Special cases of lower previsions and their use in statistics

I. Couso D. Dubois

IRIT-CNRS, Université Paul Sabatier

31062 TOULOUSE FRANCE

Outline

1. Set-valued representations of ignorance
2. Capacity-based uncertainty theories and their links to imprecise probability
3. Practical representations
4. Statistics with interval data

(From the simplest to the more complex representations of uncertainty)

Motivation for going beyond probability

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

Example

- **Variability**: daily quantity of rain in Toulouse
  - 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!
  - Information 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.

What do set-valued data mean?

- A set can represent
  - the precise description of an actual object (**ontic set**) : a region in an image.
  - or incomplete information about an ill-known entity (**epistemic set**) : interval containing an ill-known birth-date.
- The ill-known entity can be
  - A constant ($x \in E$)
  - or a random variable ($P\_x \in \{P: P(E) = 1\}$).

Set-Valued Representations of Partial Knowledge

- An ill-known quantity x is represented as a disjunctive set, i.e. a subset E 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
  - **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)

BOOLEAN POSSIBILITY THEORY

Natural set functions under incomplete information*:*

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

- Event A is possible if $A \cap E \neq \emptyset$ (logical consistency)

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

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

- Event A is sure if $E \subseteq A$ (logical deduction)

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

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

$$N(A) = 1 - \Pi(A^c) : N(A) = 1 \text{ iff } \Pi(A^c) = 0$$

$$N(A) \leq \Pi(A)$$

This corresponds to a fragment of a modal logic (KD)

Representations of uncertainty due to incompleteness

- *More expressive than epistemic sets (pure intervals or classical logic), and Boolean possibility theory*
- *Less demanding than single probability distributions*
- *Explicitly allows for missing information*
- *Allows for addressing the same problems as probability.*

Possibility Theory

(Shackle, 1961, Zadeh, 1978)

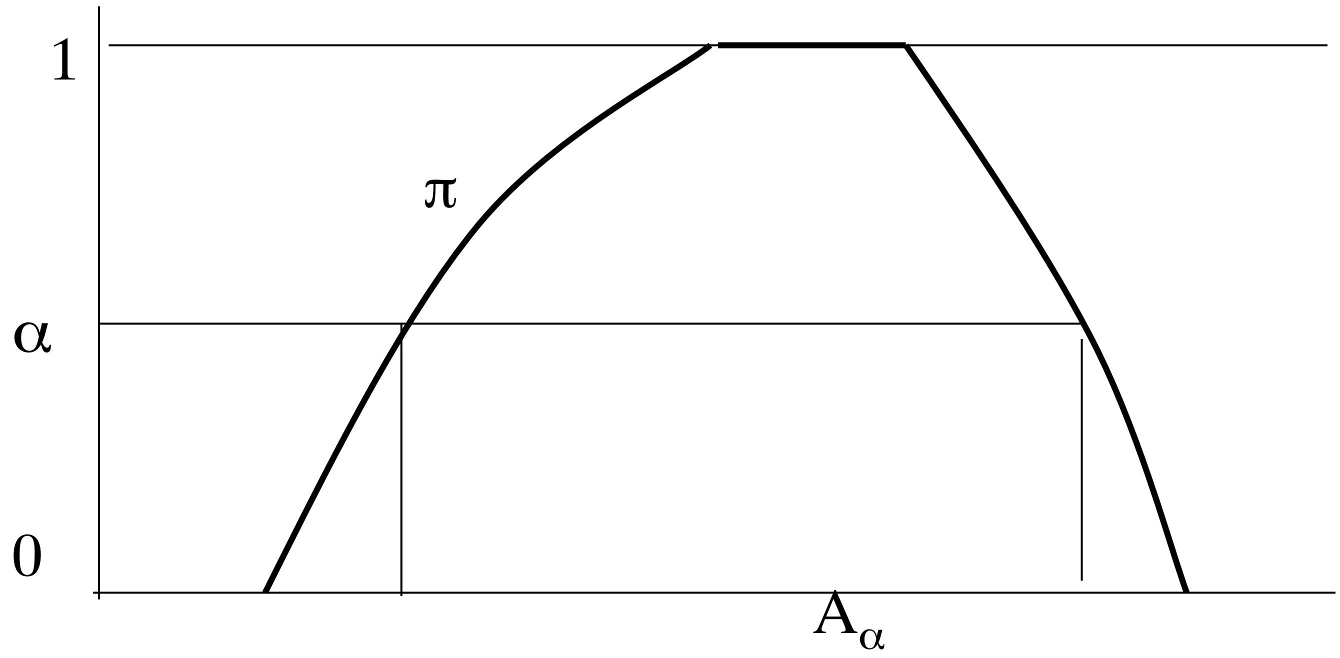
- A piece of incomplete information "$x \in E$" admits of *degrees* of possibility: $E \subseteq S$ is a (normalized) fuzzy set : $\mu_E : S \to [0, 1]$
- $\mu_E(s)$ = Possibility($x = s$) = $\pi_x(s)$ in $[0, 1]$
- $\pi_x(s)$ is the degree of plausibility of $x = s$
- **Conventions:** $\pi_x(s) = 1$ for some value 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 family of nested epistemic sets

In the continuous case: α *= Poss (x not in* A_α*)*



FUZZY INTERVAL

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 function μ_{OLD} is interpreted as a possibility distribution on possible birth dates (Zadeh).

- **Nested intervals** $E_1, E_2, \ldots E_n$ **with confidence levels**

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

- $\Pi(A) = \max_{s \in A} \pi(s)$:

  to what extent A is consistent with π

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

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

- $N(A) = 1 - \Pi(A^c) = \min_{s \notin A} 1 - \pi(s)$:

  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 (finite case)

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

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

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

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

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

Comparing information states

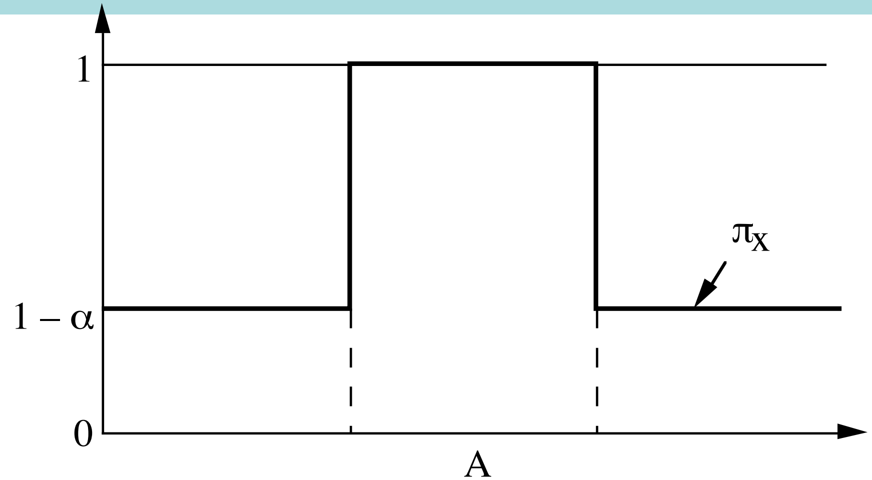
- **π' more specific than π in the wide sense if and only if $\pi' \leq \pi$**

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: $\pi(s) = 1, \forall s \in S$

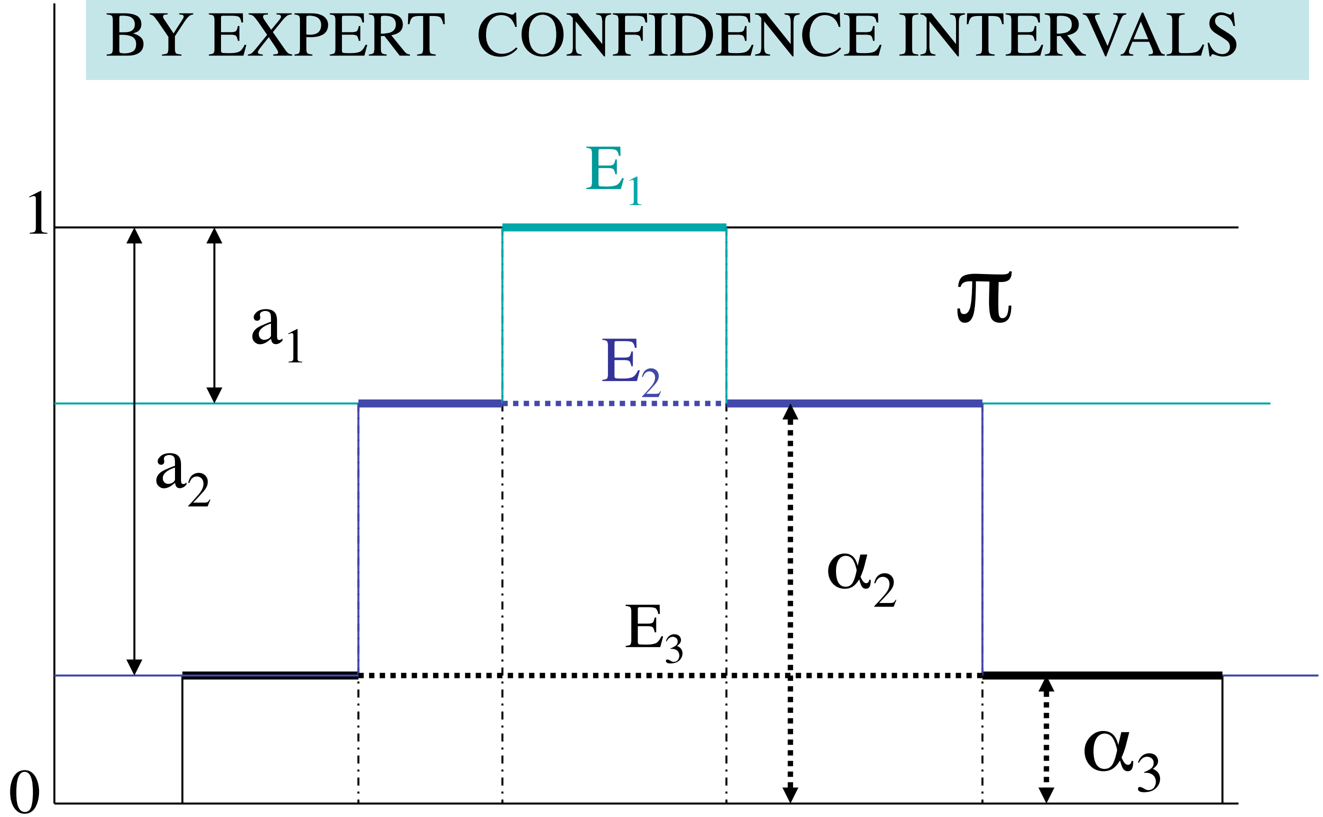
- **Principle of least commitment** (minimal specificity): In a given information state, any value not proved impossible is supposed to be possible : maximise possibility degrees.

Certainty-qualification



- Attaching a degree of certainty α to event A
- It means $N(A) \geq \alpha \Leftrightarrow \Pi(A^c) = \sup_{s \notin A} \pi(s) \leq 1 - \alpha$
- The least informative π sanctioning $N(A) \geq \alpha$ is :
  - $\pi(s) = 1$ if $s \in A$ and $1 - \alpha$ if $s \notin A$
- In other words: $\pi(s) = \max(\mu_A, 1 - \alpha)$

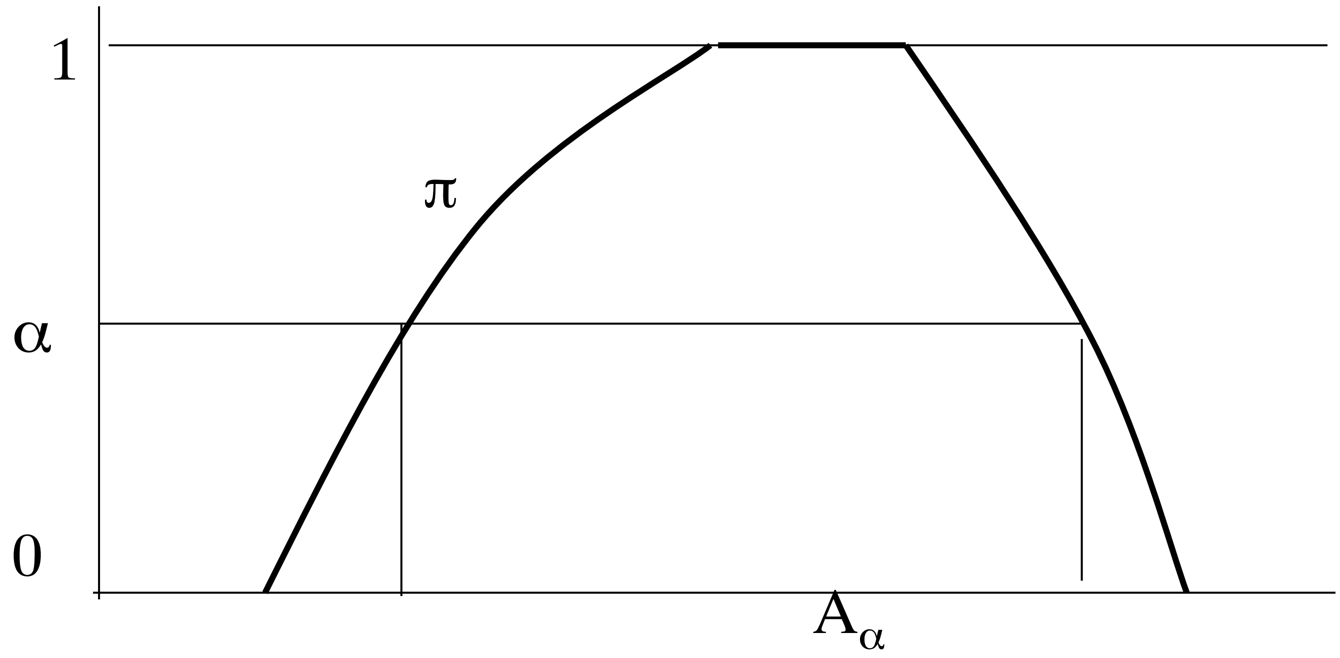
POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



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

At the limit with an infinity of nested intervals

$N(A_\alpha) \geq 1 - \alpha$, α in $(0, 1]$



FUZZY INTERVAL

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 = $1 - \Pi(A)$.

- 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

Qualitative vs. quantitative possibility theories

- **Qualitative**:
  - **comparative**: A complete pre-ordering $\geq_\pi$ on S   A well-ordered partition of S: 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

Quantitative possibility theory

- **Membership functions of fuzzy sets**
  - Natural language descriptions pertaining to numerical universes (fuzzy numbers)
  - Results of fuzzy clustering

  Semantics: metrics, proximity to prototypes
- **Imprecise probability**
  - Random experiments with imprecise nested outcomes
  - Possibility distributions encode special convex probability sets

  Semantics: frequentist, or subjectivist (gambles)...

Blending intervals and probability

- Representations that refine Boolean possibility theory and account for both variability and incomplete knowledge must combine 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

GRADUAL REPRESENTATIONS OF UNCERTAINTY using capacities

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 $\mathcal{E}$ to [0,1] such that
  - $g(\emptyset) = 0 \quad ; \quad g(S) = 1$
  - **monotony** : if A $\subseteq$ B (=A implies B) then g(A) $\leq$ g(B)
- g(A) quantifies the confidence of an agent in proposition A.
- g is a Choquet capacity

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.

Imprecise probability theory

- A state of information is represented by a family $\mathcal{P}$ of probability distributions over a set X.
- *For instance: incomplete knowledge of a frequentist probabilistic model :* $\exists P \in \mathcal{P}$.
- 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 $\{P(A), P \geq P_*\}$
- $\{P(A), P \geq P_*\}$ is convex (credal set).

WHY REPRESENTING INFORMATION BY PROBABILITY FAMILIES ?

Often probabilistic information is incomplete:

- Expert opinion (fractiles, intervals with confidence levels)
- Subjective estimates of support, mode, etc. of a distribution
- Parametric model with incomplete information on parameters (partial subjective information on mean and variance)
- Parametric model with confidence intervals on parameters due to a small number of observations

WHY REPRESENTING INFORMATION BY PROBABILITY FAMILIES ?

- In the case of generic (frequentist) information using a family of probabilistic models, rather than selecting a single one, enables to account for incompleteness and variability.
- In the case of subjective belief: distinction between
  - not believing a proposition ($P_*(A)$ and $P_*(A^c)$ low)
  - and believing its negation ($P_*(A^c)$ high).

Subjectivist view (Peter Walley)

- *A theory that handles convex probability sets*
  - $P_{low}(A)$ is the highest acceptable price for buying a bet on singular 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.
  - These prices may differ (no exchangeable bets)
- **Rationality** conditions:
  - **No sure loss** : $\{P \geq P_{low}\}$ not empty
  - **Coherence**: $P_*(A) = \inf\{P(A), P \geq P_{low}\} = P_{low}(A)$
- *Convex probability sets (**credal sets**) are actually characterized by lower expectations of real-valued functions (gambles), not just events.*

Capacity-based lower probabilities

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

$$P_*(A) + P_*(B) \leq P_*(A \cup B)$$

- One may require the 2-monotony property for Cr:

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

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

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

Cr is then called a convex capacity.

Coherence and deductive closure

- Suppose the knowledge is of the form of a consistent set $\mathcal{B}$ of assertions ϕ_i of the form

  « x in E_i » $i = 1, \ldots, n$ (interpreted as $N(E_i) = 1$)

- The set of consequences of $B = \{\phi_i \; i = 1, \ldots, n\}$ is $\mathcal{C}(\mathcal{B}) = \{\phi | \; \mathcal{B} \models \phi\}$ (deductively closed)
- Define a Boolean necessity function N* such that $N^*(A) = 1$ iff $\phi =$ « x in A » in $\mathcal{C}(\mathcal{B})$ iff $E = \cap_{i = 1, \ldots, n} E_i \subseteq A$

Coherence and deductive closure

- If the knowledge $\mathcal{B}$ is viewed as the credal set $\{P: P(E_i) = 1, i = 1, \ldots, n\}$ then the coherent lower probability induced by its natural extension is the Boolean necessity function N*, obtained from the deductive closure $\mathcal{C}(\mathcal{B})$, which is another example of coherent lower probability.
- **Conclusion** Coherence generalizes deductive closure, and a consequence of B is a formula whose set of models has lower probability 1.

Random sets

- A probability distribution *m* on the family of non-empty subsets of a set S.
- A positive weighting of non-empty subsets: mathematically, **a random set** :

$$\sum_{E \in \mathcal{F}} m(E) = 1$$

- m : *mass function*.
- *focal sets* : $E \in \mathcal{F}$ with $m(E) > 0$.

Disjunctive random sets

- m(E) = probability that the most precise description of the available information is of the form "$x \in E$" for epistemic set E.

  It is the probability of [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 P(E)**

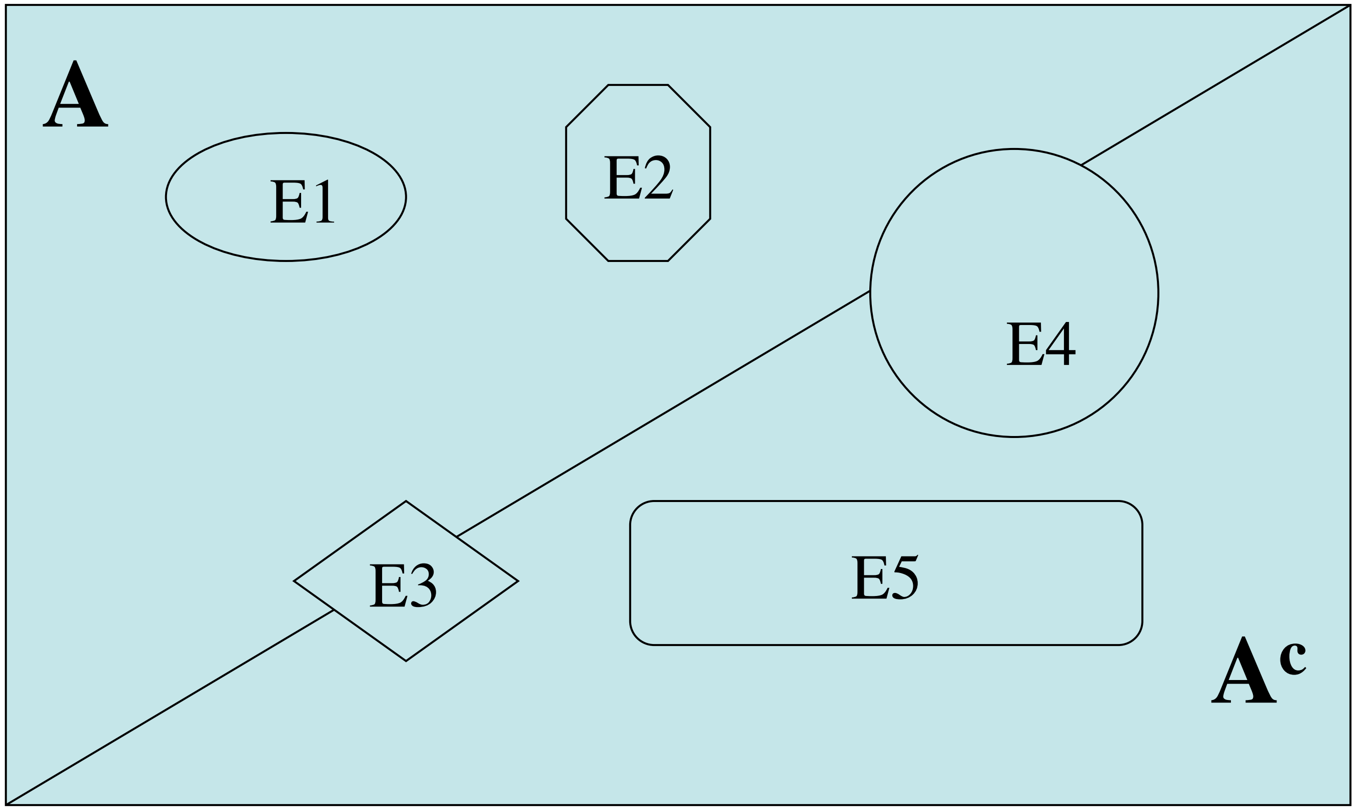
Basic set functions from random sets

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

Example : Bel(A) = m(E1) + m(E2)

Pl(A) = m(E1) + m(E2) + m(E3) + m(E4)

= 1 – m(E5) = 1 – Bel(A^c)



Random disjunctive sets vs. imprecise probabilities

- The set $\mathcal{P}_{\mathrm{bel}} = \{P \geq \mathrm{Bel}\}$ is coherent: Bel is a special case of lower probability
- Bel is ∞-monotone (super-modular at any order)
  - Order 3: $\mathrm{Bel}(A \cup B \cup C) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B) + \mathrm{Bel}(C) - \mathrm{Bel}(A \cap B) - \mathrm{Bel}(A \cap C) - \mathrm{Bel}(B \cap C) + \mathrm{Bel}(A \cap B \cap C)$, 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 (Moebius transform)
  - ***However m is positive iff g is a belief function***

PARTICULAR CASES

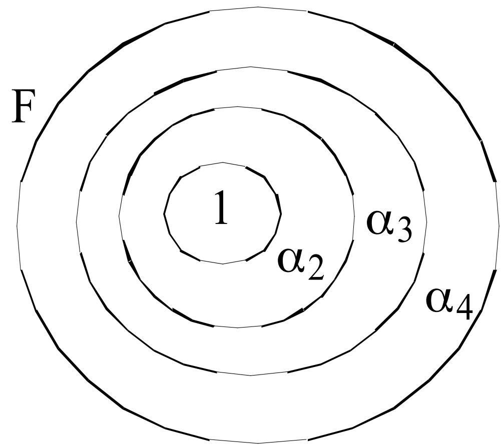
- INCOMPLETE INFORMATION:

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

- *TOTAL IGNORANCE :* $m(S) = 1$:
  - *For all* $A \neq S, \emptyset$, $Bel(A) = 0$, $Pl(A) = 1$
- PROBABILITY: if $\forall i, E_i$ = singleton $\{s_i\}$ (hence disjoint focal sets)
  - Then, *for all* A, $Bel(A) = Pl(A) = P(A)$
  - *Hence precise + scattered information*
- POSSIBILITY THEORY : the opposite 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

From possibility to random sets



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

- Given π, *construct a basic probability assignment (SHAFER)*

  let $m_i = \alpha_i - \alpha_{i+1}$ then $m_1 + \ldots + m_n = 1$,

  with focal sets = cuts $A_i = \{s, \pi(s) \geq \alpha_i\}$

  $Bel(A) = \sum_{Ai \subseteq A} m_i = N(A)$; $Pl(A) = \Pi(A)$

- Conversely, $\pi(s) = \sum_{i:\ s \in Ai} m_i$ (one point-coverage function)

  $= Pl(\{s\})$.

- ***Only in the consonant case can m be recalculated from π***

Canonical examples

- **Objectivist** : Frequentist modelling of a collection of incomplete observations (imprecise statistics) :
- **Uncertain subjective information:**
  - **Unreliable testimonies** (Shafer's book) : human-originated singular information
- **Unreliable sensors** : the quality/precision of the information depends on the ill-known sensor state.

Random sets as epistemic sets of random variables

- **Dempster model** : Indirect information (induced from a probability space).
- All we know about a random variable x with range S, based on a sample space $(\Omega, \mathcal{A}, P)$ is based on 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 $x(\omega) \in \Gamma(\omega)$**

$$m(E) = \sum \{P(\{\omega\}) : E = \Gamma(\omega)\} \quad \forall\ \omega \text{ in } \Omega$$

(finite case.)

Consult for more

- **Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables**

An Introduction for Ph.D. Student and Practitioners

By Inés Couso, Didier Dubois, Luciano Sanchez

SpringerBriefs in Applied Sciences and Technology, 2014

- Inés Couso, Didier Dubois, Statistical Reasoning with Set-Valued Information: Ontic vs. Epistemic Views. Int. J. Approximate Reasoning, 2014

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

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

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

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

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

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

- $\mu_F(x)$ is an upper bound of the probability that x is elected. It gathers responses of those who *did not give up voting* for x

- $\mathrm{Bel}(\{x\})$ 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 "$x \in E$ =[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.

Unreliable testimony with lies

- « 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».
  - *John may lie* (probability q):
  - $E = [60, 70]$

- Modeling
  - John is competent and does not lie : $m(E) = (1 - p)(1 - q)$,
  - John is competent and lies $m(E^c) = (1 - p)q$.
  - John is incompetent and is boasting : $m(S) = p$

Dempster vs. Shafer-Smets

- A disjunctive random set can represent
  - *Uncertain singular evidence* (unreliable testimonies): m(E) = subjective probability pertaining to the truth of testimony E.
    - Degrees of belief directly modelled by Bel : no appeal to an underlying probability.

  (Shafer, 1976 book; Smets)

  - *Imprecise statistical evidence*: m(E) = frequency of imprecise observations of the form E and Bel(E) is a lower probability
    - A multiple-valued mapping from a probability space to a space of interest representing an ill-known random variable.
    - Here, belief functions are explicitly viewed as lower probabilities

  (Dempster intuition)

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

Example of conjunctive random sets

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) $= \sum\{m(E) : A \subseteq E\} = Q(A)$: commonality function in belief function theory
- **Example**: « x speaks English » means « at least English »
- The belief function is not meaningful here while the commonality makes sense, contrary to the disjunctive set case.

POSSIBILITY AS UPPER PROBABILITY

- Given a numerical possibility distribution π, define
  $\mathcal{P}(\pi) = \{P \mid P(A) \leq \Pi(A) \text{ for all } A\}$

- Then, generally it holds that

  $\Pi(A) = \sup \{P(A) \mid P \in \mathcal{P}(\pi)\}$;

  $N(A) = \inf \{P(A) \mid P \in \mathcal{P}(\pi)\}$
- So N and P are special cases of coherent lower and upper probabilities
- So π is a very simple representation of a credal set (convex family of probability measures)

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 $P(A| B) = \lambda(B)$ is the likelihood of "$x \in B$" then λ should be a capacity (monotonic with inclusion):

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

It implies $\lambda(B) = \max_{x \in B} \lambda(x)$ if no prior probability is available for x.

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 logic KD) :
The language of *Disjunctive sets*

Language difficulties

- *Imprecise probability, belief functions and possibility theory are in fact not fully mutually consistent:*
  - *Concepts that make sense for credal sets, may be hard to interpret in terms of Moebius transforms or possibility distributions and conversely*
  - *Simplified representations help us cut down computation costs (possibility distributions and simple belief functions)*

Practical representations

- Fuzzy intervals
- Probability intervals
- Probability boxes
- Generalized p-boxes
- Clouds

Some are special random sets some not.

Probability intervals (De Campos, Moral)

- **Probability intervals** = a finite collection L of imprecise assignments $[l_i , u_i]$ attached to elements s_i of a finite set S.
- A collection $L = \{[l_i , u_i]\ i = 1,\dots\ n\}$ induces the family $\mathcal{P}_L = \{P\colon l_i \leq P(\{s_i\}) \leq u_i\}$.
- A probability interval model L is **coherent** in the sense of Walley if and only if
  - $\sum_{j \neq i} l_j + u_i \leq 1$ and $1 \leq \sum_{j \neq i} u_j + l_i$
- Lower/upper probabilities on events are given by
  - $P_*(A) = \max(\Sigma_{si\in A}\ l_i\ ;\ 1 - \Sigma_{si\notin A}\ u_i)$;
  - $P^*(A) = \min(\Sigma_{si\in A}\ u_i\ ;\ 1 - \Sigma_{si\notin A}\ l_i)$
- P_* is a 2-monotone Choquet capacity (De Campos and Moral)

From probabilistic confidence sets to possibility distributions

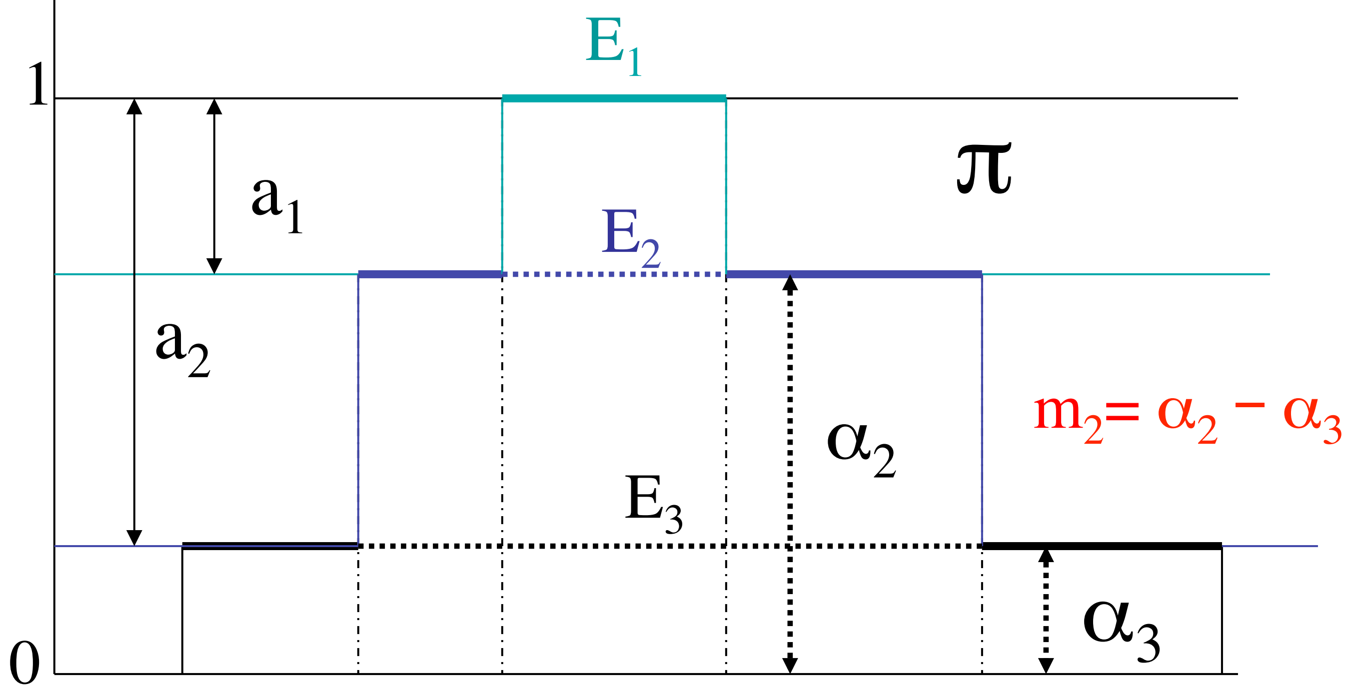
- Let $E_1, E_2, \ldots E_n$ be a nested family of sets
- A set of confidence levels $a_1, a_2, \ldots a_n$ in $[0, 1]$
- Consider the set of probabilities

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

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

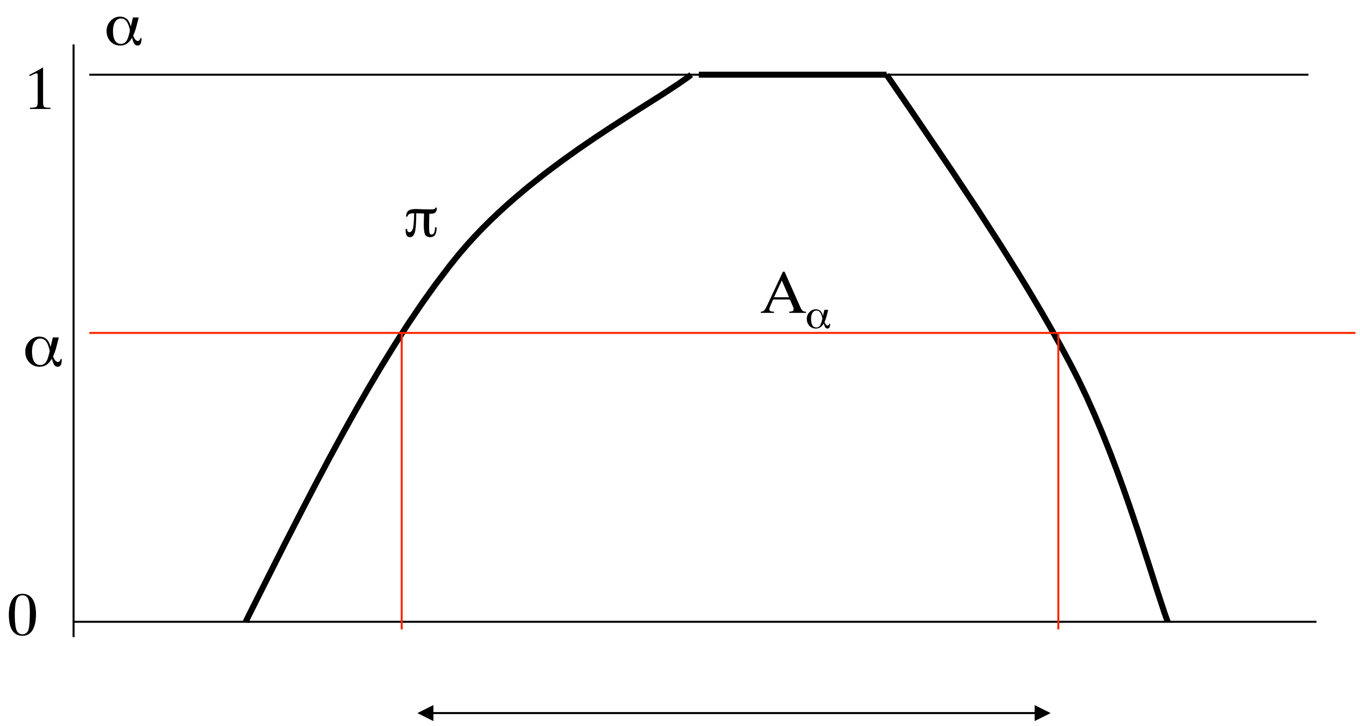
$$\pi(x) = \min_{i = 1, \ldots n} \max(\mu_{E_i}(x), 1 - a_i)$$

POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



A possibility distribution can be obtained from any family of nested confidence sets and defines the credal set

$\{P \colon P(A_\alpha) \geq 1 - \alpha, \alpha \in (0, 1]\}$



FUZZY INTERVAL: $N(A_\alpha) = 1 - \alpha$

Possibilistic view of probabilistic inequalities

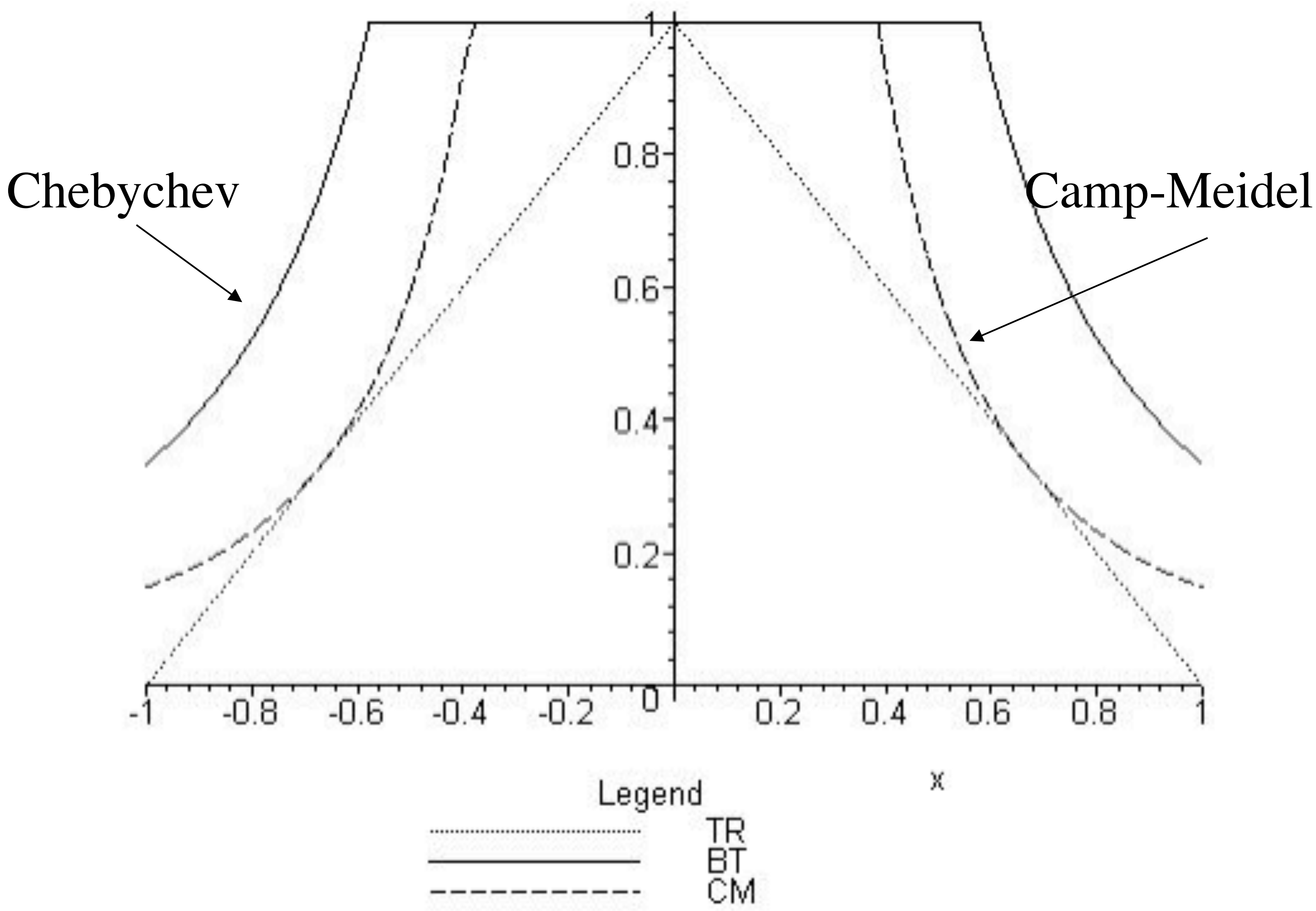
Probabilistic inequalities can be used for knowledge representation:

- Chebyshev inequality defines a possibility distribution that dominates *any* density with given mean and variance.
- Choosing sets $[x^{mean} - k\sigma, x^{mean} + k\sigma]$, k >0

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

is equivalent to writing

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


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

Possibilistic view of probabilistic inequalities 2

Probabilistic inequalities can be used for knowledge representation:

- Choosing mode, bounded support $[x_*, x^*]$ and sets E_α of the form

$$[x^{mode} - (1-\alpha)(x^{mode} - x_*),\ x^{mode} + (1-\alpha)(x^* - x^{mode})]$$

- *$P(V \in E_\alpha) \geq 1 - \alpha$ is equivalent to defining a triangular fuzzy interval (TFI)*

$$\pi(x^{mode} - (1-\alpha)(x^{mode} - x_*)) = \pi(x^{mode} + (1-\alpha)(x^* - x^{mode})) = \alpha$$

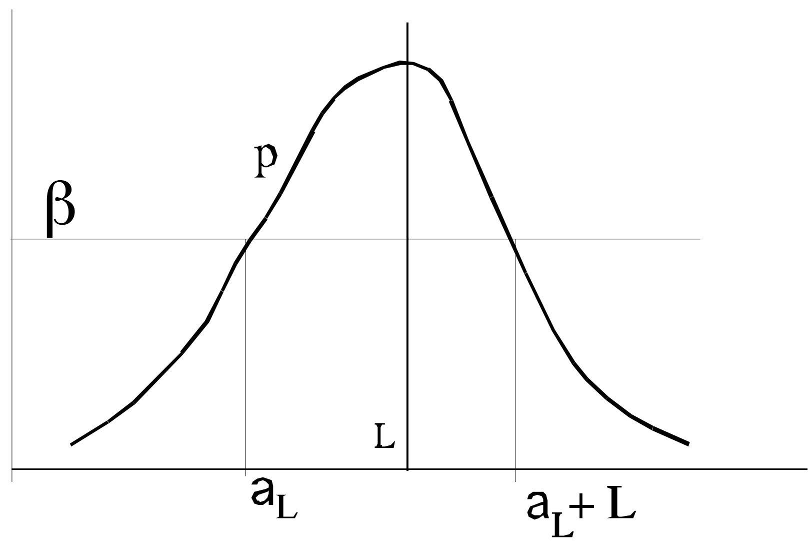
A TFN defines a possibility distribution that dominates *any* unimodal density with the same mode and bounded support as the TFN.

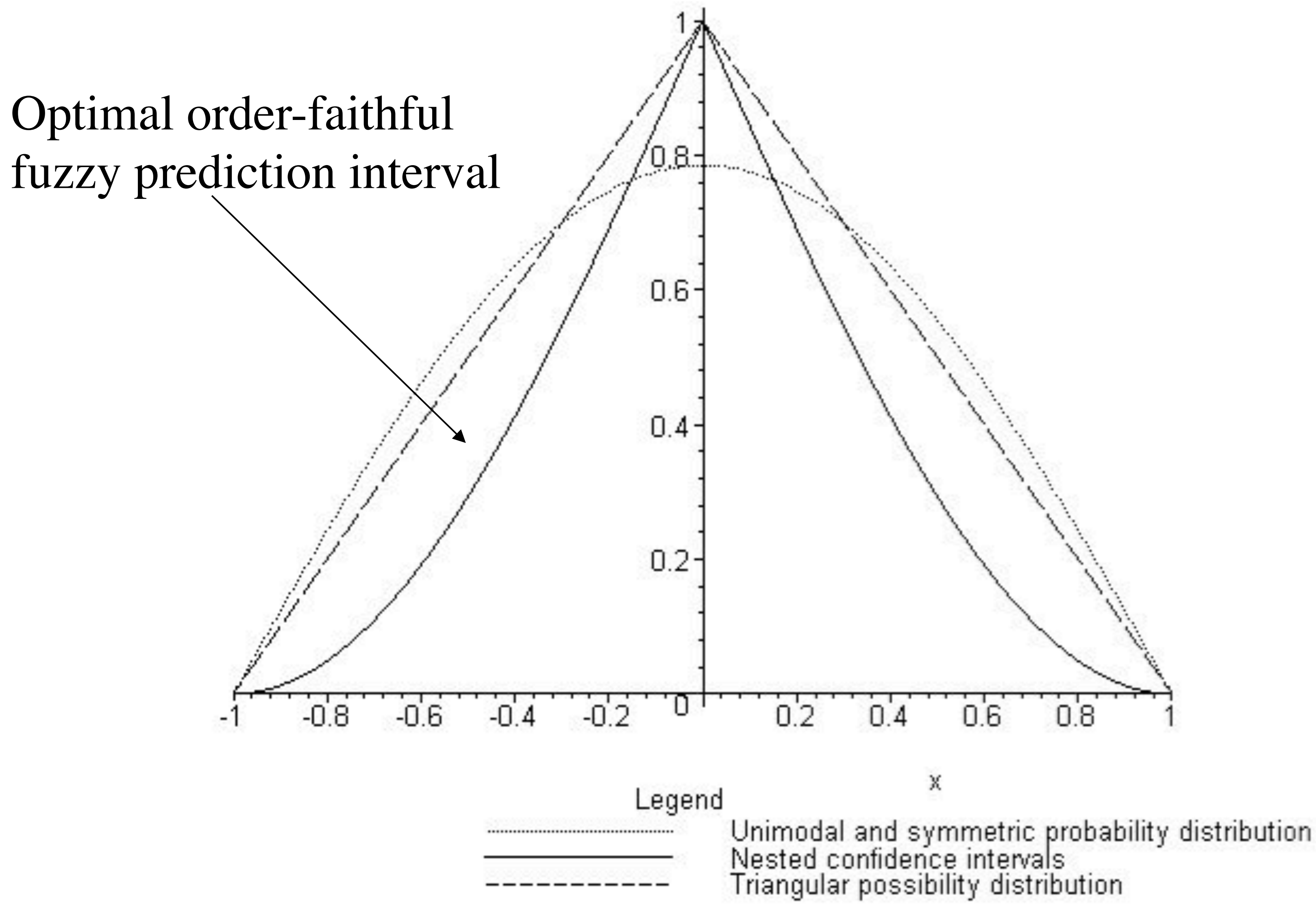
Optimal order-faithful fuzzy prediction intervals

- The interval $I_L = [a_L, a_L + L]$ of fixed length L with maximal probability is of the form $\{x, p(x) \geq \beta\}$
- The most narrow prediction interval with probability α is of the form $\{x, p(x) \geq \beta\}$
- So the most natural (narrow) possibility counterpart of p is

  $\pi_p(a_L) = \pi_p(a_L + L) =$
  $1 - P(I_L = \{x, p(x) \geq \beta\})$.

Such that $\Pi(A) \geq P(A)$ for all




Optimal order-faithful fuzzy prediction interval
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
Unimodal and symmetric probability distribution
Nested confidence intervals
Triangular possibility distribution

Applications of the prob->pos transform

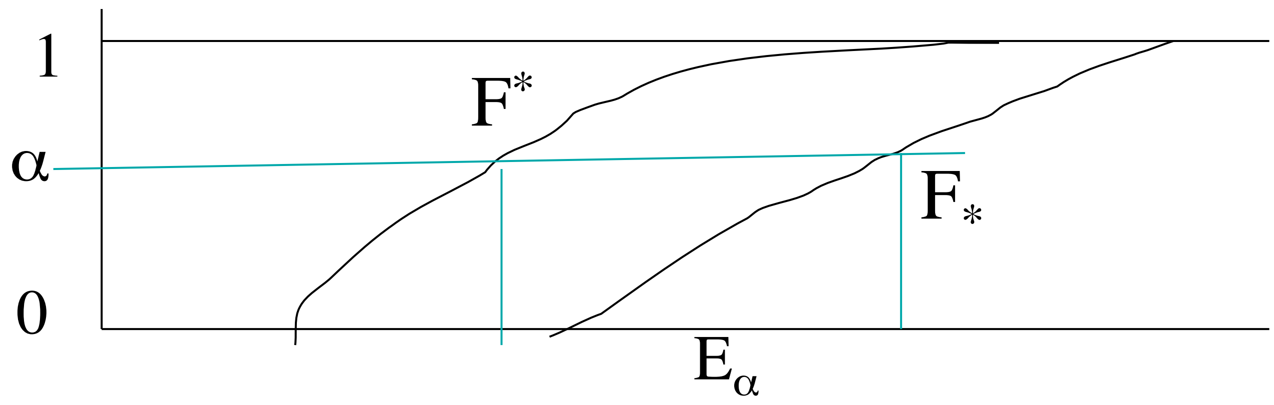
- Extraction of most narrow confidence of prediction intervals for all confidence levels
- Representing insufficient statistical data by a simple credal set.
- Comparing pdfs according to their dispersions (entropy) :

$$\pi_p \geq \pi_q \textit{ implies } Ent(p) \leq Ent(q)$$

(it works even for densities with infinite variance)

Probability boxes

- A set $\mathcal{P} = \{P\colon F^* \geq P \geq F_*\}$ 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 lower bounds induce the same ordering*.



Probability boxes from possibility distributions

- $F^*(a) = \Pi_M((-\infty, a]) = \pi(a)$ if $a \leq m$
  $= 1$ otherwise.
- $F_*(a) = N_M((-\infty, a]) = 0$ if $a < m^*$
  $= 1 - \lim_{x \downarrow a} \pi(x)$ otherwise

• *Representing families of probabilities by fuzzy intervals is more precise than with the corresponding pairs of PDFs:* $\mathcal{P}(\pi)$ is a proper subset of $\mathcal{P} = \{P: F^* \geq P \geq F_*\}$

- Not all P in $\mathcal{P}$ are such that $\Pi \geq P$

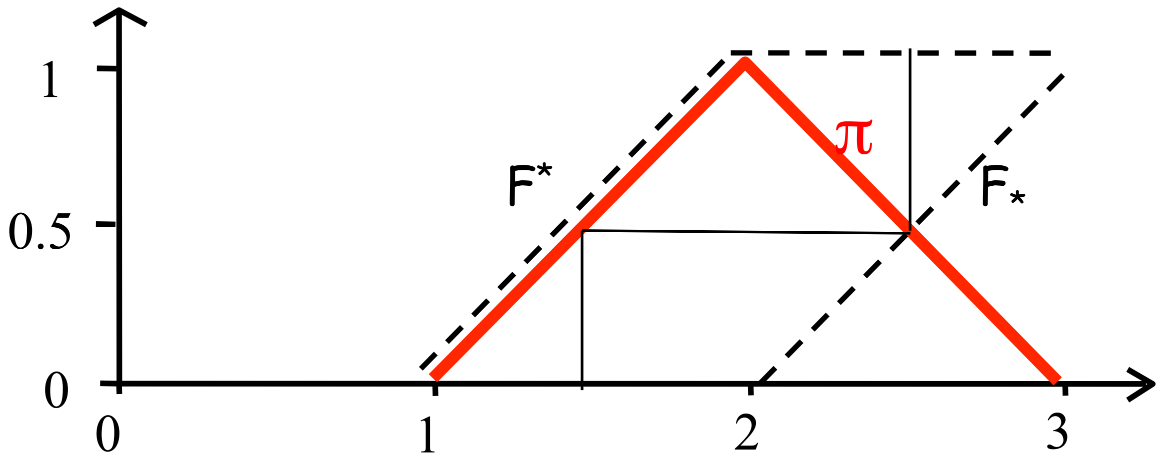
P-boxes vs. fuzzy intervals

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

$P(\{1.5, 2.5\}) = 1 > \Pi(\{1.5, 2.5\}) = 0.5$



Cumulative distributions

- A Cumulative distribution function F
  $F(x) = P(\{X \leq x\})$

of a probability function P can be viewed as a possibility distribution dominating P since the sets $\{X \leq x\}$ are nested

- in particular, $\sup\{F(x), x \in A\} \geq P(A)$
- Fuzzy intervals can be viewed as cumulative distribution functions with different kinds of nested sets as $\{X \leq x\}$

Generalized p-boxes

- Consider nested confidence intervals $E_1, E_2, \ldots E_n$ each with two probability bounds α_i and β_i such that

  $\mathcal{P} = \{\alpha_i \leq P(E_i) \leq \beta_i \text{ for } i = 1, \ldots, n\}$
- It comes down to two possibility distributions

  π (from $\alpha_i \leq P(E_i)$)

  and π_c (from $P(E_i^c) \geq 1 - \beta_i$)
- $\pi(x) = \min_{i = 1, \ldots n} \max (\mu_{Ei}(x), 1 - \alpha_i)$
- $\pi_c(x) = \min_{i = 1, \ldots n} \max (1 - \mu_{Ei}(x), \beta_i)$

We get a p-box if $E_i = \{x \leq a_i\}$

Generalized p-boxes

- Since $\alpha_i \leq \beta_i$, distributions π and π_c are such that
  - $\pi(x) \geq 1 - \pi_c(x) = \delta(x) = \max_{i = 1, \ldots n} \min (\mu_{Ei}(x), 1- \beta_i)$
  - and **π is comonotonic with δ** (they induce the same order of values x).

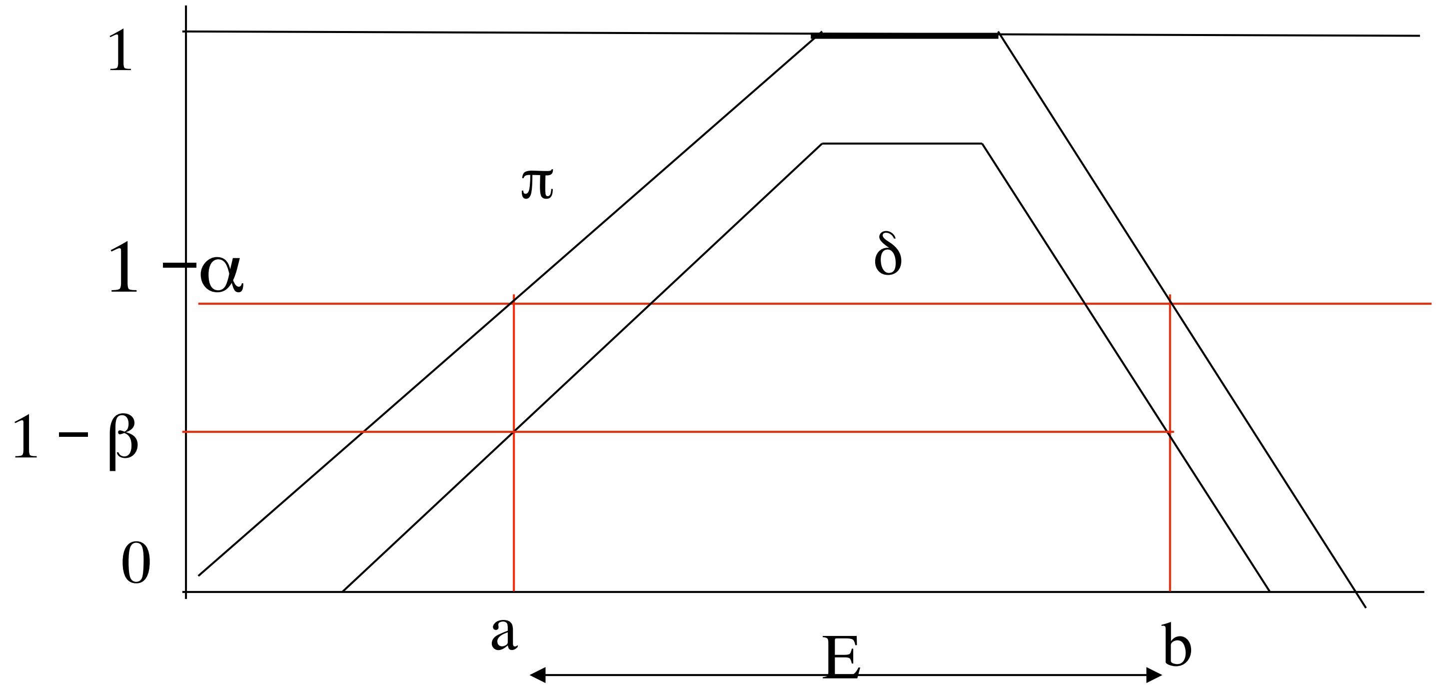
Credal set : $\mathcal{P} = \mathcal{P}(\pi) \cap \mathcal{P}(\pi_c)$

- **Theorem**: a generalized p-box is a belief function (random set) with focal sets

$$\{x: \pi(x) \geq \alpha\} \setminus \{x: \delta(x) > \alpha\}$$

If $\delta(x) = 0$: usual possibility distribution

$$\pi(a) = \pi(b) = 1 - \alpha \ ;$$
$$\delta(a) = \delta(b) = 1 - \beta$$



Generalized p-box

Elementary example of a generalized p-box

- All that is known is that P(E) in [a, b] on a finite set E of S
- It corresponds to the belief function :
- $m(E) = a$; $m(E^c) = 1 - b$; $m(S) = b - a$.
- The two possibility distributions :
  - $\pi(s) = 1$ if s in E; $1-a$ otherwise.
  - $\pi_c(s) = 1$ if s in E^c; b otherwise.
- The generalized p-box $(\pi_{1,}\ 1 - \pi_c)$

From generalized p-boxes to clouds

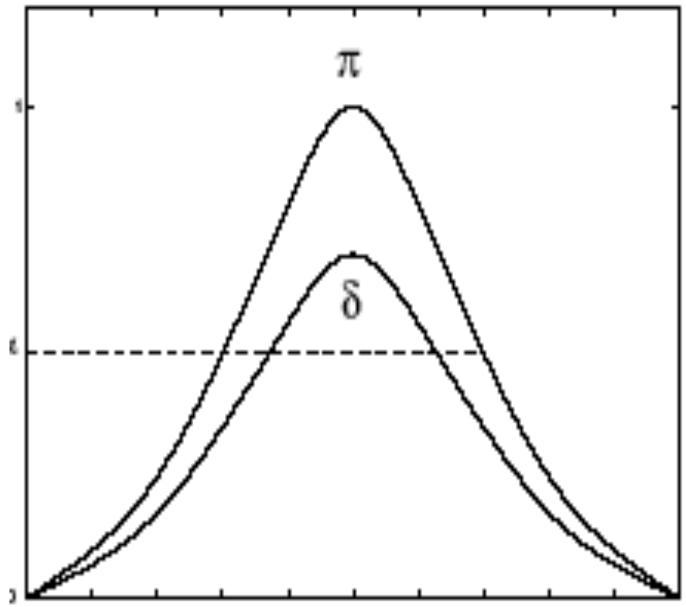


Fig 1.A Comonotonic cloud

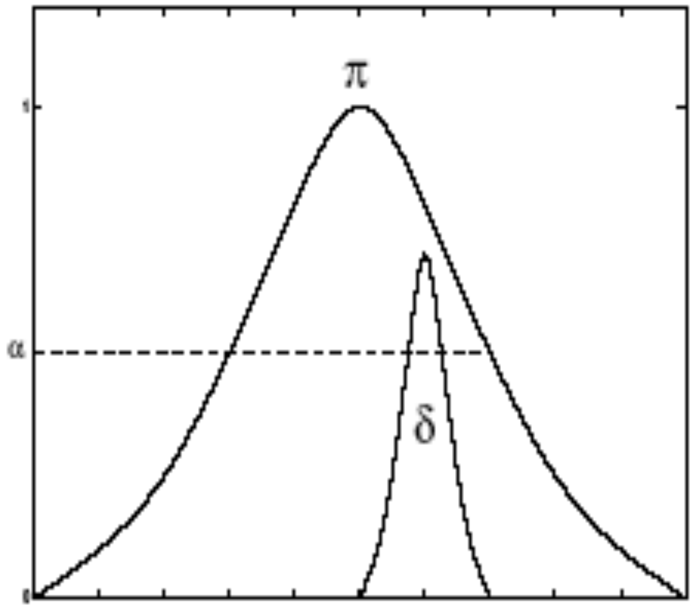


Fig 1.B Non-comonotonic cloud

How useful are these representations:

- Can help elicitating credal sets from data or experts, and summarizing outputs of an imprecise probability method.
- Usual 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

Relationships between representations

- Generalized p-boxes are special random sets that generalize BOTH p-boxes and possibility distributions
- Clouds extend G. P-boxes but induce lower probabilities that are not even 2-monotonic.
- Probability intervals are not comparable to generalized p-boxes: they induce lower probabilities that are 2-monotonic

Important pending theoretical issues

- Comparing representations in terms of **informativeness**.
- **Conditioning** : several definitions for several purposes in the various special cases.
- **Independence notions**: distinguish between epistemic and objective notions.
- Find a general setting for **information fusion** operations (e.g. Dempster rule of combination).

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

- *1. Using contour functions*:

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

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

- Corresponds to the specificity ordering in the consonant case
- Degree of imprecision

$$|m| = \sum_{E} m(E) * |E| = \sum_{s \in S} \pi(s)$$

- $\pi_1 \leq \pi_2$ implies $|m_1| \leq |m_2|$

Comparing belief functions in terms of informativeness

- 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

$$\mathcal{P}(m) = \{P \geq Bel\}:$$

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

Comparing belief functions in terms of informativeness

- 3. *Comparing commonality functions*:
  m_1 is more Q-informative that m_2 iff

$$m_1 \subseteq_Q m_2 \text{ iff } Q_1 \le Q_2$$

where $Q(A) = \sum_{A \subseteq Ei} m(E_i)$

- There are larger focal sets for m_2 than for m_1
- A typical information ordering for belief functions.

Specialisation

- 4. *m_1 is more specialised than m_2 iff*
  - Any focal set of m_1 is included in at least one focal set of m_2
  - Any focal set of m_2 contains at least one focal set of m_1
  - There is a stochastic matrix W that shares masses of focal sets of m_2 among focal sets of m_1 that contain them:
- $m_2(E) = \sum_{F \subseteq E} w(E, F)\, m_1(F)$

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
- In the consonant case : all orderings collapse to $m_1 \subseteq_{cf} m_2$ ($\pi_1 \leq \pi_2$).

Example

- $S = \{a, b, c\}$; $m_1(ab) = 0.5$, $m_1(bc) = 0.5$;
- $m_2(abc) = 0.5$, $m_2(b) = 0.5$
- $\mathbf{m_2} \subset_{\mathbf{Pl}} \mathbf{m_1}$: $Pl_1(A) = Pl_2(A)$

  but $Pl_2(ac) = 0.5 < Pl_1(ac) = 1$
- $\mathbf{m_1} \subset_{\mathbf{Q}} \mathbf{m_2}$: $Q_1(A) = Q_2(A)$

  but $Q_1(ac) = 0 < Q_2(ac) = 0.5$
- And contour functions are equal : a/0.5, b/1, c/0.5
- **Neither** $\mathbf{m_1} \sqsubseteq_{\mathbf{s}} \mathbf{m_2}$ **nor** $\mathbf{m_2} \sqsubseteq_{\mathbf{s}} \mathbf{m_1}$ holds
- ***Not comparable % specialisation***

Next step:

- To be continued with interval data statistics