UNCERTAINTY THEORIES: A UNIFIED VIEW

D. Dubois

IRIT-CNRS, Université Paul Sabatier

31062 TOULOUSE FRANCE

Outline

1. Variability vs ignorance
2. Set-valued representations of ignorance
3. Capacity-based uncertainty theories and their links to imprecise probability
4. Practical representations
5. Information, conditioning and fusion

Origins of uncertainty

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

Example

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

Knowledge vs. evidence

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

The roles of probability

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

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

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

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 :**
  - Frequentist probability is generic knowledge (statistics from a population)
  - Subjective probability pertains to singular events (degrees of belief)

Constructing beliefs

- Belief in the occurrence of a particular event may derive from its statistical probability: the **Hacking principle**:
  - Generic knowledge = probability distribution P
  - $belief_{NOW}(A) = Freq_{POPULATION}(A)$: 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*

SUBJECTIVE PROBABILITIES
(Bruno de Finetti, 1935)

- p_i = *belief degree* of an agent on the (next) 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 banker finds price p_i is too low
- **Why a belief state is a single distribution** ($\sum_j p_j = 1$):
  - 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

Bayesian probability

- **Bayesian postulate** : any state of knowledge can be represented by a single probability distribution:
  - Either via an exchangeable betting procedure
  - Or by comparison with an urn of a given composition
- Not to do it is considered to be irrational (sure money loss, Dutch book argument)

Why the unique distribution assumption?

- **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*
- The exchangeable betting framework enforces unique elementary probability assessments that 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.
- *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 one 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*

Language sensitiveness 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

LIMITATIONS OF BAYESIAN PROBABILITY FOR THE REPRESENTATION OF IGNORANCE

- ***Ignorance****: identical belief in any event different from the sure or the impossible ones*

- *A single probability cannot represent ignorance*: except on a 2-element set, the function $g(A) = 1/2 \ \forall A \neq S, \emptyset$, is NOT a probability measure.
- In the *life on other planets* example: 6 possible contingent events that cannot have the same probability.
- Function g is monotonic under inclusion : a capacity.

Ellsberg Paradox

- Savage claims that rational decision-makers choose acts according to expected utility *EU(a)* with respect to a subjective probability : *a better than b iff* $EU(a) > EU(b)$
- 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 : $EU(a) = u_a(R)p_R + u_a(W)p_W + u_a(B)p_B$
- *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.

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*.

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)

What do set-valued data mean?

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

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 is a simple modal logic (KD)

Find a representation of uncertainty due to incompleteness

- *More expressive than 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.*

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
  - $g(\emptyset) = 0$; $g(S) = 1$
  - **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) ≤ Pl(A) "certain implies plausible"
  - $Pl(A) = 1 - Cr(A^c)$ duality certain/plausible
  - If Pl = Cr then it is P.

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 : $\mu_E : S \rightarrow [0, 1]$
- $\mu_E(s)$ = 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)

Improving expressivity of incomplete information representations

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

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

POSSIBILITY AND NECESSITY OF AN EVENT

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

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

  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

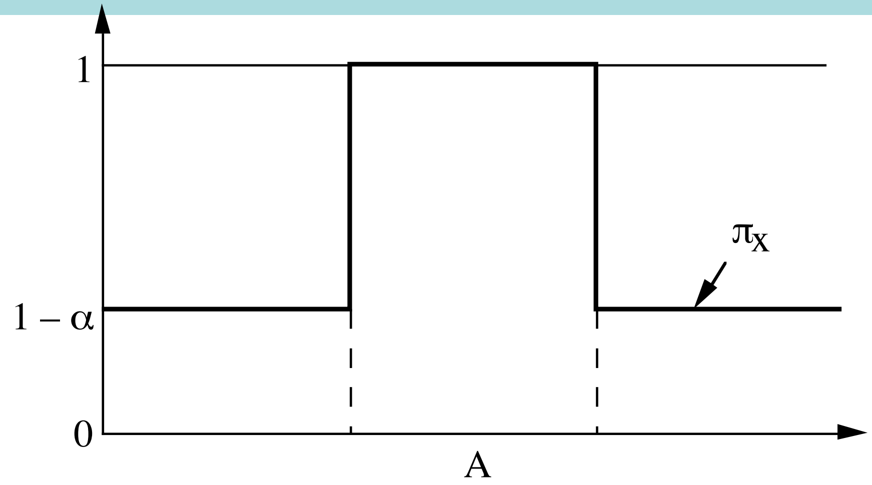
- **π' more specific than π in the wide sense if and only if $\pi' \leq \pi$**

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

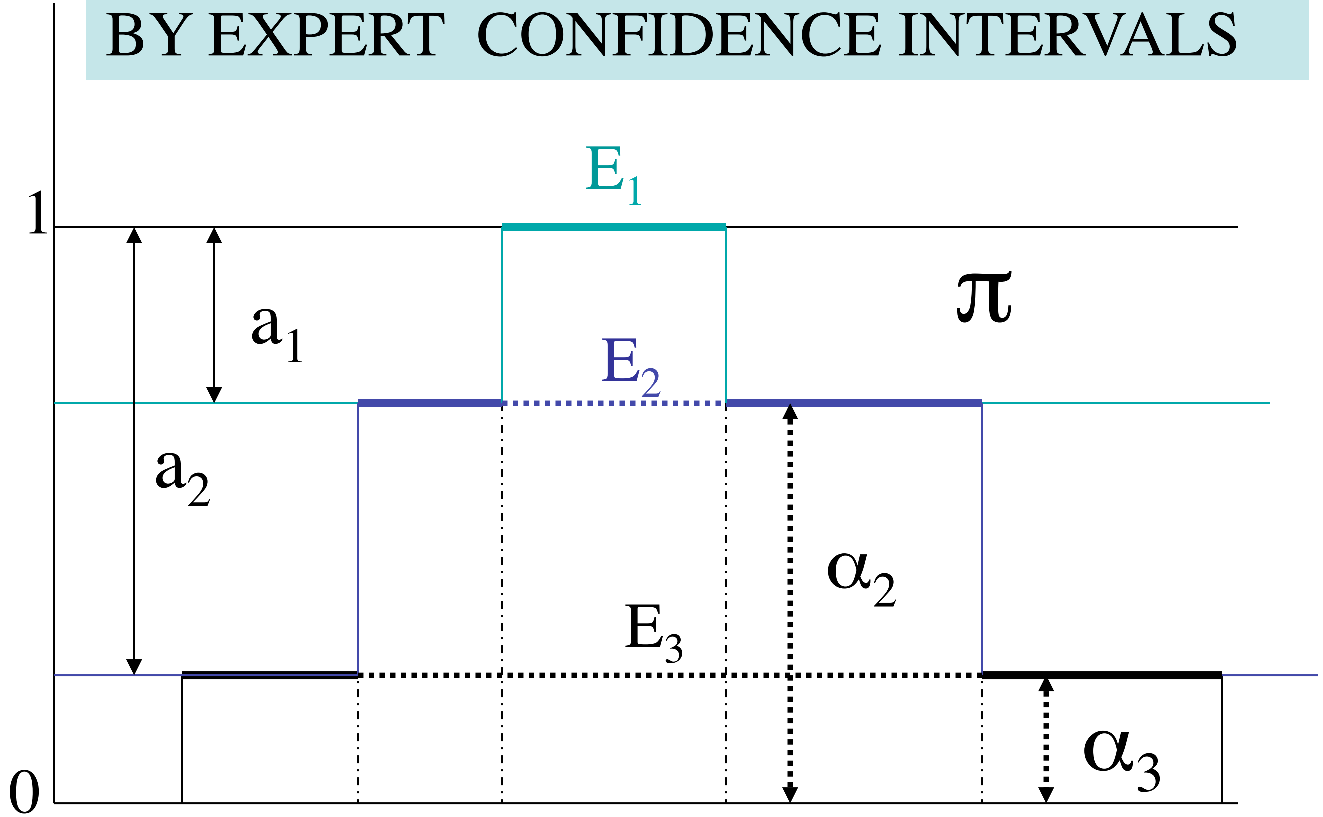
- **Principle of least commitment** (minimal specificity): In a given information state, any value not proved impossible is supposed to be possible : maximise possibility degrees.

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: $\pi(s) = \max(\mu_A, 1 - \alpha)$

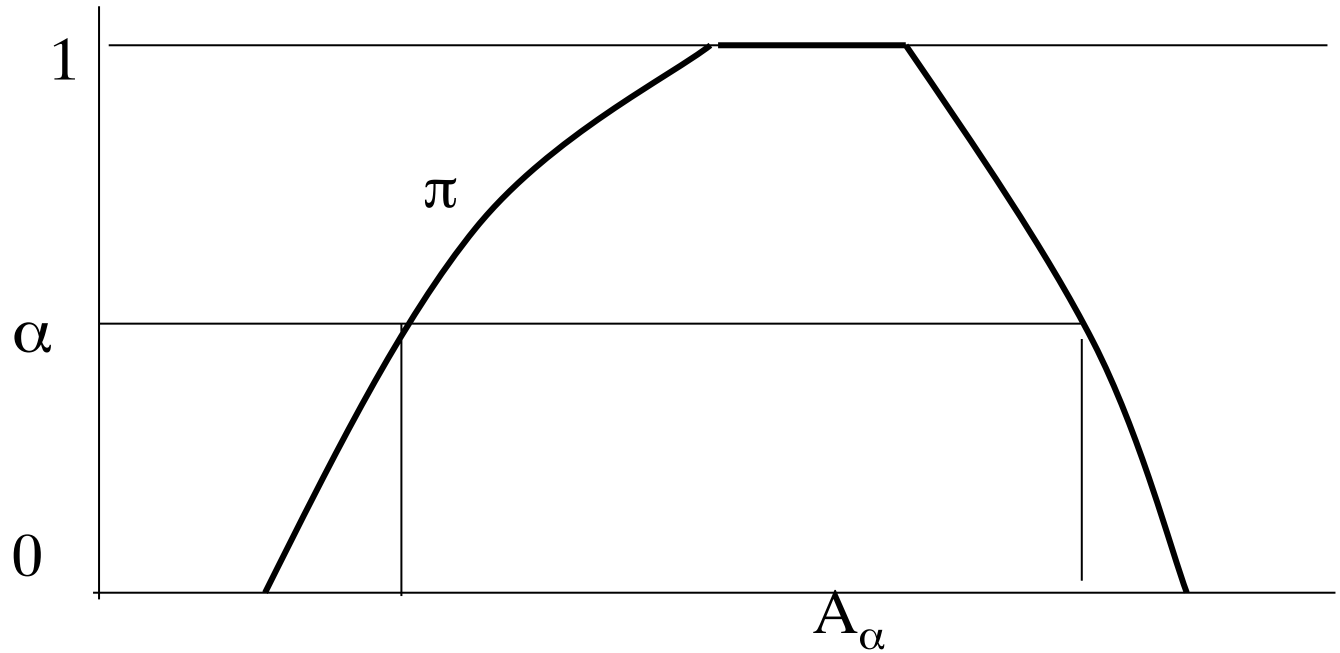
POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



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

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

One must indicate where the numbers come from.

All theories agree on the fundamental maxitivity axiom

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

Theories diverge on the conditioning operation

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 outcomes
  - Special convex probability sets

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

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"
  - = *probability(only knowing "x $\in$ E" and nothing else)*
  - It is the portion of probability mass hanging over elements of E without being allocated.
- **DO NOT MIX UP m(E) and P(E)**

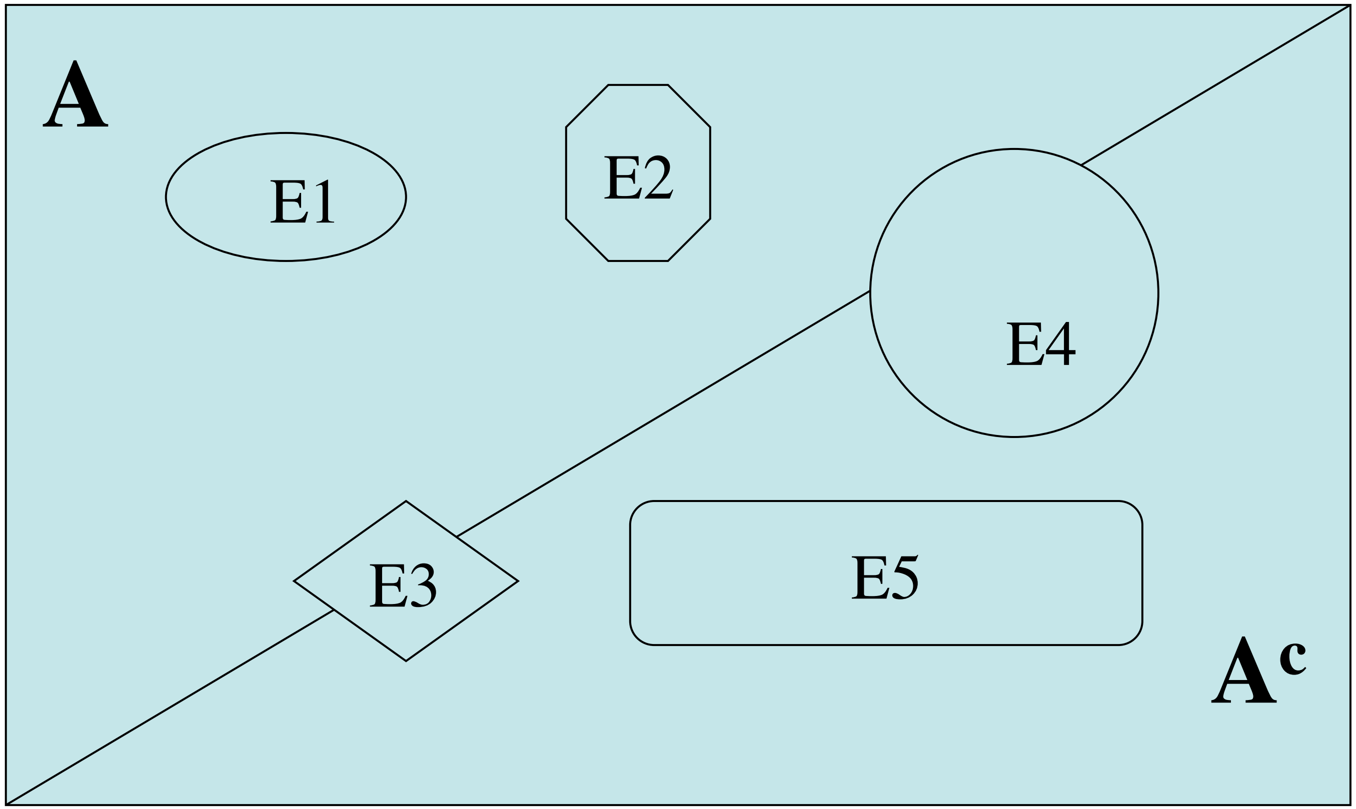
Basic set functions from random sets

- **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)



PARTICULAR CASES

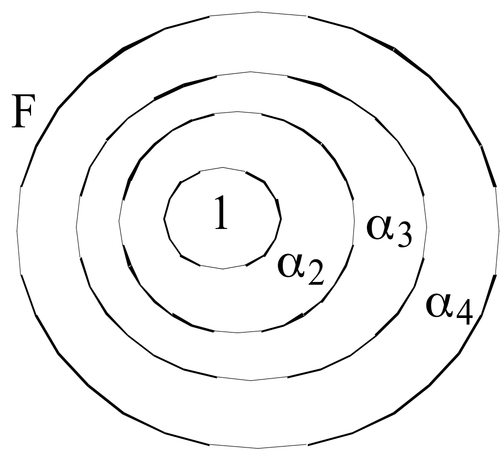
- INCOMPLETE INFORMATION:

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

- *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



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

  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) =* Π*(A)*

What can disjunctive random sets model ?

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

Canonical examples

- **Objectivist** : Frequentist modelling of a collection of incomplete observations (imprecise statistics) : incomplete generic information
- **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.

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]; Prob(Knowing "$x \in E$ =[60, 70]") = 1 − p.

- 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 generic belief function: imprecise observations in an opinion poll

- **Question** : who is your preferred candidate
  in $C = \{a, b, c, d, e, f\}$???
  - **To a population** $\Omega = \{1, \ldots, i, \ldots, n\}$ of n persons.
  - **Imprecise responses r =** « $x(i) \in E_i$ » **are allowed**
  - No opinion ($r = C$) ; « left wing » $r = \{a, b, c\}$;
  - « right wing » $r = \{d, e, f\}$;
  - a moderate candidate : $r = \{c, d\}$
- **Definition of mass function**:
  - $m(E) = \text{card}(\{i, E_i = E\})_{/}n$
  - = 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
- Bel({x}) gathers responses of those who claim they will vote for x and no one else.

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.

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).

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).

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*.
- They satisfy super-additivity: if $A \cap B = \emptyset$ then

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

- One may require the 2-monotony property:

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

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

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

Cr is then called a convex capacity.

Random disjunctive sets vs. imprecise probabilities

- The set $\mathcal{P}_{bel} = \{P \geq Bel\}$ is coherent: Bel is a special case of lower probability
- Bel is ∞-monotone (super-additive at any order)
  - Order 3: $Bel(A \cup B \cup C) \geq Bel(A) + Bel(B) + Bel(C) - Bel(A \cap B) - Bel(A \cap C) - Bel(B \cap C) + Bel(A \cap B \cap C)$, etc.
- For any set function, the solution m to the set of equations $\forall A \subseteq X$ $g(A) = \sum_{E_i \subseteq A,\, E_i \neq \emptyset} m(E_i)$

  is unique (Moebius transform)
  - ***However m is positive iff g is a belief function***

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):

  $\{x\} \subseteq B$ 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 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), θ ∈ [0, 1]

It yields θ = k/n.

Coherence and deductive closure

- Suppose the knowledge is of the form of a consistent set of assertions ϕ_i of the form « x in E_i » i = 1, …,n. ($N(E_i) = 1$)
- The set of consequences of $\{\phi_i\ i = 1, \ldots, n\}$ is deductively closed (under inclusion and conjunction)
- It defines a Boolean necessity function N corresponding to all assertions « x in A » where $E = \cap_{i = 1, \ldots, n} E_i \subseteq A$ (iff $N(A) = 1$)

Coherence and deductive closure

- If the knowledge is viewed as a credal set $\{P: P(E_i) = 1, i = 1, \ldots, n\}$ then the coherent lower probability induced by its natural extension is a Boolean necessity function N
- **Conclusion** Coherence generalizes deductive closure, interpreting a consequence as a formula with lower probability 1

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*

Practical representations

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

Some are special random sets some not.

From confidence sets to possibility distributions

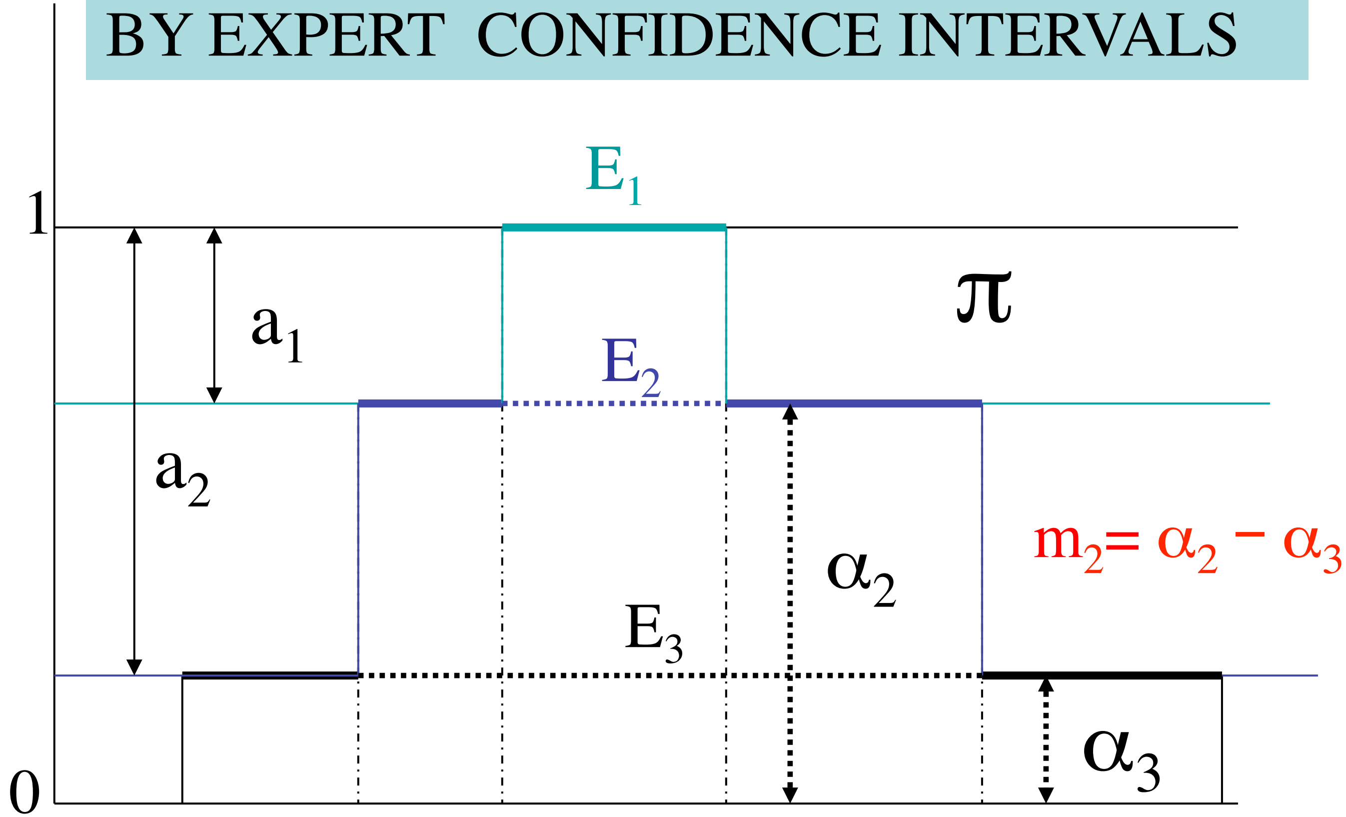
- Let $E_1, E_2, \ldots E_n$ be a nested family of sets
- A set of confidence levels $a_1, a_2, \ldots a_n$ in [0, 1]
- Consider the set of probabilities

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

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

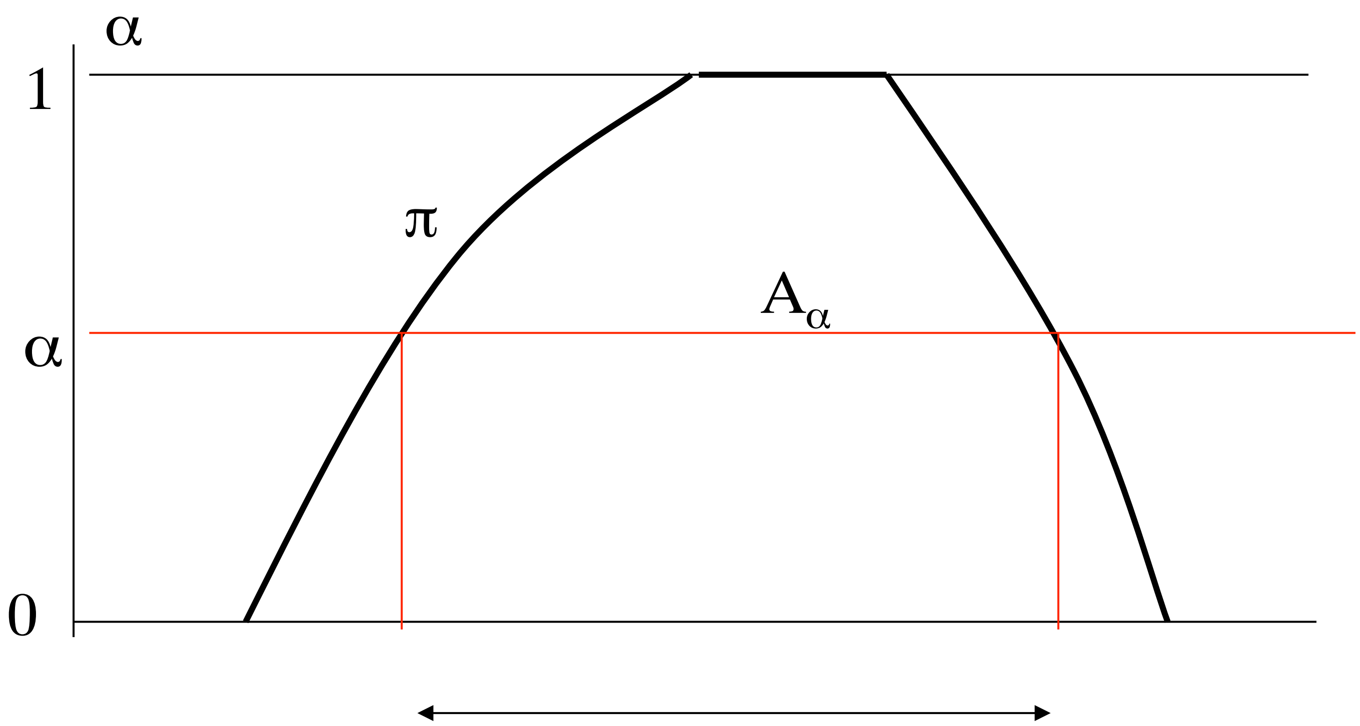
$$\pi(x) = \min_{i = 1, \ldots n} \max (\mu_{Ei}(x), 1 - a_i)$$

POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS



A possibility distribution can be obtained from any family of nested confidence sets :

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



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

Possibilistic view of probabilistic inequalities

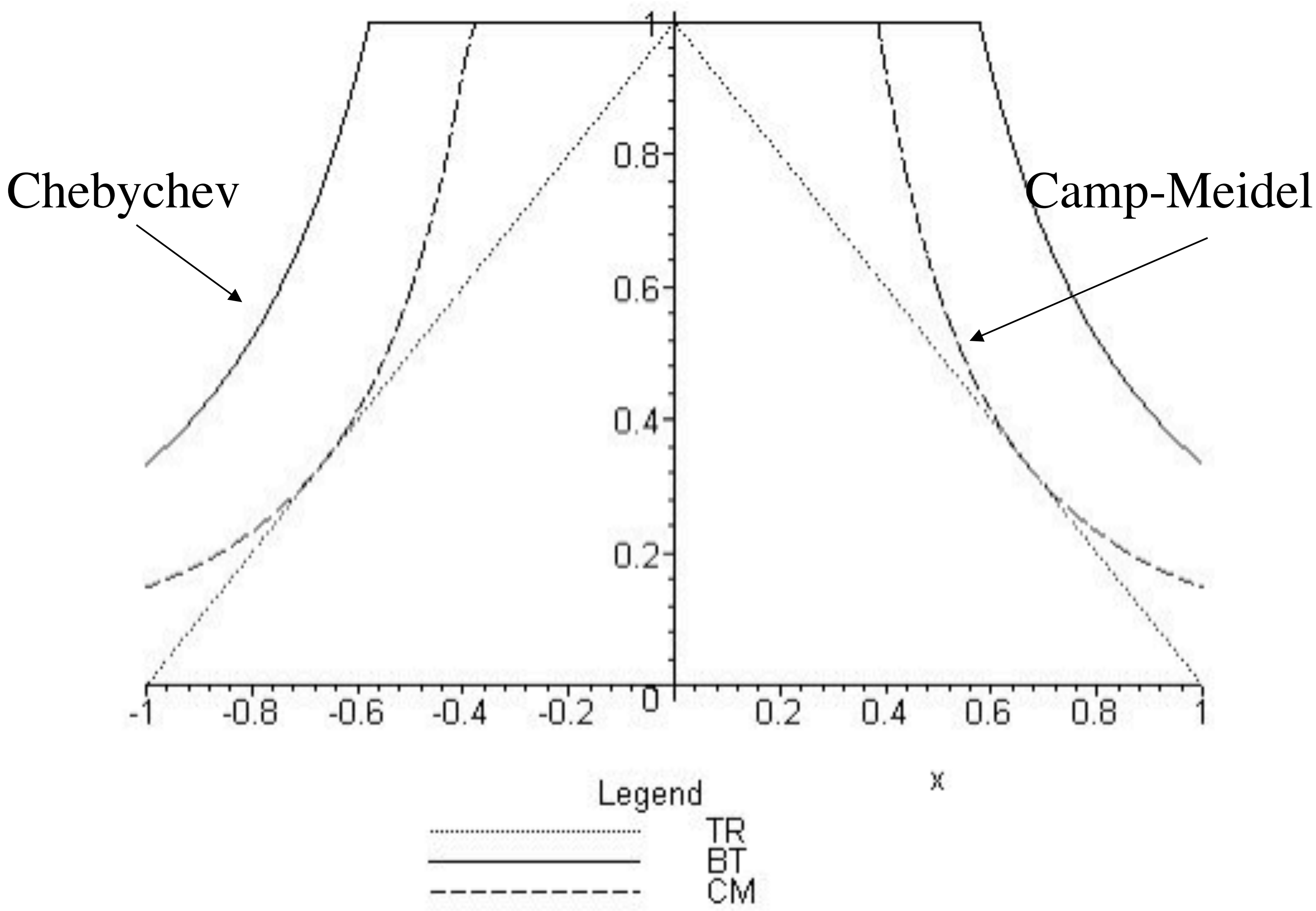
Probabilistic inequalities can be used for knowledge representation:

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

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

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

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


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 and sets E_α of the form

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

- A triangular fuzzy number (TFN) defines a possibility distribution that dominates *any* unimodal density with the same mode and bounded support as the TFN.
- *$P(V \in E_\alpha) \geq 1 - \alpha$ is equivalent to writing*

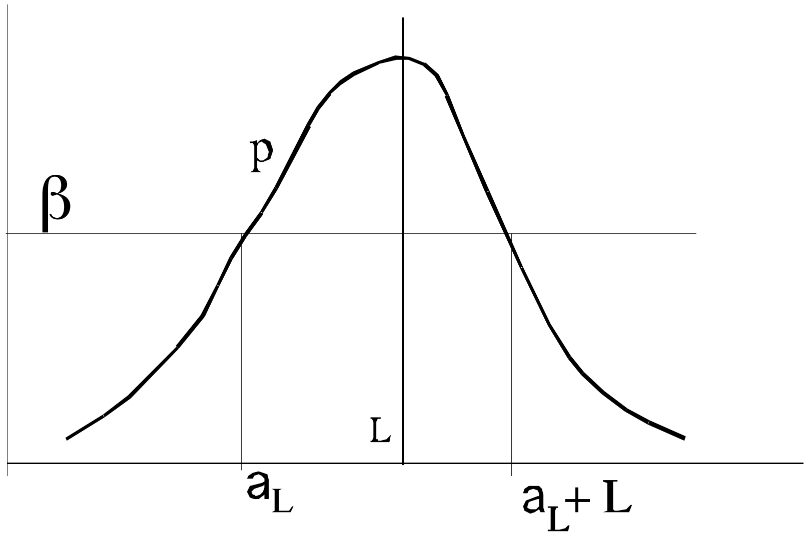
  $$\pi(x^{mode} - (1-\alpha)(x^{mode} - x_*))$$

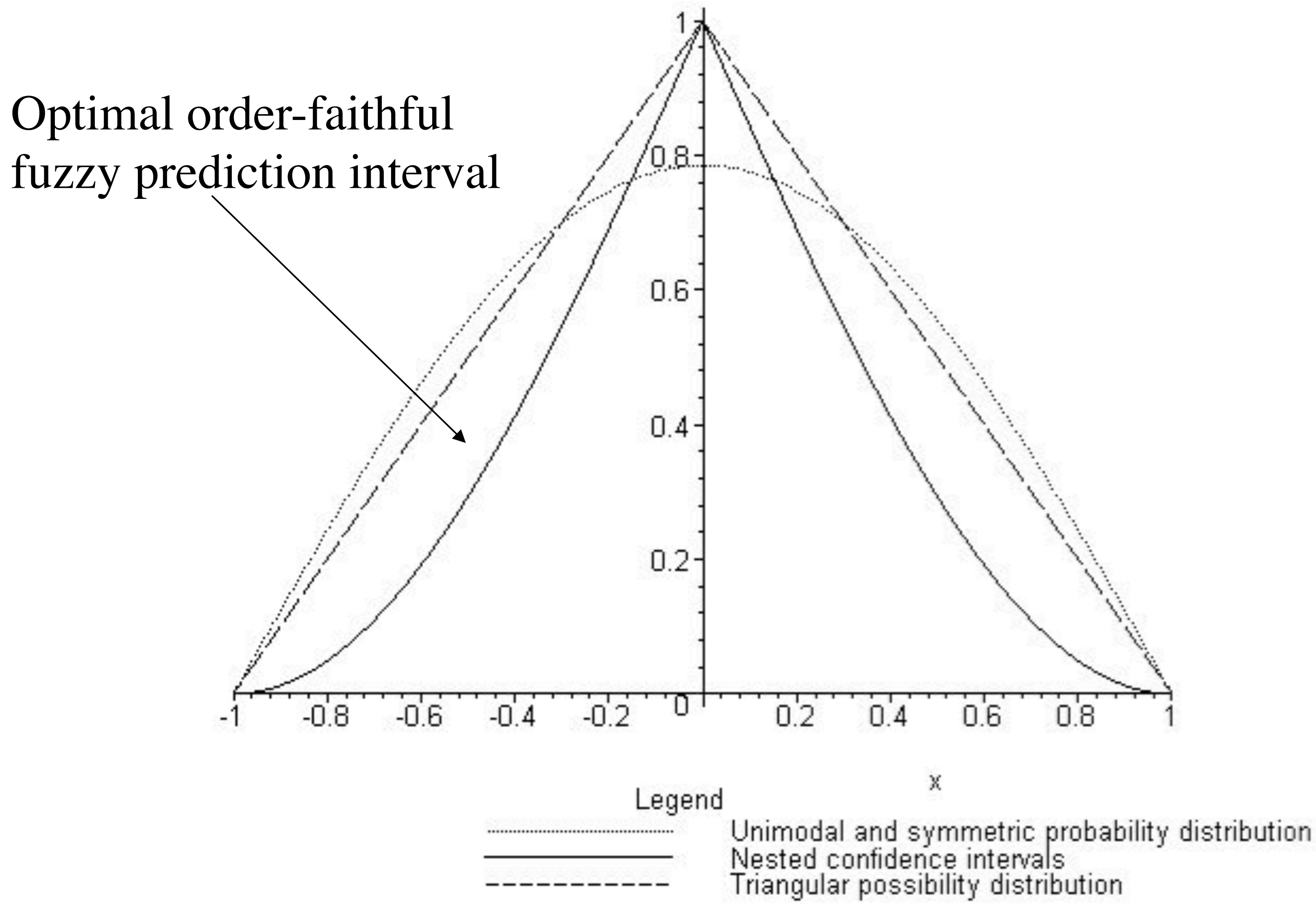
  $$= \pi(x^{mode} + (1-\alpha)(x^* - x^{mode})) = \alpha$$

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 possibilistic counterpart of p is when

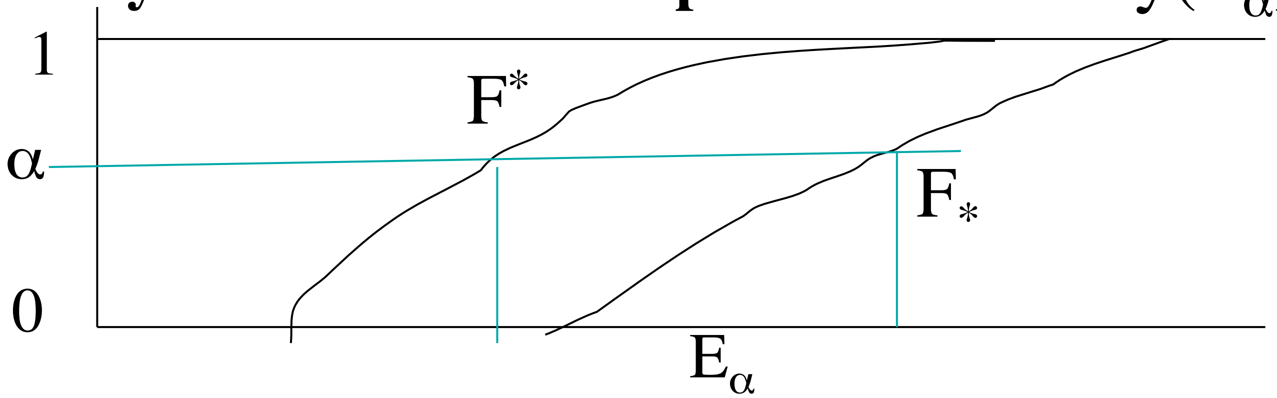
  $\pi^*(a_L) = \pi^*(a_L + L) =$
  $1 - P(I_L = \{x, p(x) \geq \beta\})$.




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

Probability boxes

- A set $\mathcal{P} = \{P: F^* \geq P \geq F_*\}$ induced by two cumulative distribution functions is called a **probability box (p-box)**,
- **A p-box is a special random interval (continuous belief function) whose upper and bounds induce the same ordering.**
- **A fuzzy interval induces a p-box** $\mathcal{P}$ **:** density(E_α) = 1



Probability boxes from possibility distributions

- *Representing families of probabilities by fuzzy intervals is more precise than with the corresponding pairs of PDFs:*
  - $F^*(a) = \Pi_M((-_\infty, a]) = \pi(a)$ if $a \le m$

    $= 1$ otherwise.
  - $F_*(a) = N_M((-_\infty, a]) = 0$ if $a < m^*$

    $= 1 - \lim_{x \downarrow a} \pi(x)$ otherwise
- $\mathcal{P}(\pi)$ is a proper subset of $\mathcal{P} = \{P: F^* \ge P \ge F_*\}$
  - Not all P in $\mathcal{P}$ are such that $\Pi \ge P$

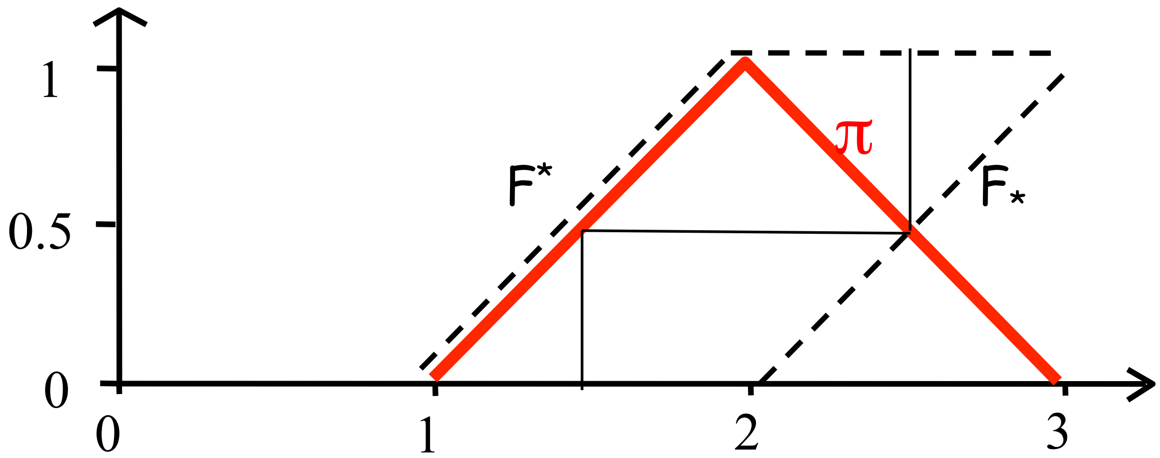
P-boxes vs. fuzzy intervals

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

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



Generalized 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, $\sup\{F(x), x \in A\} \geq P(A)$
- Choosing any order relation $\leq_R$

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

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*}) associated to the same order relation.

$$\mathcal{P} = \{P: F_R^* \geq P \geq F_{R*}\}$$

- Consider $y \leq_R x$ iff $|y - a| \geq |x - a|$ (distance to a value)
- Then $\pi(y) = F_R^*(y) \geq \delta(y) = F_{R*}(y)$
- It comes down to considering nested confidence intervals $E_1, E_2, \ldots E_n$ each with two probability bounds α_i and β_i such that

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

Generalized p-boxes

- It comes down to two possibility distributions

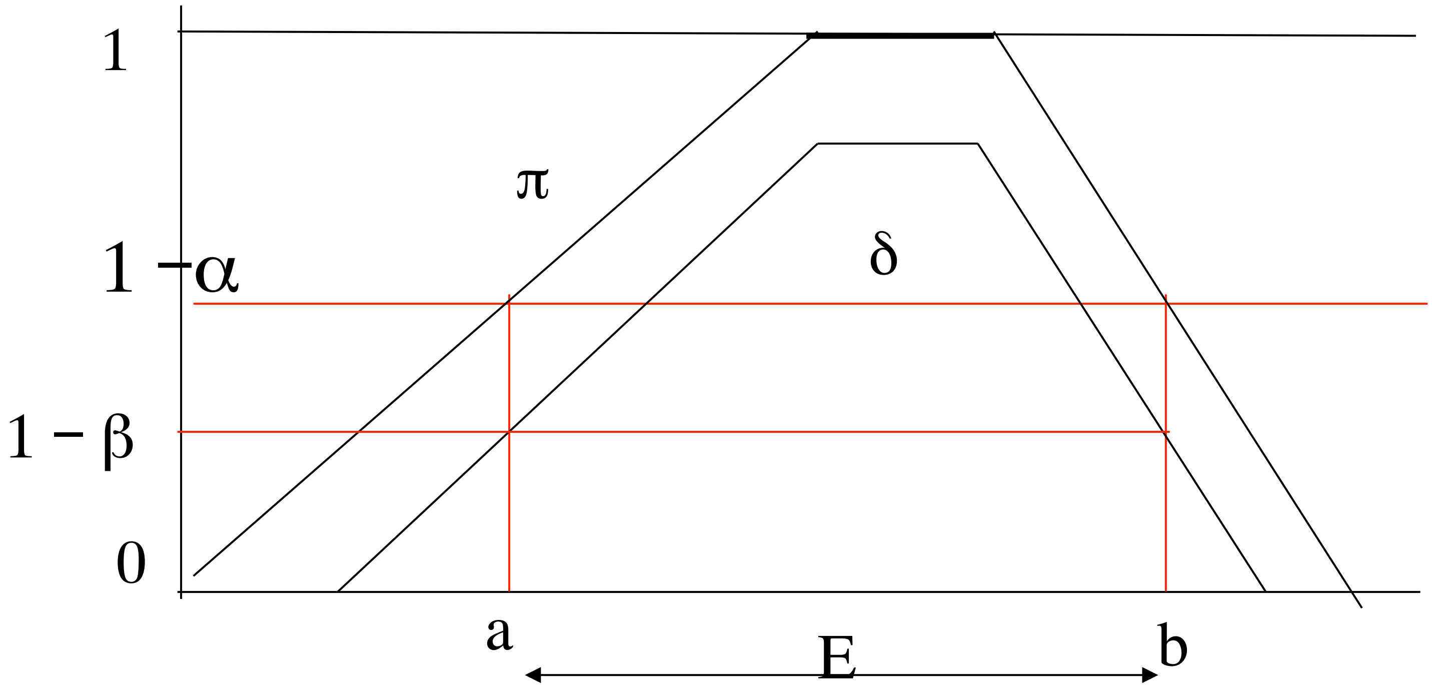
  π (from $\alpha_i \leq P(E_i)$) and π_c (from $P(E_i) \leq \beta_i$)
- Distributions π and π_c are such that $\pi \geq 1 - \pi_c = \delta$ and **π is comonotonic with δ** (they induce the same order on the referential according to $\leq_R$).
- $\mathcal{P} = \mathcal{P}(\pi) \cap \mathcal{P}(\pi_c)$
- **Theorem**: a generalized p-box is a belief function (random set) with focal sets

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

Elementary example of a generalized p-box

- All that is known is that P(E) in [a, b] on a finite set S
- It corresponds to the belief function :
  - $m(E) = a$; $m(E^c) = 1 - b$; $m(S) = b - a$.
- The two possibility distributions :
  - $\pi_1(s) = 1$ if s in E; $1-a$ otherwise.
  - $\pi_2(s) = 1$ if s in E^c; b otherwise.
- The generalized p-box $(\pi_1, 1-\pi_2)$

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



Generalized p-box

From generalized p-boxes to clouds

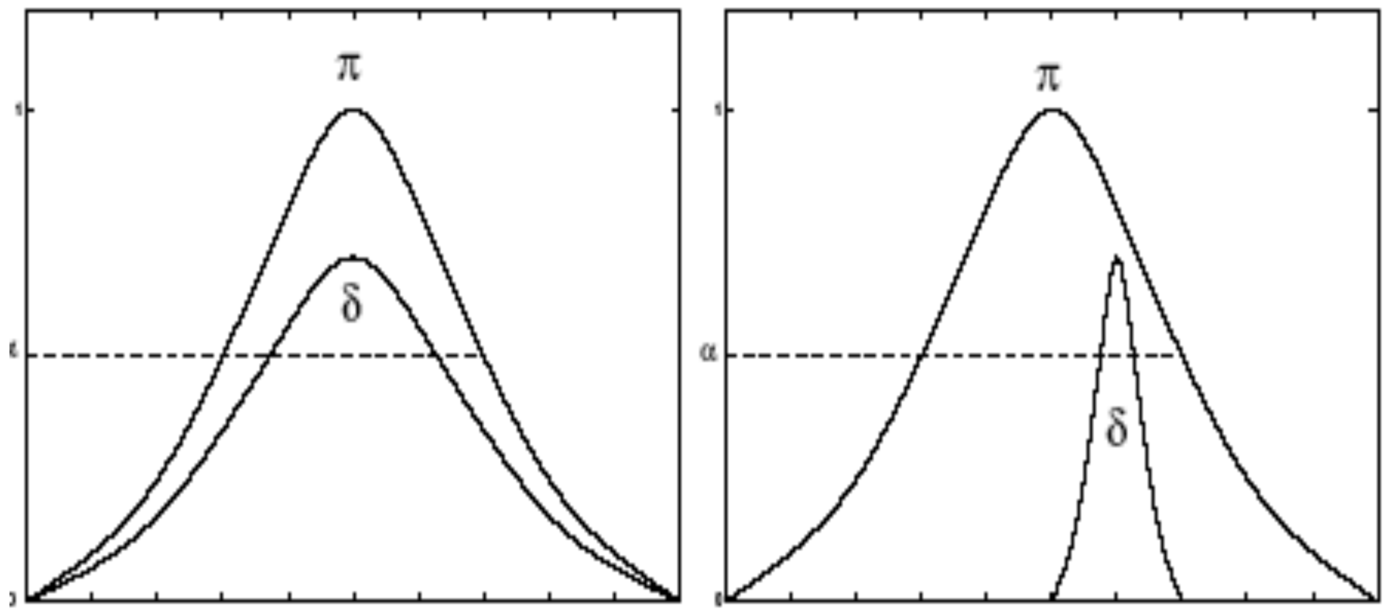


Fig 1.A Comonotonic cloud Fig 1.B Non-comonotonic cloud

CLOUDS

- Neumaier (2004) proposed a generalized interval as a pair of distributions ($\pi \geq \delta$) on a referential representing the family of probabilities $\mathcal{P}$ =
  $\{P, \text{s. t. } P(\{x: \delta(x) > \alpha\}) \leq \alpha \leq P(\{x: \pi(x) \geq \alpha\})\ \forall \alpha > 0\}$
- Distributions π and $1-\delta$ are possibility distributions such that $\mathcal{P} = \mathcal{P}(\pi) \cap \mathcal{P}(1-\delta)$
- It does not 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 π
  - Increasing π: only one probability measure p (π = cumulative distribution of p)

Probability intervals

- 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 $\mathcal{P}_L = \{P: l_i \leq P(\{s_i\}) \leq u_i\}$.
- 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)

How useful are these representations:

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

Relationships between representations

- Generalized p-boxes are special random sets that generalize BOTH p-boxes and possibility distributions
- Clouds extend GP-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.
- **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

- Using contour functions:

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

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

- Using belief or plausibility functions :

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

iff $Bel_1 \geq Bel_2$

It corresponds to comparing credal sets $\mathsf{P}(m)$:

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

Specialisation

- m_1 is more specialised than m_2 if and only if
  - 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

- $m_1 \sqsubseteq_s m_2$ implies $m_1 \sqsubseteq_{Pl} m_2$ implies $m_1 \sqsubseteq_{cf} m_2$
- Typical information ordering for belief functions : $m_1 \sqsubseteq_s m_2$ iff $Q_1 \leq Q_2$
- $m_1 \sqsubseteq_s m_2$ implies $m_1 \sqsubseteq_Q m_2$ implies $m_1 \sqsubseteq_{cf} m_2$
- However $m_1 \sqsubseteq_{Pl} m_2$ and $m_1 \sqsubseteq_Q m_2$ are not comparable and can contradict each other
- In the consonant case : all orderings collapse to $m_1 \sqsubseteq_{cf} m_2$

Example

- $S = \{a, b, c\}$; $m_1(ab) = 0.5$, $m_1(bc) = 0.5$;
- $m_2(abc) = 0.5$, $m_2(b) = 0.5$
- $m_1 \subseteq_s m_2$ nor $m_2 \subseteq_s m_1$ hold
- $m_2 \subset_{Pl} m_1 : Pl_1(A) = Pl_2(A)$
  but $Pl_2(ac) = 0.5 < Pl_1(ac) = 1$
- $m_1 \subset_Q m_2 : Q_1(A) = Q_2(A)$
  but $Q_1(ac) = 0 < Q_2(ac) = 0.5$
- And contour functions are equal : a/0.5, b/1, c/0.5

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.)

- <u>Truth-table of « A|C » viewed as a connective</u>
  - 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 X of $A \cap C = X \cap C$ form the set

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

- 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.

- Semantic entailment: $A|C \models B|D$ 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.

- Quasi-conjunction (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 $A \mid C \in 2^S \times (2^S - \{\emptyset\})$ 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 : $f(x, y) = x/(1+x-y)$

Using conditional probability

- **Prediction** :Querying a generic probability based on sure singular information:
  - P represents generic information (statistics over a population),
  - C represents singular evidence (variable instantiation for a case x at hand)
  - The relative frequency $P(B|C)$ is used as the degree of belief that $x \in C$ satisfies B.

Using conditional probability

- **Revision** of a subjective probability
  - P(A) represents singular information, an agent's prior belief on what is the current state of the world (that a birth date $x \in A$…).
  - C represents an additional sure information about the value of x : $x \in C$ for sure.
  - P(A|C) represents the agent's posterior belief that $x \in A$.

Conditioning a credal set

- *Let $\mathcal{P}$ be a credal set representing generic information and C an event*
- *The two types of tasks lead to different processing :*

1. Prediction *: C represents available singular facts:*
compute the degree of belief in A in context C as
$\mathrm{Cr}(A \mid C) = \mathrm{Inf}\{P(A \mid C), P \in \mathcal{P}, P(C) > 0\}$ (Walley).

2. Revision *: C represents a set of universal truths;*
Add P(C) = 1 to the set of conditionals $\mathcal{P}$.
$\mathrm{Cr}(A||C) = \mathrm{Inf}\{P(A)\ P \in \mathcal{P}, P(C) = 1\}$
If P(C) = 1 is incompatible with $\mathcal{P}$, use maximum likelihood (Gilboa and Schmeidler):
$\mathrm{Cr}(A||C) = \mathrm{Inf}\{P(A|C)\ P \in \mathcal{P}, P(C) \text{ maximal}\}$

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*
- ***Prediction*** by querying on context A ***:*** *Find the most narrow interval for P(C|A) (Linear programming):*

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

  - *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, Prediction does not.**

CONDITIONING RANDOM SETS AS IMPRECISE PROBABILISTIC INFORMATION

- A disjunctive random set ($\mathcal{F}$, m) representing background knowledge is equivalent to a special set of probabilities

  $\mathcal{P} = \{P: \forall A, P(A) \geq Bel(A)\}$.

- Querying this information based on evidence C comes down to performing a sensitivity analysis on the conditional probability $P(\cdot|C)$
  - $Bel_C(A) = \inf \{P(A|C): P \in \mathcal{P}, P(C) > 0\}$
  - $Pl_C(A) = \sup \{P(A|C): P \in \mathcal{P}, P(C) > 0\}$

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

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

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

where $Bel_C(A) = 1 - Pl_C(A^c)$

- *We can do it by focusing generic knowledge (the mass function) on the part of the population that satisfies C.*
- Can be done by transferring portions α_E of m(E) inside the conditioning event C:
  - If $E \subseteq C$ then $\alpha_E = 1$
  - If $E \subseteq C^c$ then $\alpha_E = 0$
  - If $E \cap C \neq \emptyset$ and $E \cap C^c \neq \emptyset$, it is not clear how much mass must be transferred to $E \cap C$.

Prediction conditioning for belief functions

- If the coefficients α_E are known for all focal sets, one can construct a conditional mass function $m_\alpha(\cdot|C)$ on C by computing

  $$m_\alpha(B) = \sum\{\alpha_E m(E): C \cap E = B\}$$

and renormalizing if $Pl_\alpha(C) < 1$

$$m_\alpha(B|C) = m_\alpha(B)/Pl_\alpha(C)$$

- Finally we compute upper and lower bounds
  - the lower belief $\inf_\alpha Bel_\alpha(A \mid C) = Bel_C(A)$
  - the upper plausibility $\sup_\alpha Pl_\alpha(A \mid C) = Pl_C(A)$.
- **We retrieve the imprecise probability conditioning**

Prediction conditioning does not enrich generic 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* $m(\{b\}) = \alpha$, $m(\{c\}) = 1 - \alpha$ (revision conditioning, see further on)

Ellsberg urn

- A bag of balls contains 1/3 red balls, the rest being black or white.
- $S = \{w, b, r\}$ and frequentist mass function : $m(r) = 1/3$, $m(\{w,b\}) = 2/3$
- **Prediction problem** : guess the colour of a ball x picked at random in the urn, knowing x is not black ($C = \{r,w\}$).

Ellsberg urn

- Before knowing anything about x, Bel(r) = Pl(r) = 1/3; Bel(w) = 0; Pl(w) = 2/3.
- After knowing it is not black :
  - $Bel_C(r) = Bel(r)/(Bel(r) + Pl(w)) = 1/3$
  - $Pl_C(r) = Pl(r)/(P(r) + Bel(w)) = 1$
  - $Bel_C(w) = Bel(w)/(Bel(r) + Pl(w)) = 0$
  - $Pl_C(w) = Pl(w)/(Bel(r) + Pl(w)) = 2/3$
- So the piece of information the *ball is not black* does not alter our beliefs about x being white or not.
- But the plausibility of the *ball being red* strongly increases. This is a loss of information.

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 $Pl(\cdot|C)$ is

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

- 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: Then 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 Gilboa and Schmeidler maximum likelihood conditioning 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

Ellsberg urn

- A bag of balls contains 1/3 red balls, the rest being black or white.
- S = {w, b, r} and frequentist mass function : m(r) = 1/3, m({w,b}) = 2/3
- **Revision problem** : guess the colour of a ball x picked at random in the urn, hearing there is no black ball in the urn (C = {r,w}).
- Then P(r) = 1/3 and P(w) = 2/3 :more information is obtained.

Decision with imprecise probability techniques

- Accept incomparability when comparing imprecise utility evaluations of decisions.
  - Pareto optimality : decisions that dominate other choices for all probability functions
  - E-admissibility : decisions that dominate other choices for at least one probability function (Walley, etc…)
- Select a single utility value that achieves a compromise between pessimistic and optimistic attitudes.
  - Select a single probability measure (Shapley value = pignistic transformation) and use expected utility (SMETS)
  - Compare lower expectations of decisions (Gilboa)
  - Generalize Hurwicz criterion to focal sets with degree of optimism (Jaffray)

Information fusion

- Dempster rule of combination in evidence theory:
  - independent sources, normalised or not
  - Does nor preserve consonance of inputs
  - No well-accepted idempotent fusion rule.
- In possibility theory : many fusion rules.
  - The minimum rule : idempotent (= minimal commitment fusion rule for consonant belief functions, not for other ones)
  - The product rule : coincides with the contour function obtained from unnormalized Dempster rule applied to consonant belief functions

Conclusion

- *There exists a coherent range of set-functions combining interval and probability for the representation of uncertainty* .
  - Imprecise probability is the proper theoretical umbrella
  - The choice between set-functions depends on how expressive it is necessary to be in a given application.
  - There exists simple practical representations of imprecise probability

Language difficulties

- *Imprecise probability, belief functions and possibility theory are not fully mutually consisten:*
  - *How to translate conditioning and fusion rules, as well as independence notions from specialised setting to imprecise probability and back.*
  - *Concepts that make sense for credal sets, may be hard to interpret in terms of Moebius transforms or possibility distributions and conversely*
  - *Can simplified representation help us cut down computation costs*
- *How to get this general non-dogmatic approach to uncertainty accepted by traditional statisticians?*

Main problems to be addressed by uncertainty theories

- **Inference:** constructing imprecise probability model from data :
  - Scarce data: Imprecise Dirichlet model (Bernard)
  - Statistics with imprecise (interval) data
- **Elicitation** of upper/ lower probabilities from experts (faithful representation of incomplete information by generalized p-boxes)
- **Uncertainty propagation** : blending interval and Monte-Carlo methods.
- Extraction of **relevant summaries** of information from computation outputs: p-boxes, possibility distribution, indices of information…
- **Prediction**: constructing beliefs from imprecise probability models on the basis of additional evidence
- **Revision** of imprecise probability models
- **Fusion** of uncertain information that account for dependent sources