# UNCERTAINTY THEORIES: A UNIFIED VIEW

D. Dubois

IRIT-CNRS, Université Paul Sabatier
31062 TOULOUSE FRANCE

Dubois@irit.fr

www.irit.fr\~Didier.Dubois

# Outline

1. Basic notions
2. Probability distributions do not account for partial knowledge
3. Blending set-valued and probabilistic representations: imprecise probabilities, belief functions and possibility theory
4. Practical representations
5. Conditioning, query-answering, revision and fusion

# UNCERTAINTY : representing graded belief.

- AN AGENT IS UNCERTAIN ABOUT A PROPOSITION IF (S)HE DOES NOT KNOW ITS TRUTH VALUE
  - **Examples**
    - The **probability** that the trip is more than one hour long is 0.7.
    - It is quite **possible** it snows to-morrow.
    - The agent has no **certainty** that Jean comes to the meeting
- HOW TO EVALUATE THE PROBABILITY, THE POSSIBILITY, THE CERTAINTY, THAT A PROPOSITION IS TRUE OR FALSE

# GRADUAL REPRESENTATIONS OF UNCERTAINTY

**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(Ø) = 0 ; g(S) = 1*
  - if A implies (= included in) B then g(A) ≤ g(B) (monotony: g is a Choquet capacity)
- 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***

# UNCERTAINTY THEORIES

- *Probability theory*: statistical, subjective
- *Set-based representations:* Reasoning about incomplete information in terms of possibility and certainty
  - *Propositional logic:* Believing = proving from a belief base.
  - *Interval analysis :* Propagation of incomplete information.
- *Possibility Theory* ordinal or numerical:
  - Tells plausible states from less plausible ones
  - use fuzzy sets of mutually exclusive values
- *Disjunctive random sets* (Dempster, or Shafer-Smets): probability on set-representations
- *Imprecise Probabilities*: the most general setting, with probability intervals.

# Probability Representations (on finite sets)

- A finite set S with n elements: A probability measure is characterized by a set of non negative weights $p_1, \ldots, p_n$, such that $\sum_{i=1,n} p_i = 1$.
  - $p_i$ = probability of state $s_i$
- **Possible meanings of a degree of probability**:
  - Counting *favourable cases* for $s_i$ over the number of possible cases assuming uniform distribution (coins, dice, cards,…)
  - *Frequencies from statistical information*: $p_i$ = limit frequency of occurrence of $s_i$ (**Objective probabilities**)
  - *Money involved in a betting scheme* (**Subjective probabilities**)

# The roles of probability

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

1. **Randomness:** capturing variability through repeated observations.
2. **Partial knowledge:** because of information is often lacking, knowledge about issues of interest is generally not perfect.

These two situations are not mutually exclusive.

# 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.
  - Bayesian approach: subjective probability

## SUBJECTIVE PROBABILITIES (Bruno de Finetti, 1935)

- $p_i$ = *belief degree* of an agent on the occurrence of $s_i$
- measured as the price of a lottery ticket with reward 1 € if state is $s_i$ in a betting game
- **Rules of the game**:
  - gambler proposes a price $p_i$
  - banker and gambler exchange roles if price $p_i$ is too low
- **Why a belief state is a single distribution**:
  - Assume player buys all lottery tickets i = 1, m.
  - If state $s_j$ is observed, the gambler gain is $1 - \sum_j p_j$
  - and $\sum_j p_j - 1$ for the banker
  - *if* $\sum p_j > 1$ gambler *always loses money ;*
  - *if* $\sum p_j < 1$ *banker exchanges roles with* gambler

# Remarks on using a single probability distribution

- **Computationally simple** : $P(A) = \sum_{s \in A} p(s)$
- **Conventions**: P(A) = 0 iff A impossible;

  P(A) = 1 iff A is certain;

  Usually P(A) = 1/2 for ignorance
- **Meaning :**
  - Objective probability is generic knowledge (statistics from a population)
  - Subjective probability is singular (degrees of belief)

- **A Bayesian network**: a set of conditional probability assessments that represent a unique distribution.

# Constructing beliefs

- Subjective probability of the occurrence of a particular event may derive from its statistical probability: the **Hacking principle**:
  - Generic knowledge = probability distribution P
  - BetP(A) = FreqP(A): equating belief and frequency
- Beliefs can be directly elicited as subjective probabilities **of singular 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

# LIMITATIONS OF BAYESIAN PROBABILITY FOR THE REPRESENTATION OF BELIEF

- *A single probability cannot represent ignorance*: except on a 2-element set, the function g(A) = 1/2 ∀A ≠ S, Ø, is NOT a probability measure.

- Subjective specification of a Bayes net imposes unnatural conditions on conditional probabilities to be assessed: complete and consistent conditional probability assessments are requested

# Why the unique distribution assumption?

- The exchangeable betting framework enforces the elementary probability assessments to sum to 1.
  - It enforces uniform probability when there is no reason to believe one outcome is more likely than another
  - ignorance and knowledge of randomness justify uniform betting rates.
- Laplace principle of insufficient reason : What is EQUIPOSSIBLE must be EQUIPROBABLE
  - *It enforces the identity between IGNORANCE and RANDOMNESS due to a symmetry assumption*
  - *Also justified by the principle of maximal entropy*
- *BASIC REMARK: Betting rates are* ***induced*** *by belief states, but are* ***not in one-to-one correspondence*** *with them.*

# Single distributions do not distinguish between incompleteness and variability

- VARIABILITY: Precisely observed random observations
- INCOMPLETENESS: Missing information
- **Example:** probability of facets of a die
  - *A fair die tested many times:* Values are known to be equiprobable
  - *A new die never tested:* No argument in favour of an hypothesis nor its contrary, but frequencies are unknown.
- *BOTH CASES LEAD TO TOTAL INDETERMINACY ABOUT THE NEXT THROW BUT THEY DIFFER AS TO THE QUANTITY OF INFORMATION*

# THE PARADOX OF IGNORANCE

- Case 1: life outside earth/ no life
  - ignorant's response 1/2 1/2
- Case 2: Animal life / vegetal only/ no life
  - ignorant's response 1/3 1/3 1/3
- They are inconsistent answers:
  - case 1 from case 2 : P(life) = 2/3 > P(no life)
  - case 2 from case 1: P(Animal life) = 1/4 < P(no life)
- **ignorance produces information *!!!!!***
- ***Uniform probabilities on distinct representations of the state space are inconsistent.***
- **Conclusion :** *a probability distribution cannot model incompleteness*

# Instability of prior probabilities

In the case of a real-valued quantity x:

- A uniform prior on [a, b] expressing ignorance about x induces a non-uniform prior for f(x) on [f(a), f(b)] if f is monotonic non-affine

*Probabilistic representation of ignorance is not scale-independent*.

- The paradox does not apply to frequentist distributions

# Ellsberg Paradox

- Savage claims that rational decision-makers choose according to expected utility with respect to a subjective probability.
- Counterexample: An Urn containing
  - 1/3 red balls ($p_R = 1/3$)
  - 2/3 black or white balls ($p_W + p_B = 2/3$)
- For the ignorant subjectivist: $p_R = p_W = p_B = 1/3$.
- Expected utility of act a: $u_a(R)p_R + u_a(W)p_W + u_a(B)p_B$

- But this is contrary to overwhelming empirical evidence about how people make decisions

# Ellsberg Paradox

1. Choose between two bets

   B1: Win 1$ if red (1/3) and 0$ otherwise (2/3)

   B2: Win 1$ if white (≤ 2/3) and 0$ otherwise

   Most people prefer B1 to B2

2. Choose between two other bets (just add 1$ on Black)

   B3: Win 1$ if red or black (≥ 1/3) and 0$ if white

   B4: Win 1 $ if black or white (2/3) and 0$ if red (1/3)

   Most people prefer B4 to B3

# Ellsberg Paradox

- Let $0 < u(0) < u(1)$ be the utilities of gain.
- If decision is made according to a subjective probability assessment for red black and white: $(1/3, p_B, p_W)$:
  - B1 > B2:

    $EU(B1) = u(1)/3 + 2u(0)/3 > EU(B2) = u(0)/3 + u(1)p_w + u(0)p_B$
  - B4 > B3:

    $EU(B4) = u(0)/3 + 2u(1)/3 > EU(G) = u(1)\,(1/3 + p_N) + u(0)p_W$

  $\Rightarrow$ (summing, as $p_B + p_N = 2/3$) $2(u(0) + u(1))/3 > 2(u(0) + u(1))/3$: CONTRADICTION!
- Such an agent cannot reason with a unique probability distribution: **Violation of the sure thing principle.**

## When information is missing, decision-makers do not always choose according to a single subjective probability

- *Plausible Explanation of Ellsberg paradox*: In the face of ignorance, the decision maker is pessimistic.

- In the first choice, agent supposes $p_w = 0$: no white ball

  $EU(B1) = u(1)/3 + 2u(0)/3 > EU(B2) = u(0)$
- In the second choice, agent supposes $p_B = 0$: no black ball

  $EU(B4) = u(0)/3 + 2u(1)/3 > EU(B3) = 2u(0)/3 + u(1)/3$

- **The agent does not use the same probability in both cases** (because of **pessimism**)**: the subjective probability depends on the proposed game.**

# Summary on expressiveness limitations of subjective probability distributions

- The Bayesian dogma that any state of knowledge can be represented by a single probability is due to the exchangeable betting framework
  - Cannot distinguish randomness from a lack of knowledge.

- Representations by single probability distributions are language- (or scale-) sensitive

- When information is missing, decision-makers do not always choose according to a single subjective probability.

# What do probabilists do when no prior information is used?

- Hypothesis testing based on likelihood functions
- Parametric estimation of probabilistic models from empirical data using maximum likelihood principle
- Extraction of confidence intervals

**But**

*Many of these techniques are not part of Kolmogorov Probability theory.*

*Confidence intervals use ad hoc thresholds.*

# 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 to be propagated by interval analysis methods
  - **Classical Logic**: good for representing incomplete symbolic (Boolean) information to be inferred from.

    E = Models of a set of propositions stated as true.

***but poorly expressive***

# Boolean belief measures from partial knowledge

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

- Event A is ***possible (plausible)*** if $A \cap E \neq \emptyset$

(logical consistency)

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

- Event A is ***certain (necessary)*** if $E \subseteq A$

(logical deduction)

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

***This is a simple modal logic (KD45) corresponding to***

**BOOLEAN POSSIBILITY THEORY**:

$\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$; $N(A \cap B) = \min(N(A), N(B))$.

# Motivation for going beyond pure probability and set representations

- **The main tools for 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*
  - Distinguish between uncertainty due to variability from uncertainty due to lack of knowledge or missing information
  - Reflect partial knowledge faithfully
  - Are more expressive that pure set representations
  - Allows for addressing the same problems as probability.
  - Can formally justify non-probabilistic statistical notions

# Blending intervals and probability

- Representations that may 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

# Possibility theory is an uncertainty theory devoted to the handling of incomplete information.

- similar to probability theory because it is based on set-functions.
- differs by the use of a pair of dual set functions (possibility and necessity measures) instead of only one.
- it is not additive and makes sense on ordinal structures.

The name "Theory of Possibility" was coined by Zadeh in 1978, who interprets fuzzy sets as possibility distributions.
Zadeh’s aim was to represent linguistic information, accounting for its incompleteness and its gradual (non-Boolean) nature.

# 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 $\mu_{OLD}$ induces a possibility distribution on possible birth dates.
- **The result of merging imprecise subjective** information summarizing opinions of one or several sources

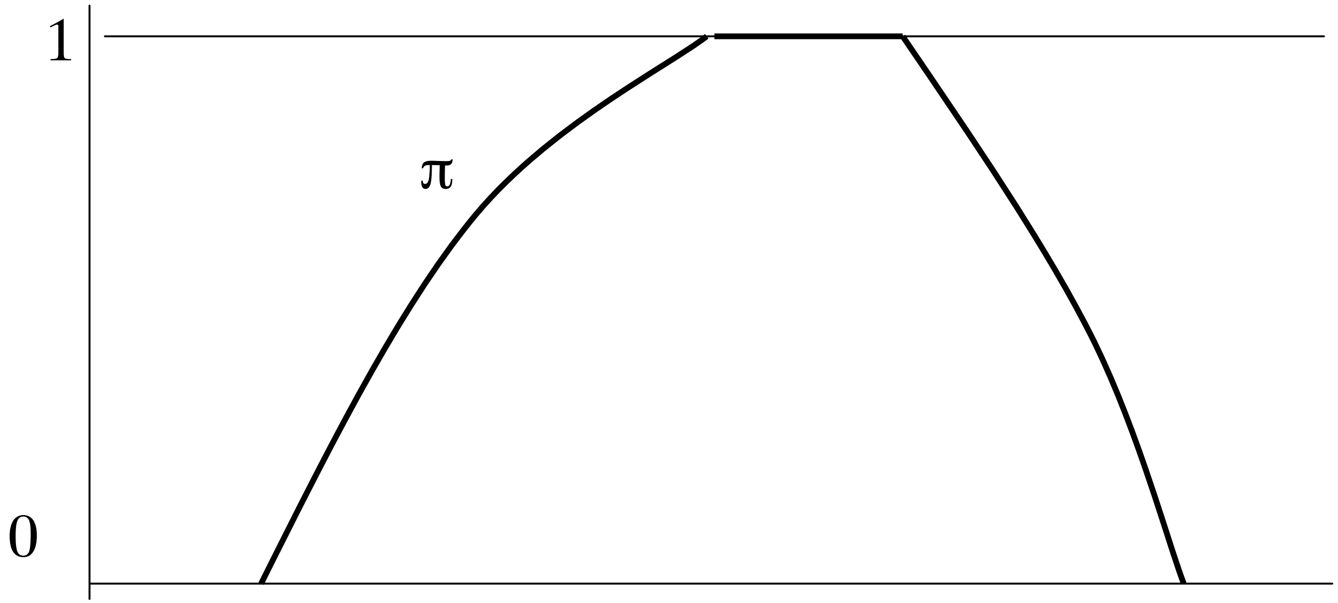
# Possibility Theory (Shackle, 1961, Zadeh, 1978)

- A piece of incomplete information "$x \in E$" admits of degrees of possibility in a plausibility scale L: *E is a (normalized) disjunctive fuzzy set*.
- L: totally ordered set of plausibility levels ([0,1], finite chain, integers,...)
- The degree of membership $\mu_E(s)$ is interpreted as the degree of plausibility that x = s and denoted $\pi_x(s)$
- A possibility distribution $\pi_x$ attached to x is a mapping from S to L : such that $\exists s, \pi_x(s) = 1$ (normalization)
- **Conventions:**

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

  $\pi_x(s) = 1$ iff x = s is normal, fully plausible, unsurprising (but no certainty)

*A possibility distribution is the representation of a state of incomplete knowledge:*
*a description of how a agent thinks the state of affairs is.*




FUZZY INTERVAL

# 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

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

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

**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

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

In other words: any possible value in information state $\pi'$ is at least as possible in information state $\pi$

*that is, $\pi'$ is more informative than $\pi$*

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

# 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 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 their interpretation and the conditioning operation

## A GENERAL SETTING FOR REPRESENTING GRADED PLAUSIBILITY AND CERTAINTY

- 2 adjoint set-functions Pl and Cr generalizing Boolean possibility Π and necessity N.
- **Conventions:**
  - Pl(A) = 0 "impossible" ;
  - Cr(A) = 1 "certain"
  - Pl(A) =1 ; Cr(A) = 0 "ignorance" **(no information)**
  - Cr(A) ≤ Pl(A) "certain implies plausible"
  - Pl(A) = 1 – Cr($A^c$) duality certain/plausible

# Imprecise probability

- A state of information is modelled by ***a credal set*:** a family $\mathcal{P}$ of probability distributions over a set X containing the true (objective) probability function.
- To each event A is attached a probability interval $[P_*(A), P^*(A)]$ such that
  - $Cr(A) = P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$
  - $Pl(A) = P^*(A) = \sup\{P(A), P \in \mathcal{P}\} = 1 - P_*(A^c)$
- $\mathcal{P}$ is generally strictly contained in the (convex) credal set $\{P: P(A) \geq P_*(A)\}$.

# Subjectivist view (Peter Walley)

- $P_{low}(A)$ is the highest acceptable price for buying 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.
- **Coherence** condition

$$P_*(A) = \inf\{P(A), P \geq P_{low}\} = P_{low}(A)$$

- *In this view, there is no « real » probability lower-bounded by $P_*(A)$, which directly represents belief.*

# BASIC PROPERTIES

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

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

- One may require the <u>2-monotony property</u>:

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

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

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

  Cr is then called a <u>convex capacity.</u>

# 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

- 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).

# Random sets and evidence theory

- A family $\mathcal{F}$ of « focal » (disjunctive) non-empty subsets of S representing
  - incomplete observations (imprecise statistics).
  - Unreliable testimonies
  - Indirect information (induced from a probability space)
- A positive weighting m of focal sets (random set) :

$$\sum_{E \in \mathcal{F}} m(E) = 1 \; ; \; m(\emptyset) = 0 \; (\textit{mass function})$$

- It is a randomized incomplete information consisting of a probability distribution on the power set of S

# Disjunctive random sets

- m(E) = probability that the most precise description of the available information is of the form "x ∈ E”
  - *= probability(only knowing "x ∈ 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)**

# Examples

- The mass m(E) may be
  - The frequency of an incomplete observation.
  - The reliability of a testimony (Shafer)
  - What we know about a random variable x with range S, based on a sample space ($\Omega$, $\mathcal{A}$, P) and a multimapping $\Gamma$ from $\Omega$ to S (Dempster):

    $m(\Gamma(\omega)) = P(\{\omega\})\ \forall\ \omega$ in $\Omega$ (finite case.)

# **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.

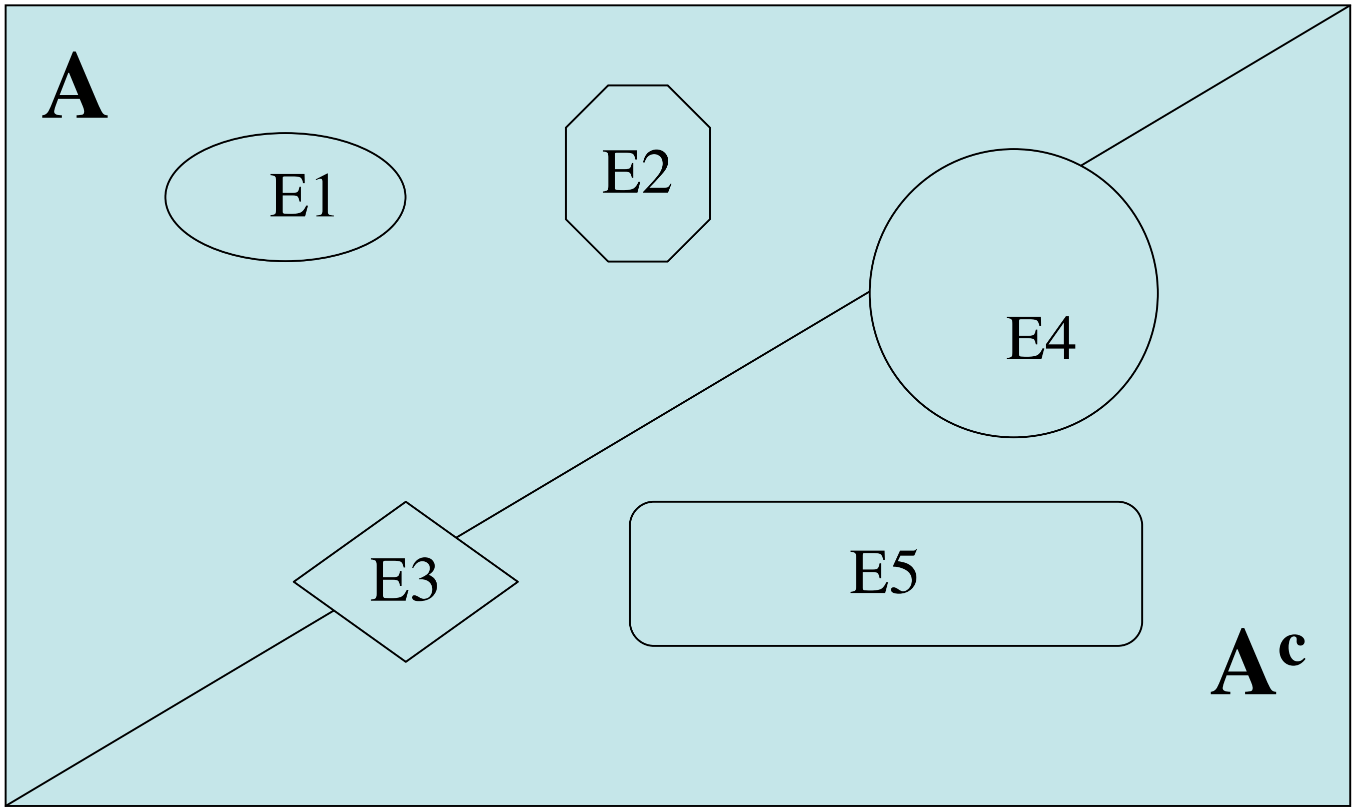
# Belief and plausibility functions

- **degree of certainty** (belief) :
  - $\mathrm{Bel}(A) = \sum_{E_i \subseteq A,\, E_i \neq \emptyset} m(E_i)$
  - total mass of information <u>implying</u> 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 <u>consistent with</u> 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($A^c$)



# 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

# Theory of evidence vs. imprecise probabilities

- The set $\mathcal{P}_{bel} = \{P \geq Bel\}$ is coherent: Bel is a special case of lower probability
- Bel is $\infty$-monotone (super-additive at any order)
- 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)

  – **It is positive iff g is a belief function**

# 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** $\Omega = \{1, \dots, i, \dots, n\}$ 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**:
  - $m(E) = \mathrm{card}(\{i, E_i = E\})/n$
  - = 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**:

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

- $\mu_F(x)$ 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.

# 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

  (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.

  (Dempster intuition)

# Quantitative possibility theory in the setting of uncertainty theories

**Possibility distributions are**

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

  *Semantics: metrics, proximity to prototypes*
- **Or simple models of imprecise probability**
  - Random experiments with consonant imprecise outcomes
  - Special convex probability sets

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

# HISTORY and TERMINOLOGY of Possibility theories

- *Numerical*
  - **Numerical impossibility measures :** Shackle's degrees of surprise (1950) ($1-\Pi$)
  - More recently Zadeh's (1978) coined the word "possibility measure": linguistic information as **fuzzy (disjunctive) sets**
  - Spohn's (ordinal conditional) **kappa functions** (integer exponents of infinitesimal probabilities)
  - Shafer's **consonant belief functions**
  - Special cases of **probability bounds** (Dubois and Prade, 1992)

# *POSSIBILITY AS EXTREME PROBABILITY*

- SPOHN's ORDINAL CONDITIONAL (KAPPA) FUNCTIONS: $\kappa(A)$ = disbelief in A
  - The higher $\kappa(A)$, the less likely.
- **Basic properties** :
  - $\kappa(A \cup B) = \min(\kappa(A),\kappa(B)) \in \mathcal{N}$ (integers)
  - $\kappa(S) = 0$
  - $\kappa(A \mid B) = \kappa(A \cap B) - \kappa(A)$ (conditioning rule)

- **<u>Probabilistic interpretation</u>** : there is some infinitesimal $\varepsilon$ such that $\kappa(A) = n \Leftrightarrow P(A) \approx \varepsilon^n$
- $P(A \cup B) \approx \varepsilon^{\kappa(A)} + \varepsilon^{\kappa(B)} \approx \varepsilon^{\min(\kappa(A), \kappa(B))}$

# *POSSIBILITY AS EXTREME PROBABILITY*

- **<u>Possibilistic interpretation of kappa functions</u>:**
- Transformation method: $\Pi_\kappa(A) = 2^{-\kappa(A)}$
  - Function $\Pi_\kappa$ is a rational-valued possibility measure on [0, 1] with $\Pi_\kappa(A) > 0$, $\forall A \neq \emptyset$, hence $\kappa(A^c) = -\mathrm{Log}_2(1 - N(A))$
  - Then, $\Pi_\kappa(A)$ represents an order of magnitude whereby $\Pi_\kappa(A) > \Pi_\kappa(B)$ indicates that B has plausibility negligible in front of A
- It yields the product conditioning rule for possibility

$$\Pi_\kappa(A \mid B) = \quad \Pi_\kappa(A \cap B)/\Pi_\kappa(B)$$

(special case of Dempster rule for belief functions)

- $\kappa(A^c) \geq n$ (integer) encodes $N(A) \geq \alpha = 1 - 2^{-\kappa(Ac)}$;

# LANDSCAPE OF UNCERTAINTY THEORIES

BAYESIAN/STATISTICAL PROBABILITY

*Randomized points*

(extreme probabilities)

**UPPER-LOWER PROBABILITIES**

***Disjunctive sets of probabilities***

KAPPA FUNCTIONS (SPOHN)

DEMPSTER UPPER-LOWER PROBABILITIES

SHAFER-SMETS **BELIEF FUNCTIONS**

*Random disjunctive sets*

*PLAUSIBILITY RANKING*

**Quantitative Possibility theory**

*Fuzzy (nested disjunctive) sets*

**Classical logic**

*Disjunctive sets*

# Possibility distributions as 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 likelihood 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 $\lambda$ should be set-monotonic:

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

It implies that by default (in the absence of other information):

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

# Maximum likelihood principle = possibility theory

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

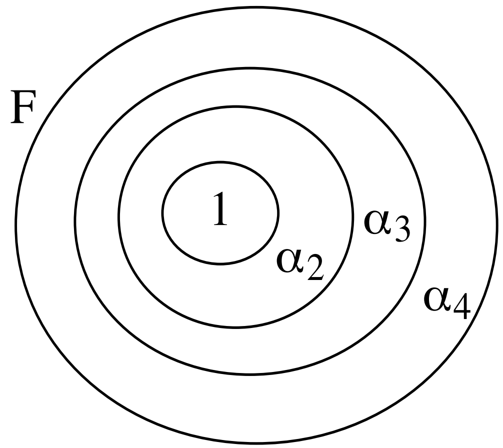
*the degree of possibility π(θ) that the probability of "head" is θ.*

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

It yields $\theta = k/n$.

(*already Shafer, 1976*).

# Random set view




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

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

# POSSIBILITY AS UPPER PROBABILITY

- Given a numerical possibility distribution $\pi$, 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 $\pi$ is a coherent representation of a family of probability measures

# From confidence sets to possibility distributions

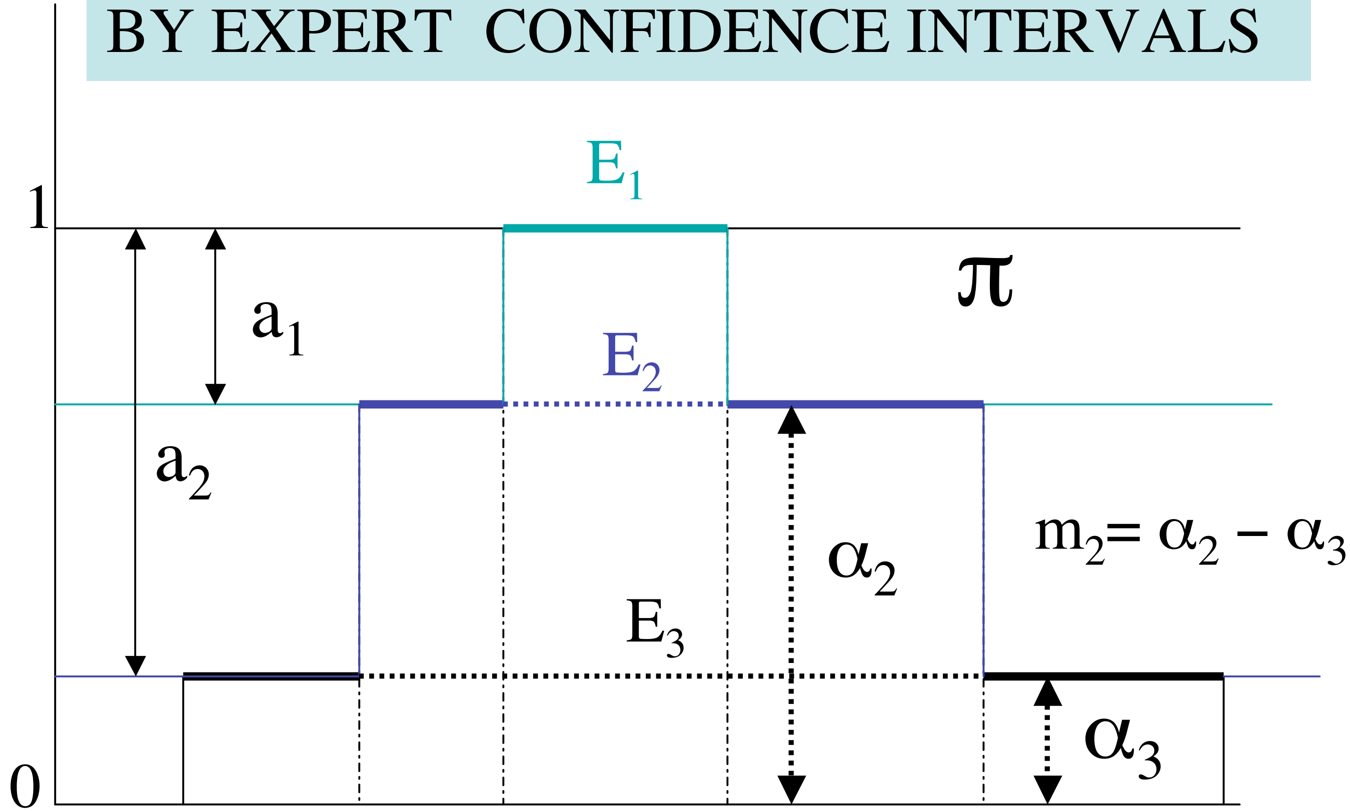
- Let $E_1, E_2, \ldots E_n$ be a nested family of sets
- A set of confidence levels $a_1, a_2, \ldots a_n$ in [0, 1]
- Consider the set of probabilities

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

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

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

# POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



# Possibilistic covering approximations

- **Given a (family of) probability distribution(s) (ill-known) probabilistic model), find a possibility measure dominating all probability measures in the family.**
  - Given an indexed set of *nested* intervals $M_\alpha = [a_\alpha, b_\alpha]$ around x*,
  - Given a probability measure P, construct a possibility distribution π such that

$$\pi(a_\alpha) = \pi(b_\alpha) = 1 - P(M_\alpha) \ (= \beta)$$

- π is a **fuzzy prediction interval** of P around x*: intervals $[a_\alpha, b_\alpha]$ are β - cuts of π that contain the unknown value with confidence level 1– β.
- Then $\Pi \geq P$.
- It is also a set of “probabilistic inequalities” for P.

# Possibility distributions generalize cumulative distributions

- <u>Particular case</u>: if M = increasing membership function

  $\pi(x) = F(x) = P((-\infty, x])$:

A Cumulative distribution function F of a probability function P can be viewed as a possibility distribution dominating P

$$\Pi(A) = \sup\{F(x), x \in A\} \geq P(A)$$

(because it corresponds to probabilities of a nested family of intervals of the form $(-\infty, x]$).

- More generally, choosing *any* order relation $\leq_R$

  $F_R(x) = P(\{y \leq_R x\})$ also induces a possibility distribution dominating P

# Possibilistic view of probabilistic inequalities

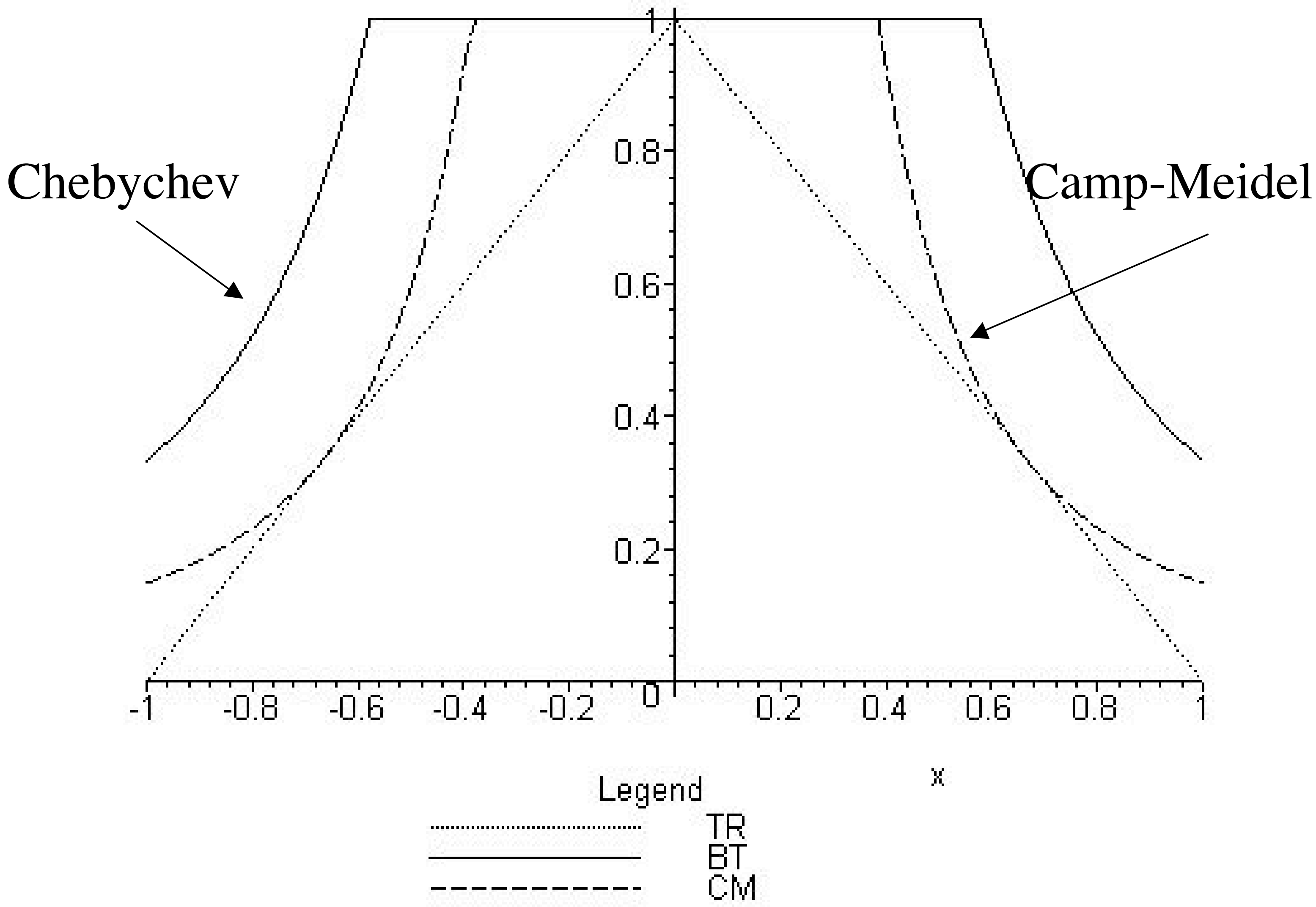
***Probabilistic inequalities can be used for knowledge representation***

- Chebyshev inequality defines a possibility distribution that dominates *any* density with given mean and variance:
  - $P(V \in [x^{mean} - k\sigma, x^{mean} + k\sigma]) \geq 1 - 1/k^2$ *defines a family of nested sets with lower probability bounds*
  - *It 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 $[x^{mode} - a, x^{mode} + a]$ as the TFN.

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

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


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

# Elementary forms of possibilistic representations exist for a long time

- **Prediction intervals**

  Replacing a statistical probability distribution by an interval A with a confidence level $c = P(A)$.

  – *It defines a possibility distribution*

$$\pi(x) = 1 \text{ if } x \in A,$$
$$= 1 - c \text{ if } x \notin A$$

  - it is clear that $\Pi \geq P$

- **Confidence intervals**: estimation of a model parameter $\theta$ in terms of available data:

  Find a data-dependent interval A that contains $\theta$ with high confidence: $P(\theta \in A) = 0.95$

# Possibilistic representation of probabilistic information

- **Why** ?
  - Simpler representation framework for uncertainty than general imprecise probability models à la Walley
  - More informative than pairs of cumulative distribution functions when they are far from each other
  - Generalizing some statistical methods or probabilistic inequalities
  - A cautious alternative interpretation of betting odds that differs from subjective probabilities
  - fusion of heterogeneous data; possibility distributions are a less restricted framework than single probability measures.

# Probability -> possibility transformations :
## BASIC PRINCIPLES FOR OBJECTIVE PROBABILITY

- **Possibility probability consistency**: $P \leq \Pi$
- **Informativity**:
  optimize information content
  (*Maximization of specificity)*

  *Additional criterion:*
- **Ordinal faithfulness:**
  - *Preserving the ordering of elementary events*

# From OBJECTIVE probability to possibility:

- **Rationale**: given a probability p, try and preserve as much information as possible with π
- Select a *most specific element* from the set $\mathcal{PI}$(P) = {Π: Π ≥ P} of possibility measures dominating P

Additionally (not compulsory)

- One may also require $\pi(x) > \pi(x')$ iff $p(x) > p(x')$
- may be weakened into:

$p(x) > p(x')$ <u>implies</u> $\pi(x) > \pi(x')$

# Optimal possibilistic representations in the finite case

•In the finite case, if $p_1 > \ldots > p_n$, the most informative possibility distribution consistent with p is defined by

$$\pi^p_i = \sum_{j=i,n} p_j \;:\; \pi^p_n = p_n;\ \pi^p_{n-1} = p_n + p_{n-1};\ \ldots..\ \pi^p_n = 1.$$

(the sum of probabilities of states with lower or equal probability)

This kind of order-faithful discrete cumulative distribution function is also known as a Lorentz curve.

•If there are equiprobable elements, order-faithfulness implies unicity of the possibility transform: equipossibility of the corresponding elements is enforced.

• *Uniform probability leads to uniform possibility.*

# ENTROPY AND TRANSFORMATIONS

- Comparing probability distributions via their possibility transforms: *distribution p is more peaked than q if and only if* $\pi^p < \pi^q$.

- Peakedness reflects lack of uncertainty of the probability distribution
- Inclusion of prediction sets of p into those of q
- Comparing the specificity of the possibility transforms

*THEOREM: if p is less peaked than q, its Shannon entropy H(p) is higher than H(q)*

*This is a consequence of mathematical results on inequalities (Hardy, Polya, Littlewood)*

# Generalized cumulative distributions

- A Cumulative distribution function F of a probability function P can be viewed as a possibility distribution dominating P

  $\sup\{F(x), x \in A\} \geq P(A)$

- The notion of cumulative distribution depends on an ordering on the space: $F_R(x) = P(X \leq_R x)$
- Choosing any order relation $\leq_R$

  $F_R(x) = P(\{y \leq_R x\})$ also induces a possibility distribution dominating P

# Cumulative distribution around the mode

- Choose the mode as $x^*$ and R the total preorder induced by the density :

  $x >_R y$ iff $p(x) > p(y)$

- Then $\{x \leq_R a\} = \{x, p(x) \leq p(a)\}$
- Then $F_R(a) = P(\{y \leq_R a\})$ is of the form

$F_R(a) = 1 - P(\{x, p(x) \geq \beta\})$ for threshold $\beta = p(a)$

It defines a possibility distribution with mode $x^*$ that dominates p.

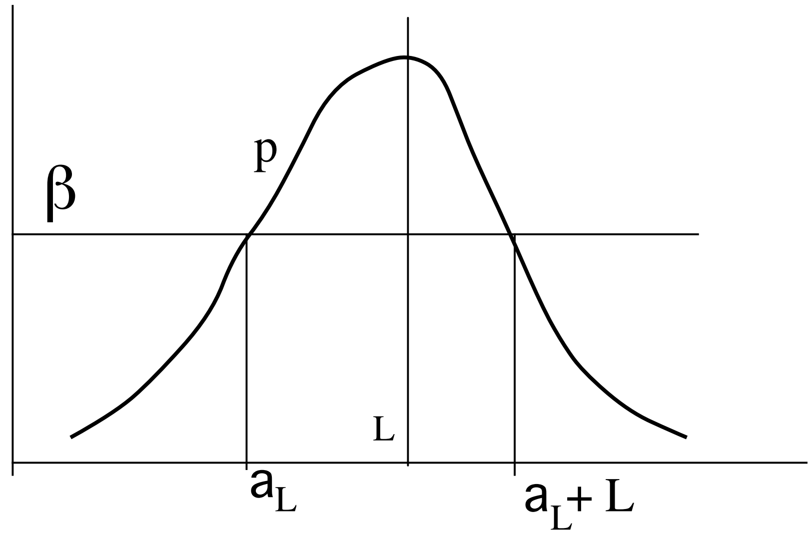
# Optimal order-faithful fuzzy prediction intervals

- the interval $I_L$ of fixed length L with maximal probability is of the form

$\{x, p(x) \geq \beta\} = [a_L, a_L + L]$

- The most narrow prediction interval of probability $\alpha$ is of the form $\{x, p(x) \geq \beta\}$
- So the most natural possibilistic counterpart of p is when

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

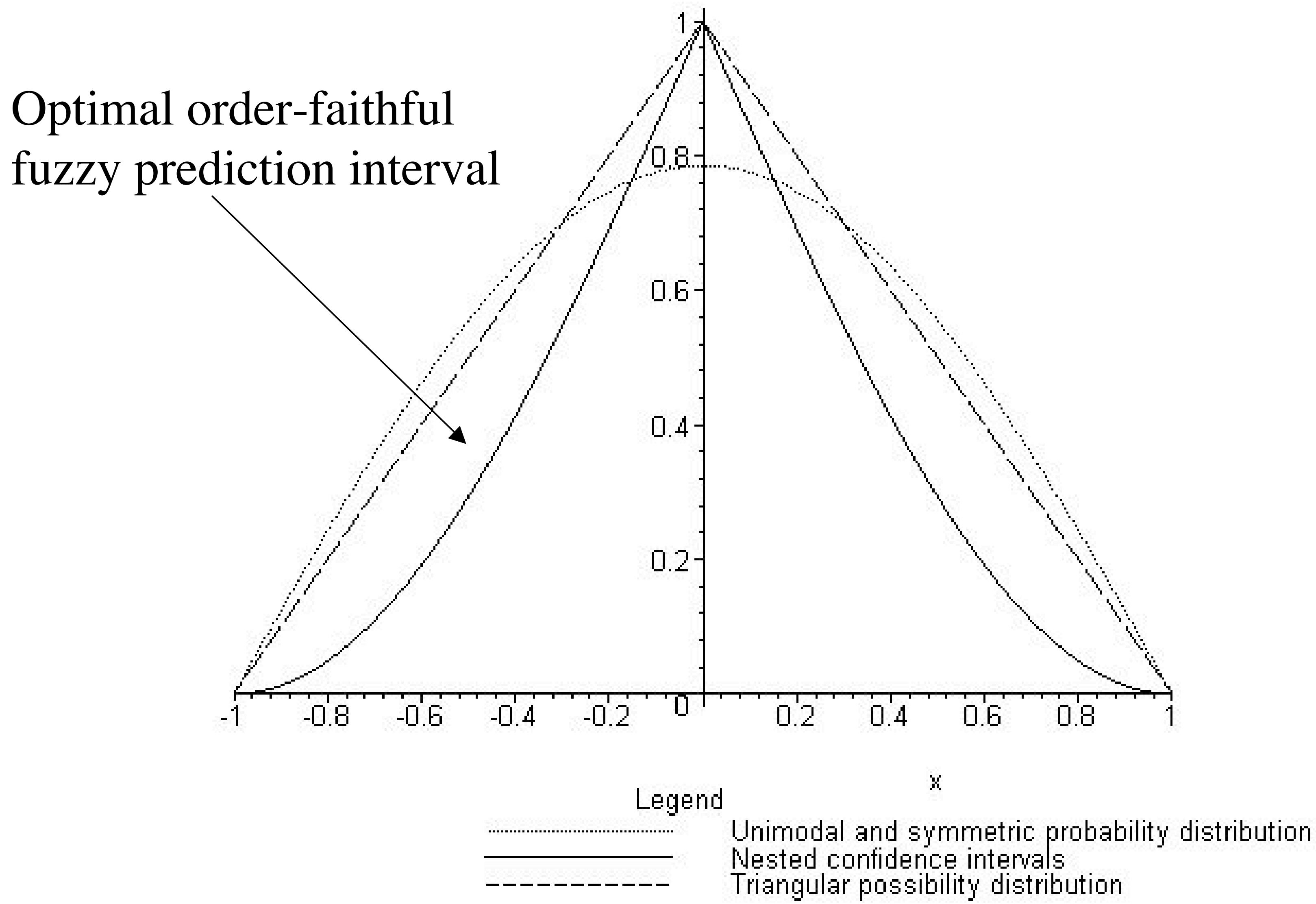


# From probability to possibility: symmetric case

- The optimal symmetric transform of the uniform probability distribution is the triangular fuzzy number
- The symmetric triangular fuzzy number (STFN) dominates any probability with unimodal symmetric density p with the same mode.
- *Moreover, its $\alpha$-cut contains the $(1-\alpha)$-confidence interval of any such p.*

*This is because for any such p, $\pi^*$ is convex on each side*

*A much tighter inequality than Chebychev for bounded support symmetric unimodal densities.*



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

# SUBJECTIVE POSSIBILITY DISTRIBUTIONS

- **Starting point** : exploit the fair betting approach to subjective probability

  (enforce "min selling price = max buying price")
- **A critique**: The agent is forced to be additive by the rules of exchangeable bets.
  - For instance, the agent provides a uniform probability distribution on a finite set whether (s)he knows nothing about the concerned phenomenon, or if (s)he knows the concerned phenomenon is purely random.
- **Idea** : It is assumed that a subjective probability supplied by an agent is only a trace of the agent's belief.

# Betting based on a belief function

- According to Smets
  - An agent has state of knowledge described by a mass function m.
  - The agent ranks decision using expected utility
- Generalized Laplace principle:
  - Select an epistemic state E with probability m(E)
  - Select an element at random in E
- The « pignistic » probability used by the agent is

$$p(s) = \sum\{m(E)/|E|, s \in E\}$$

- It is the Shapley value of the belief function Bel, and the center of gravity of its credal set.

# SUBJECTIVE POSSIBILITY DISTRIBUTIONS

- **Assumption 1:** Beliefs can be modelled by belief functions
  - (masses m(A) summing to 1 assigned to subsets A).
- **Assumption 2:** The agent uses a probability function P induced by his or her beliefs, using the pignistic transformation (Smets, 1990).
- **Method** : reconstruct the underlying belief function from the probability provided by the agent by choosing among the isopignistic ones (the ones yielding pignistic probability P).

# SUBJECTIVE POSSIBILITY DISTRIBUTIONS

– *There are clearly several belief functions with a prescribed Shapley value.*

- Consider the **least informative of those**, in the sense of a non-specificity index (expected cardinality of the random set)

$$I(m) = \sum_{A \subseteq \Omega} m(A) \cdot card(A).$$

- RESULT : The least informative belief function whose Shapley value is p *is unique and consonant.*

# SUBJECTIVE POSSIBILITY DISTRIBUTIONS

- The least specific belief function in the sense of maximizing I(m) is characterized by

$$\pi_i = \Sigma_{j=1,n} \min(p_j, p_i).$$

- It is a probability-possibility transformation, previously suggested in 1983: *This is the unique possibility distribution whose pignistic (Laplacean) probability is p*.
- It gives results that are less specific than the optimal fuzzy prediction interval approach to objective probability.

# Practical representation issues

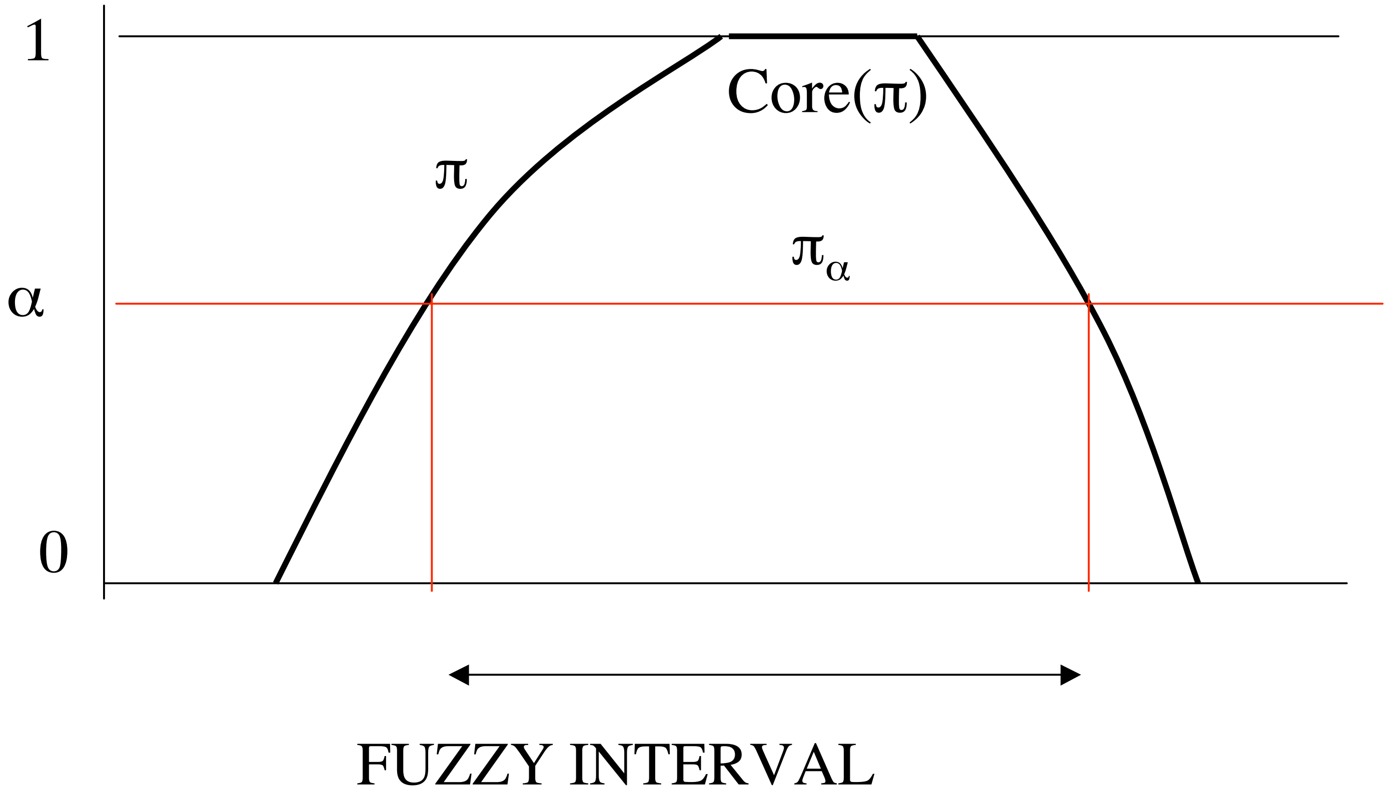
- Lower probabilities are difficult to represent ($2^{|S|}$ values): The corresponding family is a polyhedron with potentially $|S|!$ vertices.
- Finite random sets are simpler but potentially $2^{|S|}$ values
- Possibility measures are simple ($|S|$ values) but sometimes not expressive enough.
- ***There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

## Main practical representations of imprecise probabilities on the reals

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

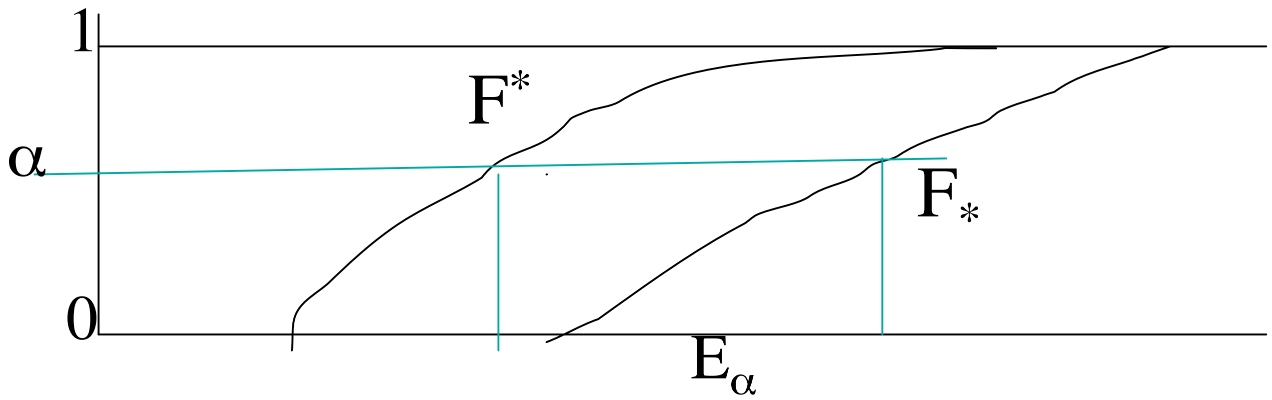
*Some are special random sets some not.*

- *Fuzzy intervals are possibility distributions that generalize intervals: they are nested random intervals*
- *They can account for probability families induced by confidence intervals, probabilistic inequalities*
- *They encompass probability families with fixed support and mode.*



# Probability boxes

- A set $\mathcal{P} = \{P: F^* \geq P \geq F_*\}$ induced by two cumulative disribution functions is called a **probability box (p-box)**,
- **A p-box is a special random interval with focal sets** $E_\alpha$ **whose upper and bounds induce the same ordering.**
- $P^*([a, b]) = F^*(b) - F_*(a)$; $P_*([a, b]) = \max(0, F_*(b) - F^*(a))$



# p-box induced by a fuzzy interval $\pi$

- **A fuzzy interval $\pi$ induces a p-box:**
  - $F^*(a) = \Pi_M((-_\infty, a]) = \pi(a)$ if $a \leq \inf \mathrm{Core}(\pi)$

    $= 1$ otherwise.
  - $F_*(a) = N_M((-_\infty, a]) = 0$ if $a < \sup \mathrm{Core}(\pi)$

    $= 1 - \lim_{x \downarrow a} \pi(x)$ otherwise
- *Representing families of probabilities by fuzzy intervals is more precise than with the corresponding pairs of PDFs:*
- $\mathcal{P}(\pi) = \{\Pi \geq P \geq N\}$ is a proper subset of $\mathcal{P} = \{P: F^* \geq F \geq F_*\}$
  - Not all P in $\{P: F^* \geq F \geq F_*\}$ are such that $\Pi \geq P$

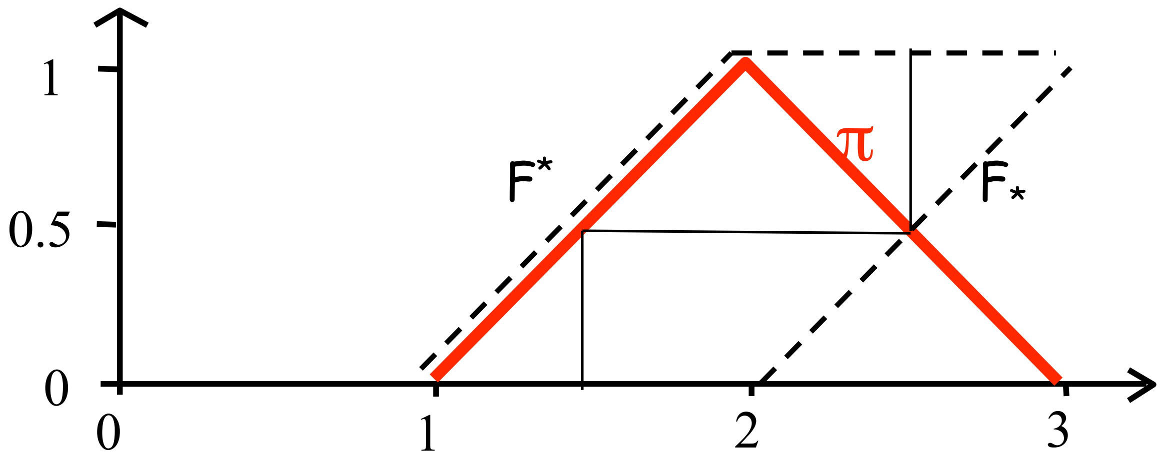
# P-boxes vs. fuzzy intervals

*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 $F_* < F < F^*$ But $P \notin \mathcal{P}(\pi)$ since

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



# Generalized p-boxes

- The notion of cumulative distribution depends on an ordering on the space: $F_R(x) = P(X \leq_R x)$
- A generalized probability box is a pair of cumulative functions $(F_R^*, F_{R*})$ with max $F^* = 1$ associated to the same order relation on a set S.
- its credal set is $\mathcal{P} = \{P: F_R^* \geq F_R \geq F_{R*}\}$
- Let $x_1 \leq_R x_2 \leq_R \ldots x_n$ and $E_i = \{x_1, \ldots, x_i\}$
- It comes down to considering nested confidence intervals $E_1, E_2, \ldots E_n$ each with two probability bounds $\alpha_i$ and $\beta_i$ such that

  $\mathcal{P} = \{ \alpha_i \leq P(E_i) \leq \beta_i$ for $i = 1, \ldots, n\}$

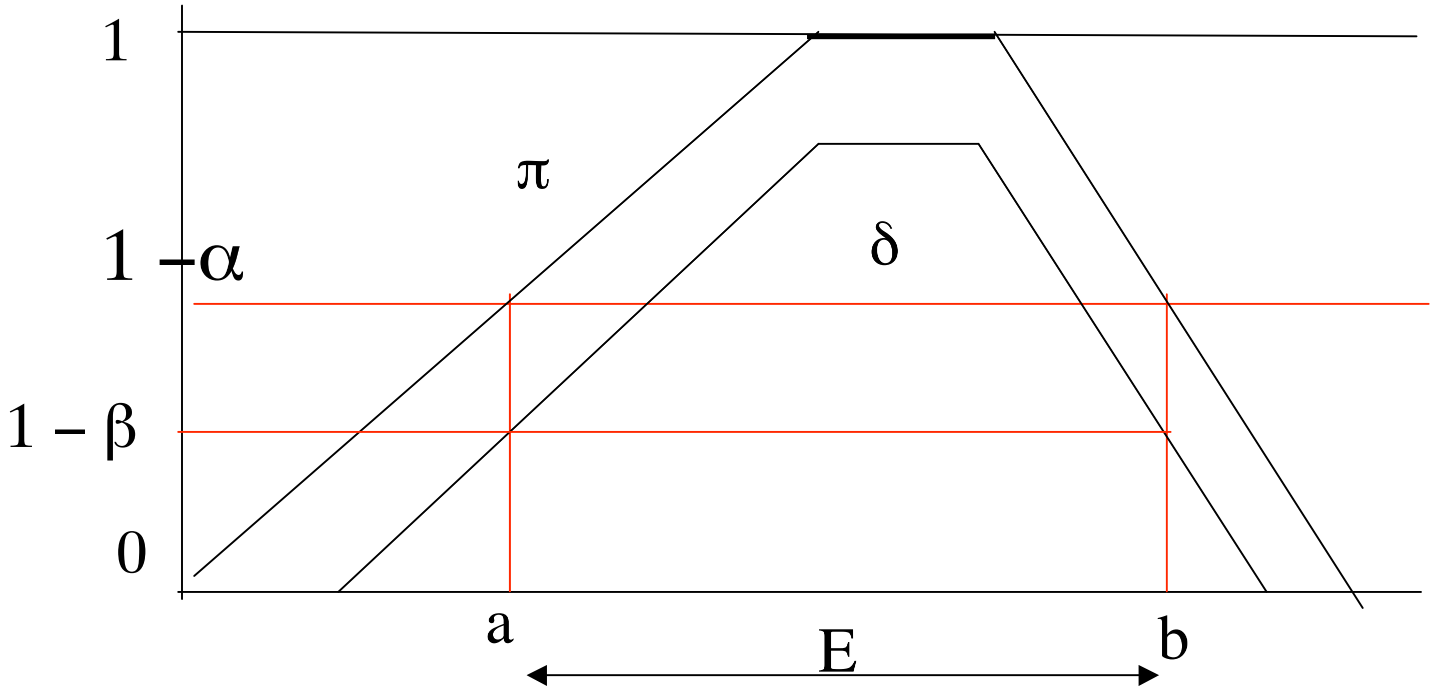
  Where $\alpha_i = F_{R*}(x_i)$ and $\beta_i = F_R^*(x_i)$.

# Generalized p-boxes and possibility theory

- A GPB comes down to two possibility distributions $\pi$ (from $\alpha_i \leq P(E_i)$) and $\pi_c$ (from $P(E_i) \leq \beta_i$ )
  $\pi(x_i) = F_R^*(x_i) = \beta_i$ and $\pi_c(x_i) = 1 - F_{R*}(x_{i-1}) = 1 - \alpha_i$
  Distributions $\pi$ and $\pi_c$ are such that $\pi \geq 1 - \pi_c = \delta$ and **$\pi$ is comonotonic with $\delta$** (they induce the same order on the referential according to $\leq_R$).
- $\mathcal{P} = \mathcal{P}(\pi) \cap \mathcal{P}(\pi_c)$
- **Theorem**: a generalized p-box is a belief function (random set) with focal sets

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

$\alpha = F_{R*}(a) = F_{R*}(b) = 1 - \pi(a) = 1 - \pi(b);$
$\beta = F_R^{\ *}(a) = F_R^{\ *}(b) = 1 - \delta(a) = 1 - \delta(b).$




Generalized p-box

# From generalized p-boxes to clouds

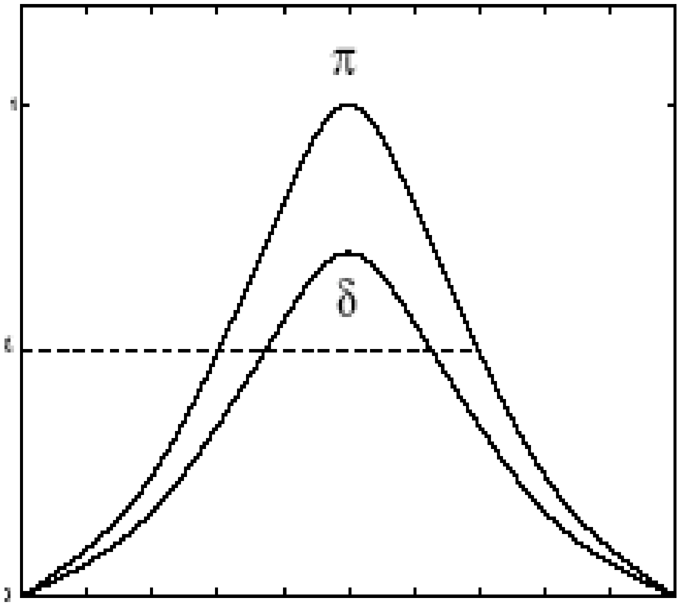



Fig 1.A Comonotonic cloud

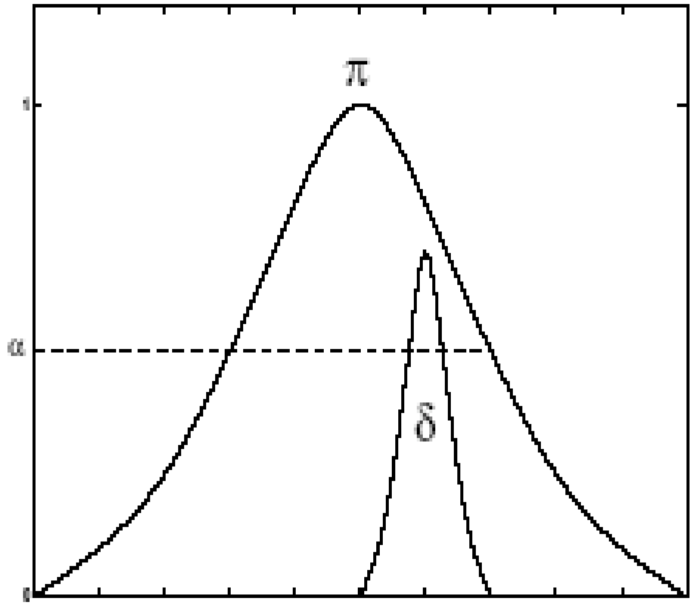



Fig 1.B Non-comonotonic cloud

# CLOUDS

- Neumaier (2004) proposed a generalized interval as a pair of distributions ($\pi \geq \delta$) on the reals representing the family of probabilities $\mathcal{P}_{cloud}$ s.t.:
  $P(\{x: \delta(x) > \alpha\}) \leq 1 - \alpha \leq P(\{x: \pi(x) \geq \alpha\})\ \forall \alpha > 0$

- Distributions $\pi$ and $1-\delta$ are possibility distributions such that $\mathcal{P}_{cloud} = \mathcal{P}(\pi) \cap \mathcal{P}(1-\delta)$
- *It does not generally correspond to a belief function, not even a convex (2-monotone) capacity*

# SPECIAL CLOUDS

- Clouds are modelled by interval-valued fuzzy sets
- **Comonotonic clouds = generalized p-boxes**
- Fuzzy clouds: $\delta = 0$; they are possibility distributions

- **Thin clouds**: $\pi = \delta$:
  - Finite case: empty
  - Continuous case: there is an infinity of probability distributions in $\mathcal{P}(\pi) \cap \mathcal{P}(1-\pi)$ for bell-shaped $\pi$
  - Increasing $\pi$: only one probability measure p ($\pi$ = cumulative distribution of p)

# Probability intervals

- **Definition**: To each element $s_i$ in S, assign a probability interval $[l_i, u_i]$, i = 1, …n.
- $\mathcal{P} = \{P, l_i \leq P(\{s_i\}) \leq u_i, i = 1, \ldots n\}$
- 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 $1 \leq \sum_{j \neq i} u_j + l_i$
- Induced upper and lower probabilities are 2-monotone capacities (De Campos and Moral)

# Main issues to be addressed about uncertainty theories

- Constructing an imprecise probability model
- Eliciting upper/ lower probabilities
- Uncertainty propagation
- Plausible reasoning: Inferring beliefs from generic information on the basis of observations
- Updating imprecise probability models
- Fusion of uncertain information
- Extracting useful information from imprecise data or from outputs of uncertainty models

# Elicitation of imprecise probabilities

- An expert is expected to provide imprecise consistent estimates rather than precise random ones
- Information requested from experts aims at singling out a unique distribution, but it actually constrains a probability family.
- There is a need to reconsider existing elicitation methods in the scope of imprecise probabilities.
- Practical representations (fuzzy intervals, p-boxes, clouds, etc.) can be useful to that end.

# Example: $\alpha \leq P(A) \leq \beta$

- As a generalized p-box
  - $F^*(x) = \beta$ if $x \in A$, and 1 if $x \notin A$
  - $F_*(x) = \alpha$ if $x \in A$, and 1 if $x \notin A$
- In terms of possibility distributions
  - $P(A) \geq \alpha$ hence $\pi_1(x) = 1$ if $x \in A$, $1-\alpha$ if $x \notin A$
  - $P(A^c) \geq 1 - \beta$, so $\pi_2(x) = 1$ if $x \notin A$, $\beta$ if $x \in A$
- As a cloud
  - $(\pi_2(x), \delta_2(x) = 1 - \pi_1(x) = \alpha$ if $x \in A$, and 0 if $x \notin A)$
  - $(\pi_1(x), \delta_1(x) = 1 - \pi_2(x) = \beta$ if $x \in A$, and 0 if $x \notin A)$
- As a random set :
  - $m(A) = \alpha$, $m(A^c) = 1 - \beta$, $m(S) = \beta - \alpha$.

# Imprecise statistical probability models

- When data is imprecise (tainted with interval-valued error), random sets are more faithful than unique probability measures
- When data is scarce it is hard to single out a probability distribution
  - imprecise Dirichlet model: the width of probability intervals varies with the number of observed data.
  - Using confidence intervals to induce imprecise parametric models

# Remarks on propagation

- Given a mathematical model y = f(x), and a credal set describing x, how to practically construct the credal set for y?
  - Random sets can be propagated using Monte-Carlo methods + interval analysis
  - Possibility distributions are stable under propagation using specific assumptions on the dependence between input variables
  - P-boxes are not stable under exact (random set) propagation
  - Clouds are propagated using approximate methods

# Hybrid possibilistic/probabilistic uncertainty propagation

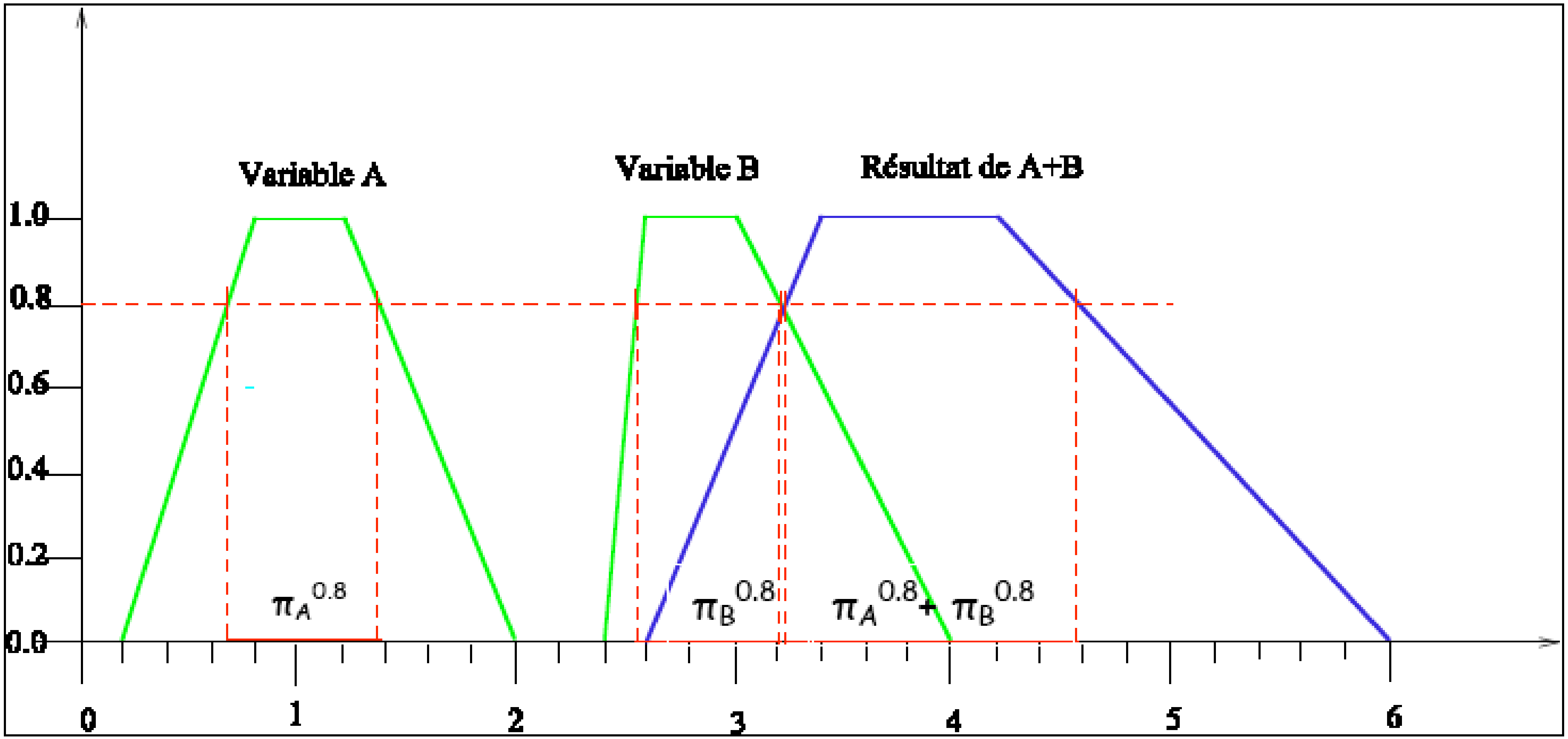
- **Formal problem**:

  Given a numerical function *f(x, y, z, …),* and some uncertain knowledge on *x, y, z,* … possibilistic ($\pi_x$), probabilistic ($p_y$) or random set-like ($v_z$)… find the resulting uncertainty on *f(x, y, z, …)*.
- **Application Context:** Evaluation of risks of potentially polluted sites for man and the environment Models simulate the transfer of pollutants from a source to a vulnerable target, for different scenarii of exposure.

# Methods for the propagation of uncertainties

- Pure Probability (Monte-Carlo)
- Pure Interval calculation.
- Pure Possibility theory (max-min extension principe).
- Hybrid method (Monte-Carlo + extension principle).
- Random set method under independence assumptions
- or without any assumed dependencies

# Pure possibilistic propagation: fuzzy interval calculations



# Fuzzy interval propagation is not always more conservative than Monte-Carlo methods

- **Beware**: if $P_A \in \mathcal{P}(\Pi_A)$ , $P_B \in \mathcal{P}(\Pi_B)$, then generally
- $P_{f(A, B)} \notin \mathcal{P}(\Pi_{f(A, B)})$ whererandom variables A and B are independent
  - $(P_{A, B} = P_A \times P_B)$,

and possibilistic A and B are non-interactive $\Pi_{A, B} = \min(\Pi_A, \Pi_B)$.

- **Exception**: sum of independent symmetric unimodal probabilistic variables vs. sum of non-interactive symmetric triangular probabilistic variables
- **Theorem** : Let *A* and *B* be two symmetric triangular fuzzy numbers. The membership function of $A \oplus B$ is the most specific possibility distribution dominating the sum (by regular convolution) of any two symmetric probability distributions having the same supports as *A* and *B* respectively

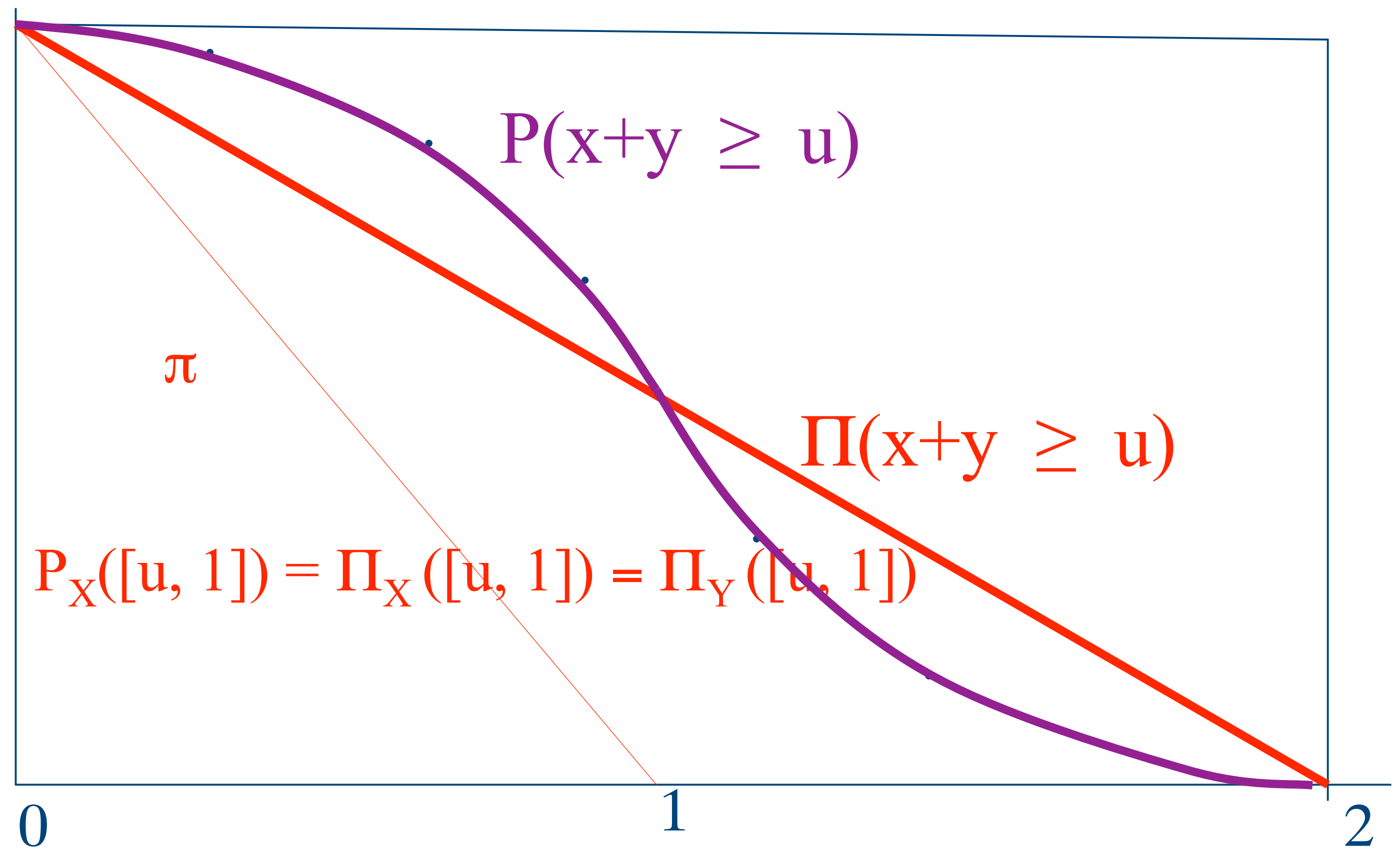
# Fuzzy vs. random arithmetic : Example

- assume X and Y have independent uniform probability densities on [0, 1], and that $\Pi_X$ and $\Pi_Y$ are linear decreasing on [0, 1], so that $\Pi_X(u) = \Pi_Y(u) = 1 - u$. Clearly,
  - $P_X([u, 1]) = P_Y([u, 1]) = 1 - u$. Hence $P_X \leq \Pi_X$ $P_Y \leq \Pi_X$
  - $P(x+y \geq u) = 1 - u^2/2$, $u \leq 1$
  - $P(x+y \geq u) = (2 - u)^2/2$, $u \geq 1$

$$\Pi(x+y \geq u) = 1 - u/2$$

- So, $\Pi(x+y \geq u) < P(x+y \geq u)$ if $u \leq 1$

# Fuzzy vs. random arithmetic



# Hybrid possibility-probability propagation

- **Formal problem**: Given a numerical function $f(x_1, \ldots x_m, y_1, \ldots y_n)$,
- assume $x_1, \ldots x_m$ are independent random variables
- assume $y_1, \ldots y_n$ are non-interactive possibilistic variables modelled by fuzzy intervals $F_1, \ldots F_n$

- Then $f(x_1, \ldots x_m, y_1, \ldots y_n)$ *is a fuzzy random variable*

# Hybrid possibility-probability propagation

- **Computation**
- Find N samples $a_1, \ldots a_m$ of $x_1, \ldots x_m$ using a Monte-Carlo method.
- For each sample, compute $f(a_1, \ldots a_m, F_1, \ldots F_n)$ using fuzzy interval computation.

- *As the output*, we get a random fuzzy interval

  $\{(C_1, v_1) \ldots (C_N, v_N)\}$

# Random set method under independence

- assume $x_1, \ldots x_m$ are independent random variables
- assume $y_1, \ldots y_n$ are *independent* possibilistic variables $F_1, \ldots F_n$ modelled by nested random sets

- Then $f(x_1, \ldots x_m, y_1, \ldots y_n)$ *is a random set*

# Random set method under independence

- **Discrete Computation (one p, one $\pi$)**
- Discretize the probability distribution as a finite disjoint random set, partitioning its support

  $E_1, \ldots E_k$ with masses $v_1(E_i)=m_i=Prob(E_i)$.
- Discretize the possibility distribution as a consonant random set, using $\alpha$-cuts

  $B_1, \ldots B_{k'}$ with masses $v_2(B_j) = n_j = \pi_j - \pi_{j+1}$.
- Compute all $f(E_i, B_j)$ using interval computation.
- for all i and j : compute mass $m_i.n_j$ and assign it to $f(E_i,B_j)$

# Monte Carlo Random Set Computation

- Find a sample $a_1, \ldots a_m$ of $x_1, \ldots x_m$ using a Monte-Carlo method.
- Pick n independent values $\alpha_1, \ldots \alpha_n$ in the unit interval and compute the corresponding $\alpha_j$-cuts $B_1, \ldots B_n$ of $F_1, \ldots F_n$
- For each sample, compute $f(a_1, \ldots a_m, B_1, \ldots B_n)$ using interval computation.

***Note that contrary to the hybrid method, the choice of the cut for each fuzzy interval is done independently.***

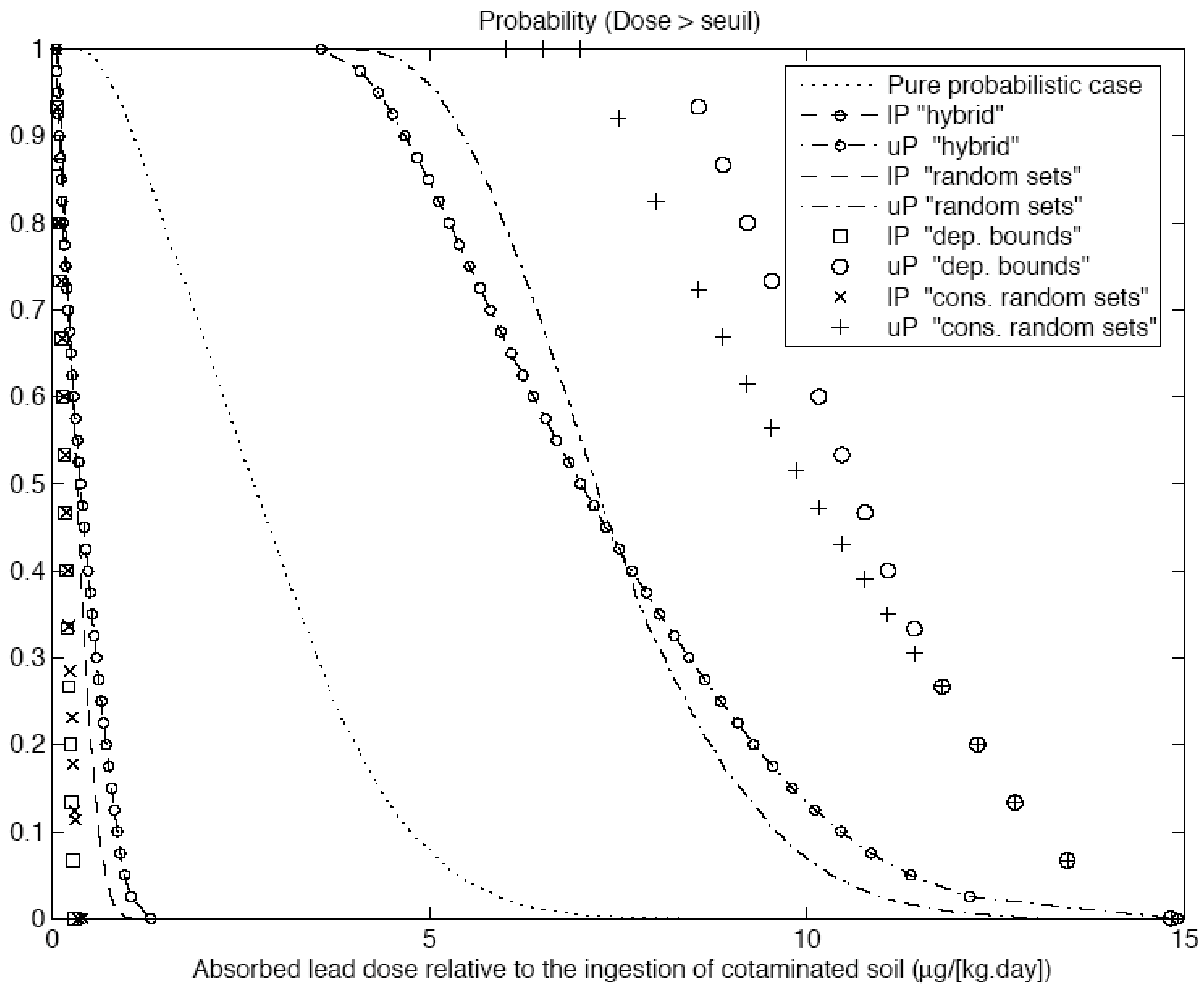
# Conservative random set method without independence

- **Idea** : Maximize and minimize P(A) for events of the form $f(x_1,\ldots x_m, y_1, \ldots y_n) \in A$, *without making any assumption on dependence between variables*.
  - In particular $f(x_1,\ldots x_m, y_1, \ldots y_n) \leq \theta$.
- **Discrete Computation (one π, one p)**
  - Like in the independent case except that the joint mass $x_{ij}$ assigned to $E_i x B_j$ is unknown, and only restricted by its marginals $v_1(E_i)$ and $v_2(B_j)$.

# Calculation method

## WITHOUT INDEPENDENCE ASSUMPTION

- The maxi(mini)mization of P(A) comes down to solving a linear programming problem.
  - Maximize $\sum x_{ij}$ such that $f(E_i, B_j) \cap A \neq \emptyset$
  - such that $\sum_j x_{ij} = v_1(E_i)$ ; $\sum_i x_{ij} = v_1(B_j)$ ; $\sum_{ij} x_{ij} = 1$.
- *One program per event ! Highly combinatorial for fine-grained discretization*.
- But very useful for computing safety PDF bounds in problems involving threshold violation.



Probability (Dose > seuil)
Pure probabilistic case
lP "hybrid"
uP "hybrid"
lP "random sets"
uP "random sets"
lP "dep. bounds"
uP "dep. bounds"
lP "cons. random sets"
uP "cons. random sets"
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
Absorbed lead dose relative to the ingestion of cotaminated soil (µg/[kg.day])

# Fuzzy random variables: three views

There are 3 understandings of a fuzzy random variable, defined by a mapping from a probability space $\Omega$ to the set $\mathcal{F}(\mathrm{R})$ of fuzzy intervals of the real line

1. A random linguistic variable: a classical r.v. whose range is a collection of fuzzy sets modelling linguistic terms (Puri and Ralescu)
2. An ill-known classical random variable whose values are imprecisely observed: a fuzzy set of classical random variables (Kruse)
3. A fuzzy mapping from the sample space to the reals, restricting a family of conditional probabilities

# Frv's as Linguistic Random Variables

- A fuzzy random variable **T**($\omega$) is a classical variable ranging on a space of functions (membership functions $\mu_{\mathbf{T}}$)
- Outcomes of random experiments are represented by vague linguistic terms.
- Linguistic scale: {small, medium, large}
- Linguistic random outcomes: P('small') = 0.25, P('medium') = 0.5 , P('large') = 0.25
- proposed by Puri and Ralescu (1986).

# Ill-known classical random variable

- The fuzzy random variable **T** represents imprecise or vague knowledge about **a classical** random variable, T: Ω → R,referred to as the original random variable.
- It defines a possibility distribution over random variables: $\pi(T)=\inf\{\boldsymbol{\mu}_{\mathbf{T}(\omega)}(T(\omega)), \omega \text{ in } \Omega\}$
- It defines a second order possibility distribution on the set of probability measures:

$$\pi(Q)= \sup\{\pi(T) : P_T = Q\}$$

# Imprecise perception of a random process

- A fuzzy mapping $\Omega \rightarrow \mathcal{F}(R)$ interpreted as a conditional possibility distribution $\pi(T \mid \omega)$ defining a family of conditional probabilities.
- If the result of the random experiment is $\omega$ then the possibility degree of t occurring in the second one is

$$\pi(t \mid \omega) = \boldsymbol{\mu}_{\mathbf{T(\omega)}}(t) \geq P(t \mid \omega).$$

- Using natural extension techniques allows to describe the available information about the probability distribution on R by means of an upper probability (E. Miranda et al.)
- It yields the family

$$\mathcal{P}_T = \{P_T, P_T(\bullet) = \int_\Omega P(\bullet \mid \omega)dP, P(\bullet \mid \omega) \leq \Pi(\bullet \mid \omega)\}$$

# Probability of events induced by a fuzzy random set

- Given a random fuzzy output $\mathcal{R} = \{(C_1,v_1) \ldots (C_N,v_N)\}$ *where* $C_i$ *is a fuzzy interval and* $v_i$ *is its frequency:*
- *Let* $\mathcal{R}_\alpha = \{(C_{\alpha 1},v_1) \ldots (C_{\alpha N},v_N)\}$ *be its* $\alpha$–cut. Then the probability P(A) lies in $[Bel_\alpha(A), Pl_\alpha(A)]$ where

$$Bel_\alpha(A) = P(\mathcal{R}_\alpha \subseteq A) \; ; \; Pl_\alpha(A) = P(\mathcal{R}_\alpha \cap A \neq \emptyset)$$

The probability of A induced by $\mathcal{R}$ is a fuzzy probability $\mathbf{P}_T(A)$ such that its $\alpha$–cut $\mathbf{P}_T(A)_\alpha = [Bel_\alpha(A), Pl_\alpha(A)]$

# Probability of events induced by a fuzzy random set

- The third view of FRV turns $\mathcal{R}$ into the random set obtained by selecting *Ci, with probability $v_i$, then the cut* $C_{\alpha i}$ with probability d$\alpha$.
- Then the probability $P_T(A)$ lies in the interval

$$[\sum_{I=1,N} v_i N_i(A), \sum_{I=1,N} v_i \Pi_i(A)]$$

**Theorem** (Baudrit Couso, Dubois): this interval coincides with the ``mean value" of the fuzzy interval $\mathbf{P}_T(A)$

*also Aumann integral of the cuts of* $\boldsymbol{P}_T(A)$ *(Ogura,Li Ralescu)*

# Empirical cumulative distributions

– *We can extract pairs of average upper and lower PDFs from* $\mathcal{R}$ :

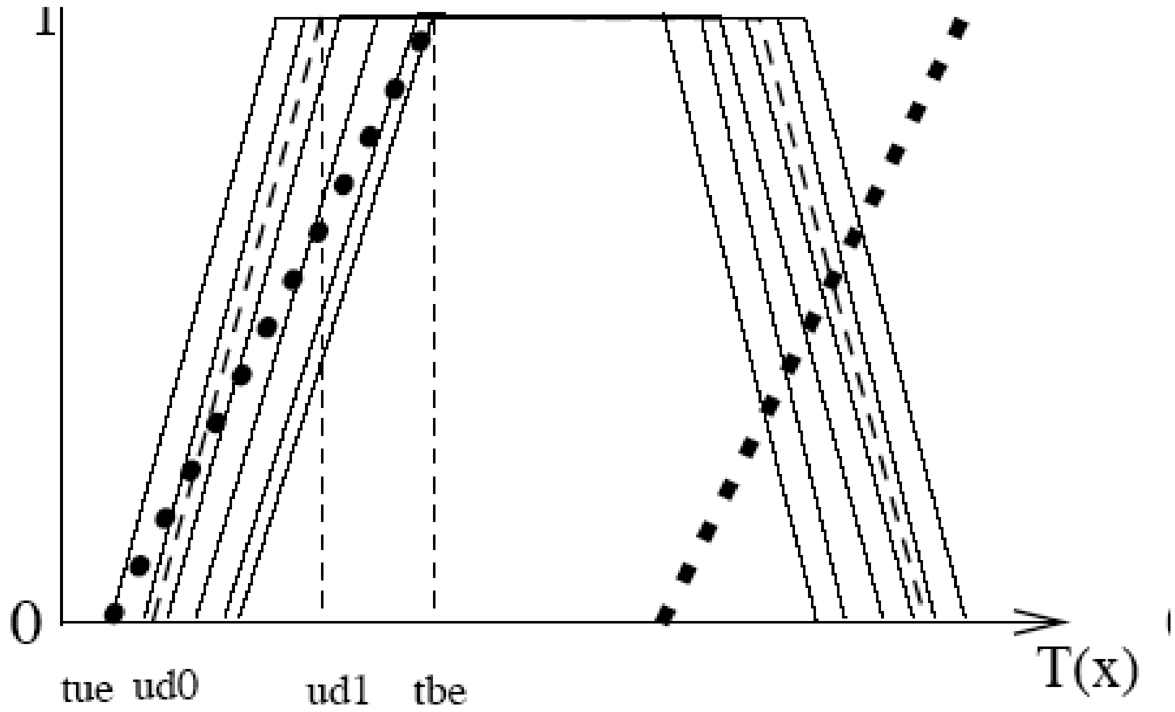
$$F^*(x) = \sum_{i=1,N} \nu_i \cdot F^*_i(x);$$

$$F_*(x) = \sum_{i=1,N} \nu_i \cdot F_{*i}(x)$$

*Where* $F^*_i(x) = \Pi_i(\,(-\infty, x])$

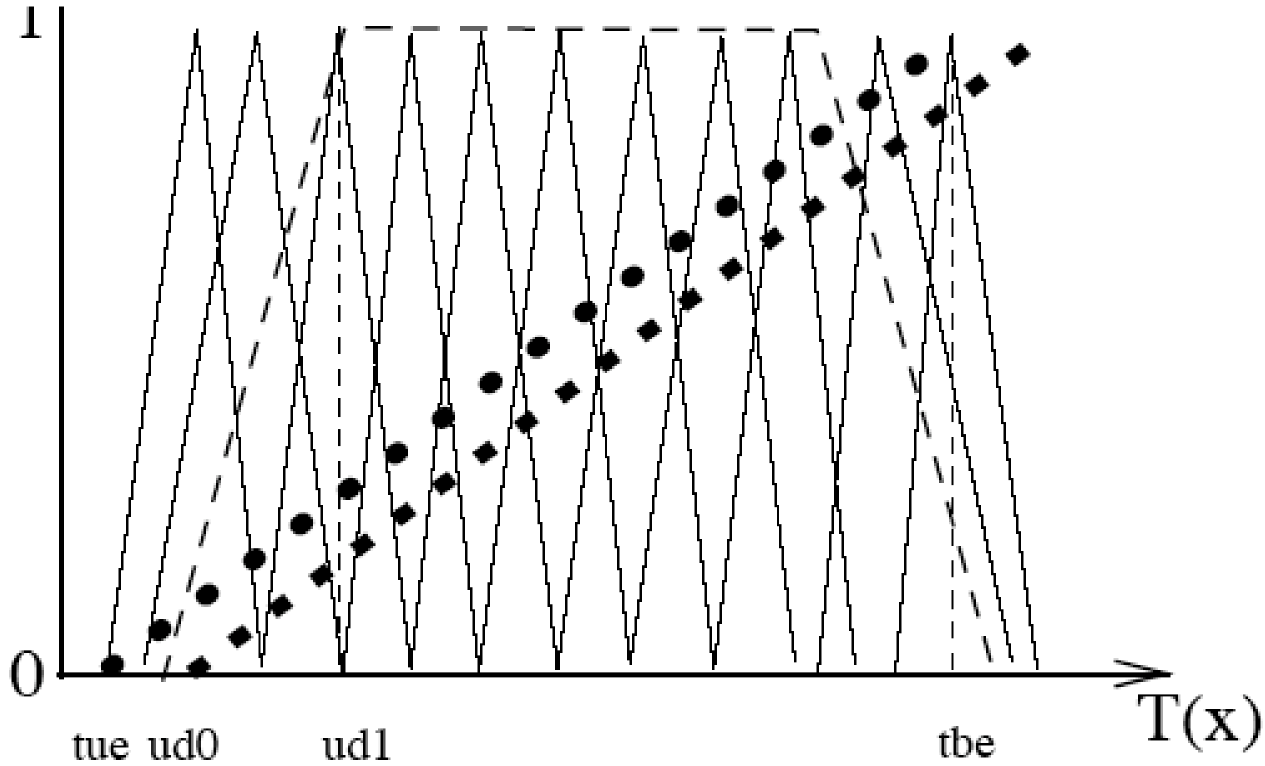
$F_{*i}(x) = \mathbf{N}_i(\,(-\infty, x])$

# Upper and lower distributions of random fuzzy outputs




small variability of the sample
Large imprecision of each fuzzy number Fi

# Upper and lower distributions of random fuzzy outputs




great variability of the sample

i Little imprecision of each fuzzy number Fi

# *Exploiting Random Fuzzy Intervals*

- Given a random fuzzy output $\mathcal{R} = \{(C_1,v_1) \ldots (C_N,v_N)\}$ *where* $C_i$ *is a fuzzy interval and* $v_i$ *is its frequency:*
  - ***Average imprecision****: compute the fuzzy average* $C = \sum_i C_i v_i$. *The average imprecision of* $\mathcal{R}$ *is the area under* $C$. *Also the width of the mean interval (bounded by upper and lower expectations of* $C$*).*
  - ***Observable Variability****: defuzzify the* $C_i$*'s (midpoint of the mean interval) and compute the standard deviation of these numbers*
  - ***Fuzzy Potential Variability****: Compute the fuzzy range of the empirical variance induced by the fuzzy intervals.*
  - ***Minimal and maximal average variability****: compute interval variance of the random set with upper probability* $\sum_{I=1,N} \nu_i \Pi_i(A)$.

# Exploitation of results

- Simple 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 and generalized p-boxes
- *However contrary to the single probability case, a p-box extracted from a credal set is only part of the information.*
- There is a need for extracting useful information indices (mean interval, potential variance, index of specificity, generalized entropy…)

# Knowledge vs. evidence

- There are two kinds of information
  - Generic knowledge
  - Singular evidence
- **Generic knowledge** pertains to a population of items, observables (e.g. laws of physics, statistical knowledge, common sense knowledge)
- **Singular evidence** pertains to a single situation (e.g. unreliable testimony, measurement)
- *Belief often pertains to singular events, and is not necessarily related to statistics.*

# SINGULAR vs. GENERIC INFORMATION

- PIECES OF EVIDENCE refer to a particular situation (measurement data, testimonies) and are singular.
  - E.g. results of medical tests on a patient
  - testimonies
  - Observations about the current state of facts
- BACKGROUND KNOWLEDGE refers to a class of situations and summarizes a set of trends
  - Laws of physics,
  - commonsense knowledge (birds fly)
  - Professional knowledge (of medical doctor),
  - Statistical knowledge

- **Generic knowledge may be tainted with exceptions, incompleteness, variability**
  - *It is not absolutely true knowledge in the mathematical sense*: *tainted with exceptions,*
  - It all comes down to considering some propositions are generally more often the case than other ones.
  - *Generic knowledge induces a normality or plausibility relation on the states of the world.*
  - *numerical* (frequentist) or *ordinal* (plausibility ranking)
  - ***In the numerical case a credal set can account for incomplete generic knowledge***
- **Observed evidence** is often made of propositions known as true about the current world.
  - It is often incomplete and can be encoded as disjunctive sets, or wff in propositional logic.
  - It delimits a reference class of situations for the case under study.
  - It can be uncertain unreliable, (subjective probability, Shafer)
  - It can be irrelevant, wrong,

# GENERIC KNOWLEDGE, EVIDENCE, BELIEFS

- An agent usually possesses three kinds of information on the world
  1. **Generic information** ***(background knowledge)*** : it pertains to a range of situations the agent is aware of.
     - **Examples** : statistics on a well-defined population commonsense knowledge (often ill-defined population)
  2. **Singular information on the current situation** ***(evidence)***
     - *Known facts (results of observations, tests, sensor measurement, testimonies)*
  3. **Beliefs about the current situation**
     - Derived from known facts and generic information

# PLAUSIBLE REASONING

- Inferring **beliefs** (plausible conclusions) on the current situation from observed evidence, using generic knowledge
  - **Example : medical diagnosis**

  Medical knowledge + test results ⇒ believed disease of the patient.
- *This mode of inference makes sense regardless of the representation*, but set-based representations are insufficient:
    - in a purely propositional setting, one cannot tell generic knowledge from singular evidence
    - in the first order logic setting there is no exception.
  - *Need more expressive settings for representing background knowledge, like probability or credal sets*
- **The basic tool for exception-tolerant inference is conditioning.**

# The belief construction problem

- *Beliefs of an agent about a situation are inferred from generic knowledge AND observed singular evidence about the case at hand.*
- **Example: Commonsense plausible inference**
  - Generic knowledge = birds fly, penguin are birds, penguins don't fly.
  - Singular observed fact = Tweety is a bird
  - Inferred belief = Tweety flies
  - Additional evidence = Tweety is a penguin
  - Inferred revised belief = Tweety does not fly

# Belief construction

- *Beliefs of an agent about a situation are derived from generic knowledge and evidence about the case.*
- **Probabilistic beliefs**: Hacking principle again
  - Uncertain singular fact = a set C = what is known about the context of the current situation.
  - Generic knowledge = probability distribution P reflecting the trends in a population (of experiments) relevant to the current situation
  - A question about the current situation: is an uncertain proposition A true in the current situation
  - $Bel_C(A) = P(A|C)$: equating belief and frequency
- **Assumption**: the current situation is typical of situations where C is true

# Conditional Probability

- **Two concepts leading to 2 definitions**:

 1. <u>derived</u> (Kolmogorov): $P(A \mid C) = \dfrac{P(A \cap C)}{P(C)}$

    requires $P(C) \neq 0$

 2. <u>primitive</u> (de Finetti): $P(A|C)$ is directly assigned a value and P is derived such that

    $P(A \cap C) = P(A|C) \cdot P(C)$.

    - Makes sense even is $P(C) = 0$

**Meaning :** $P(A \mid C)$ is the probability of A

*if C represents all that is hypothetically known on the situation*

# THE MEANING OF CONDITIONAL PROBABILITY

- P(A|C) : probability of a conditional event « A in epistemic context C » (when C is all that is known about the situation).
- *It is NOT the probability of A, if C is true*.
- **Counter-example** :
  - Uniform Probability on {1, 2, 3, 4, 5}
  - P(Even |{1, 2, 3}) = P(Even |{3, 4, 5}) = 1/3
  - Under a classical logic interpretation :
    - From « if result ∈ {1, 2, 3} then P(Even) = 1/3 »
    - And« if result ∈ {3, 4, 5} then P(Even) = 1/3 »
    - Then (classical inference) : P(Even) = 1/3 unconditionally!!!!!
  - **But of course: P(Even) = 2/5.**
- So, conditional events A|C should be studied as single entities (De Finetti).

# The nature of conditional probability

- In the frequentist settting a conditional probability P(A|C) is a relative frequency.
- It can be used to represent the weight of rules of the form « generally, if C then A » understood as « Most C's are A's » with exceptions

In logic a rule « if C then A » is represented by material implication $C^c \cup A$ that rules out exceptions

- *But the probability of a material conditional is not a conditional probability!*
- *What is the entity* A|C *whose probability is a conditional probability???*

***A conditional event!!!!***

# Material implication: the raven paradox

- Testing the rule « all ravens are black » viewed as $\forall x, \neg Raven(x) \vee Black(x)$
- Confirming the rule by finding situations where the rule is true.
  - Seeing a black raven confirms the rule
  - Seeing a white swan also confirms the rule.
  - But only the former is an example of the rule.

# 3-Valued Semantics of conditionals

- A rule « if C then A » shares the world into 3 parts
  - **Examples:** interpretations where $A \cap C$ is true
  - **Counterexamples:** interpretations where $A^c \cap C$ is true
  - **Irrelevant cases:** interpretations where C is false

*Rules « all ravens are black » and « all non-black birds are not ravens » have the same exceptions (white ravens), but different examples (black ravens and white swans resp.)*

- Truth-table of « A|C » viewed as a connective
  - Truth(A|C) = T if truth(A)= truth(C) = T (
  - Truth(A|C) = F if truth(A)=T and truth(C) = F
  - Truth(A|C) = I if truth(C)= F

  Where I is a 3d truth value expressing « irrelevance »:

  I = T: $A \cup C^c$ ; I = F: $A \cap C$ .

# A conditional event is a pair of nested sets

- The solutions of $A \cap C = X \cap C$ form the set

$$\{X: A \cap C \subseteq X \subseteq A \cup C^c\}$$

- It defines the symbolic Bayes-like equation:

$$A \cap C = (A|C) \cap C.$$

- The models of a conditional $A|C$ can be represented by the pair $(A \cap C, A \cup C^c)$, an interval in the Boolean algebra of subsets of S
- The set $A \cup C^c$ representing material implication contains the « non-exceptions » to the rule (the complement of $A \cap C^c$).

# Semantics for three-valued logic of conditional events.

- <u>Semantic entailment</u>: $A|C \models B|D$ iff

  $A \cap C \subseteq B \cap D$ and $C^c \cup A \subseteq D^c \cup B$

*B|D has more examples and less counterexamples than A|C.*

In particular $A|C \models A|B \cap C$ is false.

- <u>Quasi-conjunction</u> (Ernest Adams):

$A|C \cap B|D = (C^c \cup A) \cap (D^c \cup B) | C \cup D$

# Probability of conditionals

P(A|C) is totally determined by

– $P(A\cap C)$ (proportion of examples)

– $P(A^c\cap C) = 1 - P(A\cup C^c)$ (proportion of counter-examples)

$$P(A|C) = \frac{P(A\cap C)}{P(A\cap C) + 1 - P(A\cup C^c)}$$

- P(A|C) is increasing with $P(A\cap C)$ and decreasing with $P(A^c\cap C)$
- If $A|C \models B|D$ then $P(A|C) \leq P(B|D)$.

# CONDITIONING NON-ADDITIVE CONFIDENCE MEASURES

- **Definition :** A conditional confidence measure $g(A \mid C)$ is a mapping from conditional events $A \mid C \in S\times(S - \{\emptyset\})$ to $[0, 1]$ such that
  - $g(A \mid C) = g(A \cap C \mid C) = g(A^c \cup C \mid C)$
  - $g_C(\cdot) = g(. \mid C)$ is a confidence measure on $C \neq \emptyset$
- Two approaches:
  - Bayes-like $g(A \cap C) = g(A \mid C) \cdot g(C)$
  - Explicit Approach $g(A \mid C) = f(g(A \cap C), g(A \cup C^c))$
    Namely : $f(x, y) = x/(1+x-y)$

# Conditioning a credal set

- *Let $\mathcal{P}$ be a credal set representing generic information and C an event*
- *Two types of processing :*

1. ***Querying** : C represents available singular facts: compute* the degree of belief in A in context C as

   $Cr(A \mid C) = Inf\{P(A \mid C), P \in \mathcal{P}, P(C) > 0\}$ (Walley).

2. ***Revision** : C represents a set of universal truths;*

   *Add P(C) = 1 to the set of conditionals $\mathcal{P}$.*

   Now we must compute $Cr(A|C) = Inf\{P(A)\ P \in \mathcal{P}, P(C) = 1\}$

   If P(C) = 1 is incompatible with $\mathcal{P}$, use maximum likelihood:

   $Cr(A|C) = Inf\{P(A|C)\ P \in \mathcal{P}, P(C) \text{ maximal }\}$

## ***Example :*** $A \rightleftarrows B \longrightarrow C$

- *$\mathcal{P}$ is the set of probabilities such that*
  - $P(B|A) \geq \alpha$ *Most A are B*
  - $P(C|B) \geq \beta$ *Most B are C*
  - $P(A|B) \geq \gamma$ *Most B are A*
- ***Querying on context A :*** *Find the most narrow interval for P(C|A) (Linear programming): we find*

$$P(C|A) \geq \alpha \cdot max(0, 1 - (1 - \beta)/\gamma)$$

  - *Note : if $\gamma = 0$, $P(C|A)$ is unknown even if $\alpha = 1$.*
- ***Revision:*** *Suppose $P(A) = 1$, then $P(C|A) \geq \alpha \cdot \beta$*
  - *Note: $\beta > max(0, 1 - (1 - \beta)/\gamma)$*
- **Revision improves generic knowledge, querying does not.**

# CONDITIONING RANDOM SETS AS IMPRECISE PROBABILISTIC INFORMATION

- A disjunctive random set ($\mathcal{F}$, m) representing background knowledge is equivalent to a set of probabilities $\mathcal{P} = \{P: \forall A, P(A) \geq Bel(A)\}$ (*NOT conversely*).
- Querying this information based on evidence C comes down to performing a sensitivity analysis on the conditional probability $P(\cdot|C)$
  - $Bel_C(A) = \inf \{P(A|C): P \in \mathcal{P}, P(A) > 0\}$
  - $Pl_C(A) = \sup \{P(A|C): P \in \mathcal{P}, P(A) > 0\}$

- **Theorem:** functions $Bel_C(A)$ and $Pl_C(A)$ are belief and plausibility functions of the form

  $$Bel_C(A) = Bel_C(C \cap A)/(Bel_C(C \cap A) + Pl_C(C \cap A^c))$$

  $$Pl_C(A) = Pl_C(C \cap A)/(Pl_C(C \cap A) + Bel_C(C \cap A^c))$$

  where $Bel_C(A) = 1 - Pl_C(A^c)$
- *This conditioning does not add information*:
- If $E \cap C \neq \emptyset$ and $E \cap C^c \neq \emptyset$ for all $E \in \mathcal{F}$, then $m_C(C) = 1$ (the resulting mass function $m_C$ expresses total ignorance on C)
  - **Example: If opinion poll yields:**
  - $m(\{a, b\}) = \alpha$, $m(\{c, d\}) = 1 - \alpha$,

  The proportion of voters for a candidate in $C = \{b, c\}$ is unknown.
  - *However if we hear a and d resign ($Pl(\{a, d\} = 0)$ then $m(\{b\}) = \alpha$, $m(\{c\}) = 1 - \alpha$* (Dempster conditioning, see further on)

# CONDITIONING UNCERTAIN SINGULAR EVIDENCE

- A mass function *m on S, represents uncertain evidence*
- A new **sure** piece of evidence is viewed as a conditioning event C

1. *Mass transfer* : for all $E \in \mathcal{F}$, m(E) moves to $C \cap E \subseteq C$
   - The mass function after the transfer is $m_t(B) = \Sigma_{E : C \cap E = B}\, m(E)$
   - But the mass transferred to the empty set may not be zero!
   - $m_t(\emptyset) = Bel(C^c) = \Sigma_{E : C \cap E = \emptyset}\, m(E)$ is the degree of conflict with evidence C
2. *Normalisation*: $m_t(B)$ should be divided by Pl(C)

   $= 1 - Bel(C^c) = \Sigma_{E : C \cap E \neq \emptyset}\, m(E)$

- ***This is revision of an unreliable testimony by a sure fact***

# DEMPSTER RULE OF CONDITIONING = PRIORITIZED MERGING

The conditional plausibility function Pl(.|C) is

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

- C surely contains the value of the unknown quantity described by m. **So $Pl(C^c) = 0$**
  - *The new information is interpreted as asserting the impossibility of $C^c$: Since $C^c$ is impossible you can change $x \in E$ into $x \in E \cap C$ and transfer the mass of focal set E to $E \cap C$.*
- *The new information improves the precision of the evidence*

- **This conditioning is different from Bayesian (Walley) conditioning**

# EXAMPLE OF REVISION OF EVIDENCE : The criminal case

- **Evidence 1** : three suspects : Peter Paul Mary
- **Evidence 2** : The killer was randomly selected man vs.woman by coin tossing.
  - So, S = { Peter, Paul, Mary}
- TBM modeling: The masses are m({Peter, Paul}) = 1/2 ; m({Mary}) = 1/2
  - Bel(Paul) = Bel(Peter) = 0. Pl(Paul) = Pl(Peter) = 1/2
  - Bel(Mary) = Pl(Mary) = 1/2
- **Bayesian Modeling:** A prior probability
  - P(Paul) = P(Peter) = 1/4; P(Mary) = 1/2

- **Evidence 3** : Peter was seen elsewhere at the time of the killing.
- **TBM**: So Pl(Peter) = 0.
  - m({Peter, Paul}) = 1/2;       $m_t$({Paul}) = 1/2
  - *A uniform probability on {Paul, Mary} results*.
- **Bayesian Modeling:**
  - P(Paul | not Peter) = 1/3; P(Mary | not Peter) = 2/3.
  - A very debatable result that depends on where the story starts. *Starting with i males and j females:*
    - P(Paul | Paul OR Mary) = j/(i + j);
    - P(Mary | Paul OR Mary) = i/(i + j)
- *Walley conditioning:*
  - Bel(Paul) = 0; Pl(Paul) = 1/2
  - Bel(Mary) = 1/2; Pl(Mary) = 1

# A second problem : MERGING UNCERTAIN EVIDENCE

- Observations about the current world may be unreliable, uncertain, inconsistent:
  - *Sensor failures, dubious testimonies*
  - Propositional logic cannot account for **unreliable** evidence
  - Probability theory alone cannot account for **incomplete** evidence
- A proper account of uncertain evidence needs to cope with incompleteness and the necessity for **merging unreliable evidence** in a flexible way, before even inferring beliefs

# The evidence merging problem

- Beliefs can be directly elicited (e.g. as subjective probabilities) about singular facts with no frequentist flavor.
- Uncertain testimonies form a collection of uncertain singular facts
- **Ascertaining singular facts enables to perform plausible inference from them, and then construct inferred beliefs**
- Ascertaining singular evidence is the result of merging uncertain singular pieces of evidence

# UNCERTAIN INFORMATION MERGING

- **Contexts**:
  - experts; sensors; images;
  - belief sets; databases; sets of propositions.
- Neither classical logic nor probability theory explain how to combine conflicting information.
- Merging beliefs differs from preference aggregation, revision.
- Theories (probability, possibility, random sets, etc...) supply connectives without explaining how to use them
- The problem is independent from the chosen representation.

# WORKING ASSUMPTIONS

- Parallel information sources
- Sources are identified, heterogeneous, dependent (humans, sensors.)
- A range of problems : informing about the value of some ill-known quantity to the identification of a scenario
- Information can be poor (intervals, linguistic), incomplete, ordinal
- No prior knowledge must be available
- Reliability of sources possibly unknown, or not quantified
- Sources supposedly refer to the same problem (non-trivial issue)

# BASIC MERGING MODES

source 1: $x \in A$

$x \in ?$ *3 basic possibilities*

source 2: $x \in B$

1. **Conjunctive merging**: $x \in A \cap B$
   - Assumption : sources are totally reliable
   - Usual in logic if no contradiction ($A \cap B \neq \emptyset$)
2. **Disjunctive merging:** $x \in A \cup B$
   - Assumption : one of the two sources is reliable
   - Imprecise but sure response : $A \cap B = \emptyset$ is allowed

# BASIC MERGING MODES

**3. Merging by counting**:

build the random set: m(A) = m(B) = 1/2.

- AMB (x) = Pl(x) = $\sum_{x \in E}$ m(E) = 1 if $x \in A \cap B$

  = $^1/_2$ if $x \in (A^c \cap B) \cup (A \cap B^c)$

  = 0 otherwise

  - It lies between conjunctive and disjunctive (but *AMB is a fuzzy set):* $A \cap B \subseteq AMB \subseteq A \cup B$
  - Assumption : Pieces of information stem from identical independent sources: confirmation effect.
  - Usual assumption in statistics with many sources and precise observations

# Extension to n sources : conflict management with incomplete information

- A set $S$ of n sources i: $x_i \in A_i$, i = 1, …, n
  - Generally inconsistent so conjunctive merging fails
  - *Significant dispersion so disjunctive merging is uninformative*
  - (there is often more than one reliable source among n)
- **Method 1** : Find maximal consistent subsets of sources $\mathcal{T}_k$: $\cap_{i \in \mathcal{T}} A_i \neq \emptyset$ but $\cap_{i \in \mathcal{T} \cup \{j\}} A_i = \emptyset$
  - Conjunctive merging of information in $\mathcal{T}_k$
  - Disjunctive merging of partial results obtained

$$X = \cup_k (\cap_{i \in \mathcal{T}k} A_i)$$

- **Method 2** : Make an assumption on the <u>number</u> of reliable sources
- *Suppose k reliable sources*
- Then pick k sources at random for conjunctive merging and then disjunctively merge obtained results

$$X = \bigcup_{\mathcal{K} \subseteq \mathcal{S} : \text{card}(K) = k} \bigcap_{i \in \mathcal{K}} A_i$$

  – Must choose $k \leq \max \{\text{card}(K), \bigcap_{i \in \mathcal{K}} A_i \neq \emptyset\}$

- **Method 3** : statistical : $m(A_i) = 1/n$ for all i.

$$\text{then } Pl(x) = \sum_{i = 1, \ldots, n} A_i(x)/n.$$

# MERGING IN POSSIBILITY THEORY:

- *Fuzzy set-theoretic operations are instrumental.*
- General case:
- source 1 → $\pi_1 = \mu_{F^1}$ source 2 → $\pi_2 = \mu_{F^2}$

1. **Conjunctive merging** $F_1 \cap F_2$
   - Assumption 1 : Nothing is assumed about dependence of sources
   - **Then, Idempotence**: no accumulation effect :

- $\pi_{\cap} = \min(\pi_1, \pi_2)$ (minimum rule)
- In agreement with the logical view of information as constraints

# Normalized conjunctive merging

- **Degree of conflict**: $1 - \max \pi_{\cap}$ if $\pi_{\cap}$ is not normalized
  - **Renormalizing :** Assumption 2: sources are reliable even if conflict.
- Assumptions 1 and 2 : $\pi_{\cap}^{*} = \min(\pi_1, \pi_2) / \max \pi_{\cap}$
  - *But then Associativity is lost*
- Assumption 3: **Independent sources**: $\pi_{*} = \pi_1 \cdot \pi_2$
  - product instead of min.
  - **Renormalizing :** $\pi = \pi_1 \cdot \pi_2 / \max \pi_{*}$
  - in agreement with the Bayesian approach.
  - *Associativity is preserved*

- **Possibilistic disjunctive merging**
  - Assumption 4: one of the sources is reliable

  $F_1 \cup F_2$: $\pi_{\cup} = \max(\pi_1,\pi_2)$ (max rule)
  - **Idempotent**: sources can be redundant.
  - Adapted for inconsistent sources ( $F_1 \cap F_2 = \emptyset$)

- **Statistical Merging** *vertical average*

$$\pi_+ = (\pi_1 + \pi_2)/2$$

  - Assumption 5: Numerous identical independent sources
  - *Generally it gives a random fuzzy set*.

# MERGING PROBABILITY DISTRIBUTIONS

- The basic connective is the *convex combination : a counting scheme*
  - $P_1 \ldots P_n$ probability distributions
  - Information sources with weights $\alpha_i$ such that $\sum \alpha_i = 1$

$$P = \sum \alpha_i P_i$$

- The only possible one with
  - $P(A) = f(P_1(A),\ldots,P_n(A)) \quad \forall A \subseteq S$
  - $f(0, 0\ldots0) = 0$ ; $f(1, 1\ldots1) = 1$
  - (invariant via projections)
- Information items come from a random source ; weights express repetition of sources: Information items are independent from each other

# Bayesian Merging

- **Idea** : there is a unique probability distribution capturing the behaviour of sources.
- **Data:**
  - $x_i$: observation of the value of quantity x by source i.
  - $P(x_1 \text{ and } x_2 \mid x)$ information about source behaviour
  - $P(x)$ prior information about the value of x
- $$P(x \mid x_1 \text{ and } x_2) = \frac{P(x_1 \text{ and } x_2 \mid x).\, P(x)}{\sum_{x'} P(x_1 \text{ and } x_2 \mid x') \cdot P(x')}$$
- (requires a lot of data )

# « Idiot Bayes »

- Usual assumption: precise observations $x_1$ and $x_2$ are conditionally independent with respect to x.
- $$P(x \mid x_1 \text{ and } x_2) = \frac{P(x_1 \mid x) \cdot P(x_2 \mid x) \cdot P(x)}{\sum_{x'} P(x_1 \mid x') \cdot P(x_2 \mid x') \cdot P(x')}$$
  - Independence assumption often unrealistic
  - Conjunctive product-based combination rule similar to possibilistic merging, if we let $P(x_i \mid x) = \pi_i(x)$
- *A likelihood function is an example of a possibility distribution*

- ***What if no prior information****?* Bayesians use Laplace principle: A uniform prior

- $$P(x \mid x_1 \text{ and } x_2) = \frac{P(x_1 \mid x) . P(x_2 \mid x)}{\sum_{x'} P(x_1 \mid x') . P(x_2 \mid x')}$$

- **Too strong** : merging likelihood functions should yield a likelihood function.

- $$\pi(x) = \frac{P(x_1 \mid x) . P(x_2 \mid x)}{\sup_{x'} P(x_1 \mid x') . P(x_2 \mid x')}$$ possibilistic merging

# Possibilistic merging with prior information

- *Bayes theorem*:

  $$\pi(u_1, u_2|u) * \pi_x(u) = \pi_x(u|u_1, u_2) * \pi(u_1, u_2).$$

  – $\pi_x(u)$ a priori information about x (uniform = ignorance)
  – $\pi(u_1, u_2|u)$: results from a merging operation F
  – $\pi(u_1, u_2) = \sup_{u \in U} \pi(u_1, u_2|u) \cdot \pi_x(u)$.
- If operation F is product:

$$\pi(u) = \frac{\pi(u_1 \mid u). \pi(u_2 \mid u) . \pi_x(u).}{\sup_{u'} \pi(u_1 \mid u'). \pi(x_2 \mid u') . \pi_x(u)}$$

- Similar to probabilistic Bayes **but** more degrees of freedom

# MERGING BELIEF FUNCTIONS

- Problem :
- source i → $(\mathcal{F}^i, m_i)$ with $\sum_{A \in \mathcal{F}^i} m_i(A) = 1$

- Demspter rule of combination : *an associative scheme* generalising Dempster conditioning
  - Step 1: $m_\cap(C) = \sum_{A \cap B = C} m_1(A).m_2(B)$

    Independent random set intersection

  Step 2: $m^*(C) = m_\cap(C)/(1 - m_\cap(Ø))$

    renormalisation

$m_\cap(Ø)$ evaluates conflict ; it is eliminated.

**Example** : S = {a, b, c, d}

| $m_2$ / $m_1$ | {c} 0.2 | {b, c, d} 0.7 | S 0.1 |
|---|---|---|---|
| {b} 0.3 | Ø 0.06 | {b} 0.21 | {b} 0.03 |
| {a, b, c} 0.5 | {c} 0.1 | {b, c} 0.35 | {a, b, c} 0.05 |
| S 0.2 | {c} 0.04 | {b, c, d} 0.14 | S 0.02 |

$m_\cap(\{b\}) = 0.21 + 0.03 = 0.24$ ; $m_\cap(\{c\}) = 0.1 + 0.04 = 0.15$

$m_\cap(S) = 0.02$ ; $m_\cap(\text{Ø}) = 0.06$

# Disjunctive merging of belief functions

$$m_{\cup}(C) = \sum_{C\,:\,A \cup B = C} m_1(A).m_2(B)$$

- *Union of independent random sets*.
- More imprecise than conjunctive merging, ever normalised.

- Moreover $Bel_{\cup}(A) = Bel_1(A).Bel_2(A)$
  - Disjunctively combining two probability distributions yields a random set.
  - *Belief functions are closed via product and convex sum*.
  - If conflict is too strong, normalized conjunctive merging provides arbitrary results and should be avoided: use another scheme like disjunctive merging.

# Conjunctive merging with disjunctive conflict management

1. **Conflict is ignorance**
   - $m_{\cap\delta}(C) = \sum_{A \cap B = C} m_1(A).m_2(B)$ if $C \neq \emptyset, S$
   - $m_{\cap\delta}(S) = \sum_{A \cap B = \emptyset} m_1(A).m_2(B) + m_1(S).m_2(S)$

     $= m_{\cap}(\emptyset) + m_{\cap}(S)$

2. **Adaptive rule**: for $C \neq \emptyset$

$$m_{\cap\delta}(C) = \sum_{A \cap B = C} m_1(A).m_2(B) + \sum_{\substack{A \cup B = C \\ A \cap B = \emptyset}} m_1(A).m_2(B)$$

*These rules are not associative*.

# Compromise merging

- **Convex combination:** generalisation of the probabilistic merging rule
- $m_\alpha(A) = \alpha \cdot m_1(A) + (1 - \alpha)\, m_2(A)$
  - $\alpha$ = relative reliability of source 1 versus source 2
- **Example** : discounting an unreliable belief function with reliability $\alpha$ close to 1: combine $m_1$ with the void belief function $m_2(S) = 1$: then
  - $m_\alpha(A) = \alpha \cdot m_1(A)$ if $A \neq S$
  - $m_\alpha(A) = \alpha \cdot m_1(S) + (1 - \alpha)$

# CONCLUSION: Belief construction for an agent

1. **Perception**: collecting evidence tainted with uncertainty
2. **Merging**: Combining new evidence with current one so as to lay bare an (incomplete) description of the current situation considered as true.
3. **Plausible inference**: Forming beliefs by applying background knowledge to current evidence

- This scheme can be implemented in various settings encompassed by imprecise probability, but
  - A set-based approach is too poor : need conditional events. Non-monotonic reasoning a la Lehmann (or qualitative possibilistic logic) is minimal requirement for step 3.
  - Bayesian probability is too rich for step 3: ever complete and consistent. Walley conditioning with imprecise probability is purely deductive and may be poorly informative.
  - Shafer-Smets or possibility theory is useful for merging uncertain evidence (step 2)

# Conclusion: the role of imprecise probability methods

- Imprecise probabilities are a natural concept for conjointly handling variability of phenomena and incomplete knowledge about them.
- Imprecise probabilities unify quantitative uncertainty theories to a large extent.
  - Some discrepancies remain, e.g. Dempster rule of combination of random sets is not interpreted in the imprecise probability setting…

# 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 represents altogether knowledge about reality and knowledge about knowledge.***

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

# Generalized p-boxes

- It comes down to two possibility distributions

  $\pi$ (from $\alpha_i \leq P(E_i)$) and $\pi_c$ (from $P(E_i) \geq 1 - \beta_i$)
- Distributions $\pi$ and $\pi_c$ are such that $\pi \geq 1 - \pi_c = \delta$ and **$\pi$ is comonotonic with $\delta$** (they induce the same order on the referential).
- 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

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