



# Decision Making With Imprecise Probabilities

Matthias Troffaes (Durham)

SIPTA Summer School 2018, 24 – 28 July 2018, Oviedo, Spain

## Outline

### Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

### Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## 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.

### Group Exercise (20 minutes, in groups of 4 or 5)

- (A) Discuss what advice could you give to the manager.
- (B) Discuss what information would allow you to make a better recommendation.
  - ▶ Appoint a group reporter.
  - ▶ 10 minutes for discussion within each group.
  - ▶ 2 minutes for each group to present their answer at the end.

## Outline

### Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

### Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

# What is Decision Making: Offshore Wind Example

## ECOLOGIST

SETTING THE ENVIRONMENTAL AGENDA SINCE 1970

CLIMATE CHANGE | FOOD + FARMING | ENERGY | HEALTH | POLITICS + ECONOMICS | SCIENCE + TECHNOLOGY | NATURAL WORLD | SOCIETY

NEWS | GREEN LIVING | CAMPAIGNING | BLOGS & COMMENTS | INTERVIEWS | MAGAZINE | VIDEO | ARCHIVE | EXTRAS | DIRECTOR

Home | Blogs and Comments | Comment

SEARCH  Go

### Offshore wind is too expensive!

Jim Platts

7th December 2013

Tweet 28 +1 0 Like 35

**Prices paid in the UK to solar and wind generators will change to favour offshore wind at the expense of the others. Jim Platt warns that the policy is doomed to failure - offshore wind is just too expensive, and likely to remain so.**

**“** Building things in the middle of the sea is very, very expensive, and that isn't going to change. **”**

In 2012 the global wind industry manufactured and installed more than 20,000 turbines generating 45GW of energy. The leading firm alone, Danish company Vestas, installed more than 6GW of this via 2,500 turbines.



Offshore wind on the rocks? Photo: Chris Lishman / Shutterstock.com.

More articles about

wind | renewables | energy | oceans |

Related Articles

- **Brazil and FIFA have failed to protect their World Cup mascot**

.....  
- **Ukraine for Youth** - see response to the **IPCC**'s

# What is Decision Making: Offshore Wind Example

## Floating turbines could cut offshore wind energy costs: study

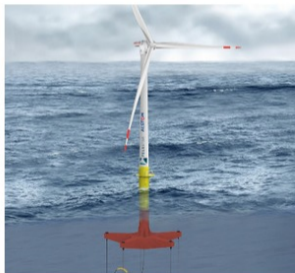
11 July 2014



Print



Email



Floating turbines could cut the cost of offshore wind power to below £85/MWh by the mid-2020s, according to an engineering design study by The Glosten Associates for the Energy Technologies Institute (ETI).

The new study revealed that the company's PelaStar tension leg floating platform (TLP) could deliver further reductions as the technology matures and is designed to provide high capacity factors in wind speeds exceeding 10m per second in water between 60m and 1,200m deep.

The UK is said to have over a third of Europe's potential offshore wind resource, which is enough to meet the power demand of the country nearly three times over.

The FEED study has shown that Glosten's PelaStar TLP design could play a major role in reducing UK offshore wind energy costs.

The company said that the TLP technology is suitable for water depths from as low as 55m up to several hundred metres.

# 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 Imprecise 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 Imprecise 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

### Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

### Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

# 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**  $\delta$ :  
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:

$U(d, x)$	$x = 0.5$	$x = 2$
$d = \text{boat}$	3	-1
$d = \text{no boat}$	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  $\delta$   
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  $\delta$  is **inadmissible** if there is a strategy  $\delta'$  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

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

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

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

3. **optimal Bayes strategy**  $\delta^*$ : **max posterior expected utility**

$$\delta^*(y) = \arg \max_d U(d|y) \quad (4)$$

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



# Review of Classical Decision Theory: Wald's Theorem

## Wald's Theorem (1939 [13])

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 robust Bayesian statistics and frequentist statistics  
sets of priors

## Plan

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

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

**Robust Decision Making (9:50am)**

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

**Robust Decision Making (9:50am)**

**Aim & Assumptions**

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Robust Decision Making: R Code Preparation

1. visit course webpage with browser:  
`http://bellman.ciencias.uniovi.es/~ssipta18/Material.html`
2. download the `improb-redux.r` file
3. start R
4. go to R console
5. clean environment: `rm(list=ls())`
6. "source" the file (File → Source R File)

# 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  $\Omega$ .
- ▶ Consider gambles as functions on  $\Omega$   
(random reward expressed in a utility scale).
- ▶ How should we choose among gambles?

Notation for any gamble  $X: \Omega \rightarrow \mathbb{R}$

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

$$\underline{P}(X) := \min_{p \in \mathcal{M}} E_p(X) \quad \text{lower prevision} \quad (6)$$

$$\overline{P}(X) := \max_{p \in \mathcal{M}} E_p(X) \quad \text{upper prevision} \quad (7)$$

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

**Robust Decision Making (9:50am)**

Aim & Assumptions

**A Very Simple Example**

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## A Very Simple Example

### Example (Machinery, Overtime, or Nothing?)

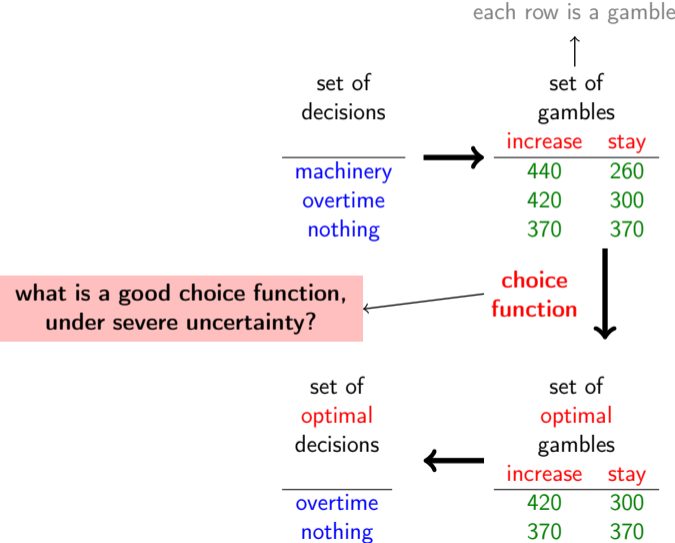
A company makes a product, and believes in increasing future demand. The manager asks you, the decision expert, whether he should buy new machinery, use overtime, or do nothing. The upcoming year, demand can either increase or remain the same.

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:

$$\mathcal{M} = \begin{array}{c|cc} & p_1 & p_2 \\ \hline \text{increase} & 0.5 & 0.8 \\ \text{stay} & 0.5 & 0.2 \end{array} \quad (\text{each column is a probability mass function})$$

What advice can we give the manager?

# A Very Simple Example: Choice





## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

**Robust Decision Making (9:50am)**

Aim & Assumptions

A Very Simple Example

**Choice Functions**

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## $\Gamma$ -Maximin

(Wald 1945 [14], Gilboa & Schmeidler 1989 [5])

### Definition ( $\Gamma$ -Maximin Optimality Criterion)

Choose any gamble whose lower prevision is maximal.

### Recipe ( $\Gamma$ -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
3. calculate minimum expectation of each gamble
4. choose decision with highest minimum expectation

matrix multiplication

minimum of each row

maximize

$$\arg \max_{d \in D} \underline{P}(X_d) \quad (8)$$



## $\Gamma$ -Maximax

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

- ▶  $\Gamma$ -maximin seems overly pessimistic; something more optimistic?

### Definition ( $\Gamma$ -Maximax Optimality Criterion)

Choose any gamble whose *upper* prevision is maximal.

### Recipe ( $\Gamma$ -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
3. calculate *maximum* expectation of each gamble
4. choose decision with highest maximum expectation

matrix multiplication

maximum of each row

maximize

$$\arg \max_{d \in D} \bar{P}(X_d) \quad (9)$$



## Interval Maximality

literature: 'interval dominance'

(Condorcet 1785 [4], Sen 1977 [10], 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

$$X \sqsupset Y \tag{10}$$

whenever

$$\underline{P}(X) > \bar{P}(Y) \tag{11}$$

$[\bar{P}(X), \underline{P}(X)]$  dominates  $[\bar{P}(Y), \underline{P}(Y)]$

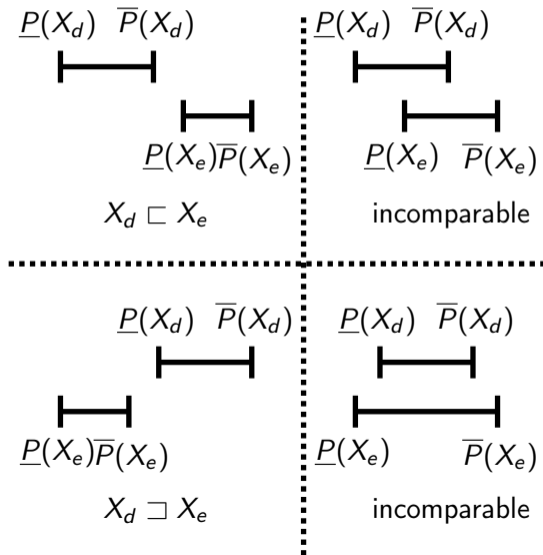
### Definition (Interval Maximality Optimality Criterion)

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

$$\{d : (\forall e \in D)(X_d \not\sqsupset X_e)\} \tag{12}$$

## Interval Maximality: Partial Ordering

□ determines a **partial ordering** between gambles



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

example:

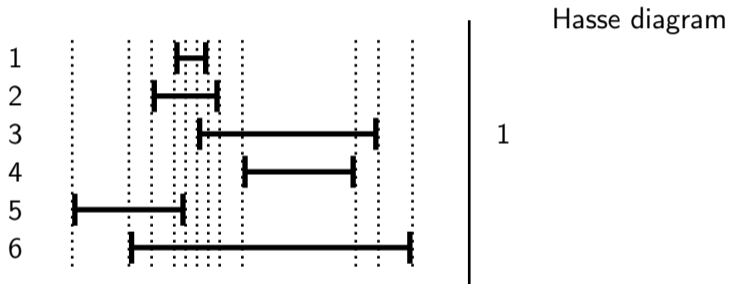




# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

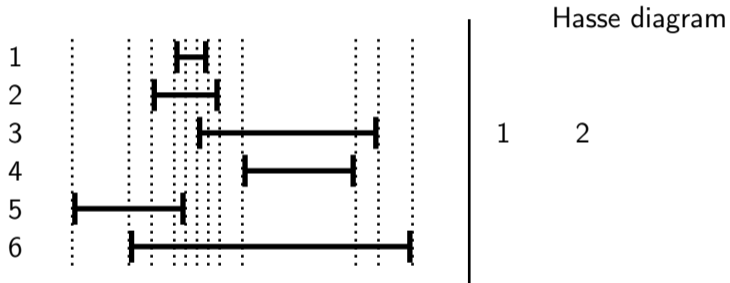
example:



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

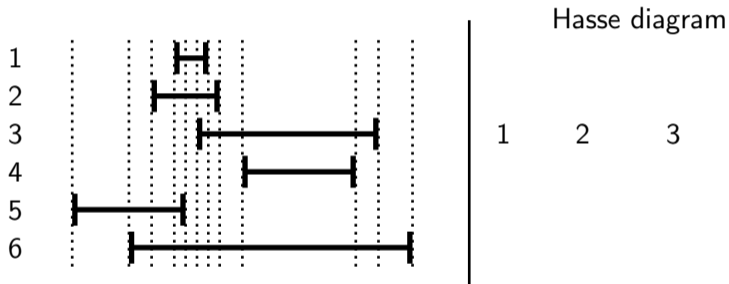
example:



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

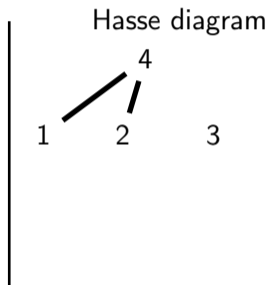
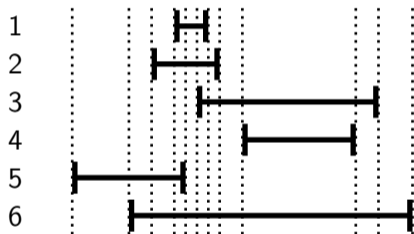
example:



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

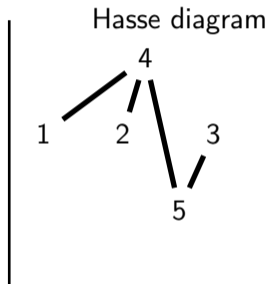
example:



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

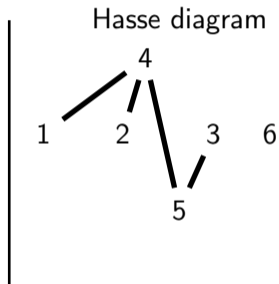
example:



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

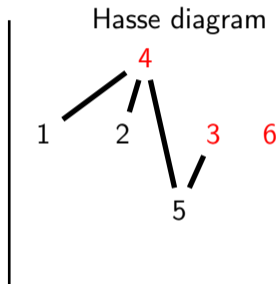
example:



# Interval Maximality: Hasse Diagram & Algorithm

maximal elements with partial ordering = **undominated** elements

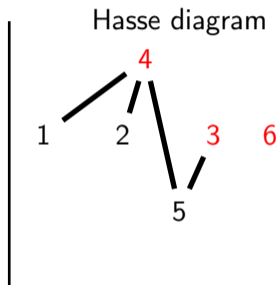
example:



# 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.*

⇒ 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

$$\left\{ d: \bar{P}(X_d) \geq \max_{e \in D} \underline{P}(X_e) \right\} \quad (13)$$

## Interval Maximality: Example

### Example (Machinery, Overtime, or Nothing)

	increase	stay	$p_1$	$p_2$	$\underline{P}$	$\bar{P}$
increase			0.5	0.8		
stay			0.5	0.2		
machinery	440	260				
overtime	420	300				
nothing	370	370				

(1)                                  (2)                                  (3) & (4)

```
pmfs = c(
  0.5, 0.5,
  0.8, 0.2)
rvars = c(
  440, 260,
  420, 300,
  370, 370)
getexpectations = getexpectationsfunc(2, pmfs)
isintervalmaximal = ismaximalfunc(getexpectations, intervalcompare)
isintervalmaximal(rvars)
```

# Robust Bayes Maximality

literature: 'maximality'

(Condorcet 1785 [4], Sen 1977 [10], Walley 1991 [15])

- ▶ 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

$$X \succ Y \tag{14}$$

whenever any of the following equivalent conditions hold:

$$(\forall p \in \mathcal{M}) (E_p(X) > E_p(Y)) \tag{15}$$

$$\underline{P}(X - Y) > 0 \tag{16}$$

(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:

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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 brownly 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

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$	
$X_1$	1	0	-1	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 brownly 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

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

for brownly points: interval maximal gambles?

# Robust Bayes Maximality: Hasse Diagram & Algorithm

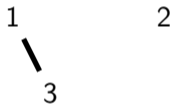
## Definition (Robust Bayes Maximality Optimality Criterion)

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

example:

Hasse diagram

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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 brownly points: interval maximal gambles?

# Robust Bayes Maximality: Hasse Diagram & Algorithm

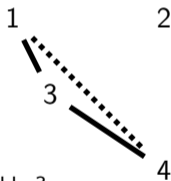
## Definition (Robust Bayes Maximality Optimality Criterion)

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

example:

Hasse diagram

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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 brownly points: interval maximal gambles?



# Robust Bayes Maximality: Hasse Diagram & Algorithm

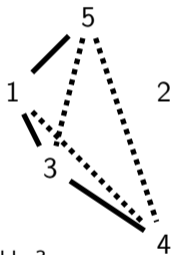
## Definition (Robust Bayes Maximality Optimality Criterion)

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 brownly points: interval maximal gambles?

# Robust Bayes Maximality: Hasse Diagram & Algorithm

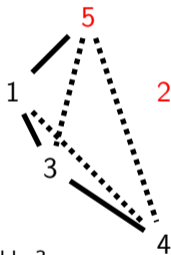
## Definition (Robust Bayes Maximality Optimality Criterion)

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 brownly points: interval maximal gambles?

# Robust Bayes Maximality: Hasse Diagram & Algorithm

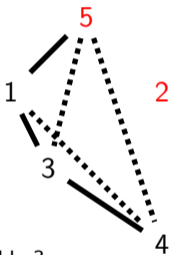
## Definition (Robust Bayes Maximality Optimality Criterion)

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

example:

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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 brown 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!

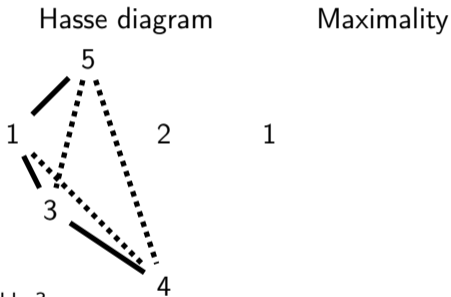
# Robust Bayes Maximality: Hasse Diagram & Algorithm

## Definition (Robust Bayes Maximality Optimality Criterion)

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

example:

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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 brownly 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: Hasse Diagram & Algorithm

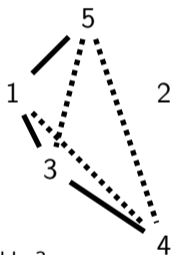
## Definition (Robust Bayes Maximality Optimality Criterion)

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

example:

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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                      2

for brownly 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: Hasse Diagram & Algorithm

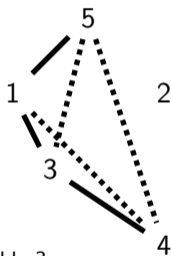
## Definition (Robust Bayes Maximality Optimality Criterion)

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

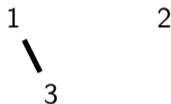
example:

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

Hasse diagram



Maximality



for brown 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: Hasse Diagram & Algorithm

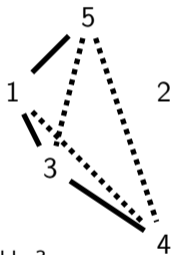
## Definition (Robust Bayes Maximality Optimality Criterion)

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

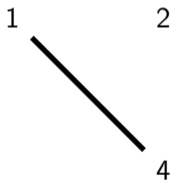
example:

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

Hasse diagram



Maximality



for brown 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!

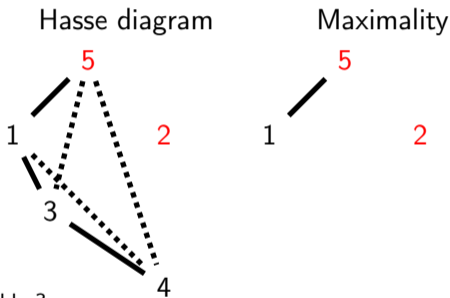
# Robust Bayes Maximality: Hasse Diagram & Algorithm

## Definition (Robust Bayes Maximality Optimality Criterion)

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

example:

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



for brown 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
3. sequentially remove all decisions  
whose expectation rows are point-wise dominated

matrix multiplication

## 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		
overtime	420	300		
nothing	370	370		

(1) (2)

```
pmfs = c(
  0.5, 0.5,
  0.8, 0.2)
rvars = c(
  440, 260,
  420, 300,
  370, 370)
getexpectations = getexpectationsfunc(2, pmfs)
isrbayesmaximal = ismaximalfunc(getexpectations, rbayescompare)
isrbayesmaximal(rvars)
```

# Robust Bayes Admissibility

literature: 'E-admissibility'

(Pascal 1662 [8], Levi 1980 [7], Berger 1984 [2], Walley 1991 [15])

- ▶ 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:

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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 [15])

- ▶ 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:

	$E_{p_1}$	$E_{p_3}$	$E_{p_3}$
$X_1$	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
3. take all decisions that achieve a maximum in some expectation column

matrix multiplication

## 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		
overtime	420	300		
nothing	370	370		

(1)                      (2) & (3)

```
pmfs = c(
  0.5, 0.5,
  0.8, 0.2)
rvars = c(
  440, 260,
  420, 300,
  370, 370)
getexpectations = getexpectationsfunc(2, pmfs)
isrbayesadmissible = isrbayesadmissiblefunc(getexpectations)
isrbayesadmissible(rvars)
```

# 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			
overtime	420	300			
nothing	370	370			

```
pmfs = c(
  0.5, 0.5,
  0.8, 0.2,
  0.65, 0.35)
rvars = c(
  440, 260,
  420, 300,
  370, 370)
getexpectations = getexpectationsfunc(2, pmfs)
isrbayesadmissible = isrbayesadmissiblefunc(getexpectations)
isrbayesadmissible(rvars)
```

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)



## Exercises

- ▶ Consider again the same very simple example. We have done additional market research, and we now know that demand will increase with probability at least 0.6, and at most 0.65. What advice can we give the manager now? Investigate with each optimality criterion.

Hint:  $\mathcal{M} =$

	$p_1$	$p_2$
increase	0.6	0.65
stay	0.4	0.35

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

**Break (11:00am)**

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

**Exercises (11:30am)**

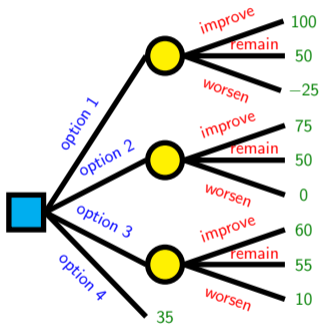
Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Exercises

- You have the option to invest some money. The market can either improve, remain, or worsen. Your set of probabilities are tabulated below. You have the choice between 4 options, summarized in the decision tree below.


$$\mathcal{M} = \begin{array}{c|cc} & p_1 & p_2 \\ \hline \text{improve} & 0.0 & 0.3 \\ \text{remain} & 0.6 & 0.3 \\ \text{worsen} & 0.4 & 0.4 \end{array}$$

Which options should you definitely not consider? First consider interval maximality, then consider robust Bayes maximality. Which of these two criteria gives the better answer?

## A ranking problem

In an environmental problem, three possible decisions can be made. The table below lists the suitability of each of the options, as a 'best estimate', and also giving lower and upper bounds:

option	best estimate	lower bound	upper bound
1	6	5	12
2	10	3	11
3	8	4	10

You may assume that there is a possibility space  $\Omega$  for this problem, and that each option  $i$  induces some gamble  $X_i$  on  $\Omega$ .

## A ranking problem (continued)

- ▶ Assume the best estimate corresponds to the (precise) expectation of  $X_i$ , so for example,  $\underline{P}(X_1) = \overline{P}(X_1) = 6$ . What is the optimal decision according to each of the decision criteria?
- ▶ We are not really sure whether we can interpret the best estimates as precise expectations, so we propose the following lower prevision, where  $E(X_i)$  is the best estimate of  $X_i$ , and  $\beta$  is a parameter between 0 and 1:

$$\underline{P} \left( \alpha_0 + \sum_{i=1}^3 \alpha_i X_i \right) := \alpha_0 + \beta \sum_{i=1}^3 \alpha_i E(X_i) \quad (17)$$

$$+ (1 - \beta) \sum_{i=1}^3 \min\{\alpha_i \min(X_i), \alpha_i \max(X_i)\} \quad (18)$$

for any values of  $\alpha_0, \dots, \alpha_3 \in \mathbb{R}$ . Try to interpret the above formula as well as the  $\beta$  parameter. Identify the optimal decisions for  $\beta = 0$ ,  $\beta = 1$ , and  $\beta = 0.7$  according to  $\Gamma$ -maximin, interval dominance, and maximality.

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Credal Classification: What is Classification?

- ▶ actual class  $c$  (unknown), attributes  $a_1, \dots, a_k$
- ▶ 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

$$d^* = \arg \max_d \sum_c U(d, c) p(c|a) \quad (19)$$

$$= \arg \max_c p(c|a) = \arg \max_c p(c, a)/p(a) \quad (20)$$

$$= \arg \max_c p(c, a) \quad (21)$$

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:

$$p(c, a) = p(c)p(a|c) = p(c) \prod_{i=1}^k p(a_i|c) \quad (22)$$

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

- ▶ maximum likelihood:

$$p(c) = \frac{n(c)}{N} \qquad p(a_i|c) = \frac{n(a_i, c)}{n(c)} \quad (23)$$

- ▶ Bayesian estimate with Dirichlet prior:

$$p(c) = \frac{n(c)+st(c)}{N+s} \qquad p(a_i|c) = \frac{n(a_i, c)+st(a_i, c)}{n(c)+st(c)} \quad (24)$$

(where  $\sum_{a_i} t(a_i, c) = t(c)$ )

# Credal Classification: The Naive Credal Classifier

## Estimation of $p(c)$ and $p(a|c)$ ?

- ▶ 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

$$\underline{p}(c, a) = \inf_t \frac{n(c) + st(c)}{N + s} \prod_{i=1}^k \frac{n(a_i, c) + st(a_i, c)}{n(c) + st(c)} \quad (25)$$

$$\geq \frac{n(c)}{N + s} \prod_{i=1}^k \frac{n(a_i, c)}{n(c) + s} = \underline{p}(c) \prod_{i=1}^k \underline{p}(a_i|c) \quad (26)$$

$$\bar{p}(c, a) = \sup_t \frac{n(c) + st(c)}{N + s} \prod_{i=1}^k \frac{n(a_i, c) + st(a_i, c)}{n(c) + st(c)} \quad (27)$$

$$\leq \frac{n(c) + s}{N + s} \prod_{i=1}^k \frac{n(a_i, c) + s}{n(c) + s} = \bar{p}(c) \prod_{i=1}^k \bar{p}(a_i|c) \quad (28)$$

# Credal Classification: The Naive Credal Classifier

## Interval Dominance

Consider the **set** of all classes  $c$  for which

$$\bar{p}(c) \prod_{i=1}^k \bar{p}(a_i|c) \geq \max_{c'} \underline{p}(c') \prod_{i=1}^k \underline{p}(a_i|c') \quad (29)$$

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

Credal Classification

# Credal Classification: Model Diagnostics

## Cross Diagnostics

- ▶ 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 Diagnostics

- ▶ 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	$c \notin \hat{c}$
determinacy	1	$ \hat{c}  = 1$
	0	$ \hat{c}  \geq 2$
single accuracy	1	$ \hat{c}  = 1$ and $\{c\} = \hat{c}$
	0	$ \hat{c}  = 1$ and $\{c\} \neq \hat{c}$
	NA	otherwise
set accuracy	1	$ \hat{c}  \geq 2$ and $c \in \hat{c}$
	0	$ \hat{c}  \geq 2$ and $c \notin \hat{c}$
	NA	otherwise
indeterminate output size	$ \hat{c} $	$ \hat{c}  \geq 2$
	NA	otherwise

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

## Breast Cancer Example: Introduction

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

### Data: 829 patients, 6 attributes

- 1 expert assessment (BI-RADS): 1 to 6
- image features:
  - 2 shape: 1 to 4
  - 3 margin: 1 to 5
  - 4 density: 1 to 4
- 5 patient age (discretized): 1 to 4
- 6 severity (actual cancer or not): 0 or 1



## Breast Cancer Example: R Code Preparation

1. visit course webpage with browser:  
`http://bellman.ciencias.uniovi.es/~ssipta18/Material.html`
2. download the `classification.r` file
3. start R
4. go to R console
5. clean environment: `rm(list=ls())`
6. "source" the file (File → Source R File)

## Breast Cancer Example: R Code Explanation

- ▶ `classifier.xxx` = returns a list containing two functions
  - ▶ `trainer` = creates a classification model from data, attribute columns, and class column
  - ▶ `tester` = calculates diagnostic(s) from a classification model, and a single data row
- ▶ `classifier.naive2(0)` = standard naive classifier
- ▶ `classifier.naive2(1)` = standard naive classifier with Laplace correction
- ▶ `classifier.credal(2)` = credal classifier with interval dominance and  $s = 2$
- ▶ `kfcv.classifier` = calculates average k-fold diagnostic from data, attribute columns, class column, and classifier

## Breast Cancer Example

```
mammo = getdata()
myclassifier = classifier.naive2(0)
model = myclassifier$trainer(mammo, 1:5, 6)
testrow = mammo[6,]
print(testrow)
print(myclassifier$tester(model, testrow))
testrow = mammo[5,]
print(testrow)
print(myclassifier$tester(model, testrow))

myclassifier = classifier.composed(
  list(classifier.naive2(0),
        classifier.naive2(1),
        classifier.credal(2)))
mammo = getdata()[1:30,]
print(kfcv.classifier(mammo, 1:5, 6, myclassifier))
print(kfcv.classifier(mammo, 2:5, 6, myclassifier))
print(kfcv.classifier(mammo, 1, 6, myclassifier))
mammo = getdata()
print(kfcv.classifier(mammo, 2:5, 1, myclassifier))
```

## Exercises

- ▶ Try to run the code for the credal classifier.
- ▶ What do you observe if you increase the amount of data that is used to train the classifier? Compare the results you get from the traditional classifier with the results you get from the credal classifier.
- ▶ What happens if you increase the  $s$  value of the credal classifier? Confirm your intuition by running the code. Can you suggest an optimal value for  $s$ ?
- ▶ How would the formulas for the credal classifier (based on interval dominance) change if the utilities were not 0–1 valued? [Hint: assume  $U(d, c) \geq 0$ .]
- ▶ Zaffalon's 1999 paper [16] discusses how the problem can be solved using robust Bayes maximality. Try to implement his algorithm in R by modifying the existing code for interval dominance. [This is a hard exercise!]

## Outline

Introduction to Decision Theory (9:00am)

A Very Simple Example (9:00am)

Example: Offshore Wind (9:20am)

Brief Review of Classical Decision Theory (9:30am)

Robust Decision Making (9:50am)

Aim & Assumptions

A Very Simple Example

Choice Functions

Exercises (10:40am)

Break (11:00am)

Exercises (11:30am)

Credal Classification (12:00pm)

Exercise: Breast Cancer Case Study (12:20am)

Lunch (1:00pm)

# 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] Amartya Sen.  
Social choice theory: A re-examination.  
*Econometrica*, 45(1):53–89, January 1977.
- [11] Matthias C. M. Troffaes.  
Decision making under uncertainty using imprecise probabilities.  
*International Journal of Approximate Reasoning*, 45(1):17–29, May 2007.  
doi:10.1016/j.ijar.2006.06.001.
- [12] Matthias C. M. Troffaes and Gert de Cooman.  
*Lower Previsions*.  
Wiley Series in Probability and Statistics. Wiley, 2014.  
URL: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470723777.html>.
- [13] 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.

# References III

- [14] Abraham Wald.  
Statistical decision functions which minimize the maximum risk.  
*The Annals of Mathematics*, 46(2):265–280, 1945.  
doi:10.2307/1969022.
- [15] Peter Walley.  
*Statistical Reasoning with Imprecise Probabilities*.  
Chapman and Hall, London, 1991.
- [16] Marco Zaffalon.  
A credal approach to naive classification.  
In G. de Cooman, F. G. Cozman, S. Moral, and P. Walley, editors, *ISIPTA '99 – Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, pages 405–414, Ghent, 1999. Imprecise Probabilities Project.