4. IMPRECISE PROBABILITY AND UNCERTAIN EVIDENCE

1. General framework and inference with imprecise probability
2. Random sets and Belief functions
3. Merging uncertain information

A GENERAL SETTING FOR REPRESENTING GRADED CERTAINTY AND PLAUSIBILITY

- 2 adjoint set-functions Pl and Cr.

- **Conventions** :
  - $Pl(A) = 0$ "impossible" ;
  - $Cr(A) = 1$ "certain"
  - $Pl(A) = 1$; $Cr(A) = 0$ "ignorance" **(no information)**
  - $Cr(A) \leq Pl(A)$ "certain implies plausible"
  - $Pl(A) = 1 - Cr(A^c)$ duality certain/plausible

Imprecise probability theory

- A state of information is represented by a family $\mathcal{P}$ of probability distributions over a set X.
- To each event A is attached a probability interval $[P_*(A), P^*(A)]$ such that
  - $P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$
  - $P^*(A) = \sup\{P(A), P \in \mathcal{P}\} = 1 - P_*(A^c)$

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

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

Imprecise probability theory

- The most general numerical approach to uncertainty.

  – **Attention !** If Cr(A), Pl(A) derive from a family $\mathcal{P}$

  $$\mathcal{P} \neq \{ P, P(A) \in [Cr(A), Pl(A)] \text{ for all } A\} \text{ ;}$$

  – Only $\mathcal{P} \subset \{ P, P(A) \in [Cr(A), Pl(A)] \text{ for all } A\}$ holds

- Equality if $Cr(A) + Cr(B) \leq Cr(A \cup B) + Cr(A \cap B)$ for all A, B (super-additivity of convex capacities)

PROBABILISTIC LOGIC (de Finetti)

- A probabilistic knowledge base is a set of weighted propositions $\mathcal{B} = \{(p_i\ \alpha_i), i = 1, \ldots, \alpha_I \in [0, 1]\}$
- $(p_i\ \alpha_i)$ means that $P(A_i) = \alpha_i$ where $A_i = [p_i]$
  = *A constraint on an unknown probability distribution on interpretations*.
- $\mathcal{B}$ is not enough to isolate a single probability distribution : it defines a probability family on interpretations
  - $\mathcal{P} = \{P: P(A_i) = \alpha_i , i = 1, \ldots\}$
- ***Inference problem:*** *What can be said about the probability another proposition p ???*
- *All that is known is* $P(p) \in [\alpha, \beta]$, generally.

Probabilistic extension of propositional logic

- *Finding the most narrow interval* $[\alpha, \beta]$ *such that* $\mathcal{B} \models P(p) \in [\alpha, \beta]$ *is a* ***non-straightforward*** *extension of classical inference.*
  - *More generally* $\mathcal{B}$ *may contain information of the form* $P(A_i) \in [\alpha_i, \beta_i]$
  - *It cannot distinguish between contingent evidence and generic knowledge*
  - *It cannot model conditional probability*
  - *Information pieces of the form* $(A_i\ \alpha_i)$ and $(A_i \Rightarrow B_i\ \beta_i)$, with material implication $\Rightarrow$ are not independent

CONDITIONAL PROBABILISTIC LOGIC

- A conditional probabilistic knowledge base is a weighted set of rules $\Delta = \{(p_i \rightarrow q_i\ \alpha_i), i = 1, \ldots\}$ where α_i lies in [0, 1].
  - **It represents Generic knowledge**
  - $(p_i \rightarrow q_i\ \alpha_i), \in \Delta$ means that $P(q_i | p_i) = \alpha_i$
  - *A probability family* $\mathcal{P}_\Delta = \{P, P(q_i | p_i) = \alpha_i , i = 1, \ldots\}$ is the semantic counterpart of Δ.
- *Mathematically it is more complex than Bayesian nets but it has the power of probability theory and leaves room to incompleteness.*

Inference with a probabilistic conditional base

- *Suppose a set of probabilistic conditionals Δ and $\mathcal{C}$ a propositional base. Let $C = [\mathcal{C}]$ be the set of models.*
- *Two types of processing :*

1. ***Querying*** *: $\mathcal{C}$ is a set of singular facts :compute* the degree of belief of A in context C as

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

2. ***Revision*** *: $\mathcal{C}$ is a set of universal truths;*

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

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

   If $P(C) = 1$ is incompatible with $\mathcal{P}_\Delta$, consider

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

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*

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

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

Imprecise probability inference extend preferential (cautious) inference

- *Infinitesimal probabilisric inference*: From set of rules Δ of the form $P(q_i | p_i) \geq 1 - \varepsilon$, prove that $P(q | p) \geq 1 - O(\varepsilon)$ from Δ where ε is an infinitesimal number.
- This is equivalent to proving $\Delta^* \models_{\forall} p \rightarrow q$ using cautious possibilistic inference (or Lehman's system P) where $\Delta^* = \{p_i \rightarrow q_i\}$ is the set of conditionals present in Δ (Adams' logic of conditionals, 1975).
- It is equivalent to the logic of conditional events
- It is equivalent to inference from set of rules Δ of the form $P(q_i | p_i) = 1$ under De Finetti's coherence approach (T. Lukasiewicz, A. Gilio).

Random sets and evidence theory

- A family $\mathcal{F}$ of « focal » (disjunctive) non-empty sets representing
  - Statistics under incomplete observations (generic)
  - Unreliable testimonies (singular)
  - Indirect information (induced from a probability space)
- A positive weighting of focal sets (a random set) :

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

- It is a randomized incomplete information

Theory of evidence

- 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)**
- In the view of Shafer (1976) and Smets the mass assignment *m* represents uncertain singular evidence on the solution of a problem

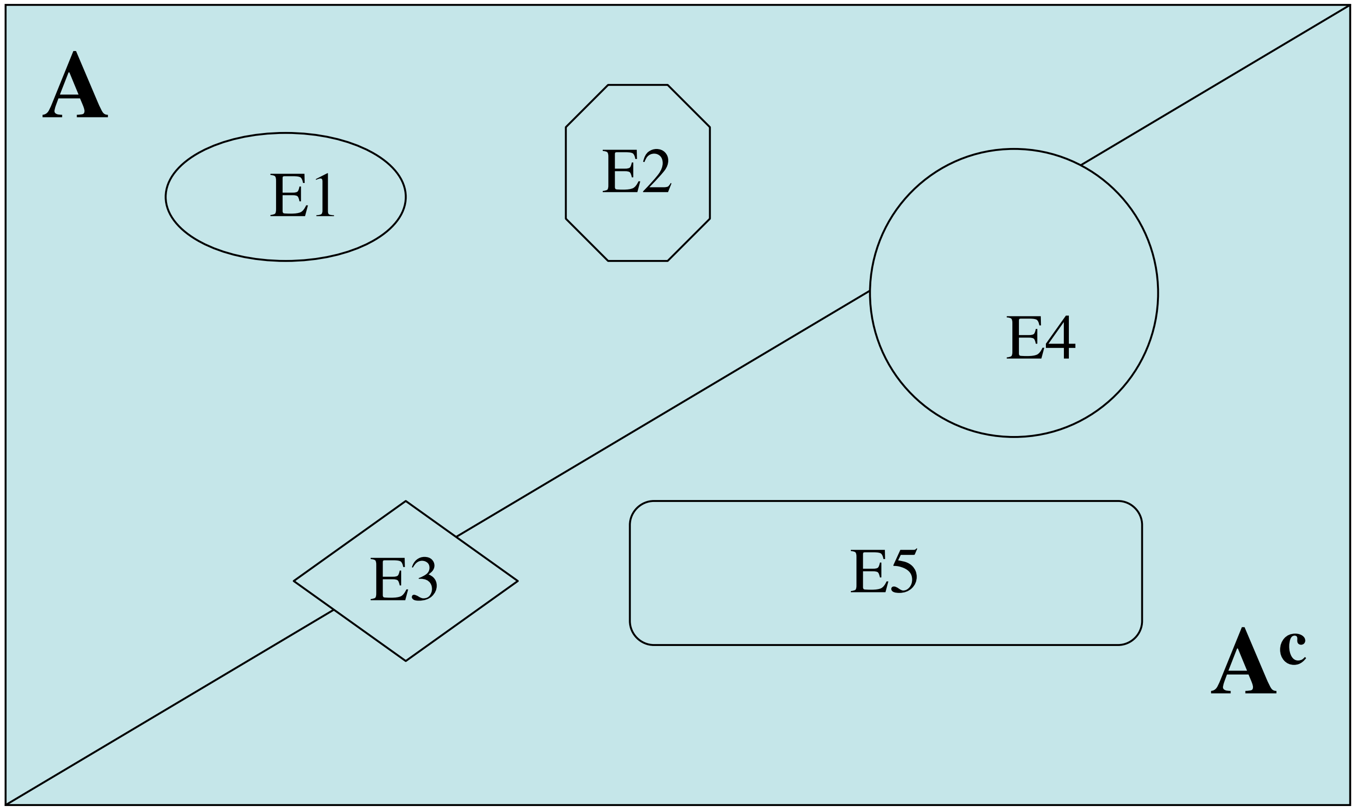
Theory of evidence

- **degree of certainty** (belief) :
    - (Function Cr denoted Bel by Shafer)
  - $\mathrm{Bel}(A) = \sum_{E_i \subseteq A,\, E_i \neq \emptyset} m(E_i)$
  - 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 (ZADEH)

  $E_1 \subseteq E_2 \subseteq E_3 \ldots \subseteq E_n$: 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

Plausible states induced by a belief function

- A random set $(\mathcal{F}, m)$ expressing available information on x induces a fuzzy set F of plausible values of x, summing for each value s, masses of all focal sets containing s.
  - $\mu_F(s) = Pl(\{s\}) = \sum \{m(E_i), E_i \in \mathcal{F}, s \in E_i\}$
  - $\exists$ s, $\mu_F(s) = 1$ iff $s \in E_i$ for all i
  - $\cap_i E_i \neq \emptyset$ (no conflict in information).
- The mass function cannot be reconstructed from function μ_F, except if
  - All E_i are disjoint : $m(E_i) = \mu_F(s)$ if $s \in E_i$
  - E_i are all nested : $Pl(\{s\})$ is a possibility distribution

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* : $m(E) = 1 - p$ and $m(S) = p$
- Equivalent to a possibility distribution
  - $\pi(s) = 1$ if $x \in E$ and $\pi(s) = p$ otherwise.

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(.|E) is

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

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

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))

THE IMPRECISE PROBABILITY VIEW (Dempster, 1967)

- A belief function on S is induced by a probability space (Ω, P) via a point to set-mapping G: $m(E_i) = p(w_i)$ if $G(w_i) = E_i$.
- Consider a selection function $\phi: \Omega \to S$ from G. For each focal set E_i assign mass $m(E_i)$ to element $\phi(w_i) \in E_i$: We get a probability P^ϕ such that

$$P^\phi(\{s\}) = \sum \{p(w_i) , \phi(w_i) = s\}$$

  - $\forall A$, $Bel(A) \leq P^\phi(A)$ and (equivalently) $P^\phi(A) \leq Pl(A)$
  - $\mathcal{P} = \{P, \forall A, Bel(A) \leq P(A)\}$ = convex hull of probabilities P^ϕ.
  - $Cr = P_*$: lower probability; $Pl = P^*$: upper probability
- $P \in \mathcal{P}$ is of the form : $P(A) = \sum P_i(A| E_i) \cdot m(E_i)$ (where P_i is any probability measure on support E_i)

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 ∞-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**

Indirect information: The unreliable watch

- **States** of the watch: $\Omega = \{OK, KO\}$;
- **Space of interest for the agent** : S = {day hours}
- **Available information**:
  1. *on the watch state*: p = Prob(KO) (small)
  2. *Logical relation between Ω and S*:
     - G(OK) = {Actual hour up to 2 mn} = $H \subseteq S$.
     - G(KO) = S (ignorance) (*broken watches may give the right time*)
- There is a probability p that agent ignores the right time : m(H) = 1 – p and m(S) = p

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**:
  - $m(E) = \mathrm{card}(\{i, E_i = E\})_/ n$
  - = 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.

CONDITIONING IMPRECISE PROBABILISTIC INFORMATION

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

- **Theorem:** functions $Bel_E(A)$ and $Pl_E(A)$ are belief and plausibility functions of the form
- $Bel_E(A) = Bel_E(E \cap A)/(Bel_E(E \cap A) + Pl_E(E \cap A^c))$
- $Pl_E(A) = Pl_E(E \cap A)/(Pl_E(E \cap A) + Bel_E(E \cap A^c))$
- $Bel_E(A) = 1 - Pl_E(A^c)$
- *They are less informative than Dempster conditioning*: If $E \cap C \neq \emptyset$ and $E \cap C^c \neq \emptyset$ for all $C \in \mathcal{F}$, then $m_E(E) = 1$ (total ignorance on E)
  - **Example: If opinion poll yields :** $m(\{a, b\}) = \alpha$, $m(\{c, d\}) = 1 - \alpha$, the proportion of voters for a candidate in $E = \{b, c\}$ is unknown.
  - *However if we hear a and d resign* $(Pl(\{a, d\} = 0)$ *then* $m(\{b\}) = \alpha$, $m(\{c\}) = 1 - \alpha$ (Dempster conditioning)

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)...
- **Order of magnitude of extreme probabilities** (Spohn functions with values on integers)

Quantitative possibility theory

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

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

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

POSSIBILITY AS UPPER PROBABILITY

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

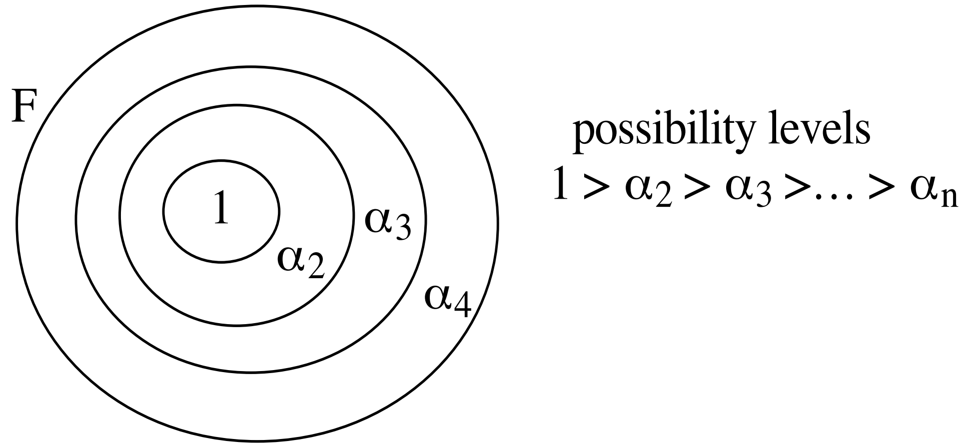
- Then, generally it holds that

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

  $$N(A) = \inf \{P(A) \mid P \in \mathcal{P}(\pi)\}$$

- So π is a faithful representation of a family of probability measures

Random set view



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

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

PLAUSIBIBILITY RANKING

Quantitative Possibility theory

Fuzzy (nested disjunctive) sets

Classical logic

Disjunctive sets

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 : $A \cap B \neq \emptyset$ is allowed

BASIC MERGING MODES

3. Merging by counting:

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

- $A\mathsf{M}B\ (x) = Pl(x) = \sum_{x \in E} m(E) = 1$ 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) :* $A \cap B \subseteq A\mathsf{M}B \subseteq A \cup B$

  – 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 $\mathcal{S}$ 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 number of reliable sources
- *Suppose k reliable sources*
- Then pick k sources at random for conjunctive merging and then disjunctively merge obtained results

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

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

- **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 $\rightarrow$ $\pi_1 = \mu_{F^1}$ source 2 $\rightarrow$ $\pi_2 = \mu_{F^2}$

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 α_i such that $\sum \alpha_i = 1$

$$P = \sum \alpha_i P_i$$

- The only possible one with
  - $P(A) = f(P_1(A),\ldots,P_n(A)) \ \forall A \subseteq S$
  - $f(0, 0\ldots0) = 0$; $f(1, 1\ldots1) = 1$
  - (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 : $m^*(C) = m_\cap(C)/(1 - m_\cap(\emptyset))$

    renormalisation

  $m_\cap(\emptyset)$ 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	$\emptyset$ 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

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

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

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
- $m_\alpha(A) = \alpha \cdot m_1(A) + (1 - \alpha)\, m_2(A)$
  - α = relative reliability of source 1 versus source 2
- **Example** : discounting an unreliable belief function with reliability α 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 applied in various settings encompassed by imprecise probability, but
  - classical logic 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 : ever complete and consistent.
  - Shafer-Smets or possibility theory is useful for merging uncertain evidence(step 2)