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, vs. representations of real sets
3. Blending set-valued and probabilistic representations: uncertainty theories
4. Comparison of uncertainty theories on
   1. Practical representations
   2. Information content
   3. Conditioning
   4. Combination
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 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 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.
Belief in the occurrence of a particular event may derive from its statistical probability: the **Hacking principle**:
- Generic knowledge = probability distribution $P$
- $\text{belief}_{\text{NOW}}(A) = \text{Freq}_{\text{POPULATION}}(A)$: 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
Using a single probability distribution to represent incomplete information is not entirely satisfactory:

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

1. **Ambiguity**: In the absence of information, how can a uniform distribution tell pure randomness and ignorance apart?

2. **Instability**: A uniform prior on $x \in [a, b]$ induces a non-uniform prior on $f(x) \in [f(a), f(b)]$ if $f$ is increasing and non-affine.

3. **Empirical refutation**: When information is missing, decision-makers do not always choose according to a single subjective probability (Ellsberg paradox).
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

$E = \text{Models of a wff } \phi \text{ stated as true.}$

This kind of information is subjective (epistemic set)
Natural set functions under incomplete information:

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

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

  \[ \Pi(A) = 1, \text{ and } 0 \text{ otherwise} \]

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

  \[ N(A) = 1, \text{ and } 0 \text{ otherwise} \]

*This is a simple modal logic (KD45)*

\[ N(A) = 1 - \Pi(A^c) \]

\[ \Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \]

\[ N(A \cap B) = \min(N(A), N(B)) \]
SETS and SETS

• Do not mix up
  – A set-valued variable $X$: the set of languages a person can speak ($A = \{\text{English (and) French}\}$) is a conjunction of values, and a real set. $X = A$ is precise
  – An ill-known point-valued variable $x$: the language this person is speaking now ($E = \{\text{English (exor) French}\}$): it is a disjunction of values, and an epistemic set (a piece of information in the head of an agent). $x \in E$ is imprecise.
Find a representation of uncertainty due to incompleteness

- More informative than sets
- 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 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
A GENERAL SETTING FOR REPRESENTING GRADED CERTAINTY AND PLAUSIBILITY

• 2 conjugate set-functions $\text{Pl}$ and $\text{Cr}$ generalizing probability $P$, possibility $\Pi$, and necessity $N$.

• Conventions:
  – $\text{Pl}(A) = 0$ "impossible" ; $\text{Cr}(A) = 1$ "certain"
  – $\text{Pl}(A) = 1$ ; $\text{Cr}(A) = 0$ "ignorance" (no information)
  – $\text{Pl}(A) - \text{Cr}(A)$ quantifies ignorance about $A$

• Postulates
  – $\text{Cr}$ and $\text{Pl}$ are monotonic under inclusion (= capacities).
  – $\text{Cr}(A) \leq \text{Pl}(A)$ "certain implies plausible"
  – $\text{Pl}(A) = 1 - \text{Cr}(A^c)$ duality certain/plausible
  – If $\text{Pl} = \text{Cr}$ then it is $P$. 
Possibility Theory
(Shackle, 1961, Zadeh, 1978)

• A piece of incomplete information "\(x \in E\)"
  admits of degrees of possibility: \(E\) is a
  (normalized) fuzzy set.

• \(\mu_E(s) = \text{Possibility}(x = s) = \pi_x(s)\)

• Conventions:
  \(\forall s, \pi_x(s)\) is the degree of plausibility of \(x = s\)
  \(\pi_x(s) = 0\) iff \(x = s\) is impossible, totally surprising
  \(\pi_x(s) = 1\) iff \(x = s\) is normal, fully plausible, unsurprising
  (but no certainty)
A pioneer of possibility theory

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

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

• The degree of surprize of an event is the degree of surprize of its least surprizing realization.

• He introduces a notion of conditional possibility
Improving expressivity of incomplete information representations

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

• Nested intervals \( E_1, E_2, \ldots E_n \) with confidence levels \( N(E_i) = a_i \):
  \[ \pi(x) = \min_{i = 1, \ldots n} \max (\mu_{E_i}(x), 1 - a_i) \]
POSSIBILITY AND NECESSITY
OF AN EVENT

How confident are we that $x \in A \subset S$? \textit{(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 $\pi$
  (= some $x \in A$ is possible)
  
  The degree of possibility \textit{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 $\pi$ implies $A$

  The degree of certainty \textit{(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* \( N(A) > 0 \Rightarrow \Pi(A) = 1 \)
Qualitative vs. quantitative possibility theories

- **Qualitative:**
  - **comparative:** A complete pre-ordering $\geq_\pi$ on $U$. A well-ordered partition of $U$: $E_1 > E_2 > \ldots > E_n$
  - **absolute:** $\pi_x(s) \in L = $ finite chain, complete lattice...

- **Quantitative:** $\pi_x(s) \in [0, 1]$, integers...

One must indicate where the numbers come from.

*All theories agree on the fundamental maxitivity axiom*

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

Theories diverge on the conditioning operation.
Imprecise probability theory

• A state of information is represented by a family $\mathcal{P}$ of probability distributions over a set $X$.

• To each event $A$ is attached a probability interval $[P_*(A), P^*(A)]$ such that
  – $P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$
  – $P^*(A) = \sup\{P(A), P \in \mathcal{P}\} = 1 - P_*(A^c)$

• Usually $\mathcal{P}$ is strictly contained in $\{P(A), P \geq P_*\}$

• For instance: incomplete knowledge of an objective probabilistic model: $\exists P \in \mathcal{P}$.
Subjectivist view (Peter Walley)

- \( P_{\text{low}}(A) \) is the highest acceptable price for buying a bet on singular event A winning 1 euro if A occurs
- \( P_{\text{high}}(A) = 1 - P_{\text{low}}(A^c) \) is the least acceptable price for selling this bet.

**Rationality** conditions:
- **No sure loss**: \( \{ P \geq P_{\text{low}} \} \) not empty
- **Coherence**: \( P_{\ast}(A) = \inf\{P(A), P \geq P_{\text{low}}\} = P_{\text{low}}(A) \)

- A theory that handles convex probability sets
- Convex probability sets are actually characterized by lower expectations of real-valued functions (gambles), not just events.
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 or conjunctive ??
Random disjunctive sets

- A focal set is an epistemic state \((x \in E)\)
- \((\mathcal{F}, m)\) is a randomized incomplete information
- \(m(E) = \) probability that the most precise description of the available information is of the form "\(x \in E\)"
  - \(=\) probability (only knowing "\(x \in E\)" and nothing else)
  - It is the portion of probability mass hanging over elements of \(E\) without being allocated.
- **DO NOT MIX UP** \(m(E)\) and \(P(E)\)
Basic set functions from random sets

• **degree of certainty** (belief):
  - \( \text{Bel}(A) = \sum_{E_i \subseteq A, E_i \neq \emptyset} m(E_i) \)
  - total mass of information implying the occurrence of \( A \)
  - (probability of provability)

• **degree of plausibility**:
  - \( \text{Pl}(A) = \sum_{E_i \cap A \neq \emptyset} m(E_i) = 1 - \text{Bel}(A^c) \geq \text{Bel}(A) \)
  - total mass of information consistent with \( A \)
  - (probability of consistency)
PARTICULAR CASES

• INCOMPLETE INFORMATION:
  \[ m(E) = 1, \ m(A) = 0, \ A \neq E \]

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

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

• POSSIBILITY THEORY:
  - the consonant case
  \( E_1 \subseteq E_2 \subseteq E_3 \ldots \subseteq E_n \): imprecise and coherent information
  - iff \( Pl(A \cup B) = \max(Pl(A), Pl(B)) \), possibility measure
  - iff \( Bel(A \cap B) = \min(Bel(A), Bel(B)) \),
    necessity measure = consonant belief function
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) = \( \text{Pl}(\{s\}) \).
- Only in the consonant case can \( m \) be recalculated from \( \pi \)
- \( \text{Bel}(A) = \sum_{F_i \subseteq A} m_i = N(A); \text{Pl}(A) = \Pi(A) \)
Random disjunctive sets vs. imprecise probabilities

- The set $P_{\text{bel}} = \{P \geq \text{Bel}\}$ 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 $\forall A \subseteq X$
  $$g(A) = \sum m(E_i)$$
  $$E_i \subseteq A, E_i \neq \emptyset$$
  is unique (Moebius transform)
  
  - However $m$ is positive iff $g$ is a belief function
Dempster original model

- Indirect information (induced from a probability space).
- What we know about a random variable \( x \) with range \( S \), based on a sample space \( (\Omega, A, P) \) and a multimapping \( \Gamma \) from \( \Omega \) to \( S \) (Dempster):
  - The meaning of the multimapping \( \Gamma \) from \( \Omega \) to \( S \):
    - if we observe \( \omega \) in \( \Omega \) then all we know is \( x(\omega) \in \Gamma(\omega) \)
  - \( m(\Gamma(\omega)) = P(\{\omega\}) \) \( \forall \omega \) in \( \Omega \) (finite case.)
  - Then a belief function is a lower probability.
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 \( \text{Bel} \): no appeal to an underlying probability.

(Shafer, 1976 book; Smets)

  – *Imprecise statistical evidence*: \( m(E) = \) frequency of imprecise observations of the form \( E \) and \( \text{Bel}(E) \) 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
Canonical examples

• **Statistical probabilities**: Frequentist modelling of a collection of incomplete observations (imprecise statistics): incomplete generic information

• **Uncertain singular information**:
  – **Unreliable testimonies** (Shafer’s book): unreliable singular information
  – **Unreliable sensors**: the quality 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 ».
  - $E = [60, 70]$; $\text{Prob}(\text{Knowing} \ x \in E = [60, 70]) = 1 - p$.
  - 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.

- *If John may lie* (probability q): $m(E) = (1 - p)(1 - q)$, $m(E^c) = (1 - p)q$. 
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** \( r = \langle x(i) \in E_i \rangle \) **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**:
  - \( m(E) = \text{card}(\{i, E_i = E\}) \frac{1}{n} \)
  - = Proportion of imprecise responses « \( x(i) \in E \) »
• The probability that a candidate in subset $A \subseteq C$ is elected is imprecise:
  $$\text{Bel}(A) \leq P(A) \leq \text{Pl}(A)$$

• There is a fuzzy set $F$ of potential winners:
  $$\mu_F(x) = \sum_{x \in E} m(E) = \text{Pl}\{\{x\}\} \text{ (contour function)}$$

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

• $\text{Bel}\{\{x\}\}$ 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 $\mathcal{L} = \{\text{Basque, Chinese, Dutch, English, French,} \ldots\}$
- $m(E)$ = % people who speak *exactly* all languages in $E$ (and not other ones)
- $\text{Prob}(x \text{ speaks } A) = \sum \{m(E) : A \subseteq E\} = Q(A)$: commonality function in belief function theory
- **Example**: « x speaks English » means « at least English »
- The belief function is not meaningful here while the commonality makes sense, contrary to the disjunctive set case.
POSSIBILITY AS UPPER PROBABILITY

• Given a numerical possibility distribution $\pi$, define
  $\mathcal{P}(\pi) = \{ P \mid P(A) \leq \Pi(A) \text{ for all } A \}$
  - Then, $\Pi$ and $N$ can be recovered
    - $\Pi(A) = \sup \{ P(A) \mid P \in \mathcal{P}(\pi) \}$;
    - $N(A) = \inf \{ P(A) \mid P \in \mathcal{P}(\pi) \}$
  - So $\pi$ is a faithful representation of a special family of probability measures

• Likewise for belief functions: $\mathcal{P}(\mathcal{m}) = \{ P \mid P(A) \leq Pl(A), \forall A \}$
  - $Pl(A) = \sup \{ P(A) \mid P \in \mathcal{P}(\mathcal{m}) \}$;
  - $Bel(A) = \inf \{ P(A) \mid P \in \mathcal{P}(\mathcal{m}) \}$

• But this view makes sense for belief functions
  - Representing statistical knowledge
  - Representing maximal buying prices as per Walley
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 $\lambda(B) = \max_{x \in B} \lambda(x)$
Maximum likelihood principle is possibility theory

• The classical coin example: $\theta$ is the unknown probability of “heads”
• Within $n$ experiments: $k$ heads, $n-k$ tails
• $P(k\text{ heads, } n-k\text{ tails} \mid \theta) = \theta^k \cdot (1 - \theta)^{n-k}$ is the degree of possibility $\pi(\theta)$ that the probability of “head” is $\theta$.

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

It yields $\theta = k/n$. 
LANDSCAPE OF UNCERTAINTY THEORIES

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

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

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

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

BOOLEAN POSSIBILITY THEORY (modal logic KD45): The language of Disjunctive sets
Practical representation issues

• Lower probabilities are difficult to represent \((2^{|S|} \text{ values})\): 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 or random sets.*
Practical representations

• Fuzzy intervals (continuous consonant bf)
• Probability intervals
• Probability boxes (continuous random intervals with comonotonic bounds)
• 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
  \[ P = \{ P, P(E_i) \geq a_i, \text{ for } i = 1, \ldots n \} \]
- Then $P$ is representable by means of a possibility measure with distribution
  \[ \pi(x) = \min_{i = 1, \ldots n} \max (\mu_{E_i}(x), 1 - a_i) \]
POSSIBILITY DISTRIBUTION INDUCED BY EXPERT CONFIDENCE INTERVALS

$\pi$

$E_1$

$E_2$

$E_3$

$a_1$

$a_2$

$a_3$

$m_2 = \alpha_2 - \alpha_3$

$\alpha_2$

$\alpha_3$
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: $N(\pi_\alpha) = 1 - \alpha$
Possibilistic view of probabilistic inequalities

• Probabilistic inequalities can be used for knowledge representation
  – Chebyshev inequality defines a possibility distribution that dominates any density with given mean and variance:
    \[ P(V \in [x^{\text{mean}} - k\sigma, x^{\text{mean}} + k\sigma]) \geq 1 - 1/k^2 \]
    is equivalent to writing
    \[ \pi(x^{\text{mean}} - k\sigma) = \pi(x^{\text{mean}} + k\sigma) = 1/k^2 \]

  – A triangular fuzzy number (TFN) defines a possibility distribution that dominates any unimodal density with the same mode and bounded support as the TFN.
Optimal order-faithful fuzzy prediction interval
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 with probability $\alpha$ is of the form $\{x, p(x) \geq \beta\}$
- So the most natural possibilistic counterpart of $p$ is when 
  $\pi^*(a_L) = \pi^*(a_L + L) = 1 - P(I_L = \{x, p(x) \geq \beta\})$. 

![Diagram showing a normal distribution with critical values $a_L$ and $a_L + L$. The diagram illustrates the relationship between the probability $p$, the critical value $\beta$, and the interval $I_L$. The area under the curve between the interval $I_L$ is shaded to represent the probability of $p(x) \geq \beta$. The critical values $a_L$ and $a_L + L$ are marked on the x-axis.]
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 $\mathcal{P}$
Probability boxes from possibility distributions

- Representing families of probabilities by fuzzy intervals is more precise than with the corresponding pairs of PDFs:
  - \( F^*(a) = \Pi_M( ( -\infty, a] ) = M(a) \text{ if } a \leq m \)
    \[ = 1 \text{ otherwise.} \]
  - \( F_*(a) = N_M( ( -\infty, a] ) = 0 \text{ if } a < m^* \)
    \[ = 1 - \lim_{x \downarrow a} M(x) \text{ otherwise} \]
- \( \mathcal{P}(\pi) \) is a proper subset of \( \mathcal{P} \)
  - Not all \( P \) in \( \{ P : F^* \geq P \geq F_* \} \) are such that \( \Pi \geq P \)
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 $F_* < F < F \notin P(\Pi)$ since $P\{1.5, 2.5\} = 1 > \Pi\{1.5, 2.5\} = 0.5$.
Generalized cumulative distributions

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

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

- Choosing any order relation $\leq_R$

  $$F_R(x) = P(\{y \leq_R x\})$$ 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 $\alpha_i$ and $\beta_i$ 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 \( \pi \) (from \( \alpha_i \leq P(E_i) \)) and \( \pi_c \) (from \( P(E_i) \leq \beta_i \)).

- 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 according to \( \leq_R \)).

- \( P = P(\pi) \cap P(\pi_c) \)

- **Theorem**: a generalized p-box is a belief function (random set) with focal sets
  \[ \{ x : \pi(x) \geq \alpha \} \setminus \{ x : \delta(x) > \alpha \} \]
From generalized p-boxes to clouds

Fig 1.A Comonotonic cloud

Fig 1.B Non-comonotonic cloud
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 \(\pi\) 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 $\pi$
  - Increasing $\pi$: only one probability measure $p$ ($\pi =$ 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 $L = \{[l_i, u_i] \mid i = 1, \ldots, n\}$ induces the family $\mathcal{P}_L = \{P: l_i \leq P(\{s_i\}) \leq u_i\}$.

• Lower/upper probabilities on events are given by
  - $P_*(A) = \max(\Sigma_{s_i \in A} l_i; 1 - \Sigma_{s_i \notin A} u_i)$;
  - $P_*(A) = \min(\Sigma_{s_i \in A} u_i; 1 - \Sigma_{s_i \notin A} l_i)$

• $P_*$ is a 2-monotone Choquet capacity
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.

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

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

  - Complete knowledge: \( \pi(s_0) = 1 \) and \( \pi(s) = 0 \) otherwise.
  - Ignorance: \( \pi(s) = 1, \forall s \in S \)
Comparing belief functions in terms of informativeness

- Using contour functions:
  \[ \pi(s) = \Pi(s) = \sum_{x \in E} m(E) \]
  \( m_1 \) is more cf-informative than \( m_2 \) iff \( \pi_1 \leq \pi_2 \)

- Using belief or plausibility functions:
  \( m_1 \) is more pl-informative than \( m_2 \) iff \( \Pi_1 \leq \Pi_2 \) iff \( \text{Bel}_1 \geq \text{Bel}_2 \)

It corresponds to comparing credal sets \( \mathcal{P}(m) \):
\( \Pi_1 \leq \Pi_2 \) if and only if \( \mathcal{P}(m_1) \subseteq \mathcal{P}(m_2) \)
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

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

- $S = \{a, b, c\}; \; m_1(ab) = 0.5, \; m_1(bc) = 0.5$
- $m_2(abc) = 0.5, \; m_2(b) = 0.5$
- $m_1 \subseteq s m_2 \text{ nor } m_2 \subseteq s m_1$ hold
- $m_2 \subset_{pl} m_1: Pl_1(A) = Pl_2(A)$
  \hspace{1cm} \text{but } Pl_2(ac) = 0.5 < Pl_1(ac) = 1$
- $m_1 \subset_{Q} m_2: Q_1(A) = Q_2(A)$
  \hspace{1cm} \text{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

• 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 $B$. 
Conditional probability

• Revising 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∈A…).
  – C represents an additional sure information about the value of x: x∈C for sure.
  – P(A|C) represents the agent’s posterior belief that x∈A.
Conditioning a credal set

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

2. **Revision**: $C$ represents a set of universal truths:
   
   Add $P(C) = 1$ to the set of conditionals $P$.
   
   Now we must compute $Cr(A \mid C) = \inf\{P(A) P \in P, P(C) = 1\}$
   
   If $P(C) = 1$ is incompatible with $P$, use maximum likelihood:
   
   $Cr(A \mid C) = \inf\{P(A\mid C) P \in P, P(C) \text{ maximal}\}$
Example: \( \begin{array}{ccc} A & \leftrightarrow & B & \rightarrow & C \end{array} \)

- \( \mathcal{P} \) 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.
A disjunctive random set \((\mathcal{F}, m)\) representing background knowledge is equivalent to a set of probabilities \(\mathcal{P} = \{P: \forall A, P(A) \geq \text{Bel}(A)\}\) (NOT conversely).

Querying this information based on evidence \(C\) comes down to performing a sensitivity analysis on the conditional probability \(P(\cdot|C)\)

- \(\text{Bel}_C(A) = \inf \{P(A|C): P \in \mathcal{P}, P(C) > 0\}\)
- \(\text{Pl}_C(A) = \sup \{P(A|C): P \in \mathcal{P}, P(C) > 0\}\)
• **Theorem:** functions $\text{Bel}_C(A)$ and $\text{Pl}_C(A)$ are belief and plausibility functions of the form

\[
\text{Bel}_C(A) = \frac{\text{Bel}(C \cap A)}{\text{Bel}(C \cap A) + \text{Pl}(C \cap A^c)}
\]

\[
\text{Pl}_C(A) = \frac{\text{Pl}(C \cap A)}{\text{Pl}(C \cap A) + \text{Bel}(C \cap A^c)}
\]

where $\text{Bel}_C(A) = 1 - \text{Pl}_C(A^c)$

• *This conditioning does not add information:*

• If $E \cap C \neq \emptyset$ and $E \cap C^c \neq \emptyset$, it is not clear how much mass must be transferred to $E \cap C$.

• If so 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* ($\text{Pl}(\{a, d\} = 0$) then $m(\{b\}) = \alpha$, $m(\{c\}) = 1 - \alpha$ (Dempster conditioning, see further on)
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 \( m_t(B) = \sum_{E: C \cap E = B} m(E) \)
   - But the mass transferred to the empty set may not be zero!
   - \( m_t(\emptyset) = \text{Bel}(C^c) = \sum_{E: C \cap E = \emptyset} m(E) \) is the degree of conflict with evidence \( C \)

3. **Normalisation**: \( m_t(B) \) should be divided by
   \[
   \text{Pl}(C) = 1 - \text{Bel}(C^c) = \sum_{E: C \cap E \neq \emptyset} m(E)
   \]

- *This is revision of an unreliable testimony by a sure fact*
DEMPSTER RULE OF CONDITIONING = PRIORITIZED MERGING

The conditional plausibility function $\text{Pl}(\cdot|C)$ is

$$ \text{Pl}(A|C) = \frac{\text{Pl}(A \cap C)}{\text{Pl}(C)} $$

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

- $C$ surely contains the value of the unknown quantity described by $m$. So $\text{Pl}(C^c) = 0$
  - *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, $S = \{ \text{Peter, Paul, Mary} \}$

- **TBM modeling**: The masses are $m(\{\text{Peter, Paul}\}) = 1/2$; $m(\{\text{Mary}\}) = 1/2$
  - $\text{Bel}(\text{Paul}) = \text{Bel}(\text{Peter}) = 0$, $\text{Pl}(\text{Paul}) = \text{Pl}(\text{Peter}) = 1/2$
  - $\text{Bel}(\text{Mary}) = \text{Pl}(\text{Mary}) = 1/2$

- **Bayesian Modeling**: A prior probability
  - $P(\text{Paul}) = P(\text{Peter}) = 1/4$; $P(\text{Mary}) = 1/2$
• **Evidence 3**: Peter was seen elsewhere at the time of the killing.

• **TBM**: So $\text{Pl}(\text{Peter}) = 0$.
  - $m(\{\text{Peter, Paul}\}) = 1/2$; $m_t(\{\text{Paul}\}) = 1/2$
  - *A uniform probability on \{Paul, Mary\} results.*

• **Bayesian Modeling:**
  - $P(\text{Paul} \mid \text{not Peter}) = 1/3$; $P(\text{Mary} \mid \text{not Peter}) = 2/3$.
  - A very debatable result that depends on where the story starts.
    
    *Starting with \(i\) males and \(j\) females:*
    - $P(\text{Paul} \mid \text{Paul OR Mary}) = j/(i + j)$;
    - $P(\text{Mary} \mid \text{Paul OR Mary}) = i/(i + j)$

• **Walley conditioning:**
  - $\text{Bel}(\text{Paul}) = 0$; $\text{Pl}(\text{Paul}) = 1/2$
  - $\text{Bel}(\text{Mary}) = 1/2$; $\text{Pl}(\text{Mary}) = 1$
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
Information fusion

• Fusion with credal sets
  – The intersection of credal sets is a credal set.
  – The lower probability bounds obtained by the intersection of credal sets induced from belief functions: \( P^*(A) = \inf\{P(A), P \geq Bel_1, P \geq Bel_2\} \) is not a belief function.
  – The set function \( g(A) = \max(Bel_1(A), Bel_2(A)) \) is not a belief function
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)
Conclusion

- There exist a coherent range of uncertainty theories combining interval and probability representations.
  - Imprecise probability is the proper theoretical umbrella
  - The choice between subtheories depends on how expressive it is necessary to be in a given application.
  - There exists simple practical representations of imprecise probability: possibility theory is the simplest, and belief functions are a good compromise between calculability and expressivity

- Discrepancies between the theories remain on conditioning, combination rules, because their language primitives differ:
  - One cannot obviously express concepts defined by probability masses using credal sets, and conversely, let alone possibility distributions.