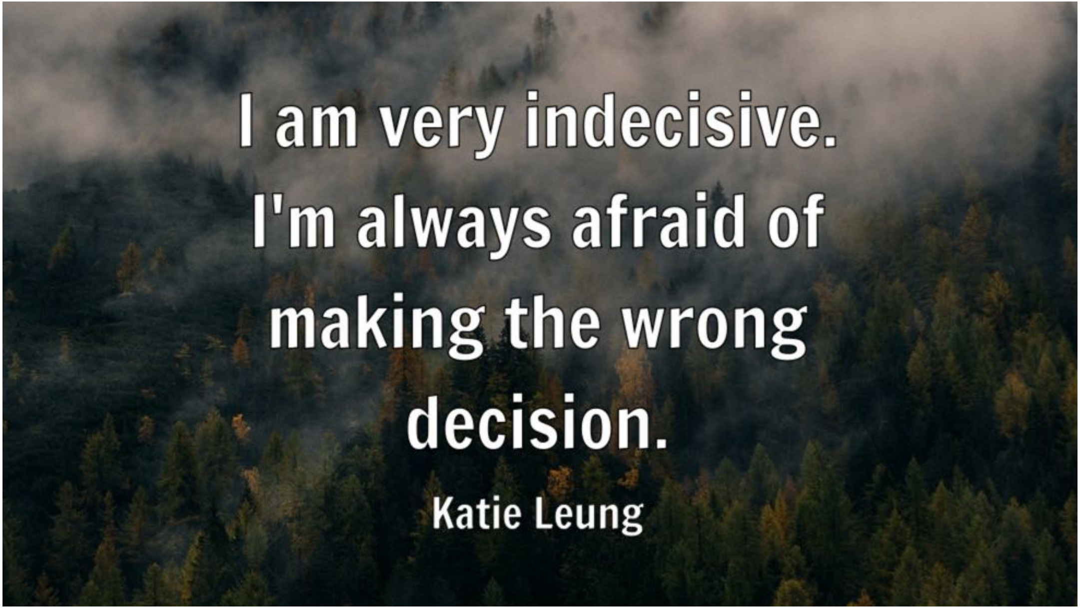
Decisions

Matthias Troffaes (Durham University, UK)

SIPTA Summer School 2024, 12–16 August 2024, Gent, Belgium


I am very indecisive. I'm always afraid of making the wrong decision.
Katie Leung

Outline

- Introduction to Decision Theory (9:00am)
  - Example: Offshore Wind (9:00am)
  - Brief Review of Classical Decision Theory (9:10am)
- Robust Decision Making (9:30am)
  - Aim & Assumptions
  - A Very Simple Example
  - Choice Functions (Break at 10:00am)
- Exercises (10:45am)
- Break (11:15am)
- Credal Classification (11:30pm)
- Exercise: Breast Cancer Case Study (12:00am)
- Lunch (12:30pm)

Outline

https://theconversation.com/
offshore-wind-farms-connected-by-an-underwater-power-grid-for-transmission-could-revolutionize-how-

Offshore wind farms connected by an underwater power grid for transmission could revolutionize how the East Coast gets its electricity

Published: July 31, 2024 1.40pm BST   Updated: August 1, 2024 2.47am BST

Each offshore wind turbine can produce large amounts of power. AP Photo/Michael Dwyer

Offshore winds have the potential to supply coastlines with massive, consistent flows of clean electricity. One study estimates wind farms just offshore could meet 11 times the projected global electricity demand in 2040.

In the U.S., the East Coast is an ideal location to capture this power, but there's a problem: getting electricity from ocean wind farms to the cities and towns that need it.

While everyone wants reliable electricity in their homes and businesses, few support the construction of the transmission lines necessary to get it there. This has always been a problem, both in the U.S. and internationally, but it is becoming an even bigger challenge as countries speed toward net-zero carbon energy systems that will use more electricity.

The U.S. Department of Energy and 10 states in the Northeast States Collaborative on Interregional Transmission are working on a potentially transformative solution: plans for an offshore electric power grid.

https://www.reuters.com/business/energy/japan-collaborate-with-us-cutting-floating-offshore-wind-costs-2024-04-11/

Japan to collaborate with US on cutting floating offshore wind costs

By **Reuters**

April 11, 2024 9:18 AM GMT+1 · Updated 4 months ago

TOKYO, April 11 (Reuters) - Japan has agreed to partner with the United States to help reduce the cost of floating offshore wind projects, the White House said in a statement released during Prime Minister Fumio Kishida's visit to Washington to meet with President Joe Biden.

Under the agreement, Japan and the U.S. would work to accelerate developments in engineering, manufacturing and other areas related to floating wind farms, the statement said. Tokyo would also contribute 120 billion yen ($784 million) to develop floating wind technology via its Green Innovation Fund.

The United States has set a goal of installing 15 gigawatts of floating offshore wind capacity by 2035 - enough to power more than 5 million homes - to help displace fossil fuel for power generation and fight climate change.

The U.S. plan also calls for cutting the cost of floating offshore wind installations operating in deep waters by more than 70% to $45 per megawatt-hour over the next decade. Floating wind power installations are typically bigger and costlier than bottom-fixed structures.

https://nypost.com/2024/07/24/opinion/nantucket-wind-farm-wreck-reveals-the-costs-of-green-energy/

Nantucket wind-farm wreck reveals the true cost of the left's green energy push

By Daniel Turner

Published July 24, 2024, 6:48 p.m. ET

135

Joe Biden's presidency isn't the only thing in freefall this summer.

On July 13, a massive blade from a wind turbine nearly as tall as the Eiffel Tower collapsed into Nantucket Sound, throwing the Massachusetts tourist destination into economic crisis at the height of the summer season and creating a political headache for the green warriors who have championed these costly and unproven boondoggles.

Yet media coverage has been notably subdued, even though miles of the island's famous ocean beaches had to be closed for days due to the dangerous debris.

After all, the Biden-Harris administration has pledged to generate 30 gigawatts of offshore wind by 2030, a national imperative they won't abandon come hell or dirty water.

For comparison, the $4 billion Nantucket wind farm currently collapsing into the Atlantic had hoped to generate a mere 0.8 gigawatts — meaning the administration's wind scheme offers 37 times as much potential economic and environmental disaster.

When the Vineyard Wind project that broke down so horrifically this month was greenlit by Biden's Department of Interior in May 2021, it was billed as "the first large-scale, offshore wind project in the United States."

What is Decision Making: Offshore Wind Example

Operations & Maintenance of Offshore Wind

30% of cost of offshore wind is operations & maintenance
= **huge** chunk of money

Types of Maintenance

- preventive (prevent future failures)
- corrective (fix after failure)

What is Decision Making: Offshore Wind Example

Decisions

criterion: minimize cost

- when to perform maintenance?
- what is a good preventive/corrective balance?

limiting factor = wind speed & wave height for boarding

Uncertainties

Enormous potential for saving costs by making accurate predictions of:

- wind & waves at different time scales
  avoid missing maintenance opportunities
  avoid costly transport when turbine cannot be boarded
- forecast failures before they happen
  cost of preventing $\lll$ cost of fixing

What is Decision Making: Offshore Wind Example

drastically different issues at different time scales:

Short Term: Optimize Actual Operations

what data on the wind farm should we collect
how to use it?

Medium Term: Business Case

how to convince investors to invest in offshore wind
may not have very much data to go from!

Long Term: Policy & Politics

should we encourage offshore, or look at other technologies?
very little data to go by, enormous uncertainty concerning future
climate change, attitude of electorate, etc.
not just about money

What is Decision Making: Offshore Wind Example

Why Use Bounded Probability for Decision Making?

- increases confidence in analysis based on sparse data
  may help at all levels/time horizons
- risk-averse industries: rare events with large impact

Why **NOT** Use Bounded Probability for Decision Making?

- computational expense
- abundant data, non-critical decisions
  standard statistical treatment works as well

Communication!

how to communicate uncertainty?
uncertainty analysis only useful if results can be communicated

Outline

Review of Classical Decision Theory: Example

Example: Visit Offshore Turbine by Boat in the Next Hour?

- parameter: average wave height X for next hour: unknown!
  assume only possible values are $x = 0.5$ and $x = 2$
- data: observation Y, say average wave height in last hour
  assume only possible values are $y = 0.5$ and $y = 2$
- decision: $d =$ take boat, or $d =$ do not take boat
- decision strategy δ:
  which decision to make based on data y?

Review of Classical Decision Theory: Example

Example: Visit Offshore Turbine by Boat in the Next Hour?

- utility function $U(d, x)$: each combination of decision & parameter leads to a different final reward value
  - can only board offshore turbine for maintenance if $X < 1$
  - taking boat costs €1000
  - doing maintenance saves €4000

  for example, expressed in units of €1000:

$U(d, x)$	$x = 0.5$	$x = 2$
$d =$ boat	3	-1
$d =$ no boat	0	0

- likelihood: probability of data given parameter $p(y\|x)$

$p(y\|x)$	$y = 0.5$	$y = 2$
$x = 0.5$	0.9	0.1
$x = 2$	0.3	0.7

- prior: probability of parameter $p(x)$ before you see the data

$p(x)$	$x = 0.5$	$x = 2$
	0.4	0.6

Review of Classical Decision Theory: Example

Frequentist Solution: Wald's Expected Utility, Admissibility

frequentist = use likelihood

1. for every possible strategy δ
   and for every possible value x of X
   calculate Wald's expected utility

   expected utility = −risk

$$U(\delta|x) := E(U(\delta(Y), x)|x) = \sum_y U(\delta(y), x)p(y|x) \quad (1)$$

2. a strategy δ is inadmissible if there is a strategy δ' such that
   $U(\delta'|x) \geq U(\delta|x)$ for all x, and
   $U(\delta'|x) > U(\delta|x)$ for at least one x

   partial ordering of strategies
3. **optimal Wald strategy**
   all admissible strategies

   maximal elements w.r.t. partial ordering

Review of Classical Decision Theory: Example

Bayesian Solution: Maximize Posterior Expected Utility

Bayesian = use posterior ($\propto$ likelihood $\times$ prior)

1. calculate the posterior

$$p(x|y) = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')} \tag{2}$$

2. for every possible observation y
   and every possible decision d
   calculate the posterior expected utility:

$$U(d|y) = E(U(d,X)|y) = \sum_x U(d,x)p(x|y) \tag{3}$$

3. **optimal Bayes strategy** δ^*: max posterior expected utility

$$\delta^*(y) = \arg\max_d U(d|y) \tag{4}$$

much easier to calculate than Wald's admissible strategies! (why?)

Review of Classical Decision Theory: Wald's Theorem

Wald's Theorem (1939 [15])

The set of Wald admissible strategies can always be recovered from a Bayesian analysis, simply by varying the prior over all possible distributions.

[Technical details omitted!]

'equivalence' of $\underbrace{\text{robust Bayesian statistics}}_{\text{sets of priors}}$ and frequentist statistics

Demonstration of Wald's Theorem

`https://colab.research.google.com/github/mcmtroffaes/sipta-school-2024-decision-notebooks/blob/main/01-classical-decision-theory.ipynb`

Plan

- develop decision making directly from sets of distributions
- look at some practical examples

Outline

Outline

Robust Decision Making: Aim & Assumptions

Can we develop a decision theory based on only partial knowledge of probabilities?

Simple setting:

- Set $\mathcal{M}$ of probability mass functions on Ω.
- Consider gambles as functions on Ω (random reward expressed in a utility scale).
- How should we choose among gambles?
- Notation:

$$E_p(X) := \sum_{\omega \in \Omega} p(\omega) X(\omega) \qquad \text{for any } p \in \mathcal{M} \tag{5}$$

$$\underline{E}(X) := \min_{p \in \mathcal{M}} E_p(X) \qquad \overline{E}(X) := \max_{p \in \mathcal{M}} E_p(X) \qquad \text{lower \& upper expectation} \tag{6}$$

Outline

A Very Simple Example

Example (Machinery, Overtime, or Nothing?)

A company makes a product, and believes in increasing future demand. The manager asks you, the decision expert, whether he should buy new machinery, use overtime, or do nothing. The upcoming year, demand can either increase or remain the same.
If we buy new machinery, then the profit at the end of the year will be 440 (in thousands of pounds) if demand increases, and 260 otherwise. Alternatively, if we use overtime, then the profit will be 420 if demand increases, and 300 otherwise. If we do nothing, profit will be 370.
According to our best current judgement, demand will increase with probability at least 0.5, and at most 0.8:

$\mathcal{M} =$

	p_1	p_2
increase	0.5	0.8
stay	0.5	0.2

(each column is a probability mass function)

What advice can we give the manager?

A Very Simple Example: Choice

each row is a gamble

set of decisions	set of gambles: increase	stay
machinery	440	260
overtime	420	300
nothing	370	370

choice function

what is a good choice function, under severe uncertainty?

set of optimal decisions	set of optimal gambles: increase	stay
overtime	420	300
nothing	370	370

Outline

- Introduction to Decision Theory (9:00am)
  - Example: Offshore Wind (9:00am)
  - Brief Review of Classical Decision Theory (9:10am)
- Robust Decision Making (9:30am)
  - Aim & Assumptions
  - A Very Simple Example
  - Choice Functions (Break at 10:00am)
- Exercises (10:45am)
- Break (11:15am)
- Credal Classification (11:30pm)
- Exercise: Breast Cancer Case Study (12:00am)
- Lunch (12:30pm)

Γ-Maximin

(Wald 1945 [16], Gilboa & Schmeidler 1989 [5])

Definition (Γ-Maximin Optimality Criterion)

Choose any gamble whose lower expectation is maximal.

Recipe (Γ-Maximin Optimality Criterion)

1. set up the table with gambles and probabilities
2. calculate the expectation of each gamble
   with respect to each probability mass function — matrix multiplication
3. calculate minimum expectation of each gamble — minimum of each row
4. choose decision with highest minimum expectation — maximize

$$\arg\max_{d \in D} \underline{E}(X_d) \tag{7}$$

Γ-Maximin: Example

Example (Machinery, Overtime, or Nothing)

	increase	stay	p_1	p_2	$\underline{E}$
increase			0.5	0.8	
stay			0.5	0.2	
machinery	440	260			
overtime	420	300			
nothing	370	370			
	(1)		(2)		(3) & (4)

Γ-Maximax

(Satia and Lave 1973 [9], probably others as well)

- Γ-maximin seems overly pessimistic; something more optimistic?

Definition (Γ-Maximax Optimality Criterion)

Choose any gamble whose *upper* expectation is maximal.

Recipe (Γ-Maximax Optimality Criterion)

1. set up the table with gambles and probabilities
2. calculate the expectation of each gamble
   with respect to each probability mass function — matrix multiplication
3. calculate *maximum* expectation of each gamble — maximum of each row
4. choose decision with highest maximum expectation — maximize

$$\arg\max_{d \in D} \overline{E}(X_d) \tag{8}$$

Γ-Maximax: Example

Example (Machinery, Overtime, or Nothing)

	increase	stay	p_1	p_2	$\overline{E}$
increase			0.5	0.8	
stay			0.5	0.2	
machinery	440	260	350	404	
overtime	420	300	360	396	
nothing	370	370	370	370	
	(1)		(2)		(3) & (4)

Break (10:00am)

Interval Maximality (10:15am)

literature: 'interval dominance'

(Condorcet 1785 [4], Sen 1977 [12], Satia and Lave 1973 [9], Kyburg 1983 [6], *many* others)

- get all reasonable options (from pessimistic to optimistic) at once?

Definition (Partial Ordering by Interval Comparison)

We say that a gamble X interval dominates Y, and write

$$X \sqsupset Y \tag{9}$$

whenever

$$\underline{E}(X) > \overline{E}(Y) \tag{10}$$

$[\overline{E}(X), \underline{E}(X)]$ dominates $[\overline{E}(Y), \underline{E}(Y)]$

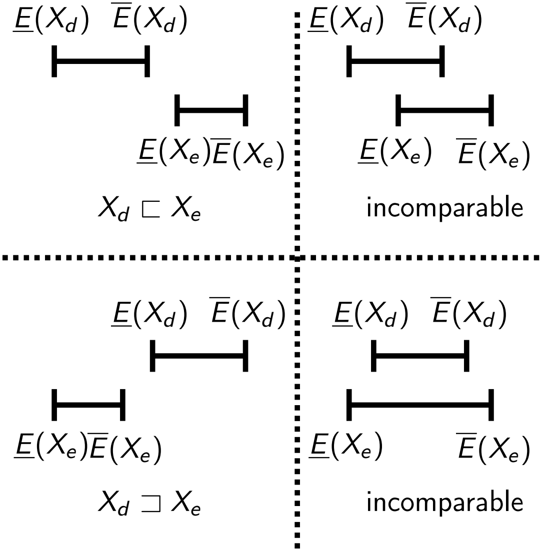
Definition (Interval Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\sqsupset$.

$$\{d \colon (\forall e \in D)(X_d \not\sqsubset X_e)\} \tag{11}$$

Interval Maximality: Partial Ordering

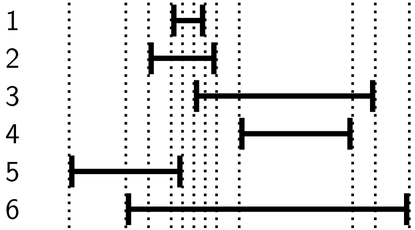
- $\sqsubset$ determines a partial ordering between gambles



Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

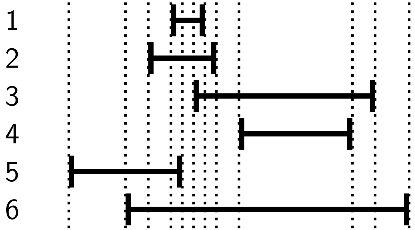
example:



Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:



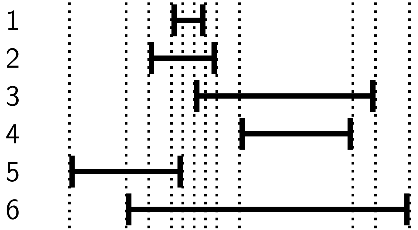
Hasse diagram

1

Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:



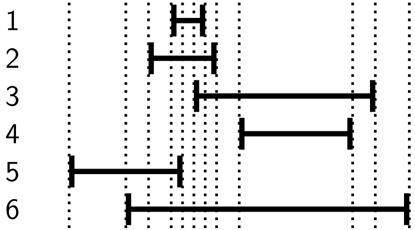
Hasse diagram

1 2

Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:

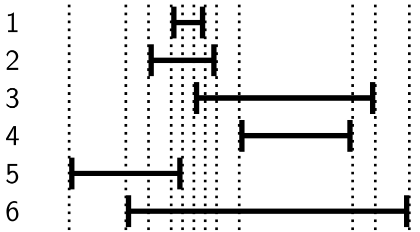


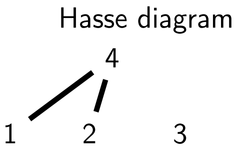
Hasse diagram

Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:

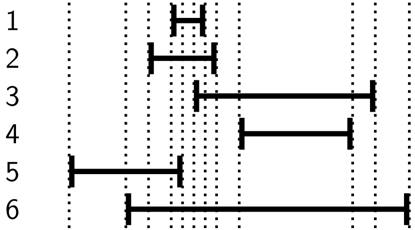


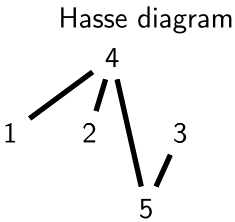


Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:

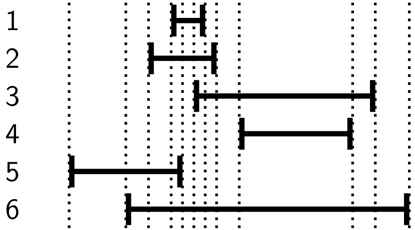




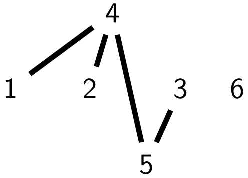
Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:



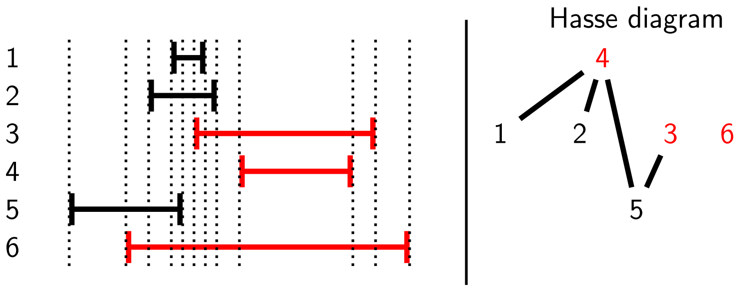
Hasse diagram



Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

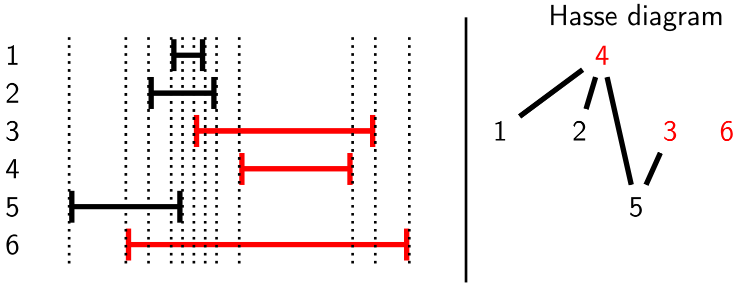
example:



Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = undominated elements

example:



Theorem

All non-interval-maximal elements are dominated by the interval that has the highest lower bound.

$\implies$ no need for Hasse diagram to find interval maximal elements

Interval Maximality: Practical Implementation

Recipe (Interval Maximality Optimality Criterion)

1. set up the table with gambles and probabilities
2. calculate the expectation of each gamble
   with respect to each probability mass function — matrix multiplication
3. calculate *minimum and maximum* expectation of each gamble
   = interval expectation — minimum & maximum of each row
4. choose the decisions whose maximum expectation
   exceeds the overall largest minimum expectation — undominated intervals

$$\left\{d\colon \overline{E}(X_d) \geq \max_{e\in D} \underline{E}(X_e)\right\} \tag{12}$$

Interval Maximality: Example

Example (Machinery, Overtime, or Nothing)

	increase	stay	p_1	p_2	$\underline{E}$	$\overline{E}$
increase			0.5	0.8		
stay			0.5	0.2		
machinery	440	260	350	404		
overtime	420	300	360	396		
nothing	370	370	370	370		
	(1)		(2)		(3) & (4)	

Robust Bayes Maximality

literature: 'maximality'

(Condorcet 1785 [4], Sen 1977 [12], Walley 1991 [17], Seidenfeld 1995 [11])

- exploits the behavioural interpretation of lower previsions
- refines interval maximality (see Exercise 3 later!)

Definition (Partial Ordering by Robust Bayesian Comparison)

We say that X robust Bayes dominates Y, and write

$$X \succ Y \tag{13}$$

whenever any of the following equivalent conditions hold:

$$(\forall p \in \mathcal{M})\,(E_p(X) > E_p(Y)) \tag{14}$$

$$\underline{E}(X - Y) > 0 \tag{15}$$

(willing to pay a small amount in order to trade Y for X)
($X - Y + \epsilon$ is desirable for some $\epsilon > 0$)

Remember, for any probability mass function p and any gamble X: $E_p(X) := \sum_{\omega \in \Omega} p(\omega) X(\omega)$

Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

for browny points: interval maximal gambles?

Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

Hasse diagram

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

1

for browny points: interval maximal gambles?

Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

Hasse diagram

1 2

for browny points: interval maximal gambles?

Robust Bayes Maximality: Hasse Diagram & Algorithm

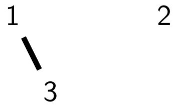
Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

Hasse diagram



for browny points: interval maximal gambles?

Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

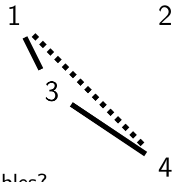
Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

for browny points: interval maximal gambles?

Hasse diagram



Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

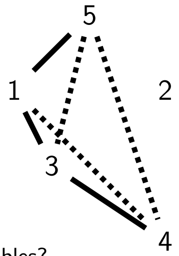
Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

for browny points: interval maximal gambles?

Hasse diagram



Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

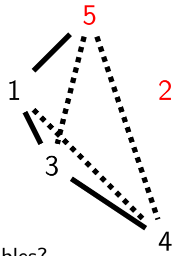
Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

for browny points: interval maximal gambles?

Hasse diagram



Robust Bayes Maximality: Hasse Diagram & Algorithm

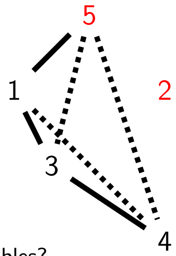
Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

Hasse diagram



for browny points: interval maximal gambles?

Theorem

Every non-maximal element is dominated by a maximal element.

holds for arbitrary partial orderings!

$\Longrightarrow$ no need for Hasse diagram to find maximal elements:
once non-maximal element removed, no need to consider further!

Robust Bayes Maximality: Hasse Diagram & Algorithm

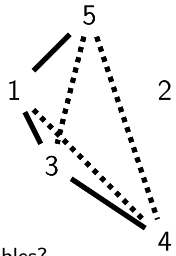
Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

Hasse diagram



Maximality

1

for browny points: interval maximal gambles?

Theorem

Every non-maximal element is dominated by a maximal element.

holds for arbitrary partial orderings!

$\Longrightarrow$ no need for Hasse diagram to find maximal elements:
once non-maximal element removed, no need to consider further!

Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

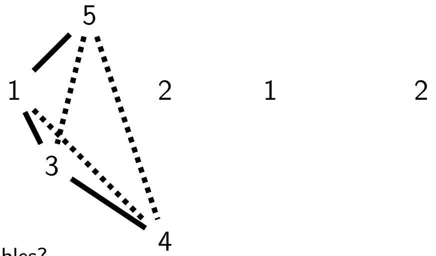
Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

for browny points: interval maximal gambles?

Hasse diagram Maximality



Theorem

Every non-maximal element is dominated by a maximal element.

holds for arbitrary partial orderings!

$\implies$ no need for Hasse diagram to find maximal elements:
once non-maximal element removed, no need to consider further!

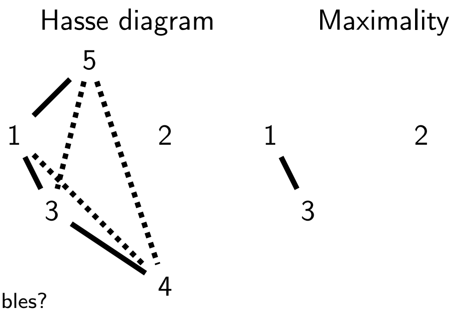
Robust Bayes Maximality: Hasse Diagram & Algorithm

Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_4	0.2	-2	-3
X_5	2	1	-0.5



for browny points: interval maximal gambles?

Theorem

Every non-maximal element is dominated by a maximal element.

holds for arbitrary partial orderings!

$\Longrightarrow$ no need for Hasse diagram to find maximal elements:
once non-maximal element removed, no need to consider further!

Robust Bayes Maximality: Hasse Diagram & Algorithm

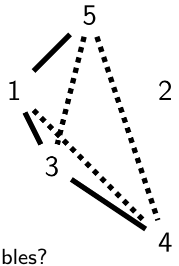
Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

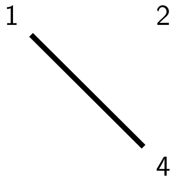
example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_5	2	1	-0.5

Hasse diagram



Maximality



for browny points: interval maximal gambles?

Theorem

Every non-maximal element is dominated by a maximal element.

holds for arbitrary partial orderings!

$\Longrightarrow$ no need for Hasse diagram to find maximal elements:
once non-maximal element removed, no need to consider further!

Robust Bayes Maximality: Hasse Diagram & Algorithm

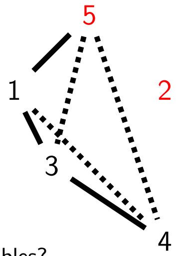
Definition (Robust Bayes Maximality Optimality Criterion)

Choose any gamble which is undominated with respect to $\succ$.

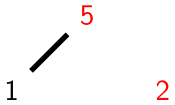
example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_2	0	0	0
X_5	2	1	-0.5

Hasse diagram



Maximality



for browny points: interval maximal gambles?

Theorem

Every non-maximal element is dominated by a maximal element.

holds for arbitrary partial orderings!

$\implies$ no need for Hasse diagram to find maximal elements: once non-maximal element removed, no need to consider further!

Robust Bayes Maximality: Practical Implementation

Recipe (Robust Bayes Maximality Optimality Criterion)

1. set up the table with gambles and probabilities
2. calculate the expectation of each gamble
   with respect to each probability mass function — matrix multiplication
3. sequentially remove all decisions
   whose expectation rows are point-wise dominated

Robust Bayes Maximality: Example

Example (Machinery, Overtime, or Nothing)

	increase	stay	p_1	p_2
increase			0.5	0.8
stay			0.5	0.2
machinery	440	260	350	404
overtime	420	300	360	396
nothing	370	370	370	370
	(1)		(2)	

Robust Bayes Admissibility

literature: 'E-admissibility'

(Pascal 1662 [8], Levi 1980 [7], Berger 1984 [2], Walley 1991 [17], Seidenfeld 2007 [10])

- refines robust Bayes maximality

Definition (Robust Bayes Admissibility Optimality Criterion)

Choose any gamble which maximizes expectation with respect to some $p \in \mathcal{M}$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

notes:

- computational challenge if $\mathcal{M}$ is large
- not invariant under convex hull operation: not enough to look at extreme points

Robust Bayes Admissibility

literature: 'E-admissibility'

(Pascal 1662 [8], Levi 1980 [7], Berger 1984 [2], Walley 1991 [17], Seidenfeld 2007 [10])

- refines robust Bayes maximality

Definition (Robust Bayes Admissibility Optimality Criterion)

Choose any gamble which maximizes expectation with respect to some $p \in \mathcal{M}$.

example:

	E_{p_1}	E_{p_2}	E_{p_3}
X_1	1	0	-1
X_2	0	0	0
X_3	0.5	-1	-2
X_4	0.2	-2	-3
X_5	2	1	-0.5

notes:

- computational challenge if $\mathcal{M}$ is large
- not invariant under convex hull operation: not enough to look at extreme points

Robust Bayes Admissibility: Practical Implementation

Recipe (Robust Bayes Admissibility Optimality Criterion)

1. set up the table with gambles and probabilities
2. calculate the expectation of each gamble
   with respect to each probability mass function — **matrix multiplication**
3. take all decisions that achieve a maximum
   in some expectation column

Robust Bayes Admissibility: Example

Example (Machinery, Overtime, or Nothing)

	increase	stay	p_1	p_2
increase			0.5	0.8
stay			0.5	0.2
machinery	440	260	350	404
overtime	420	300	360	396
nothing	370	370	370	370
	(1)		(2) & (3)	

Robust Bayes Admissibility: Extreme Points Issue

Example (Machinery, Overtime, or Nothing)

	increase	stay	p_1	p_2	p_3
increase			0.5	0.8	0.65
stay			0.5	0.2	0.35
machinery	440	260	350	404	377
overtime	420	300	360	396	378
nothing	370	370	370	370	370

Outline

Exercises

- Visit `https://github.com/mcmtroffaes/sipta-school-2024-decision-notebooks`
- Click on the google colab link in the readme on the bottom of the page.
- Complete any exercises from the first two notebooks (classical decision theory, robust decision making).
- It is advised to make a copy of the notebooks in google so that you can save your work.

Outline

Outline

Credal Classification: What is Classification?

- actual class c (unknown), attributes $a_1, \ldots, a_k$
- decided class d
- $U(d, c)$ utility for deciding class is d if real class is c
  typical choice: $U(d, c) = 1$ if $d = c$ and $U(d, c) = 0$ if $d \neq c$
- aim: choose the best class given attributes

$$d^* = \arg\max_d \sum_c U(d, c) p(c|a) \tag{16}$$

$$= \arg\max_c p(c|a) = \arg\max_c p(c, a)/p(a) \tag{17}$$

$$= \arg\max_c p(c, a) \tag{18}$$

Open issues:

- How do we estimate the probabilities?
- Dealing with scarce data?
- Dealing with missing data?

Credal Classification: The Naive Bayes Classifier

Naive Bayes Classifier

Assume attributes are independent conditional on class:

$$p(c, a) = p(c)p(a|c) = p(c)\prod_{i=1}^{k} p(a_i|c) \tag{19}$$

Estimation of $p(c)$ and $p(a_i|c)$?

- maximum likelihood:

$$p(c) = \frac{n(c)}{N} \qquad p(a_i|c) = \frac{n(a_i,c)}{n(c)} \tag{20}$$

- Bayesian estimate with Dirichlet prior and hyperparameters t:

$$p_t(c) = \frac{n(c)+st(c)}{N+s} \qquad p_t(a_i|c) = \frac{n(a_i,c)+st(a_i,c)}{n(c)+st(c)} \tag{21}$$

(where $\sum_c t(c) = 1$, $\sum_{a_i} t(a_i, c) = t(c)$, $t(c) > 0$, and $t(a_i, c) > 0$)

Credal Classification: The Naive Credal Classifier

Estimation of $p(c)$ and $p(a_i|c)$?

- robust Bayesian estimate with imprecise Dirichlet model:
  as with Bayesian estimate but with
  sensitivity analysis over all possible $t(c)$ and $t(a_i, c)$
- Bounds for probabilities/expected utilities via optimisation.
- Use any of the decision criteria we discussed (interval dominance, robust Bayes maximality, robust Bayes admissibility, ...)

Case that we will study here:

- Simple approximate probability intervals.
- Interval dominance criterion.

Credal Classification: The Naive Credal Classifier

Bounds

$$\underline{p}(c, a) = \inf_t p_t(c, a) = \inf_t \frac{n(c) + st(c)}{N + s} \prod_{i=1}^{k} \frac{n(a_i, c) + st(a_i, c)}{n(c) + st(c)} \tag{22}$$

$$\geq \underbrace{\frac{n(c)}{N + s}}_{\underline{p}(c)} \prod_{i=1}^{k} \underbrace{\frac{n(a_i, c)}{n(c) + s}}_{\underline{p}(a_i|c)} \tag{23}$$

$$\overline{p}(c, a) = \sup_t p_t(c, a) = \sup_t \frac{n(c) + st(c)}{N + s} \prod_{i=1}^{k} \frac{n(a_i, c) + st(a_i, c)}{n(c) + st(c)} \tag{24}$$

$$\leq \underbrace{\frac{n(c) + s}{N + s}}_{\overline{p}(c)} \prod_{i=1}^{k} \underbrace{\frac{n(a_i, c) + s}{n(c) + s}}_{\overline{p}(a_i|c)} \tag{25}$$

Credal Classification: The Naive Credal Classifier

Interval Dominance

Consider the set of all classes c for which

$$\overline{p}(c)\prod_{i=1}^{k}\overline{p}(a_i|c) \geq \max_{c'}\underline{p}(c')\prod_{i=1}^{k}\underline{p}(a_i|c') \tag{26}$$

classifier can return multiple classes if it is unsure about the probabilities!

Credal Classification

Credal Classification: Model Diagnostics

Cross Validation

- divide entire data set into two parts (not necessarily equal in size):
  training data & test data
- use training data to create the model (i.e. lower and upper probabilities)
- classify every item in the test data, and calculate a diagnostic (discrepancy, predictive ability, ...)
- average out the diagnostic

k-Fold Cross Validation

- divide entire data set into k equal parts
- do cross validation k times
  each time using k'th part as test data and remaining parts as training data
- average out diagnostics across all k runs

Credal Classification: Diagnostics [3]

- c = actual class
- $\hat{c}$ = set of predicted classes

name	value	condition
accuracy	1	$c \in \hat{c}$
	0	$c \notin \hat{c}$
single accuracy	1	$\lvert\hat{c}\rvert = 1$ and $\{c\} = \hat{c}$
	0	$\lvert\hat{c}\rvert = 1$ and $\{c\} \neq \hat{c}$
	NA	otherwise
set accuracy	1	$\lvert\hat{c}\rvert \geq 2$ and $c \in \hat{c}$
	0	$\lvert\hat{c}\rvert \geq 2$ and $c \notin \hat{c}$
	NA	otherwise
indeterminate output size	$\lvert\hat{c}\rvert$	$\lvert\hat{c}\rvert \geq 2$
	NA	otherwise
determinacy	1	$\lvert\hat{c}\rvert = 1$
	0	$\lvert\hat{c}\rvert \geq 2$

Outline

Breast Cancer Case Study: Introduction

- data publicly available from `http://archive.ics.uci.edu/ml/datasets/mammographic+mass`
- issue: low predictive power of expert mammogram interpretation (BI-RADS)
- solution: use computer image analysis! can we quantify value of such automation?

Data: 830 patients, 6 attributes

1 expert assessment (BI-RADS): 1 to 6

2 patient age (discretized): 0+, 45+, 55+, or 75+

– image features:

  3 shape: 1 to 4

  4 margin: 1 to 5

  5 density: 1 to 4

6 severity (actual cancer or not): 0 or 1

Breast Cancer Case Study: Exercises

- Visit https://github.com/mcmtroffaes/sipta-school-2024-decision-notebooks
- Click on the google colab link in the readme on the bottom of the page.
- Complete any exercises from the third notebook (credal classification).
- It is advised to make a copy of the notebook in google so that you can save your work.

Outline

References I

[1] Thomas Augustin, Frank P. A. Coolen, Gert De Cooman, and Matthias C. M. Troffaes, editors.
Introduction to Imprecise Probabilities.
Wiley Series in Probability and Statistics. Wiley, 2014.
URL: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470973811.html.

[2] James O. Berger.
The robust Bayesian viewpoint.
In J. B. Kadane, editor, *Robustness of Bayesian Analyses*, pages 63–144. Elsevier Science, Amsterdam, 1984.

[3] Giorgio Corani and Marco Zaffalon.
Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2.
Journal of Machine Learning Research, 9(4):581–621, 2008.

[4] Marquis de Condorcet.
Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix.
L'Imprimerie Royale, Paris, 1785.

[5] Itzhak Gilboa and David Schmeidler.
Maxmin expected utility with non-unique prior.
Journal of Mathematical Economics, 18(2):141–153, 1989.

[6] H. E. Kyburg.
Rational belief.
Technical report, University of Rochester, 1983.

[7] Isaac Levi.
The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability, and Chance.
MIT Press, Cambridge, 1980.

References II

[8] Blaise Pascal.
Pensées.
Maxi-Livres, Paris, 2001.
Unfinished work, published posthumously from collected fragments. First incomplete edition: Port-Royal, 1670. First complete reproduction: Michaut, Basle, 1896.

[9] Jay K. Satia and Jr. Roy E. Lave.
Markovian decision processes with uncertain transition probabilities.
Operations Research, 21(3):728–740, 1973.

[10] Teddy Seidenfeld, Mark Schervish, and Jay Kadane.
Coherent choice functions under uncertainty.
In Gert de Cooman, Jiřina Vejnarová, and Marco Zaffalon, editors, *ISIPTA'07: Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications*, pages 385–394, Prague, July 2007. Charles University, Faculty of Mathematics and Physics.

[11] Teddy Seidenfeld, Mark J. Schervish, and Jay B. Kadane.
A representation of partially ordered preferences.
The Annals of Statistics, 23:2168–2217, 1995.

[12] Amartya Sen.
Social choice theory: A re-examination.
Econometrica, 45(1):53–89, January 1977.

[13] Matthias C. M. Troffaes.
Decision making under uncertainty using imprecise probabilities.
International Journal of Approximate Reasoning, 45(1):17–29, May 2007.
arXiv:1807.03705, doi:10.1016/j.ijar.2006.06.001.

References III

[14] Matthias C. M. Troffaes and Gert de Cooman.
Lower Previsions.
Wiley Series in Probability and Statistics. Wiley, May 2014.
doi:10.1002/9781118762622.

[15] Abraham Wald.
Contributions to the theory of statistical estimation and testing hypotheses.
The Annals of Mathematical Statistics, 10(4):299–326, December 1939.
doi:10.1214/aoms/1177732144.

[16] Abraham Wald.
Statistical decision functions which minimize the maximum risk.
The Annals of Mathematics, 46(2):265–280, 1945.
doi:10.2307/1969022.

[17] Peter Walley.
Statistical Reasoning with Imprecise Probabilities.
Chapman and Hall, London, 1991.