Belief functions and uncertainty theories

Sébastien Destercke

Heudiasyc, CNRS Compiegne, France

BFAS 2017 School

Lecture goal/content

What you will find in this talk

- An overview of different "mainstream" uncertainty theories
- Short discussion about comparing informative contents
- Short discussion about conditioning (and partly fusion)
- Short discussion about independence modelling

What you will not find in this talk

- A deep and exhaustive study of a particular topic
- Elements about statistical learning or decision (due to time limitations)

Proposal for ̀practical" session

- Ask additional (prepared) questions about this talk
- Solve exercises shown during this presentation

Plan

1 Introductory elements

2 Uncertainty theories and representation

3 Comparing informative contents

4 Conditioning in uncertainty theories

5 Independence in uncertainty theories

Generic vs singular quantity

A quantity of interest S can be

- **Generic**, when it refers to a population, or a set of situations.

Generic quantity example

The distribution of mother tongue within European population

- **Singular**, when it refers to an individual or a peculiar situation

Singular quantity example

The mother tongue of Thierry

Ontic and epistemic information [2]

An item of information $\mathcal{I}$ possessed by an agent about S can be

- **Ontic**, if it is a faithful, perfect representation of S

Ontic information example

A set S representing the exact set of languages spoken by Thierry

- **Epistemic**, if it is an imperfect representation of S

Epistemic information example

A set E containing the mother tongue of Thierry

Everything is possible

We can have

- **Ontic** information about a **singular** quantity: the hair colour of a suspect; the mother tongue of someone
- **Epistemic** information about a **singular** quantity: the result of the next dice toss; the set of possible mother tongues
- **Ontic** information about a **generic** quantity: the exact distribution of pixel colours in an image
- **Epistemic** information about a **generic** quantity: an interval about the frequency of French persons higher than 1.80 m

Uncertainty definition

Uncertainty: when our information $\mathcal{I}$ does not characterize the quantity of interest S with certainty

$\rightarrow$ **In this view, uncertainty is necessarily epistemic**

Can concern both:

- Singular information
  - items in a data-base, values of some logical variables, time before failure of **a** component
- Generic information
  - parameter values of classifiers/regression models/probability distributions, time before failure of component**s**, truth of a logical sentence ("birds fly")

Uncertainty main origins [1, Ch. 3]

- **Variability** of a population applied to a peculiar, singular situation

Variability example

The result of one dice throw when knowing the probability of each face

- **Imprecision and incompleteness** due to partial information about the quantity S

Imprecision example

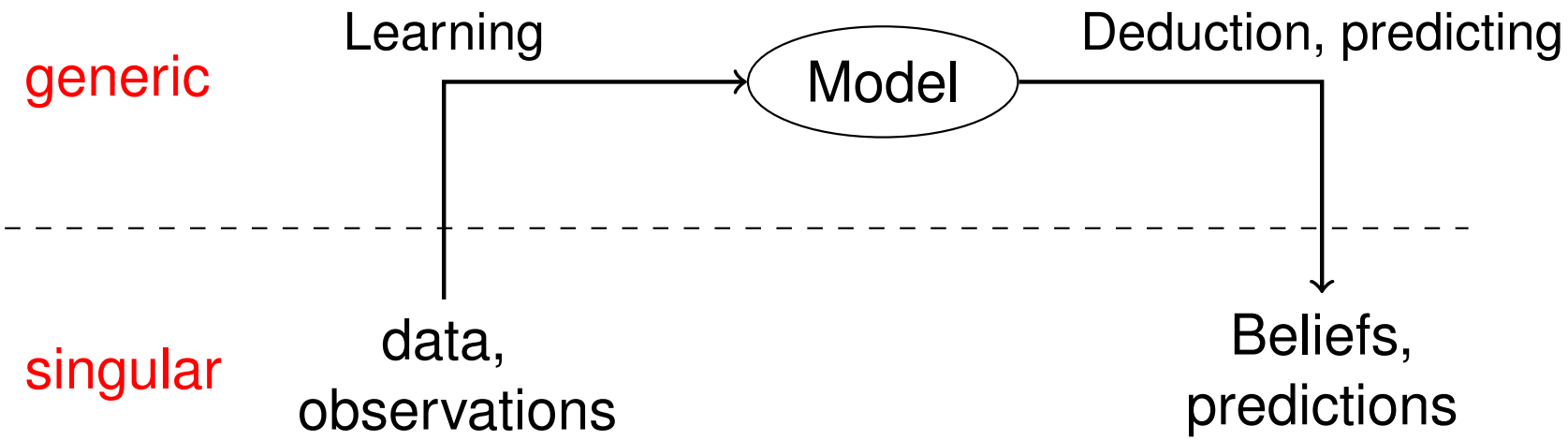
Observing limited sample of the population, describing suspect as "young"

- **Conflict** between different sources of information (data/expert)

Conflict example

Two redundant data base entries with different information for an attribute, two sensors giving different measurements of a quantity

Handling uncertainty



Common problems in one sentence

- **Learning**: use singular information to estimate generic information (induction in logical sense)
- **Prediction**: interrogate model and observations to deduce information on quantity of interest ($\sim$ inference/deduction in logical sense)
- **Information revision**: merge new information with old one
- **Information fusion**: merge multiple information pieces about same quantity

Plan

1 Introductory elements

2 Uncertainty theories and representation

3 Comparing informative contents

4 Conditioning in uncertainty theories

5 Independence in uncertainty theories

Section goals

- Introduce main ideas of theories
- Provide elements about links between them
- Illustrate how to get uncertainty representations within each

Basic framework

Quantity S with possible **exclusive** states $\mathcal{S} = \{s_1, \ldots, s_n\}$

- $\mathcal{S}$: data feature, model parameter, ...

Basic tools

A confidence degree $\mu : 2^{|\mathcal{S}|} \to [0, 1]$ is such that

- $\mu(A)$: confidence $S \in A$
- $\mu(\emptyset) = 0$, $\mu(\mathcal{S}) = 1$
- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

Uncertainty modelled by 2 degrees $\underline{\mu}, \overline{\mu} : 2^{|\mathcal{S}|} \to [0, 1]$:

- $\underline{\mu}(A) \leq \overline{\mu}(A)$ (monotonicity)
- $\underline{\mu}(A) = 1 - \overline{\mu}(A^c)$ (duality)

Probability

Basic tool

A probability distribution $p : \mathcal{S} \to [0, 1]$ from which

- $\underline{\mu}(A) = \overline{\mu}(A) = \mu(A) = \sum_{s \in A} p(s)$
- $\mu(A) = 1 - \mu(A^c)$: auto-dual

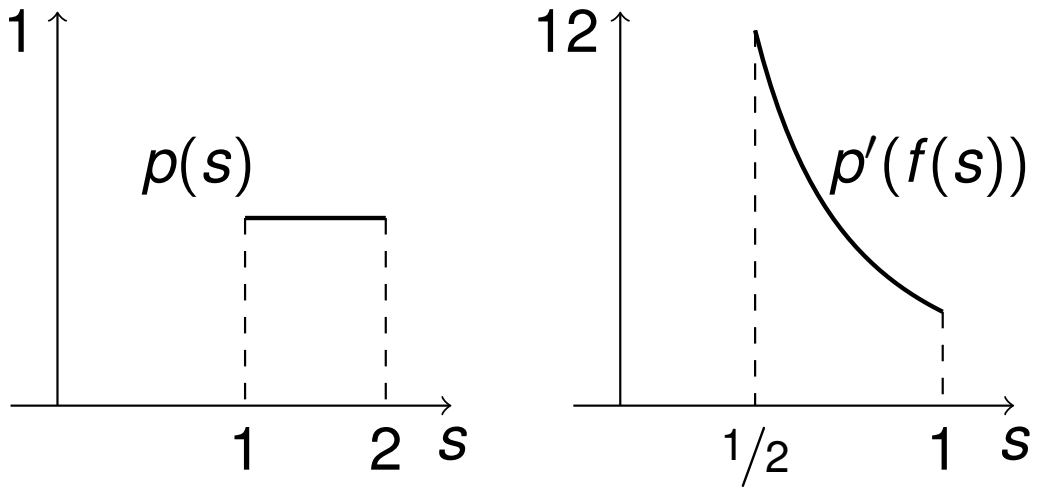
Main interpretations

- **Frequentist [5] :** $\mu(A)$= number of times A observed in a population
  - ▷ only applies to generic quantities (populations)
- **Subjectivist [3] :** $\mu(A)$= price for gamble giving 1 if A happens, 0 if not
  - ▷ applies to both singular and generic quantities

Probability as a model of (partial) ignorance

The assumption ignorance=uniform probability has some issues

- Assume we know nothing about $S \in [1,2]$, then ignorance is $p(s) \sim \mathcal{U}[1,2]$
- Yet, if we consider the variable $1/s$, change of variable induce non-uniform probability over $[1/2, 1]$



$\rightarrow$ "mathematically right", but model of ignorance should be insensitive to variable changes

A discrete example

Assume now that we want to model our ignorance about life in space:

- a first choice is to take $\mathcal{S} = \{Life, NoLife\}$, obtaining

$$p_1(Life) = p_1(NoLife) = 1/2$$

- a second choice is to take $\mathcal{S} = \{VegLife, AnimLife, NoLife\}$, obtaining

$$p_2(VegLife) = p_2(AnimLife) = p_2(NoLife) = 1/3$$

which is different from p_1, or

$$p_2(VegLife) = p_2(AnimLife) = 1/4,\ p_2(NoLife) = 1/2$$

if we start from p_1, which is no longer uniform

$\rightarrow$ Uniformity cannot be maintained through refinement.

$\rightarrow$ In general, probabilistic modelling does not allow to have a model where $P(A)$ (belief in A) is identical for any A.

Sets

Basic tool

A set $E \subseteq \mathcal{S}$ with true value $S \in E$ from which

- $E \subseteq A \rightarrow \underline{\mu}(A) = \overline{\mu}(A) = 1$ (certainty truth in A)
- $E \cap A \neq \emptyset, E \cap A^c \neq \emptyset \rightarrow \underline{\mu}(A) = 0, \overline{\mu}(A) = 1$ (ignorance)
- $E \cap A = \emptyset \rightarrow \underline{\mu}(A) = \overline{\mu}(A) = 0$ (truth cannot be in A)

$\underline{\mu}, \overline{\mu}$ are binary $\rightarrow$ limited expressiveness

Classical use of sets:

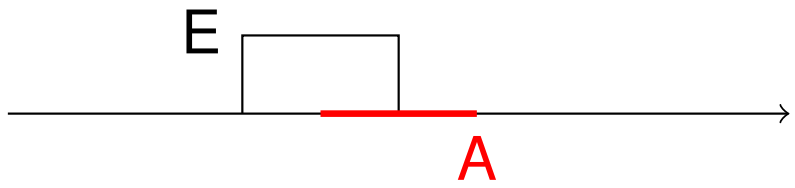
- Interval analysis [4] (E is a subset of $\mathbb{R}$)
- Propositional logic (E is the set of models of a KB)

Other cases: robust optimisation, decision under risk, ...

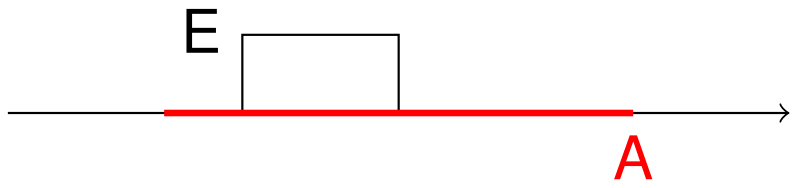
Example

Assume that it is known that pH value $E \in [4.5, 5.5]$, then

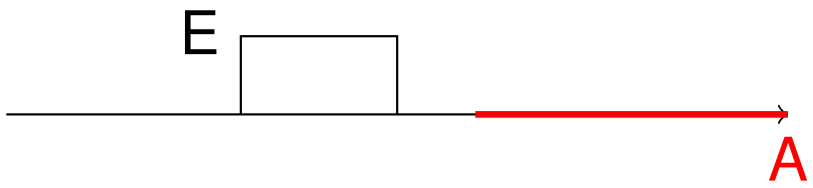
- if $A = [5, 6]$, then $\underline{\mu}(A) = 0, \overline{\mu}(A) = 1$



- if $A = [4, 7]$, then $\underline{\mu}(A) = \overline{\mu}(A) = 1$



- if $A = [6, 9]$, then $\underline{\mu}(A) = \overline{\mu}(A) = 0$



In summary

Probabilities . . .

- (+) very informative quantification (do we need it?)
- (-) need lots of information (do we have it?)
- (-) if not enough, requires a choice (do we want to do that?)
- use probabilistic calculus (convolution, stoch. independence, . . .)

Sets . . .

- (+) need very few information
- (-) very rough quantification of uncertainty (Is it sufficient for us?)
- use set calculus (interval analysis, Cartesian product, . . .)

$\rightarrow$ Need for **frameworks bridging these two**

Possibility theory

Basic tool

A distribution $\pi : \mathcal{S} \to [0, 1]$, usually with s_i such that $\pi(s_i) = 1$, from which

- $\overline{\mu}(A) = \max_{s \in A} \pi(s)$ (Possibility measure)
- $\underline{\mu}(A) = 1 - \overline{\mu}(A^c) = \min_{s \in A^c}(1 - \pi(s))$ (Necessity measure)

Sets E captured by $\pi(s) = 1$ if $s \in E$, 0 otherwise

Interval/set as special case

The set E can be modelled by the possibility distribution π_E such that

$$\pi_E(s) = \begin{cases} 1 & \text{if } s \in E \\ 0 & \text{else} \end{cases}$$

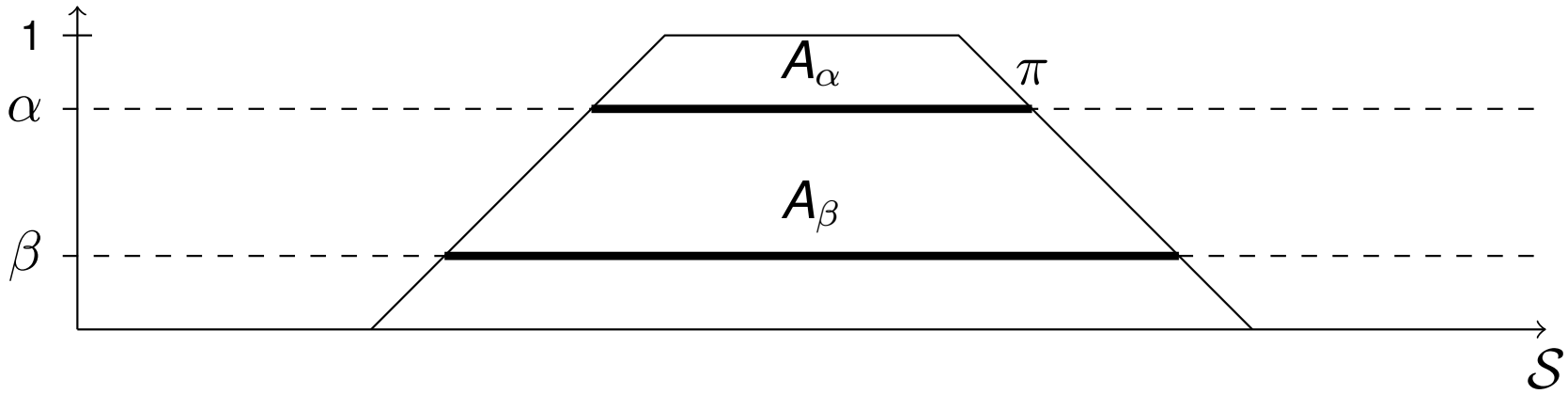
A nice characteristic: Alpha-cut [7]

Definition

$$A_\alpha = \{s \in \mathcal{S} | \pi(s) \geq \alpha\}$$

- $\underline{\mu}(A_\alpha) = 1 - \alpha$
- If $\beta \leq \alpha$, $A_\alpha \subseteq A_\beta$

Simulation: draw $\alpha \in [0, 1]$ and associate A_α



$\Rightarrow$ Possibilistic approach ideal to model **nested structures**

Two interpretations

$[\underline{\mu}, \overline{\mu}]$ as

Subjective confidence degrees (**possibility theory**) [10]

- $\underline{\mu}(A)$ measures how much it is certain that the true value is in A, $\overline{\mu}(A)$ how much it is plausible.
- $\underline{\mu}$ is not, in general, interpreted as an imprecise description of a precise measure μ
- The underlying quantity of interest is often of singular nature (has a unique, true value)

Bounds of an **ill-known probability** $\mu \Rightarrow \underline{\mu} \leq \mu \leq \overline{\mu}$ [11]

- $[\underline{\mu}, \overline{\mu}]$ is an imprecise description of a precise measure μ
- The underlying quantity of interest is assumed to be generic

A basic distribution: simple support

A set E of most plausible values

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

Two interesting cases:

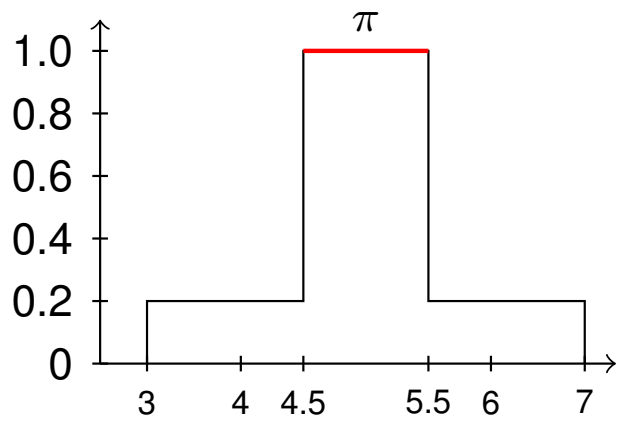
- Expert providing most plausible values E
- E set of models of a formula ϕ

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

- confidence degrees over nested sets [39]
- hierarchical knowledge bases [35]

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

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



A basic distribution: simple support

A set E of most plausible values

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

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 [39]
- hierarchical knowledge bases [35]

variables p, q

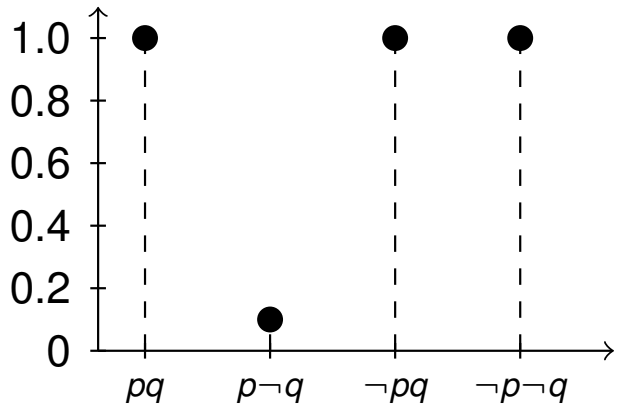
$$\Omega = \{pq, \neg pq, p\neg q, \neg p\neg q\}$$

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

($\sim$ "almost certain")

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

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

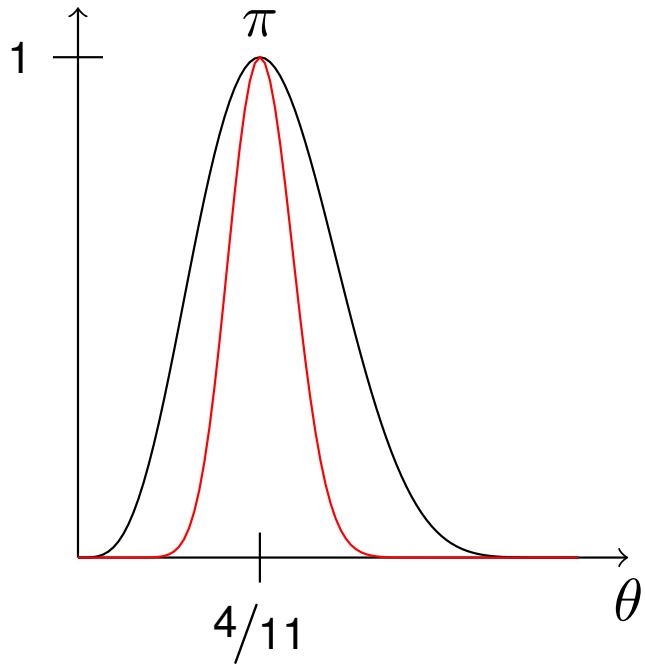


Normalized likelihood as possibilities [9] [28]

$$\pi(\theta) = \mathcal{L}(\theta|x) / \max_{\theta \in \Theta} \mathcal{L}(\theta|x)$$

Binomial situation:

- $\theta =$ success probability
- x number of observed successes
- x= 4 succ. out of 11
- x= 20 succ. out of 55



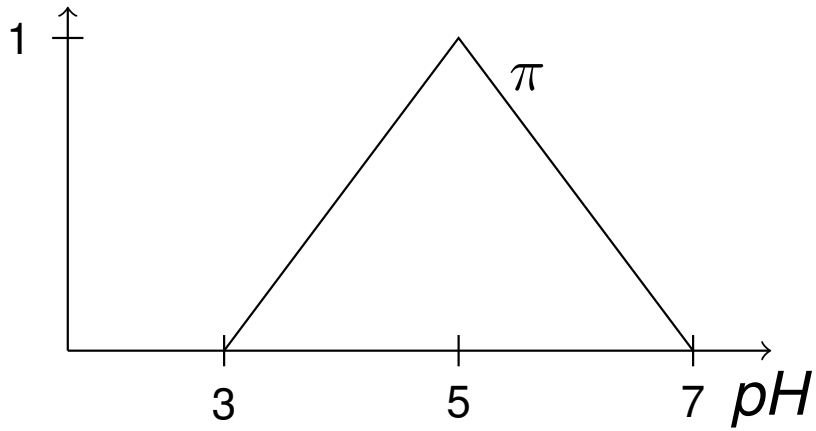
Partially specified probabilities [27] [34]

Triangular distribution: $[\underline{\mu}, \overline{\mu}]$ encompass all probabilities with

- mode/reference value M
- support domain $[a, b]$.

Getting back to *pH*

- $M = 5$
- $[a, b] = [3, 7]$



Other examples

- Statistical inequalities (e.g., Chebyshev inequality) [34]
- Linguistic information (fuzzy sets) [30]
- Approaches based on nested models

Possibility: limitations

$$\underline{\mu}(A) > 0 \Rightarrow \overline{\mu}(A) = 1$$
$$\overline{\mu}(A) < 1 \Rightarrow \underline{\mu}(A) = 0$$

$\Rightarrow$ interval $[\underline{\mu}(A), \overline{\mu}(A)]$ with one trivial bound

Does not include probabilities as special case:

$\Rightarrow$ possibility and probability at odds

$\Rightarrow$ respective calculus hard (sometimes impossible?) to reconcile

Going beyond

Extend the theory

$\Rightarrow$ by complementing π with a lower distribution δ ($\delta \leq \pi$) [12], [33]

$\Rightarrow$ by working with interval-valued possibility/necessity degrees [6]

$\Rightarrow$ by working with sets of possibility measures [13]

Use a more general model

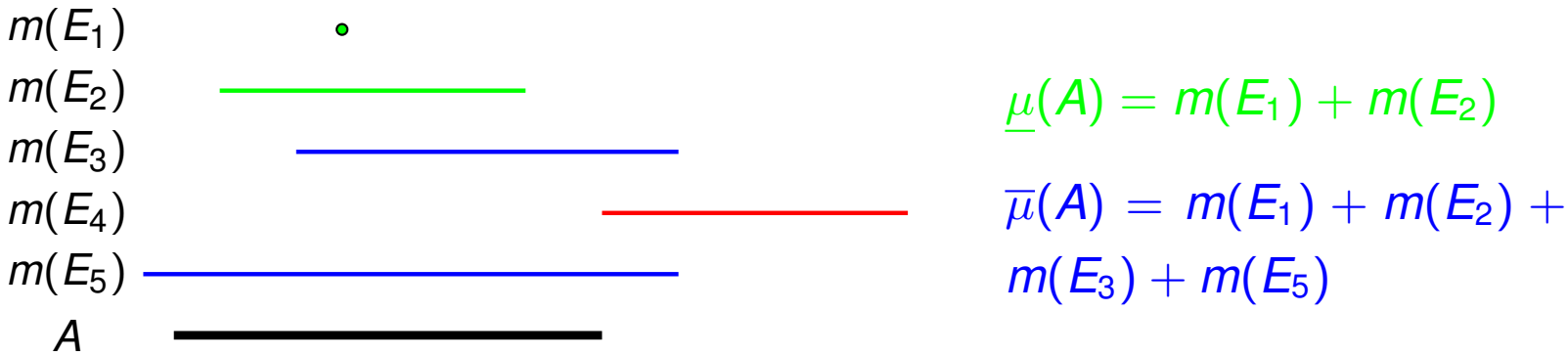
$\Rightarrow$ Random sets and belief functions

Random sets and belief functions

Basic tool

A positive distribution $m : 2^{|\mathcal{S}|} \to [0,1]$, with $\sum_E m(E) = 1$ and usually $m(\emptyset) = 0$, from which

- $\overline{\mu}(A) = \sum_{E \cap A \neq \emptyset} m(E)$ (Plausibility measure)
- $\underline{\mu}(A) = \sum_{E \subseteq A} m(E) = 1 - \overline{\mu}(A^c)$ (Belief measure)



$[\underline{\mu}, \overline{\mu}]$ as

- subjective confidence degrees of **evidence theory** [18], [19], [16]
- bounds of an **ill-known probability** $\mu \Rightarrow \underline{\mu} \leq \mu \leq \overline{\mu}$

A characteristic of belief functions

Complete monotonicity

If $\underline{\mu}$ is a belief measure if and only if it satisfies the inequality

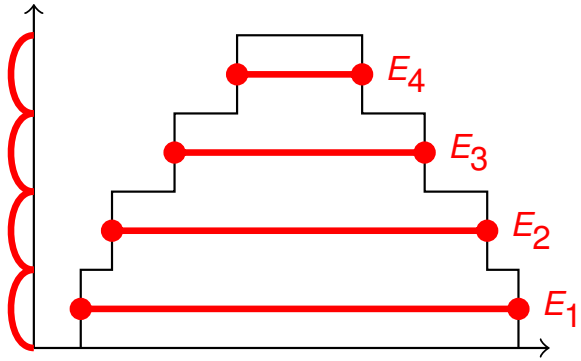
$$\underline{\mu}(\cup_{i=1}^{n} A_i) \geq \sum_{\mathcal{A} \subseteq \{A_1,\ldots,A_n\}} (-1)^{|\mathcal{A}|+1} \underline{\mu}(\cap_{A_i \in \mathcal{A}} A_i)$$

Simply the exclusion/inclusion principle with an equality

special cases

Measures $[\underline{\mu}, \overline{\mu}]$ include:

- Probability distributions: mass on atoms/singletons
- Possibility distributions: mass on nested sets



$\rightarrow$ "simplest" theory that includes both sets and probabilities as special cases!

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" $\{N, F, D, M, O\} \rightarrow m(\mathcal{S}) = 0.15$

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

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

...

P-box [36]

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

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

- Percentiles by experts;
- Kolmogorov-Smirnov bounds;

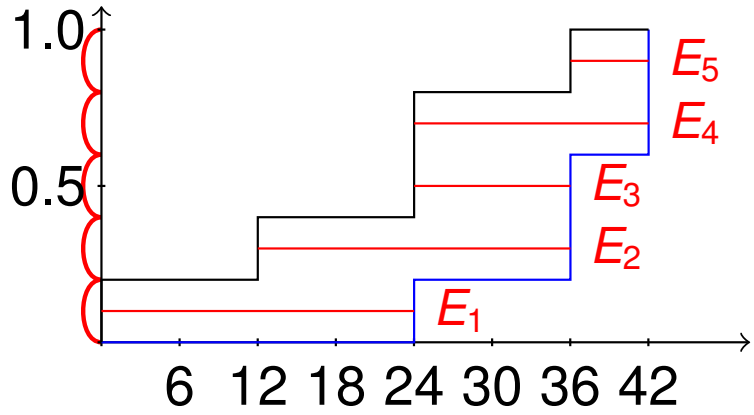
Can be extended to any pre-ordered space [32], [40] $\Rightarrow$ multivariate spaces!

Expert providing percentiles

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

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

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

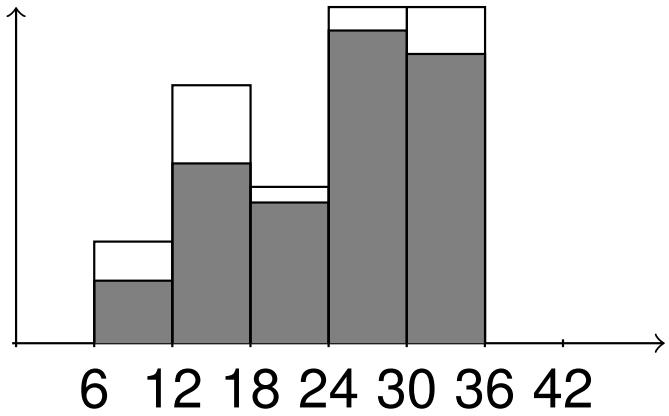


Other means to get random sets/belief functions

- Extending modal logic: probability of provability [20]
- Parameter estimation using pivotal quantities [17]
- Statistical confidence regions [31]
- Modify source information by its reliability [38]
- ...

Limits of random sets

- Not yet satisfactory extension of Bayesian/subjective approach
- Still some items of information it cannot model in a simple way, e.g.,
  - probabilistic bounds over atoms s_i (imprecise histograms, ...) [29] ;
  - comparative assessments such as $2P(B) \leq P(A)$ [37]



Imprecise probabilities

Basic tool

A set $\mathcal{P}$ of probabilities on $\mathcal{S}$ or an equivalent representation

- $\overline{\mu}(A) = \sup_{P \in \mathcal{P}} P(A)$ (Upper probability)
- $\underline{\mu}(A) = \inf_{P \in \mathcal{P}} P(A) = 1 - \overline{\mu}(A^c)$ (Lower probability)

Note: lower/upper bounds on events alone cannot model any convex $\mathcal{P}$

$[\underline{\mu}, \overline{\mu}]$ as

- subjective lower and upper betting rates [25]
- bounds of an **ill-known probability measure** $\mu \Rightarrow \underline{\mu} \leq \mu \leq \overline{\mu}$ [21] [26]

Some basic properties

Avoiding sure loss and coherence

Given some bounds $\underline{\mu}(A)$ over every event $A \subseteq \mathcal{S}$, we say that

- $\underline{\mu}$ avoids sure loss iff

$$\mathcal{P}(\underline{\mu}) = \{\mu : \underline{\mu} \leq \mu \leq \overline{\mu}\} \neq \emptyset$$

- $\underline{\mu}$ is coherent iff for any A, we have

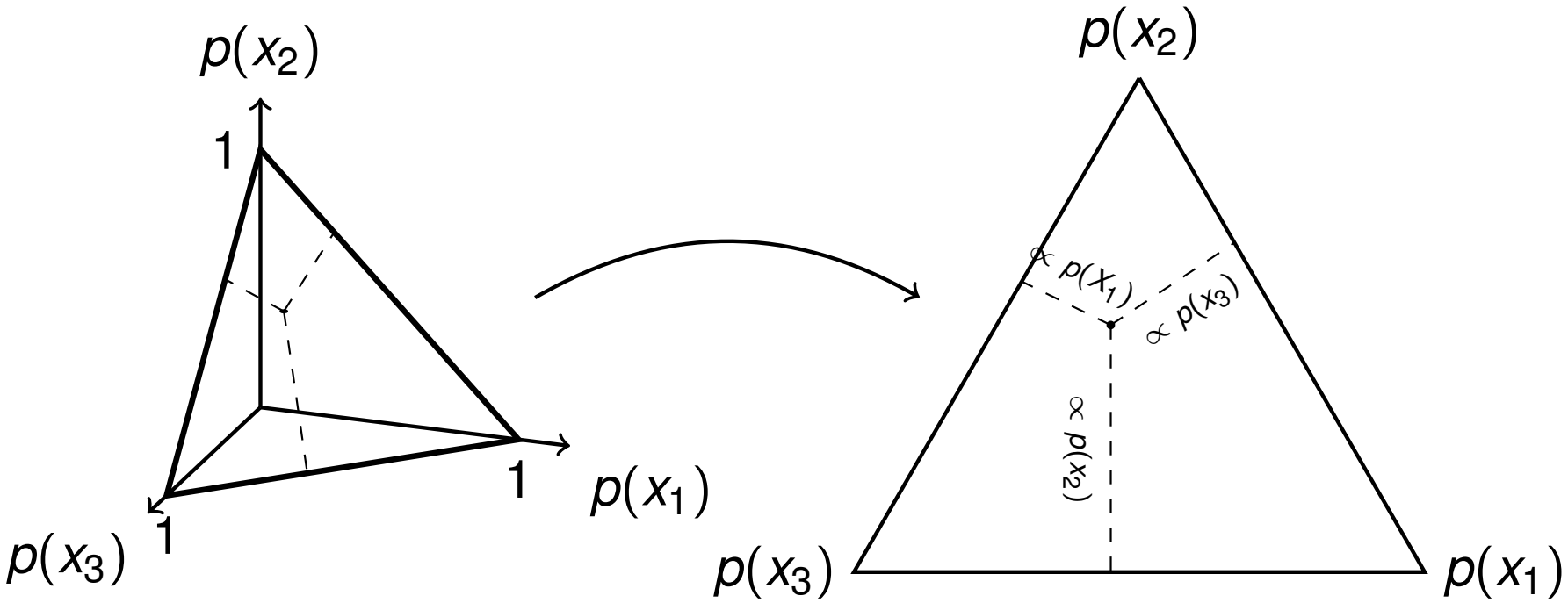
$$\inf_{\mu \in \mathcal{P}(\underline{\mu})} \mu(A) = \underline{\mu}(A)$$

An example of $\underline{\mu}$ incurring sure loss

Consider $\mathcal{S} = \{s_1, s_2, s_3\}$, then $\underline{\mu}(\{s_i, s_j\}) = 1/2$ for any distinct i, j incurs sure loss (**exercise: show it**)

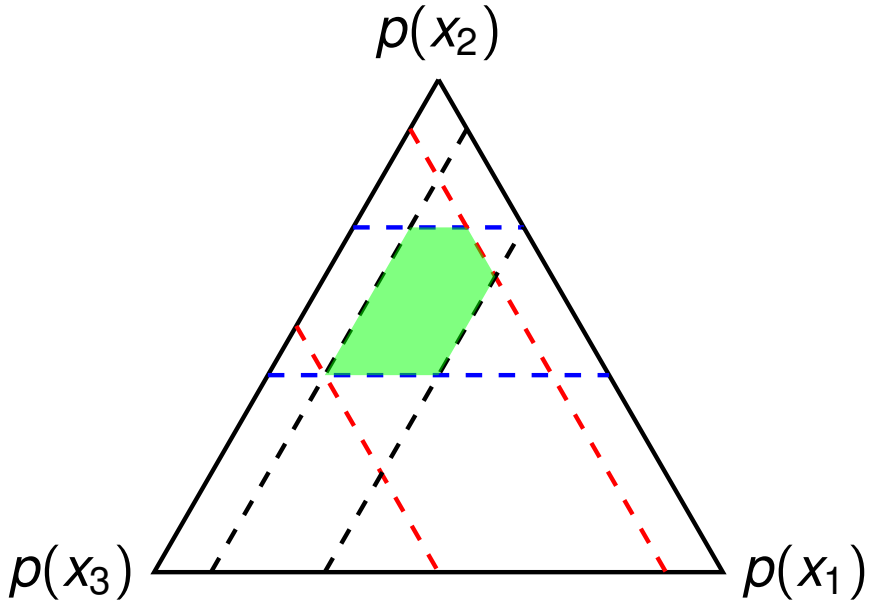
Illustrative example

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



Illustrative example

$p(x_1) \in [0.1, 0.3]$, $p(x_2) \in [0.4, 0.7]$, $p(x_3) = [0.1, 0.5]$



$\rightarrow$ these constraints still induce a belief function **(Exercise: show it)**

$\rightarrow$ but not all constraints on singletons induce a belief function **(Exercise: find an example)**

Means to get Imprecise probabilistic models

- Include all representations mentioned so far . . .
- . . . and a couple of others
  - probabilistic comparisons
  - density ratio-class
  - expectation bounds
  - . . .
- fully coherent extension of Bayesian approach

$$\mathcal{P}(\theta|x) = L(\theta|x)\mathcal{P}(\theta)$$

  $\rightarrow$ often easy for "conjugate prior" [24]
- make probabilistic logic approaches imprecise [23, 22]

A crude summary

Possibility distributions

- +: very simple, natural in many situations (nestedness), extend set-based approach
- -: at odds with probability theory, limited expressiveness

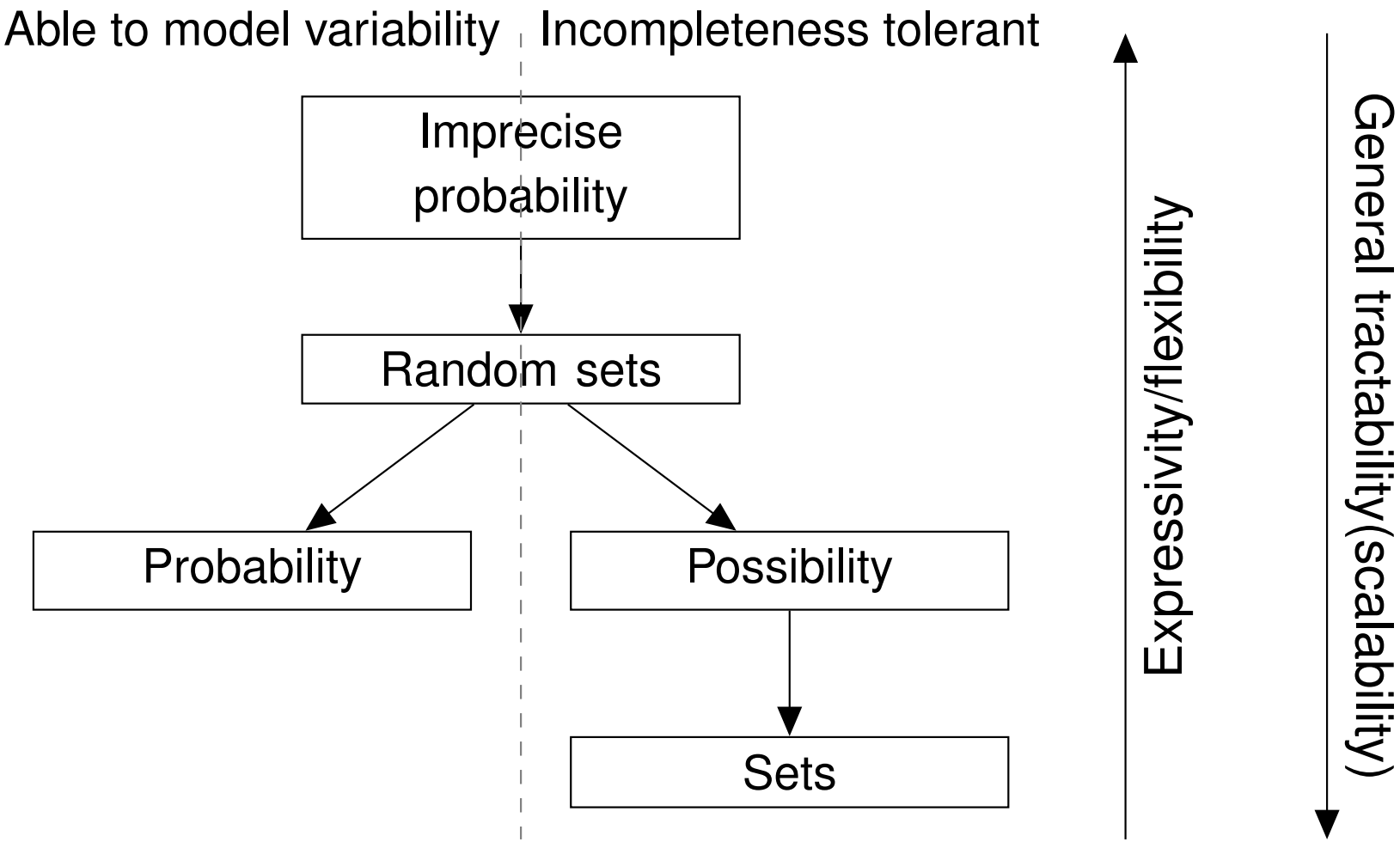
Random sets

- +: include probabilities and possibilities, include many models used in practice
- -: general models can be intractable, limited expressiveness

Imprecise probabilities

- +: most consistent extension of subjective probabilistic approach, very flexible
- -: general models can be intractable

A not completely accurate but useful picture



Plan

1 Introductory elements

2 Uncertainty theories and representation

3 Comparing informative contents

4 Conditioning in uncertainty theories

5 Independence in uncertainty theories

Introduction

Main question

Given two pieces of information $\underline{\mu}_1, \underline{\mu}_2$, is one more informative than the others? How can we answer?

Examples of use

- **Least commitment principle**: given multiple models satisfying given constraints, pick the most conservative one
  - Partial elicitation
  - Revision
  - Inverse Pignistic
- **(Outer)-approximation**: Pick a model $\underline{\mu}_2$ simpler than $\underline{\mu}_1$ (e.g., generic belief mass into possibility), ensuring that $\underline{\mu}_2$ does not add information to $\underline{\mu}_1$.

A natural notion: set inclusion

A set $A \subseteq \mathcal{S}$ is **more informative** than $B \subseteq \mathcal{S}$ if

$$A \subseteq B \Leftrightarrow A \sqsubseteq B$$

- Propositional logic: A more informative if A entails B
- Intervals: A includes all values of B, is more precise than B

$\Rightarrow$ extends this notion to other uncertainty theories

Extensions to other models

Denoting $\underline{\mu}_A, \underline{\mu}_B$ the uncertainty models of sets A, B, we do have

$$A \sqsubseteq B \Leftrightarrow \underline{\mu}_A(C) \leq \underline{\mu}_B(C) \text{ for any } C \subseteq \mathcal{S}$$

Derivations of $\underline{\mu}_1 \leq \underline{\mu}_2$ in different frameworks

- Possibility distributions: $\pi_1 \sqsubseteq \pi_2 \Leftrightarrow \pi_1 \geq \pi_2$
- Belief functions: $m_1 \sqsubseteq m_2 \Leftrightarrow Bel_1 \leq Bel_2$ (plausibility inclusion, there are others [42])
- Probability sets: $\underline{P}_1 \sqsubseteq \underline{P}_2 \Leftrightarrow \mathcal{P}_1 \subseteq \mathcal{P}_2$

Inclusion: interest and limitations

- +: very natural way to compare informative content
- -: only induces a partial order between information models

Example

Consider the space $\mathcal{S} = \{a, b, c\}$ and the following mass functions:

$$m_1(\{b\}) = 0.3, m_1(\{b, c\}) = 0.2, m_1(\{a, b, c\}) = 0.5$$

$$m_2(\{a\}) = 0.2, m_2(\{b\}) = 0.3, m_2(\{c\}) = 0.3, m_2(\{a, b, c\}) = 0.2$$

$$m_3(\{a, b\}) = 0.3, m_3(\{a, c\}) = 0.3, m_3(\{a\}) = 0.4$$

We have $m_2 \sqsubseteq m_1$, but m_3 incomparable with $\sqsubseteq$ **(exercise: show it)**

$\Rightarrow$ ok theoretically, but leads in practice to non-uniqueness of solutions

Numerical assessment of informative content

- For probabilities, distinct μ_1 and μ_2 always incomparable
- A solution, associate to each μ a number $I(\mu)$, i.e., entropy

$$I(\mu) = -\sum_{x \in \mathcal{S}} p(x) ln(p(x))$$

and declare that $\mu_1 \sqsubseteq \mu_2$ if $I(\mu_1) \leq I(\mu_2)$.

- This can be extended to other theories, where we can ask

$$\underline{\mu_1} \leq \underline{\mu_2} \Rightarrow I(\mu_1) \geq I(\mu_2)$$

Measure I should be consistent with inclusion

Plan

1 Introductory elements

2 Uncertainty theories and representation

3 Comparing informative contents

4 Conditioning in uncertainty theories

5 Independence in uncertainty theories

Three probabilistic use of conditional and conditioning

Focusing: querying generic knowledge about singular situation

- P: generic knowledge (usually about population)
- $P(|C)$: what we know from P in the singular context C

Revising: modifying current knowledge/belief with new information

- P: knowledge or belief (generic or singular)
- $P(|C)$: we learn that C is certainly true $\rightarrow$ how should we modify our knowledge/belief

Learning: using observation to infer generic knowledge

- P: beliefs about the parameter
- $P(|C)$: modified beliefs once we observe C ($\simeq$ multiple singular observations)

Focusing and revising in probabilities

In probability, upon learning C, the revised/focused knowledge is

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A \cap C)}{P(A \cap C) + P(A^c \cap C)}$$

coming down to the use of Bayes rule of conditioning in both cases.

Focusing

- Observing C does not modify our generic knowledge/beliefs
- We may lose information $\rightarrow$ the more C is specific, the less our general knowledge applies to it
- The consistency of generic knowledge/beliefs should be preserved (C cannot contradict it, only specify to which case it should apply)
- If we observe later $A \subseteq C$, we should start over from generic knowledge

Focusing in uncertainty theories

Focusing with belief functions

- Given initial Bel, this gives

$$Bel(A||C) = \frac{Bel(A \cap C)}{Bel(A \cap C) + Pl(A^c \cap C)}$$

$$Pl(A||C) = \frac{Pl(A \cap C)}{Pl(A \cap C) + Bel(A^c \cap C)}$$

We can have $Bel(A||C) < Bel(A) \leq Pl(A) < Pl(A||C)$ ("loss" of information).

- Can be interpreted as a sensitivity analysis of Bayes rule:

$$\underline{\mu}(A||C) = \inf\{P(A|C) : P \in \mathcal{P}, P(C) > 0\}$$

Revision

- Observing C modifies our knowledge and belief
- Observing C refines our beliefs and knowledge, that should become more precise
- If we observe later $A \subseteq C$, we should start from the modified knowledge (we may ask for operation to be order-insensitive)
- C is a new knowledge, that may be partially inconsistent with current belief/knowledge

Revision in uncertainty theories

Focusing with belief functions

- Given initial Pl, this gives

$$Pl(A|C) = \frac{Pl(A \cap C)}{Pl(C)} \Rightarrow Bel(A|C) = 1 - Pl(A^c|C)$$

  We necessarily have $Pl(A|C) < Pl(A)$ (refined information)
- Can be interpreted Bayes rule applied to most plausible situations:

$$\underline{\mu}(A||C) = \inf\{P(A|C) : P \in \mathcal{P}, P(C) = \overline{\mu}(C)\}$$

Revision as prioritized fusion

When $Pl(C) = 1$ and C precise observation

- $Pl(A|C)$= result of conjunctive combination rule
- $\mathcal{P}_{|C} = \mathcal{P} \cap \{P : P(C) = 1\}$

$\rightarrow$ can be interpreted as a fusion rule where C has priority. If $Pl(C) < 1$, interpreted as not fully consistent with knowledge.

Case where observation C is uncertain and inconsistent with knowledge.

- Minimally change $\underline{\mu}$ to be consistent with $C \rightarrow$ in probability, Jeffrey's rule (extensions to other theories exist)
- Not a symmetric fusion process, new information usually has priority ($\neq$ from usual belief fusion rules)!

A small example: focusing

The hotel provides the following plates for breakfast

a=Century egg, b=Rice, c=Croissant, d=Raisin Muffin

In a survey about their choices, respondent gave the reply

$$m(\{a,b\}) = \alpha,\ m(\{c,d\}) = 1-\alpha$$

Applying focusing **(Exercice: compute whole belief)**

We learn that customer C does not like eggs nor raisins ($C = \{b,c\}$), what can we tell about him choosing Rice?

$$Bel(b||C) = {}^{0}/_{(0+1-\alpha)} = 0, \quad Pl(b||C) = {}^{\alpha}/_{(\alpha+0)} = 1$$

We lose all our previous knowledge, and cannot tell anything.

A small example: revision

The hotel provides the following plates for breakfast

a=Century egg, b=Rice, c=Croissant, d=Raisin Muffin

In a survey about their choices, respondent gave the reply

$$m(\{a,b\}) = \alpha,\ m(\{c,d\}) = 1-\alpha$$

Applying revision **(Exercice: compute whole belief)**

We learn that suppliers no longer have eggs nor raisins ($C = \{b,c\}$), what is the proportion of rice we should buy to satisfy customers?

$$Pl(b|C) = \alpha/1 = \alpha, \quad Bel(b|C) = 1 - Pl(\{a,c,d\}|C) = \alpha$$

Our knowledge is now fully precise!

A short word on fusion

Let π_1, π_2 be possibility distributions, m_1, m_2 their mass functions and $\mathcal{P}_1, \mathcal{P}_2$ their associated probability sets. Then, in general, the conjunctive combinations

- $\pi_\cap(x) = \pi_1(x) \cdot \pi_2(x)$ (possibility)
- $m_\cap(A) = \sum_{C \cap B = A} m_1(C) m_2(B)$ (Dempster's)
- $\mathcal{P}_\cap = \mathcal{P}_1 \cap \mathcal{P}_2$ (Imprecise probabilities)

will provide different results ($\mathcal{P}_\cap$ may not even induce a belief function)

Plan

1 Introductory elements

2 Uncertainty theories and representation

3 Comparing informative contents

4 Conditioning in uncertainty theories

5 Independence in uncertainty theories

On independence and interpretation

Use of Independence $\sim$ facilitate computations in multi-variate problems

But what does it mean for two quantities X_1, X_2 to be independent, when they are:

- **generic**: in this case, X_1, X_2 associated to distributions μ_1, μ_2 and stochastic independence may apply $\rightarrow$ in uncertainty theories, sensitivity analysis of an "objective" concept.
- **singular**: X_1, X_2 are supposed to have one true value. Independence here is "subjective", as it purely concerns our beliefs and not how the values of X_1 or X_2 can affect each others.

In singular case, much less clear how it should be modelled and even measured?

Two formal views of independence

Consider two variables X_1, X_2 taking values in spaces $\mathcal{S}_1$, $\mathcal{S}_2$.
Independence/irrelevance statements of X_1, X_2 can take two forms:

- Compositional, meaning that we only need $\underline{\mu}_1, \underline{\mu}_2$ on $\mathcal{S}_1$, $\mathcal{S}_2$ to build a joint model over $\mathcal{S}_1 \times \mathcal{S}_2$

$$\underline{\mu}_{1,2} = \underline{\mu}_1 \otimes \underline{\mu}_2$$

  where $\otimes$ is some symmetric operator to be defined
- Informational irrelevance, where X_1 is said irrelevant to X_2 if the joint model satisfy that

$$\underline{\mu}_{1,2}(\cdot|X_1 \in A) = \underline{\mu}_2$$

  for any $A \subseteq \mathcal{S}$.

Additional aspects

Compositional independence:

- aim at easily building joint models from marginal ones;
- is by nature symmetric;

Informational irrelevance (of X_1 to X_2)

- aims at expressing that learning about X_1 does not change our beliefs about X_2;
- may have non-unique solutions $\rightarrow$ least-commitment;
- depends strongly on how conditioning is defined;
- is by nature asymmetric.

The case of probabilities

Compositional independence:

$$\mu_{1,2}(x_1, x_2) = \mu_1(x_1)\mu_2(x_2)$$

Information irrelevance

$$\mu_{1,2}(x_2|A) = \mu_2(x_2) \Rightarrow \mu_{1,2}(x_1, x_2) = \mu_1(x_1)\mu_2(x_2)$$

Both notions lead to stochastic independence $\rightarrow$ no longer true for more general uncertainty theories

Different extensions

Random set independence/Belief function independence:

$$m_{1,2}(A \times B) = m_1(A)m_2(B)$$

different from stochastic independence for all probabilities $\underline{\mu}_i \leq \mu_i$

Example **(Exercise: work fully detailed example)**

Consider $\mathcal{S}_1 = \{a, b\}$, $\mathcal{S}_2 = \{b, c\}$ with

$$m_1(\{a\}) = 0.3, m_1(\{b\}) = 0.5, m(\mathcal{S}_1) = 0.2$$

$$m_2(\{c\}) = m_2(\{d\}) = 0.4, m(\mathcal{S}_2) = 0.2$$

For event $A = (a, d) \cup (b, c)$ we have

$$Bel_{1,2}(A) = m_{1,2}((a, d)) + m_{1,2}((b, c)) = 0.32$$

Different extensions

Random set independence/Belief function independence:

$$m_{1,2}(A \times B) = m_1(A)m_2(B)$$

different from stochastic independence for all probabilities $\underline{\mu}_i \leq \mu_i$

Example **(Exercise: work fully detailed example)**

Consider $\mathcal{S}_1 = \{a, b\}$, $\mathcal{S}_2 = \{b, c\}$ with

$$m_1(\{a\}) = 0.3, m_1(\{b\}) = 0.5, m(\mathcal{S}_1) = 0.2$$

$$m_2(\{c\}) = m_2(\{d\}) = 0.4, m(\mathcal{S}_2) = 0.2$$

For event $A = (a, d) \cup (b, c)$, we have

$$\mu_{1,2}(A) = \mu_1(a)\mu_2(d) + \mu_1(b)\mu_2(c)$$

whose lowest value for all $\underline{\mu}_i \leq \mu_i$ is 0.46

Conclusions: unity among diversity

- Imprecision and uncertainty require more than probabilities and sets
- Different theories (possibility, belief, imprecise probability) propose to do so
- Most can be interpreted as models for generic or singular information
- Most consider similar problems (conditioning, independence) with different approaches
- Importance to understand the links and differences, and to clarify contexts in which tools are adequate

References I

General bibliography

[1] Denis Bouyssou, Didier Dubois, Henri Prade, and Marc Pirlot.
Decision Making Process: Concepts and Methods.
John Wiley & Sons, 2013.

[2] Inés Couso and Didier Dubois.
Statistical reasoning with set-valued information: Ontic vs. epistemic views.
International Journal of Approximate Reasoning, 55(7):1502–1518, 2014.

[3] B.de Finetti.
Theory of probability, volume 1-2.
Wiley, NY, 1974.
Translation of 1970 book.

[4] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter.
Applied Interval Analysis.
London, 2001.

[5] R. von Mises.
Probability, Statistics and Truth.
Dover books explaining science. Dover Publications, 1981.

Possibility theory

[6] Salem Benferhat, Julien Hué, Sylvain Lagrue, and Julien Rossit.
Interval-based possibilistic logic.
In *IJCAI*, pages 750–755, 2011.

[7] I. Couso, S. Montes, and P. Gil.
The necessity of the strong alpha-cuts of a fuzzy set.
Int. J. on Uncertainty, Fuzziness and Knowledge-Based Systems, 9:249–262, 2001.

References II

[8] G. de Cooman and D. Aeyels.
Supremum-preserving upper probabilities.
Information Sciences, 118:173–212, 1999.

[9] D. Dubois, S. Moral, and H. Prade.
A semantics for possibility theory based on likelihoods,.
Journal of Mathematical Analysis and Applications, 205(2):359 – 380, 1997.

[10] D. Dubois and H. Prade.
Possibility Theory: An Approach to Computerized Processing of Uncertainty.
Plenum Press, New York, 1988.

[11] D. Dubois and H. Prade.
When upper probabilities are possibility measures.
Fuzzy Sets and Systems, 49:65–74, 1992.

[12] Didier Dubois and Henri Prade.
An overview of the asymmetric bipolar representation of positive and negative information in possibility theory.
Fuzzy Sets and Systems, 160(10):1355–1366, 2009.

[13] Didier Dubois, Henri Prade, and Agnès Rico.
Representing qualitative capacities as families of possibility measures.
International Journal of Approximate Reasoning, 58:3–24, 2015.

[14] G. L. S. Shackle.
Decision, Order and Time in Human Affairs.
Cambridge University Press, Cambridge, 1961.

[15] L.A. Zadeh.
Fuzzy sets as a basis for a theory of possibility.
Fuzzy Sets and Systems, 1:3–28, 1978.

References III

Random sets and belief functions

[16] A.P. Dempster.
Upper and lower probabilities induced by a multivalued mapping.
Annals of Mathematical Statistics, 38:325–339, 1967.

[17] Ryan Martin, Jianchun Zhang, and Chuanhai Liu.
Dempster-shafer theory and statistical inference with weak beliefs.
Technical report, 2008.

[18] G. Shafer.
A mathematical Theory of Evidence.
Princeton University Press, New Jersey, 1976.

[19] P. Smets.
The transferable belief model and other interpretations of dempster-shafer's model.
In *Proc. of the Sixth Annual Confernce on Uncertainty in Artifical Intelligence*, pages 375–384, 1990.

[20] Philippe Smets.
Probability of provability and belief functions.
Logique et Analyse, 133-134:177–195, 1991.

Imprecise probability

[21] J. O. Berger.
An overview of robust Bayesian analysis.
Test, 3:5–124, 1994.
With discussion.

[22] Gert de Cooman.
Belief models: An order-theoretic investigation.
Ann. Math. Artif. Intell., 45(1-2):5–34, 2005.

References IV

[23] Pierre Hansen, Brigitte Jaumard, Marcus Poggi de Aragão, Fabien Chauny, and Sylvain Perron.
Probabilistic satisfiability with imprecise probabilities.
Int. J. Approx. Reasoning, 24(2-3):171–189, 2000.

[24] Erik Quaeghebeur and Gert de Cooman.
Imprecise probability models for inference in exponential families.
In *ISIPTA*, pages 287–296, 2005.

[25] P. Walley.
Statistical reasoning with imprecise Probabilities.
Chapman and Hall, New York, 1991.

[26] Kurt Weichselberger.
The theory of interval-probability as a unifying concept for uncertainty.
International Journal of Approximate Reasoning, 24(2–3):149 – 170, 2000.

Practical representations

[27] C. Baudrit and D. Dubois.
Practical representations of incomplete probabilistic knowledge.
Computational Statistics and Data Analysis, 51(1):86–108, 2006.

[28] M. Cattaneo.
Likelihood-based statistical decisions.
In *Proc. 4th International Symposium on Imprecise Probabilities and Their Applications*, pages 107–116, 2005.

[29] L.M. de Campos, J.F. Huete, and S. Moral.
Probability intervals: a tool for uncertain reasoning.
I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 2:167–196, 1994.

References V

[30] G. de Cooman and P. Walley.
A possibilistic hierarchical model for behaviour under uncertainty.
Theory and Decision, 52:327–374, 2002.

[31] T. Denoeux.
Constructing belief functions from sample data using multinomial confidence regions.
I. J. of Approximate Reasoning, 42:228–252, 2006.

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

[33] S. Destercke, D. Dubois, and E. Chojnacki.
Unifying practical uncertainty representations: II clouds.
Int. J. of Approximate Reasoning (in press), pages 664–677, 2008.

[34] D. Dubois, L. Foulloy, G. Mauris, and H. Prade.
Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities.
Reliable Computing, 10:273–297, 2004.

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

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

[37] Enrique Miranda and Sébastien Destercke.
Extreme points of the credal sets generated by comparative probabilities.
Journal of Mathematical Psychology, 64:44–57, 2015.

References VI

[38] Frédéric Pichon, Didier Dubois, and Thierry Denoeux.
Relevance and truthfulness in information correction and fusion.
Int. J. Approx. Reasoning, 53(2):159–175, 2012.

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

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

Information comparison

[41] J. Abellan and S. Moral.
A non-specificity measure for convex sets of probability distributions.
International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 8:357–367, 2000.

[42] D. Dubois and H. Prade.
A set-theoretic view on belief functions: logical operations and approximations by fuzzy sets.
Int. J. of General Systems, 12:193–226, 1986.

[43] D. Dubois and H. Prade.
Properties of measures of information in evidence and possibility theory.
Fuzzy sets and systems, 24:161–182, 1987.

[44] R.R. Yager.
Entropy and specificity in a mathematical theory of evidence.
I. J. of Systems, 9:249–260, 1983.

Conditioning

References VII

[45] Didier Dubois and Henri Prade.
Focusing vs. belief revision: A fundamental distinction when dealing with generic knowledge.
In *ECSQARU 97*, pages 96–107. Springer, 1997.

[46] Ronald Fagin and Joseph Halpern.
A new approach to updating beliefs.
In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 6, pages 347–374. North-Holland, Amsterdam, 1991.

[47] Jean-Yves Jaffray.
Bayesian updating and belief functions.
IEEE Transactions on Systems, Man and Cybernetics, 22:1144–1152, 1992.

[48] Henry E. Kyburg Jr.
Bayesian and non-bayesian evidential updating.
Artif. Intell., 31(3):271–293, 1987.

Independence

[49] I. Couso and S. Moral.
Independence concepts in evidence theory.
International Journal of Approximate Reasoning, 51:748–758, 2010.

[50] I. Couso, S. Moral, and P. Walley.
A survey of concepts of independence for imprecise probabilities.
Risk Decision and Policy, 5:165–181, 2000.