Positioning belief functions among uncertainty theories

D. Dubois

IRIT-CNRS, Université Paul Sabatier
31062 TOULOUSE FRANCE

Outline

1. Variability vs incomplete information
2. Limited expressiveness of probability
3. Set-valued representations
4. Blending set-valued and probability
    1. Imprecise probability
    2. Belief functions and random sets
    3. Numerical possibility theory
5. Comparison of uncertainty theories on
    1. Practical representations: fuzzy intervals, p-boxes, probability intervals.
    2. Information content

Origins of uncertainty

- The variability of observed natural phenomena : **randomness.**
  - Coins, dice…: what about the outcome of the next throw?
- The lack of information: **incompleteness**
  - because of information is often lacking, knowledge about issues of interest is generally not perfect.
- Conflicting testimonies or reports: **inconsistency**
  - The more sources, the more likely the inconsistency

Example

- **Frequentist**: daily quantity of rain in Toulouse
  - Represents variability: it may change every day
  - It is objective: can be estimated through statistical data
- **Incomplete information** : Birth date of Brazilian President
  - It is not a variable: it is a constant!
  - It is subjective: Most may have a rough idea (an interval), a few know precisely, some have no idea.
  - Statistics on birth dates of other presidents do not help much.
- **Inconsistent information** : several conflicting testimonies
  - Too much information
  - cannot deduce everything like in math…

Knowledge vs. evidence

- There are two kinds of information that help us make decisions in the course of actions:
  - Generic knowledge: repeable events
    - pertains to a population of observables (e.g. statistical knowledge)
    - Describes a general trend based on objective data
    - Tainted with exceptions
    - Deals with observed frequencies or ideas of typicality
  - Singular evidence: unique events
    - Consists of direct information about the current world.
    - pertains to a single situation
    - Can be unreliable, uncertain (e.g. unreliable testimony)

The roles of probability

Probability theory is generally used for representing two types of phenomena:

1. **Randomness:** capturing variability through repeated observations.
2. **Belief:** describes a person's opinion on the occurrence of a singular event.

As opposed to frequentist probability, subjective probability that models unreliable evidence is not necessarily related to statistics.

Measuring belief of unique events by probability

- Using statistical probability: the **Hacking principle**:
  - Generic knowledge = probability distribution P
  - $\text{belief}_{\text{NOW}}(A) = \text{Freq}_{\text{POPULATION}}(A)$: equating belief and frequency

- Direct elicitation as subjective probabilities **of non-repeatable events** with no frequentist flavor
  - frequencies may not be available nor known
  - non repeatable events.

- But *a single subjective probability distribution cannot distinguish between uncertainty due to variability and uncertainty due to lack of knowledge*

SUBJECTIVE PROBABILITIES
(Bruno de Finetti, 1935)

- A set S of mutually exclusive, exhaustive alternatives (states of nature, affairs, possible worlds).
- P(A)= *belief degree* of an agent on the (next) occurrence of A
- measured as the price You accept to buy a lottery ticket, with reward 1 € if A occurs
- You must accept to sell it at the same price if requested.

SUBJECTIVE PROBABILITIES (Bruno de Finetti, 1935)

- **Why a belief state is a single probability distribution** ($\sum_j p_j = 1$):
  - Assume You buy all lottery tickets for {i}, i = 1, m.
  - If state s_j is observed, the gain is $1 - \sum_j p_j$
  - it is $\sum_j p_j - 1$ if you have sold the tickets
  - *if* $\sum p_j > 1$ You *always lose money if you buy;*
  - *if* $\sum p_j < 1$ *You always lose money if you sell*

So, You must provide degrees of belief such that $\sum_j p_j = 1$, in order to be rational.

Using a single probability distribution to represent incomplete information is not entirely satisfactory:

The betting behavior setting of Bayesian subjective probability enforces a representation of partial ignorance based on single probability distributions.

1. **Ambiguity :** In the absence of information, how can a uniform distribution tell pure randomness and ignorance apart ?
2. **Instability :** A uniform prior on $x \in [a, b]$ induces a non-uniform prior on $f(x) \in [f(a), f(b)]$ if f is increasing and non-affine.
3. **Empirical refutation**: When information is missing, decision-makers do not always choose according to a single subjective probability (Ellsberg paradox).

Set-Valued Representations of Partial Knowledge

- An ill-known quantity x is represented as a *disjunctive* set, i.e. a subset $E \subseteq S$ of *mutually exclusive values*, one of which is the real one.
- Pieces of information of the form $x \in E$
  - **Intervals** E = [a, b]: good for representing incomplete numerical information: birth-date $\in$ [1940-1970]
  - **Classical Logic**: good for representing incomplete symbolic (Boolean) information

    E = Models of a wff ϕ stated as true.

This kind of information is subjective (epistemic set)

SETS and SETS

- Do not mix up
  - **A set-valued variable X**: the set of languages a person can speak (A = {English (and) French})

    a conjunction of values, and a real set.

    X = A is precise
  - **An ill-known point-valued variable x** : the birth nationality of this person

    (E = {English (exor) French})

    a disjunction of values, and an epistemic set (a piece of information in the head of an agent).

    $x \in E$ *is imprecise.*

BOOLEAN POSSIBILITY THEORY

Natural set functions under incomplete information:

If all we know is that $x \in E$ *then*

- Event A is possible if $A \cap E \neq \emptyset$

(logical consistency)

$\Pi(A) = 1$, and 0 otherwise

- Event A is sure if $E \subseteq A$

(logical deduction)

$N(A) = 1$, and 0 otherwise

Characteristic properties

$$N(A) = 1 - \Pi(A^c)$$

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B));$$

$$N(A \cap B) = \min(N(A), N(B)).$$

PROPERTIES

- $N(A) \leq \Pi(A)$
- A is surely true : $N(A) = 1$ (hence $\Pi(A) = 1$)
- A is surely false : $N(A^c) = 1$ (i.e., $\Pi(A^c) = 1$)
- A is unknown : $N(A) = N(A^c) = 0$

  $(\Pi(A) = \Pi(A^c) = 1)$

- $E_1 \subseteq E_2$: E_1 is less informative than E_2

Motivation for going beyond probability and epistemic logic

- Distinguish between uncertainty due to variability from uncertainty due to lack of knowledge or missing information.
- **Traditional tools to representing uncertainty are**
  - **Probability distributions :** good for expressing variability, but information demanding
  - **Sets**: good for representing incomplete information, but often crude modal representation of uncertainty
- *Find representations that allow for both aspects of uncertainty*.

Find a representation of uncertainty due to incompleteness

- *More informative than sets and epistemic logic*
- *Less demanding than single probability distributions*
- *Explicitly allows for missing information*
- *Allows for addressing the same problems as probability.*

GRADUAL REPRESENTATIONS OF UNCERTAINTY

Belief is a matter of degree !

- **Family of propositions or events $\mathcal{E}$ forming a Boolean Algebra**
  - S, Ø are events that are certain and ever impossible respectively.
- **A confidence measure** g : a function from $\mathcal{E}$ in [0,1] such that
  - $g(\emptyset) = 0 \quad ; \quad g(S) = 1$
  - if A implies (= included in) B then $g(A) \leq g(B)$ (monotony)
- g(A) quantifies the confidence of an agent in proposition A.

BASIC PROPERTIES OF CONFIDENCE MEASURES

- **$g(A \cup B) \geq \max(g(A), g(B))$;**
- **$g(A \cap B) \leq \min(g(A), g(B))$**
- It includes :
  - probability measures : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - possibility measures $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$
  - necessity measures $N(A \cap B) = \min(N(A), N(B))$

- ***The two latter functions do not require a numerical setting***

A GENERAL SETTING FOR REPRESENTING GRADED CERTAINTY AND PLAUSIBILITY

- 2 conjugate set-functions Pl and Cr generalizing probability P, possibility Π, and necessity N.
- **Conventions** :
  - Pl(A) = 0 "impossible" ; Cr(A) = 1 "certain"
  - Pl(A) =1 ; Cr(A) = 0 "ignorance" **(no information)**
  - Pl(A) - Cr(A) quantifies ignorance about A
- **Postulates**
  - Cr and Pl are monotonic under inclusion (= capacities).
  - $Cr(A) \leq Pl(A)$ "certain implies plausible"
  - $Pl(A) = 1 - Cr(A^c)$ duality certain/plausible
  - If Pl = Cr then it is P.

Canonical examples of (Cr, Pl) pairs

- Accounting for both variability and incomplete knowledge requires **(Cr, Pl) pairs**
- Obtained by combining probability and sets.
  - Sets of probabilities : imprecise probability theory
  - Random(ised) sets : Dempster-Shafer theory
  - Fuzzy sets: numerical possibility theory
- Each event has a degree of belief (certainty) and a degree of plausibility, instead of a single degree of probability

Possibility Theory

(Shackle, 1961, Zadeh, 1978)

- A piece of incomplete information "$x \in E$" admits of *degrees* of possibility:
- E is a (normalized) fuzzy set.
- $\mu_E(s)$ = Possibility(x = s) = $\pi_x(s)$
- **Conventions:**

  $\forall s, \pi_x(s)$ is the degree of plausibility of x = s

  $\pi_x(s) = 0$ iff x = s is impossible, totally surprising

  $\pi_x(s) = 1$ iff x = s is normal, fully plausible, unsurprising

  (but no certainty)

A pioneer of possibility theory

- In the 1950's, **G.L.S. Shackle** called "degree of potential surprize" of an event its degree of impossibility = $1 - \Pi(A)$.

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

- Potential surprize is valued on a disbelief scale, namely a positive interval of the form $[0, y^*]$, where y^* denotes the absolute rejection of the event to which it is assigned, and 0 means that nothing opposes to the occurrence of A.

- The degree of surprize of an event is the degree of surprize of its least surprizing realization.
- He introduces a notion of conditional possibility

Improving expressivity of incomplete information representations

- *What about the birth date of the president?*
- **partial ignorance with ordinal preferences :** May have reasons to believe that 1933 > 1932 ≡ 1934 > 1931 ≡ 1935 > 1930 > 1936 > 1929
- **Linguistic information** described by fuzzy sets: " he is old " : membership μ_{OLD} is interpreted as a possibility distribution on possible birth dates.

- **Nested intervals** $E_1, E_2, \ldots E_n$ **with confidence levels** $N(E_i) = a_i$:
  - $\pi(x) = \min_{i=1,\ldots n} \max(\mu_{Ei}(x), 1 - a_i)$

POSSIBILITY AND NECESSITY OF AN EVENT

How confident are we that $x \in A \subset S$? (*an event A occurs*)
given a possibility distribution on S

- $\Pi(A) = \max_{s \in A} \pi(s)$: plausibility as consistency with π

  to what extent A is consistent with π

  (= some $x \in A$ is possible)

  The degree of possibility *that* $x \in A$

- $N(A) = 1 - \Pi(A^c)$

  $= \min_{s \notin A} 1 - \pi(s)$: certainty as entailment from π

  to what extent no element outside A is possible

  = to what extent π *implies A*

  The degree of certainty (necessity) that $x \in A$

Basic properties

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B));$$

$$N(A \cap B) = \min(N(A), N(B)).$$

Mind that most of the time :

$$\Pi(A \cap B) < \min(\Pi(A), \Pi(B));$$

$$N(A \cup B) > \max(N(A), N(B)$$

Example: Total ignorance on A and B = A^c

Corollary $N(A) > 0 \Rightarrow \Pi(A) = 1$

Comparing information states

- **π' more specific than π in the wide sense if and only if $\pi' \leq \pi$**

Any possible value according to π' is at least according to π :
π' is more informative than π

  - COMPLETE KNOWLEDGE: The most specific ones
    - $\pi(s_0) = 1$; $\pi(s) = 0$ otherwise
  - IGNORANCE: $\pi(s) = 1, \forall\ s \in S$

- **Principle of least commitment** (minimal specificity): In a given information state, any value not proved impossible is supposed to be possible : maximise possibility degrees.

Qualitative vs. quantitative possibility theories

- **Qualitative**:
  - **comparative**: A complete pre-ordering $\geq_\pi$ on U A well-ordered partition of U: E1 > E2 > … > En
  - **absolute:** $\pi_x(s) \in L$ = finite chain, complete lattice...
- **Quantitative**: $\pi_x(s) \in [0, 1]$, integers...

One must indicate where the numbers come from.

All theories agree on the fundamental maxitivity axiom

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$$

Theories diverge on the conditioning operation

Imprecise probability theory

- A state of information is represented by a family $\mathcal{P}$ of probability distributions over a set X.
- To each event A is attached a probability interval $[P_*(A), P^*(A)]$ such that
  - $P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$
  - $P^*(A) = \sup\{P(A), P \in \mathcal{P}\} = 1 - P_*(A^c)$
- Usually $\mathcal{P}$ is strictly contained in $\{P(A), P \geq P_*\}$
- *For instance: incomplete knowledge of a FREQUENTIST probabilistic model :* $\exists P \in \mathcal{P}$.

Subjectivist view (Peter Walley)

- $P_{low}(A)$ is the highest acceptable buying price for a bet on event A winning 1 euro if A occurs
- $P^{high}(A) = 1 - P_{low}(A^c)$ is the least acceptable price for selling this bet.
- *A theory of convex probability (credal) sets*
- **Rationality** conditions for credal sets:
  - **No sure loss** : $\{P \geq P_{low}\}$ not empty
  - **Coherence**: $P_*(A) = \inf\{P(A), P \geq P_{low}\} = P_{low}(A)$
- Credal sets are characterized by lower expectations of real-valued functions (maximal gamble buying prices), not just events.

Capacity-based lower probabilities

- Coherent lower probabilities are important examples of certainty functions. *The most general numerical approach to uncertainty*.
- They satisfy super-additivity: if $A \cap B = \emptyset$ then

$$Cr(A) + Cr(B) \leq Cr(A \cup B)$$

- One may require the 2-monotony property:

$$Cr(A) + Cr(B) \leq Cr(A \cup B) + Cr(A \cap B)$$

  – ensures non-empty **coherent** credal set:

$$\{P: P(A) \geq Cr(A)\} \neq \emptyset .$$

Cr is then called a convex capacity.

Random sets

- A positive weighting of non-empty subsets: mathematically, **a random set** $\mathcal{R}$ on S:

$$\sum_{E \subseteq S} m(E) = 1 \; ; \; m(\emptyset) = 0$$

- A probability distribution *m* on the family of non-empty subsets of a set S.
  - m : *mass function* : $m(E) = Prob(\mathcal{R} = E)$.
  - *focal sets* : E with $m(E) > 0$:
- ***Interpretation problem****: disjunctive or conjunctive ???*

Basic set functions from random sets

- **Probabilities of containment** :
  - $P(\mathcal{R} \subseteq A) = \sum_{E_i \subseteq A,\, E_i \neq \emptyset} m(E_i)$
  - $P(A \subseteq \mathcal{R}) = \sum_{A \subseteq E_i,\, E_i \neq \emptyset} m(E_i)$
- **Probability of hitting**:
  - $P(A \cap \mathcal{R} \neq \emptyset) = \sum_{E_i \cap A \neq \emptyset} m(E_i)$

Three views of random sets

- A set-valued standard random variable: *conjunctive random set*. **Matheron**
- A disjunctive set of random variables: *epistemic uncertainty about a random variable* (statistical) **Dempster**
- A subjective probability over possible epistemic sets : *unreliable testimony*
  (non-statistical) **Shafer book, Smets**

Example of conjunctive random set

Experiment on linguistic capabilities of people :

- **Question** to **a population** $\Omega = \{1, \ldots, i, \ldots, n\}$ of n persons: which languages can you speak ?
- **Answers** : Subsets in $\mathcal{L} = \{$Basque, Chinese, Dutch, English, French,.…$\}$?
- m(E) = % people who speak *exactly* all languages in E (and not other ones)
- Prob(x speaks A) $= \sum\{m(E) : A \subseteq E\} = P(A \subseteq \mathcal{R})$

  = Q(A) : commonality function in belief function theory
- **Example**: « x speaks English » means « at least English »
- Belief functions are not meaningful here while the commonality makes sense, contrary to the disjunctive set case.

Random disjunctive sets

- A focal set is an epistemic state ($x \in E$)
- ($\mathcal{F}$, m) is a randomized incomplete information
- m(E) = probability that the most precise description of the available information is of the form "$x \in E$"

  = *probability (only knowing "$x \in E$" and nothing else)*

  – It is the portion of probability mass hanging over elements of E without being allocated.
- **DO NOT MIX UP m(E) and Prob(E) :**

$$m(E) = Prob(\{E\})$$

Belief and Plausibility from Epistemic Random Sets

- **degree of certainty** (belief) :
  - $\mathrm{Bel}(A) = \sum_{E_i \subseteq A,\, E_i \neq \emptyset} m(E_i) = P(\mathcal{R} \subseteq A)$
  - total mass of information <u>implying</u> the occurrence of A
  - *(probability of provability)*
- **degree of plausibility** : $P(A \cap \mathcal{R} \neq \emptyset)$
  - $\mathrm{Pl}(A) = \sum_{E_i \cap A \neq \emptyset} m(E_i) = 1 - \mathrm{Bel}(A^c) \geq \mathrm{Bel}(A)$
  - total mass of information <u>consistent with</u> A
  - *(probability of consistency)*

PARTICULAR CASES

- INCOMPLETE INFORMATION:

  $m(E) = 1, m(A) = 0,\ A \neq E$

- *TOTAL IGNORANCE :* $m(S) = 1$:
  - *For all* $A \neq S, \emptyset$, *Bel(A) = 0, Pl(A) = 1*

- PROBABILITY: if $\forall i, E_i$ = singleton $\{s_i\}$ (disjoint focal sets) :
  Hence precise + scattered information
  - Then, *for all* A, Bel(A) = Pl(A) = P(A)

- POSSIBILITY THEORY : the consonant case
  $E_1 \subseteq E_2 \subseteq E_3 \ldots \subseteq E_n$: imprecise and coherent information
  - iff $Pl(A \cup B) = \max(Pl(A), Pl(B))$, possibility measure
  - iff $Bel(A \cap B) = \min(Bel(A), Bel(B))$,
    necessity measure = consonant belief function

Dempster vs. Shafer-Smets

- A disjunctive random set can represent
  - *Imprecise statistical information* (Dempster intuition)
    m(E) = frequency of imprecise observations of the form E
    - an ill-known random variable.
    - Bel(E) is a lower probability
  - *Uncertain singular evidence* (Shafer, 1976; Smets)
    m(E) = subjective probability that testimony E is reliable.
    - Degrees of belief directly modelled by Bel : no appeal to an underlying probability.

In all cases, a focal set E is a set of mutually exclusive values and does not represent a real set-valued entity

Canonical examples

- **Statistical probabilities** : Frequentist modelling of a collection of incomplete observations (imprecise statistics) : incomplete generic information (Dempster)
- **Uncertain singular information:**
  - **Unreliable testimonies** (Shafer's book) : unreliable singular information
  - **Unreliable sensors** (Smets): the quality of the information depends on the ill-known sensor state.

Dempster original model

- Indirect information (induced from a probability space).
- What we know about a random variable x with range S, based on a sample space $(\Omega, \mathcal{A}, P)$ and a multimapping Γ from Ω to S (Dempster):
- *The meaning of the multimapping Γ from Ω to S* :
  - if we observe ω in Ω then all we know is $x(\omega) \in \Gamma(\omega)$
- $m(\Gamma(\omega)) = P(\{\omega\})$ $\forall$ ω in Ω (finite case.)
- Then a belief function is a lower probability.

SELECTION FUNCTIONS

- A *selection function* of the multimapping Γ from Ω to S is a random variable x such that $\forall\ \omega$ in $\Omega, x(\omega) \in \Gamma(\omega)$
- Each random variable x induces a probability P_x on S from $(\Omega, \mathcal{A}, P)$: $P_x(A) = P(x^{-1}(A))$.
- The set $\mathfrak{P}_\Gamma = \{P_x : \forall\ \omega$ in $\Omega, x(\omega) \in \Gamma(\omega)\}$ of probabilities induced by Γ is not convex
- $Bel(A) = \sum_{\Gamma(\omega) \subseteq A} P(\omega)$

$$= \min\{P_x(A): P_x \in \mathfrak{P}_\Gamma\}$$

Example of generic belief function: imprecise observations in an opinion poll

- **Question** : who is your preferred candidate
  in $C = \{a, b, c, d, e, f\}$???
  - **To a population** $\Omega = \{1, \ldots, i, \ldots, n\}$ of n persons.
  - **Imprecise responses r =** « $x(i) \in E_i$ » **are allowed**
  - No opinion ($r = C$) ; « left wing » $r = \{a, b, c\}$;
  - « right wing » $r = \{d, e, f\}$;
  - a moderate candidate : $r = \{c, d\}$

- **Definition of mass function**: $m(E) = card(\{i, E_i = E\}) / n$
  = Proportion of imprecise responses « $x(i) \in E$ »
  uniform distribution on Ω, $E_i = \Gamma(i)$,

- *The probability that a candidate in subset* $A \subseteq C$ *is elected is imprecise :*

$$\text{Bel}(A) \leq P(A) \leq \text{Pl}(A)$$

- **There is a fuzzy set F of potential winners**:

$$\mu_F(x) = \sum_{x \in E} m(E) = \text{Pl}(\{x\}) \text{ (contour function)}$$

- $\mu_F(x)$ is an upper bound of the probability that x is elected. It gathers responses of those who *did not give up voting* for x
- Bel({x}) gathers responses of those who claim they will vote for x and no one else.

Example of uncertain evidence : Unreliable testimony (SHAFER-SMETS VIEW)

- « John tells me the president is between 60 and 70 years old, but there is some chance (*subjective* probability p) he does not know and makes it up».
  - E =[60, 70]; Prob(Knowing "x∈ E =[60, 70]") = 1 − p.
  - With probability p, John invents the info, so *we know nothing (Note that this is different from a lie)*.
- We get a *simple support belief function :*

$$m(E) = 1 - p \quad \text{and} \quad m(S) = p$$

- Equivalent to a possibility distribution
  - $\pi(s) = 1$ if $x \in E$ and $\pi(s) = p$ otherwise.
- *If John may lie* (probability q): $m(E) = (1 - p)(1 - q)$, $m(E^c) = (1 - p)q$.

Random disjunctive sets vs. imprecise probabilities

- The set $\mathcal{P}_{\text{bel}} = \{P \geq \text{Bel}\}$ is coherent: Bel is a special case of lower probability
- The set of probabilities $\mathfrak{P}_\Gamma$ induced by selection functions (Dempster model) is but a subset of $\mathcal{P}_{\text{bel}}$.
  - In the election example, $\mathfrak{P}_\Gamma$ is finite.
- However $\mathcal{P}_{\text{bel}}$ is the convex hull of $\mathfrak{P}_\Gamma$:

$$\text{Bel}(A) = \inf \{P(A): P_x \in \mathcal{P}_{\text{bel}}\}$$
$$= \min\{P_x(A): P_x \in \mathfrak{P}_\Gamma\}$$
$$\text{Pl}(A) = \sup \{P(A): P_x \in \mathcal{P}_{\text{bel}}\}$$
$$= \inf\{P_x(A): P_x \in \mathfrak{P}_\Gamma\}$$

Random disjunctive sets vs. imprecise probabilities

- Bel is ∞-monotone (super-additive at any order)
  - Order 3: $Bel(A \cup B \cup C) \geq Bel(A) + Bel(B) + Bel(C) - Bel(A \cap B) - Bel(A \cap C) - Bel(B \cap C) + Bel(A \cap B \cap C)$, etc.
- For any set function, the solution m to the set of equations $\forall A \subseteq X \; g(A) = \sum_{E_i \subseteq A, E_i \neq \emptyset} m(E_i)$ is unique (m is the Moebius transform of)
  - ***However m is positive iff g is a belief function***

Probability intervals

- **Definition**: To each element s_i in S, assign a probability interval $[l_i, u_i]$, i = 1, …n.
- $\mathcal{P} = \{P, l_i \leq P(\{s_i\}) \leq u_i , i = 1, \ldots n\}$ is the credal set.
- A probability interval model L is **coherent** in the sense of Walley if and only if all bounds u_j , l_i are attained iff
  - $\sum_{j \neq i} l_j + u_i \leq 1$ and $1 \leq \sum_{j \neq i} u_j + l_i$
  - In particular $\sum_j l_j \leq 1 \leq \sum_j u_j$

    Then $l_i = \inf\{P(\{s_i\}): P \text{ in } \mathcal{P}\}$; $u_i = \sup\{P(\{s_i\}): P \text{ in } \mathcal{P}\}$
- Induced upper and lower probabilities are 2-monotone capacities (De Campos and Moral), NOT BELIEF FUNCTIONS
- $P_*(A) = \max(\sum_{j \text{ in } A} l_j , \sum_{j \text{ not in } A} u_j)$

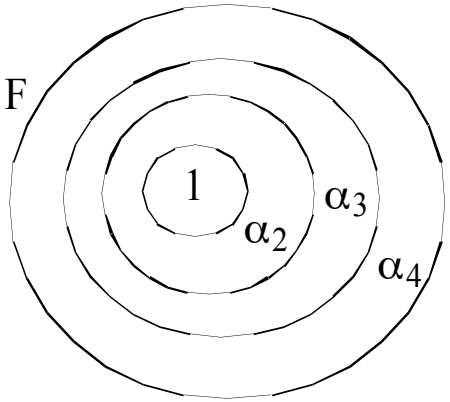
Example

- On $\{1, 2, 3\}$:
  - p_1 in $[0.1, 0.3]$
  - p_2 in $[0.2, 0.4]$
  - p_3 in $[0.3, 0.5]$
- Coherence checking:
  - $0.1 + 0.2 + 0.5 \leq 1 \leq 0.3 + 0.4 + 0.3$
  - $0.1 + 0.4 + 0.3 \leq 1 \leq 0.3 + 0.2 + 0.5$
  - $0.3 + 0.2 + 0.3 \leq 1 \leq 0.1 + 0.4 + 0.5$

Example

- Lower probabilities
  - $P_*(\{1, 2\}) = \max(0.3, 1- 0.5) = 0.5$
  - $P_*(\{2, 3\}) = \max(0.5, 1- 0.3) = 0.7$
  - $P_*(\{1, 3\}) = \max(0.4, 1- 0.4) = 0.6$
- Moebius transforms
  - $m(1) = 0.1$; $m(2) = 0.2$; $m(3) = 0.3$;
  - $m(\{1, 2\}) = 0.5 - 0.1 - 0.2 = 0.2$
  - $m(\{2, 3\}) = 0.7 - 0.2 - 0.3 = 0.2$
  - $m(\{1, 3\}) = 0.6 - 0.1 - 0.3 = 0.2$
  - $m(\{1, 2, 3\}) = 1 -$ other masses $= -0.2$

From possibility to belief functions



possibility levels
$1 > \alpha_2 > \alpha_3 > \ldots > \alpha_n$

- Let $m_i = \alpha_i - \alpha_{i+1}$ then $m_1 + \ldots + m_n = 1$,
  with focal sets = cuts $E_i = \{s, \pi(s) \geq \alpha_i\}$

  A consonant basic probability assignment (SHAFER)
- $\pi(s) = \sum_{i:\, s \in Ei} m_i$ (one point-coverage function) = $Pl(\{s\})$.
- ***Only in the consonant case can m be recalculated from π***
- *$Bel(A) = \sum_{Ei \subseteq A} m_i = N(A)$; $Pl(A) = \Pi(A)$*

POSSIBILITY AS UPPER PROBABILITY

- Given a numerical possibility distribution π, define
  $P(\pi) = \{P \mid P(A) \leq \Pi(A) \text{ for all } A\}$
  - Then, Π and N can be recovered
    - $\Pi(A) = \sup \{P(A) \mid P \in P(\pi)\}$;
    - $N(A) = \inf \{P(A) \mid P \in P(\pi)\}$
  - So π is a faithful representation of a special family of coherent probability measures
- Likewise for belief functions : $P(m) = \{P \mid P(A) \leq Pl(A), \forall A\}$
    - $Pl(A) = \sup \{P(A) \mid P \in P(m)\}$;
    - $Bel(A) = \inf \{P(A) \mid P \in P(m)\}$

  - Representing statistical knowledge
  - Representing maximal buying prices as per Walley

From confidence sets to possibility distributions

- Let $E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n$ be a nested family of sets
- A set of confidence levels $a_1, a_2, \ldots a_n$ in $[0, 1]$
- Consider the set of probabilities

$$\mathcal{P} = \{P, P(E_i) \geq a_i, \text{ for } i = 1, \ldots n\}$$

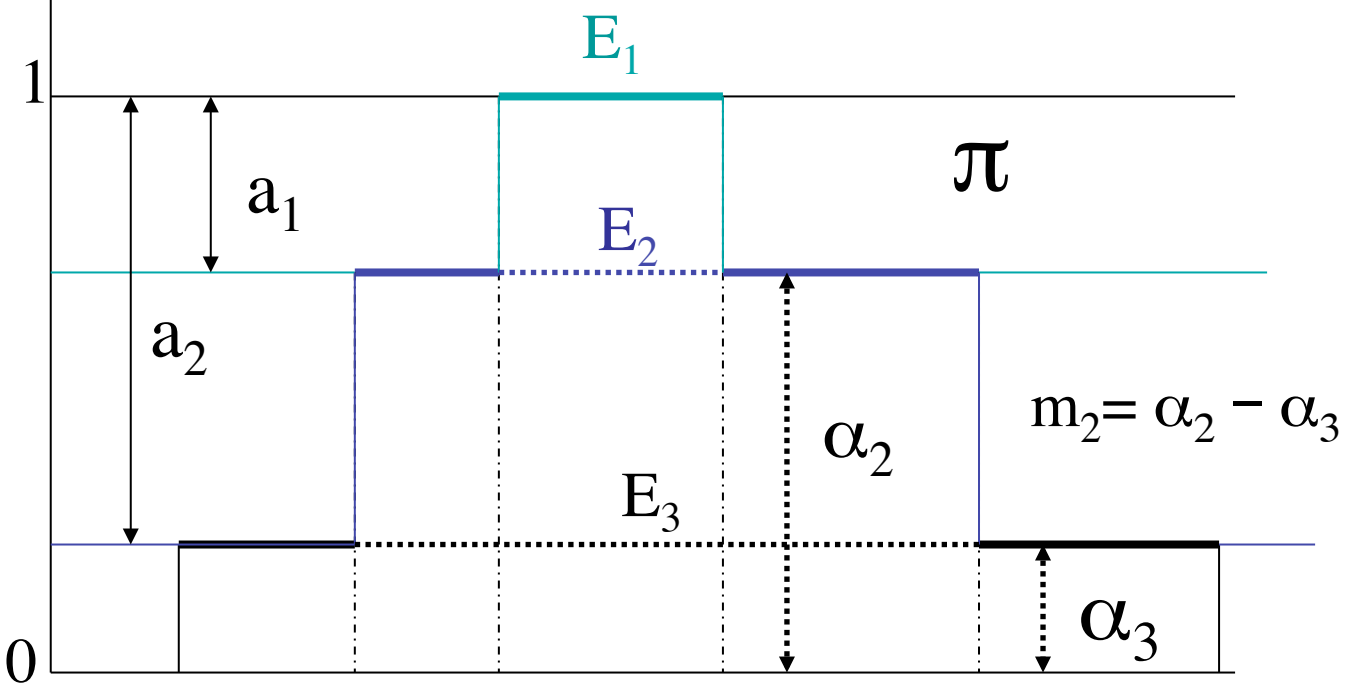
- Then $\mathcal{P}$ is representable by means of a possibility measure with distribution

$$\pi(x) = \min_{i = 1, \ldots n} \max (\mu_{Ei}(x), 1 - a_i)$$

A consonant belief function with mass

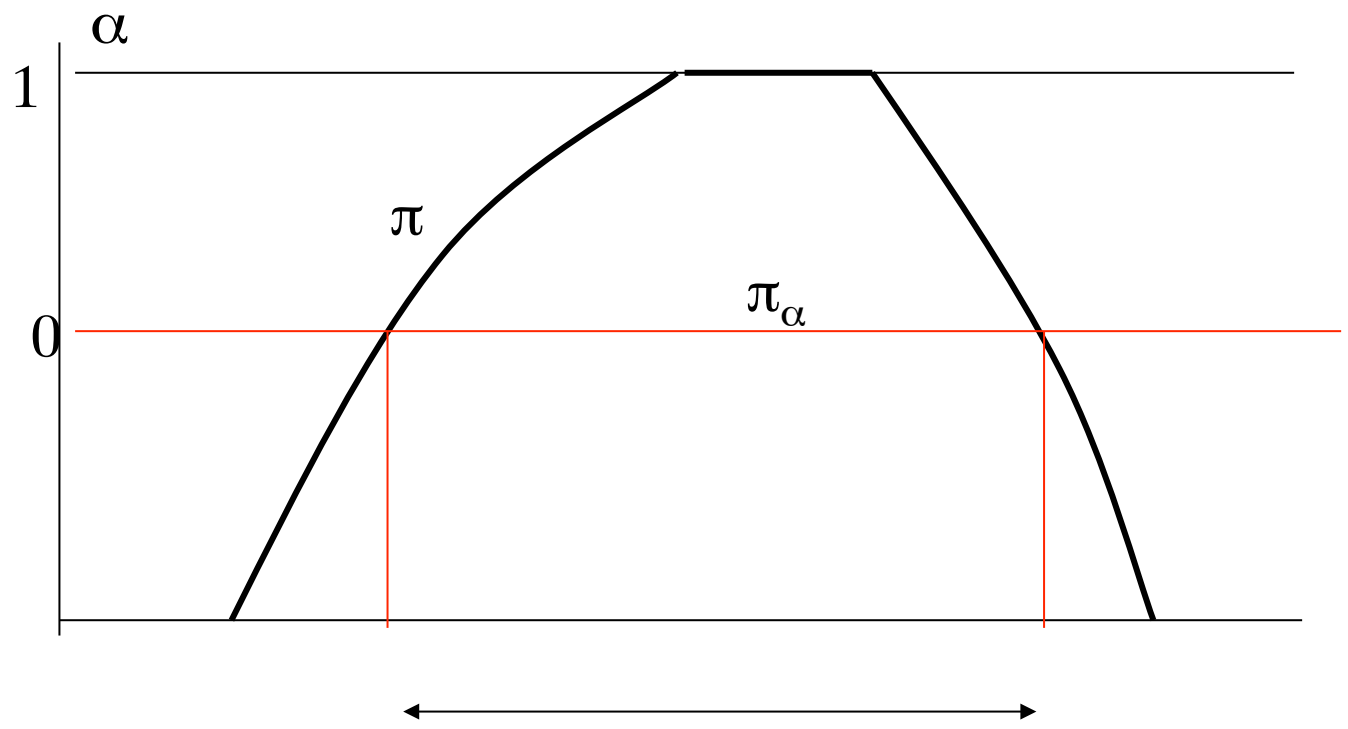
$$m(E_i) = a_i - a_{i+1}$$

POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



A fuzzy interval corresponds to a :

- *convex set of probabilities: {P:* $P(A_\alpha) \geq 1 - \alpha, \alpha \in (0, 1]\}$
- *continuous belief function ([0, 1], B,* λ*) -> /R,*



FUZZY INTERVAL: $\Gamma(\alpha) = \pi_\alpha$

Possibilistic view of probabilistic inequalities

- Probabilistic inequalities can be used for representing consonant belief functions
  - Chebyshev inequality defines a possibility distribution that dominates *any* density with given mean and variance:

$P(V \in [x^{mean} - k\sigma, x^{mean} + k\sigma]) \geq 1 - 1/k^2$ *is equivalent to writing*

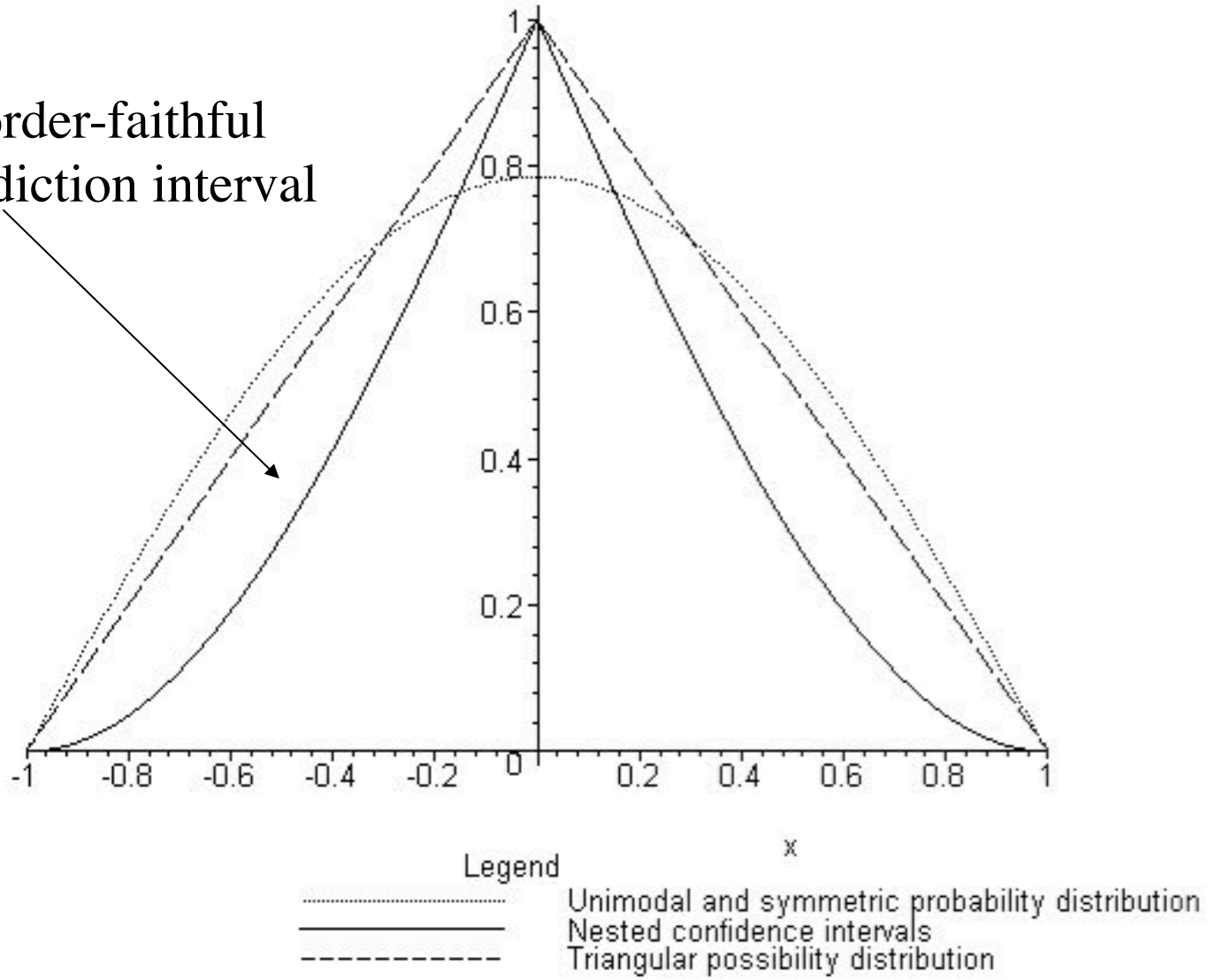
$\pi(x^{mean} - k\sigma) = \pi(x^{mean} + k\sigma) = 1/k^2$

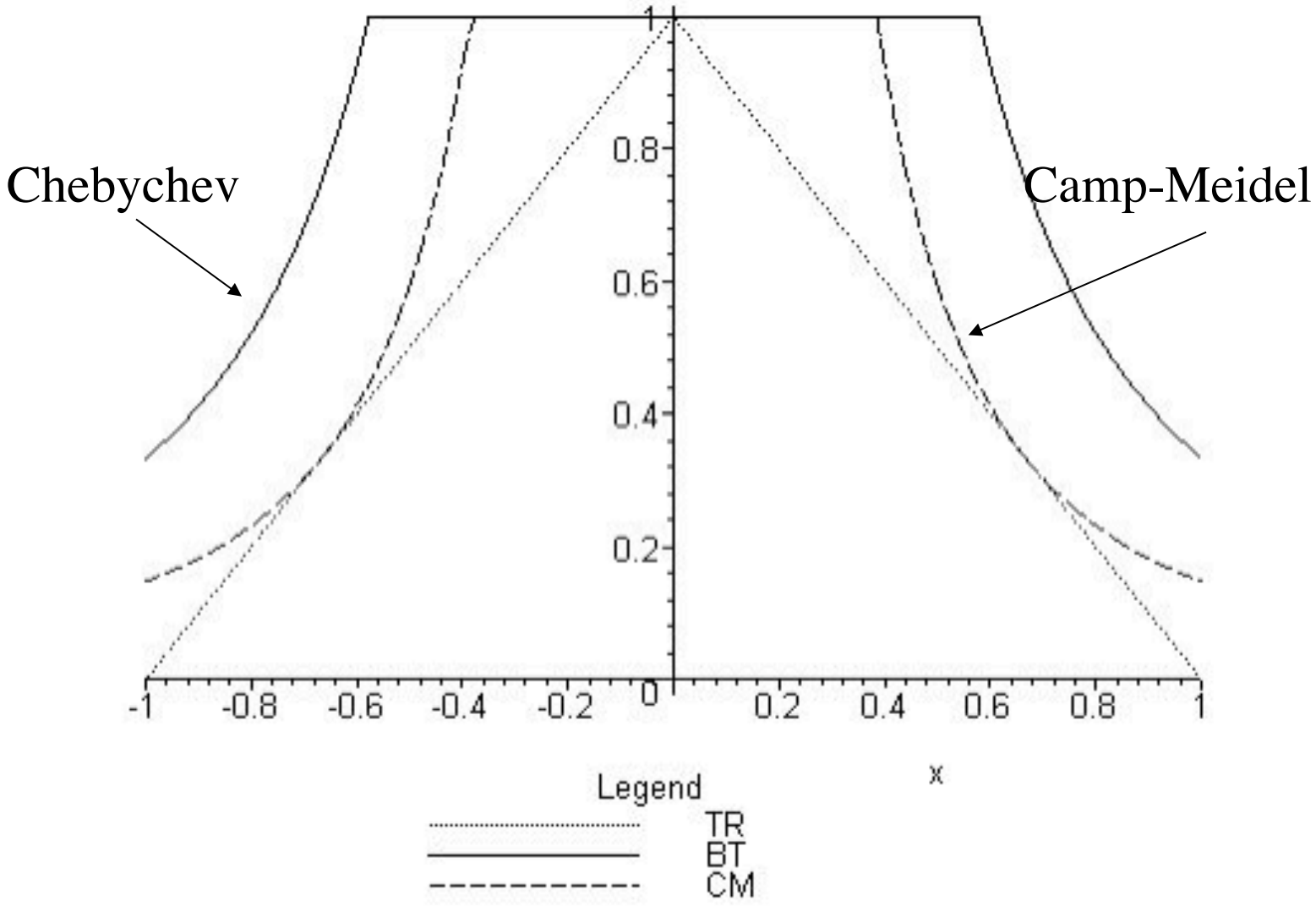
  - A triangular fuzzy number (TFN) defines a possibility distribution that dominates *any* unimodal density with the same mode and bounded support as the TFN (related to Gauss inequality, 1823).

$P(V \in [x, 2x^{mode} - x]) \geq |x^{mode} - x|/a$ *is equivalent to writing*

$\pi(x) = 1 - |x^{mode} - x|/a$ *on* $[x^{mode} - a, x^{mode} + a]$:

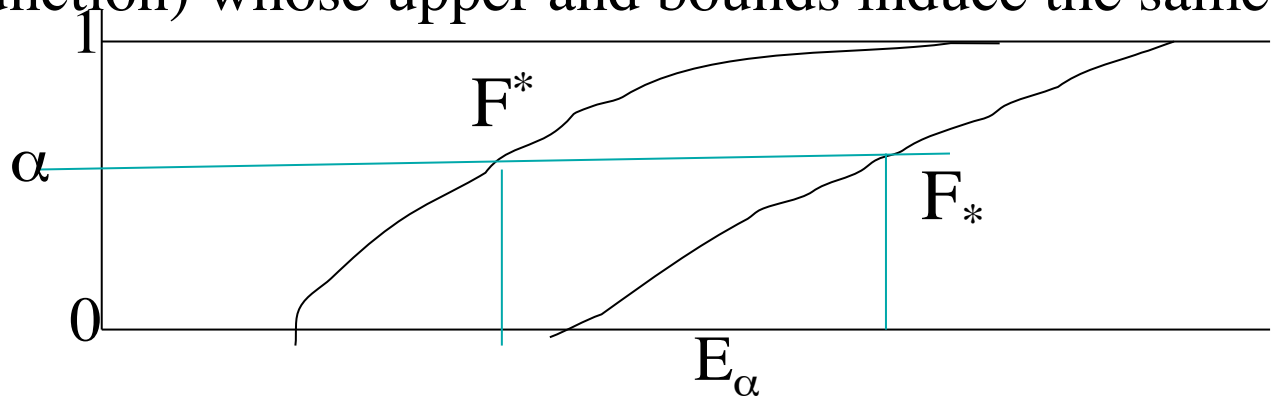
Optimal order-faithful
fuzzy prediction interval




Chebychev
Camp-Meidel
1
0.8
0.6
0.4
0.2
0
-1
-0.8
-0.6
-0.4
-0.2
0.2
0.4
0.6
0.8
1
x
Legend
TR
BT
CM

Probability boxes

- Let be F_P be the cumulative distribution of P : F_P (a) = P($-\infty$, a]
- A set $\mathcal{P} = \{P: F^* \geq F_P \geq F_*\}$ induced by two cumulative disribution functions is called a **probability box (p-box)**,
- A p-box is a special random interval (continuous belief function) whose upper and bounds induce the same ordering.



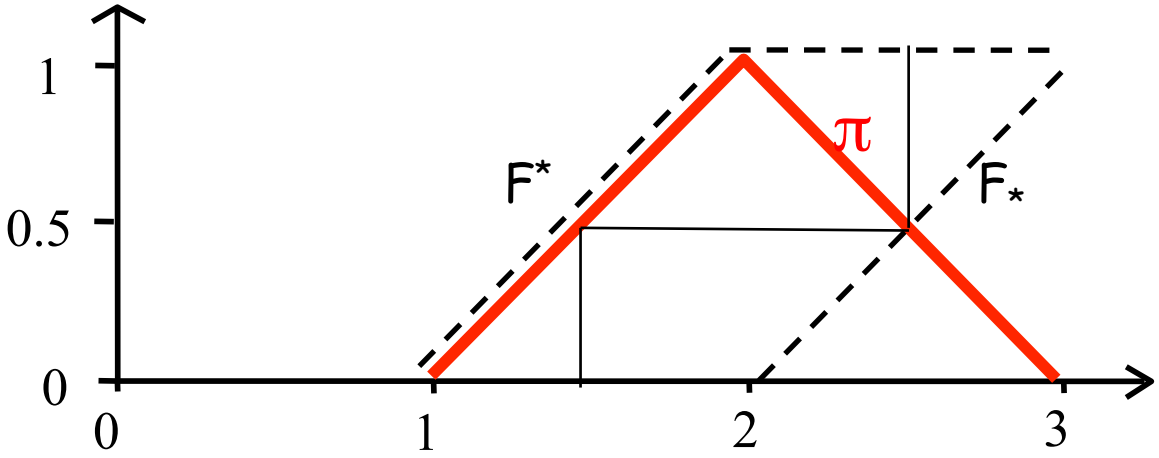
Probability boxes from possibility distributions

- Given a fuzzy interval π
  - $F^*(a) = \Pi_M((-\infty, a]) = \pi(a)$ if $a \leq m$
    $= 1$ otherwise.
  - $F_*(a) = N_M((-\infty, a]) = 0$ if $a < m^*$
    $= 1 - \lim_{x \downarrow a} \pi(x)$ otherwise
- $\mathcal{P}(\pi)$ is a proper subset of $\mathcal{P} = \{P: F^* \geq F_P \geq F_*\}$
  - Not all P in $\mathcal{P}$ are such that $\Pi \geq P$

Representing families of probabilities by fuzzy intervals is more precise than with the corresponding pairs of PDFs

P-boxes vs. fuzzy intervals: counterexample

A triangular fuzzy number with support [1, 3] and mode 2.
Let P be defined by P({1.5})=P({2.5})=0.5.
Then $F_* < F < F^* \; P \notin P(\Pi)$ since
$P(\{1.5, 2.5\}) = 1 > \Pi(\{1.5, 2.5\}) = 0.5$



Practical representations of imprecise probabilities

- Fuzzy intervals (continuous consonant bf)
- Probability intervals (2-monotone only)
- Probability boxes (continuous random intervals with comonotonic bounds)

How useful are these representations:

- P-boxes can address questions about threshold violations ($x \geq a$??),
- not questions of the form $a \leq x \leq b$
- The latter questions are better addressed by possibility distributions or generalized p-boxes

LIKELIHOOD FUNCTIONS

- **Likelihood functions** $\lambda(x) = P(A|\, x)$ behave like possibility distributions when there is no prior on x, and $\lambda(x)$ is used as the likekihood of x.
- It holds that $\lambda(B) = P(A|\, B) \leq \max_{x \in B} P(A|\, x)$
- If $P(A|\, B) = \lambda(B)$ then λ should be set-monotonic:

$\{x\} \subseteq B$ implies $\lambda(x) \leq \lambda(B)$

It implies $\lambda(B) = \max_{x \in B} \lambda(x)$

Then it induces a (consonant) belief function

Maximum likelihood principle is possibility theory

- The classical coin example: θ is the unknown probability of "heads"
- Within n experiments: k heads, n-k tails
- P(k heads, n-k tails | θ) = $\theta^k \cdot (1-\theta)^{n-k}$ is

the degree of possibility $\pi(\theta)$ that the probability of "head" is θ.

In the absence of other information the best choice is the one that maximizes $\pi(\theta)$, $\theta \in [0, 1]$

It yields $\theta = k/n$.

LANDSCAPE OF UNCERTAINTY THEORIES

BAYESIAN/STATISTICAL PROBABILITY: the language of *unique* probability distributions (*Randomized points)*

UPPER-LOWER PROBABILITIES : the language of *disjunctive convex sets of probabilities, and lower expectations*

SHAFER-SMETS BELIEF FUNCTIONS: The language of Moebius masses (*Random disjunctive sets)*

QUANTITATIVE POSSIBILITY THEORY : The language of possibility distributions (*Fuzzy (nested disjunctive) sets)*

BOOLEAN POSSIBILITY THEORY (modal doxastic logic KD) : The language of *Disjunctive sets*

Important pending theoretical issues

- Relative **informativeness**.
- **Conditioning** : several definitions for several purposes.
- **Independence notions**: distinguish between epistemic and objective notions.
- Find a general setting for **information fusion** operations (e.g. Dempster rule of combination).
- **Uncertainty propagation :** Monte-carlo+ interval analysis (not possible for imprecise probabilities)

Comparing belief functions in terms of informativeness

- **Consonant case** : relative specificity.

π' more specific (more informative) than π in the wide sense if and only if $\pi' \leq \pi$.

(any possible value in information state π' is at least as possible in information state π)

- Complete knowledge: $\pi(s_0) = 1$ and $= 0$ otherwise.
- Ignorance: $\pi(s) = 1, \forall\ s \in S$

Comparing belief functions in terms of informativeness

- *Using contour functions*:

$$\pi(s) = Pl(s) = \sum_{x \in E} m(E)$$

m_1 is more cf-informative that m_2 iff $\pi_1 \leq \pi_2$

- *Using belief or plausibility functions* :

m_1 is more pl-informative that m_2 iff $Pl_1 \leq Pl_2$

iff $Bel_1 \geq Bel_2$

It corresponds to comparing credal sets $P(m)$:

$Pl_1 \leq Pl_2$ if and only if $P(m_1) \subseteq P(m_2)$

Comparing commonalities

- M_1 is **more Q-informative** than m_2 iff $Q_1 \leq Q_2$

where $Q(A) = \sum_{A \subseteq E} m(E)$ (antimonotonic)

- Denoted by $m_1 \subseteq_Q m_2$
- *Q is all the larger as m has large focal sets (so m is less informative).*
- Typical information ordering for belief functions used by Smets

Specialisation

- m_1 is *more specialised* than m_2 if and only if There is a joint mass function w(E, F) whose marginals are $m_1(E)$ and $m_2(F)$, such that each w(E, F) = 0 if E is not included in F.
- we have that
  - $m_2(E) = \sum_{F \subseteq E} w(E, F)\, m_1(F)$
  - Any focal set of m_1 is inside at least one focal set of m_2
  - Any focal set of m_2 contains at least one focal set of m_1
  - masses of focal sets of m_2 are shared among focal sets of m_1 that lie in them.

Results

- $m_1 \subseteq_s m_2$ implies $m_1 \subseteq_{Pl} m_2$ implies $m_1 \subseteq_{cf} m_2$
- $m_1 \subseteq_s m_2$ implies $m_1 \subseteq_Q m_2$ implies $m_1 \subseteq_{cf} m_2$
- *However $m_1 \subseteq_{Pl} m_2$ and $m_1 \subseteq_Q m_2$ are not comparable and can contradict each other*
- $m_1 \subseteq_{Pl} m_2$ and $m_2 \subseteq_Q m_1$ imply $\pi_1 = \pi_2$ $(m_1 =_{cf} m_2)$
- In the consonant case : all orderings collapse to $m_1 \subseteq_{cf} m_2$

Example

- $S = \{a, b, c\}$;
- $m_1(ab) = 0.5, m_1(bc) = 0.5$;
- $m_2(abc) = 0.5, m_2(b) = 0.5$
- Neither $m_1 \subseteq_s m_2$ nor $m_2 \subseteq_s m_1$ hold
- $m_2 \subset_{Pl} m_1 : Pl_1(A) = Pl_2(A)$
  but $Pl_2(ac) = 0.5 < Pl_1(ac) = 1$
- $m_1 \subset_Q m_2 : Q_1(A) = Q_2(A)$
  but $Q_1(ac) = 0 < Q_2(ac) = 0.5$
- And contour functions are equal : a/0.5, b/1, c/0.5

Conclusion

- *There exist a coherent range of set-functions in uncertainty theories combining interval and probability representations.*
  - Imprecise probability is the most general setting
  - The choice between subtheories depends on how expressive it is necessary to be in a given application.
  - There exists simple practical representations of belief functions, like possibility distributions and p-boxes (not interval probability masses!)
  - possibility theory is the simplest framework, and belief functions are a good compromise between calculability and expressivity

Uncertainty theories : distinct languages

- *Discrepancies between the theories remain on conditioning, combination rules, because their language primitives differ*:
  - Possibility theory : fuzzy sets of possible values
  - Belief functions : Moebius masses
  - Imprecise probabilities : Convex probability sets
- *It is not always clear how to express notions of one theory in the language of another :* One cannot obviously express concepts defined by Moebius masses using credal sets, and conversely, let alone possibility distributions.

Practical representation issues

- Lower probabilities are difficult to represent ($2^{|S|}$ values): The corresponding family is a polyhedron with potentially |S|! vertices.
- Finite random sets (belief functions) are simpler but potentially $2^{|S|}$ values
- Possibility measures are simple (|S| values) but sometimes not expressive enough.
- ***There is a need for simple and more expressive representations of imprecise probabilities or random sets.***

Conclusion: the nature of imprecise models

- ***Imprecise modeling is unusual***.
  - In classical approaches, a probabilistic model is an approximate but precise representation of variability.
- In contrast, an imprecise model is of higher order, *hence is not objective*.

It models reality and knowledge (of an agent) about reality : reality must just be consistent with the imprecise representation, which depends on the agent.

- There is a need to reconsider the foundations of systems analysis in this perspective.

Some references

- B. De Finetti. Theory of probability. Wiley, New York, 1974.
- G. L.S. Shackle. Expectations in Economics, Cambridge, University Press, UK, 1949.
- G. Shafer. A Mathematical Theory of Evidence. Princeton Univ. Press, 1976.
- L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1:3–28,1978.
- D. Dubois and H. Prade. Théorie des Possibilités, Masson, Paris, 1987 (translation: Possibility theory., Plenum, New York, 1988).
- P. Walley. Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, 1991
- P. Smets and R. Kennes. The transferable belief model. Artificial Intelligence, 66:191–234,1994.
- G. Coletti and R. Scozzafava. Probabilistic Logic in a Coherent Setting. Kluwer Academic Pub, 2002.
- T. Augustin, F. Coolen, G. de Cooman, M. Troffaes (Eds.) Introduction to Imprecise Probabilities., Wiley, 2014.

Speaker References

- D. Dubois, H. Prade. Formal representations of uncertainty. Decision-making Process- Concepts and Methods. D. Bouyssou, et al. (Eds.), ISTE London & Wiley, Chap. 3, p. 85-156, 2009.
- S. Destercke D.Dubois, E. Chojnacki, Unifying practical uncertainty representations – Part I: Generalized p-boxes. International Journal of Approximate Reasoning, 49, Issue 3, 2008, 649-663; Part II: Clouds. International Journal of Approximate Reasoning, 49, 2008, 664-677.
- I. Couso, D.Dubois, L. Sanchez. *Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables*, Springer, SpringerBriefs in Computational Intelligence, 2014.
- I. Couso, D. Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views (with discussion). International Journal of Approximate Reasoning,Vol. 55 N. 7, p. 1502-1518, 2014.
- S. Destercke, D.Dubois (2014) Other uncertainty theories based on capacities. In : Introduction to Imprecise Probabilities. Thomas Augustin, Frank Coolen, Gert de Cooman, Matthias Troffaes (Eds.), Wiley, p. 93-113