$	−4	8	14



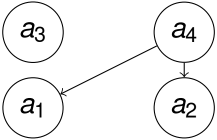
$$\underline{P}(a_4 - a_1) = 0.5 \cdot -4 + 0.5 \cdot 8 + 0 \cdot -6 = 2$$

In the example, difference dominance give $a_4 \succ a_2, a_4 \succ a_1$

Example

Agitated is the most likely state ($p(x_2) \geq p(x_1)$ and $p(x_2) \geq p(x_3)$ + $p(x_i) \geq 0$ + $\sum p(x) = 1$)

	x_1	x_2	x_3
a_2	-2	8	0
a_4	8	8	4
$a_4 - a_2$	6	0	4



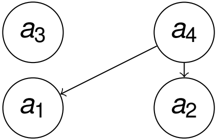
$$\underline{P}(a_4 - a_2) \geq 0 \text{ because of Pareto-dominance}$$

In the example, difference dominance give $a_4 \succ a_2, a_4 \succ a_1$

Example

Agitated is the most likely state ($p(x_2) \geq p(x_1)$ and $p(x_2) \geq p(x_3)$ + $p(x_i) \geq 0$ + $\sum p(x) = 1$)

	x_1	x_2	x_3
a_3	1	5	10
a_4	8	8	4
$a_4 - a_3$	7	3	-6
$a_3 - a_4$	-7	-3	6



$$\underline{P}(a_4 - a_3) = 0 \cdot 7 + 0.5 \cdot 3 + 0.5 \cdot -6 = -1.5 \text{ and } \underline{P}(a_3 - a_4) = -5$$

In the example, difference dominance give $a_4 \succ a_2, a_4 \succ a_1$

References I

[1] K. J. Arrow and L. Hurwicz.
An optimality criterion for decision-making under ignorance.
Uncertainty and expectations in economics, pages 1–11, 1972.

[2] S. Destercke, D. Dubois, and E. Chojnacki.
Unifying practical uncertainty representations: I generalized p-boxes.
Int. J. of Approximate Reasoning, 49:649–663, 2008.

[3] D. Dubois and H. Prade.
Possibilistic logic: a retrospective and prospective view.
Fuzzy Sets and Systems, 144(1):3 – 23, 2004.

[4] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz.
Constructing probability boxes and dempster-shafer structures.
Technical report, Sandia National Laboratories, 2003.

[5] S. Sandri, D. Dubois, and H. Kalfsbeek.
Elicitation, assessment and pooling of expert judgments using possibility theory.
IEEE Trans. on Fuzzy Systems, 3(3):313–335, August 1995.

[6] L. J. Savage.
The theory of statistical decision.
Journal of the American Statistical association, 46(253):55–67, 1951.

[7] M. C. M. Troffaes and S. Destercke.
Probability boxes on totally preordered spaces for multivariate modelling.
Int. J. Approx. Reasoning, 52(6):767–791, 2011.

References II

[8] A. Wald.
Statistical decision functions.
1950.