Introduction to belief functions

Thierry Denœux¹

¹Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 6599)
http://www.hds.utc.fr/~tdenoeux

Spring School BFTA 2011
Autrans, April 4-8, 2011
Contents of this lecture

1. Context, position of belief functions with respect to classical theories of uncertainty.
3. Some more advanced concepts: least commitment principle, cautious rule, multidimensional belief functions.
In science and engineering we always need to reason with partial knowledge and uncertain information (from sensors, experts, models, etc.).

Different kinds of uncertainty:

- **Aleatory uncertainty** induced by the variability of entities in populations and outcomes of random (repeatable) experiments. Example: drawing a ball from an urn. Cannot be reduced;
- **Epistemic uncertainty**, due to lack of knowledge. Example: inability to distinguish the color of a ball because of color blindness. Can be reduced.

Classical frameworks for reasoning with uncertainty:

1. Probability theory;
2. Set-membership approach.
Probability theory can be used to represent:

- **Aleatory uncertainty**: probabilities are considered as *objective* quantities and interpreted as *frequencies* or limits of frequencies;
- **Epistemic uncertainty**: probabilities are *subjective*, interpreted as *degrees of belief*.

Main objections against the use of probability theory as a model epistemic uncertainty (Bayesian model):

- Inability to represent *ignorance*;
- Not a plausible model of how people *make decisions* based on weak information.
Principle of Indifference (PI): in the absence of information about some quantity $X$, we should assign equal probability to any possible value of $X$.

The wine/water paradox:

There is a certain quantity of liquid. All that we know about the liquid is that it is composed entirely of wine and water, and the ratio of wine to water is between $1/3$ and $3$. What is the probability that the ratio of wine to water is less than or equal to $2$?
Let $X$ denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. According to the PI, $X \sim U_{[1/3,3]}$. Consequently:

$$P(X \leq 2) = (2 - 1/3)/(3 - 1/3) = 5/8.$$ 

Now, let $Y = 1/X$ denote the ratio of water to wine. Similarly, we only know that $Y \in [1/3, 3]$. According to the PI, $Y \sim U_{[1/3,3]}$. Consequently:

$$P(X \leq 2) = P(Y \geq 1/2)$$

$$= (3 - 1/2)/(3 - 1/3) = 15/16.$$
Suppose you have an urn containing **30 red balls** and **60 balls, either black or yellow**. You are given a choice between two gambles:

- **A**: You receive 100 euros if you draw a **red ball**;
- **B**: You receive 100 euros if you draw a **black ball**.

Also, you are given a choice between these two gambles (about a different draw from the same urn):

- **C**: You receive 100 euros if you draw a **red or yellow ball**;
- **D**: You receive 100 euros if you draw a **black or yellow ball**.

Most people **strictly prefer A to B**, hence $P(\text{red}) > P(\text{black})$, but they **strictly prefer D to C**, hence

$$P(\text{black}) + P(\text{yellow}) > P(\text{red}) + P(\text{yellow})$$

$$\Rightarrow P(\text{black}) > P(\text{red}).$$
Set-membership approach

- Partial knowledge about some variable $X$ is described by a set of possible values $E$ (constraint).
- Example:
  - Consider a system described by the equation
    \[ y = f(x_1, \ldots, x_n; \theta) \]
    where $y$ is the output, $x_1, \ldots, x_n$ are the inputs and $\theta$ is a parameter.
  - Knowing that $x_i \in [\underline{x}_i, \overline{x}_i]$, $i = 1, \ldots, n$ and $\theta \in [\underline{\theta}, \overline{\theta}]$, find a set $\mathbb{X}$ surely containing $x$.
- Advantage: **computationally simpler** than the probabilistic approach in many cases (interval analysis).
- Drawback: no way to express doubt, **conservative** approach.
Theory of belief functions

- Alternative theories of uncertainty:
  - Possibility theory (Zadeh, 1978; Dubois and Prade 1980’s-1990’s);
  - Imprecise probability theory (Walley, 1990’s);

- The theory of belief functions extends both the Set-membership approach and Probability Theory:
  - A belief function may be viewed both as a generalized set and as a non additive measure.
  - The theory includes extensions of probabilistic notions (conditioning, marginalization) and set-theoretic notions (intersection, union, inclusion, etc.)
Outline

1 Basics
   - Belief representation
   - Information fusion
   - Decision making

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions
Outline

1 Basics
   - Belief representation
   - Information fusion
   - Decision making

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions
Mass function

Definition

- Let $X$ be a variable taking values in a finite set $\Omega$ (frame of discernment).
- **Mass function:** $m : 2^\Omega \to [0, 1]$ such that
  \[
  \sum_{A \subseteq \Omega} m(A) = 1.
  \]
- Every $A$ of $\Omega$ such that $m(A) > 0$ is a focal set of $m$.
- $m$ is said to be normalized if $m(\emptyset) = 0$. This condition may be required or not.
A murder has been committed. There are three suspects: \( \Omega = \{ \text{Peter, John, Mary} \} \).

A witness saw the murderer going away in the dark, and he can only assert that it was man. How, we know that the witness is drunk 20% of the time.

This piece of evidence can be represented by

\[
m(\{ \text{Peter, John} \}) = 0.8,
\]

\[
m(\Omega) = 0.2
\]

The mass 0.2 is not committed to \( \{ \text{Mary} \} \), because the testimony does not accuse Mary at all!
A mass function \( m \) on \( \Omega \) may be viewed as arising from:

- A set \( \Theta = \{ \theta_1, \ldots, \theta_r \} \) of interpretations;
- A probability measure \( P \) on \( \Theta \);
- A multi-valued mapping \( \Gamma : \Theta \rightarrow 2^\Omega \).

Meaning: under interpretation \( \theta_i \), the evidence tells us that \( X \in \Gamma(\theta_i) \), and nothing more. The probability \( P(\{\theta_i\}) \) is transferred to \( A_i = \Gamma(\theta_i) \).

\( m(A) \) is the probability of knowing only that \( X \in A \), given the available evidence.
Mass functions

Special cases

- Only one focal set:

  \[ m(A) = 1 \text{ for some } A \subseteq \Omega \]

  → categorical (logical) mass function (∼ set). Special case: \( A = \Omega \), vacuous mass function, represents total ignorance.

- All focal sets are singletons:

  \[ m(A) > 0 \Rightarrow |A| = 1 \]

  → Bayesian mass function (∼ probability mass function).

- A mass function can thus be seen as
  - a generalized set;
  - a generalized probability distribution.
Belief and plausibility functions

Definitions

$$\Omega$$

$$A$$

$$B_1$$

$$B_2$$

$$B_3$$

$$B_4$$

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B),$$

$$pl(A) \geq bel(A), \quad \forall A \subseteq \Omega.$$
Interpretations:
- $\text{bel}(A)$ = degree to which the evidence supports $A$.
- $\text{pl}(A)$ = upper bound on the degree of support that could be assigned to $A$ if more specific information became available.

Special case: if $m$ is Bayesian, $\text{bel} = \text{pl}$ (probability measure).
We observe that

\[ \text{bel}(A \cup B) \geq \text{bel}(A) + \text{bel}(B) - \text{bel}(A \cap B) \]

\[ \text{pl}(A \cup B) \leq \text{pl}(A) + \text{pl}(B) - \text{bel}(A \cap B) \]

bel and pl are non additive measures.
Let $X$ denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. This is modeled by the categorical mass function $m_X$ such that $m_X([1/3, 3]) = 1$. Consequently:

$$\text{bel}_X([2, 3]) = 0, \quad \text{pl}_X([2, 3]) = 1.$$  

Now, let $Y = 1/X$ denote the ratio of water to wine. All we know is that $Y \in [1/3, 3]$. This is modeled by the categorical mass function $m_Y$ such that $m_Y([1/3, 3]) = 1$. Consequently:

$$\text{bel}_Y([1/3, 1/2]) = 0, \quad \text{pl}_Y([1/3, 1/2]) = 1.$$
Relations between $m$, $bel$ et $pl$

- Relations:
  
  $$bel(A) = pl(Ω) - pl(\overline{A}), \quad ∀A \subseteq Ω$$

  $$m(A) = \begin{cases} 
  \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} bel(B), & A \neq \emptyset \\
  1 - bel(Ω), & A = \emptyset 
  \end{cases}$$

- $m$, $bel$ et $pl$ are thus three equivalent representations of
  - a piece of evidence or, equivalently,
  - a state of belief induced by this evidence.
Assume that the focal sets of $m$ are nested:
$A_1 \subset A_2 \subset \ldots \subset A_r \rightarrow m$ is said to be consonant.

The following relations hold:

$$pl(A \cup B) = \max (pl(A), pl(B)), \quad \forall A, B \subseteq \Omega.$$ 

$pl$ is this a possibility measure, and $bel$ is the dual necessity measure.

The possibility distribution is the contour function:

$$\pi(x) = pl(\{x\}), \quad \forall x \in \Omega.$$ 

The theory of belief function can thus be considered as more expressive than possibility theory.
Outline

1 Basics
   - Belief representation
   - Information fusion
   - Decision making

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions
Let $m_1$ and $m_2$ be two mass functions on $\Omega$ induced by two independent items of evidence.

### Definitions

1. **Unnormalized Dempster’s rule**

   \[(m_1 \bigtriangleup m_2)(A) = \sum_{B \cap C = A} m_1(B)m_2(C)\]

2. **Normalized Dempster’s rule**

   \[(m_1 \bigtriangledown m_2)(A) = \begin{cases} \frac{(m_1 \bigtriangleup m_2)(A)}{1-K_{12}} & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases}\]

   \[K_{12} = (m_1 \bigtriangleup m_2)(\emptyset): \text{ degree of conflict.}\]
Dempster’s rule
Example

- We have $m_1(\{Peter, John\}) = 0.8$, $m_1(\Omega) = 0.2$.

- New piece of evidence: a blond hair has been found. There is a probability 0.6 that the room has been cleaned before the crime $\rightarrow m_2(\{John, Mary\}) = 0.6$, $m_2(\Omega) = 0.4$.

<table>
<thead>
<tr>
<th></th>
<th>${Peter, John}$</th>
<th>$\Omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${John, Mary}$</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>${John}$</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>${John, Mary}$</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>$\Omega$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.32</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Let \((\Theta_1, P_1, \Gamma_1)\) and \((\Theta_2, P_2, \Gamma_2)\) be the multi-valued mappings associated to \(m_1\) and \(m_2\).

If \(\theta_1 \in \Theta_1\) and \(\theta_2 \in \Theta_2\) both hold, then \(X \in \Gamma_1(\theta_1) \cap \Gamma_2(\theta_2)\).

If the two pieces of evidence are independent, then this happens with probability \(P_1(\{\theta_1\})P_2(\{\theta_2\})\).

The normalized rule is obtained after conditioning on the event \(\{(\theta_1, \theta_2)|\Gamma_1(\theta_1) \cap \Gamma_2(\theta_2) \neq \emptyset\}\).
Dempster’s rule
Properties

- Commutativity, associativity. Neutral element: $m_{\Omega}$.
- Generalization of intersection: if $m_A$ and $m_B$ are categorical mass functions, then

$$m_A \cap m_B = m_{A \cap B}$$

- Generalization of probabilistic conditioning: if $m$ is a Bayesian mass function and $m_A$ is a categorical mass function, then $m \oplus m_A$ is a Bayesian mass function that corresponding to the conditioning of $m$ by $A$.
- Notations for conditioning (special case):

$$m \cap m_A = m(\cdot | A), \quad m \oplus m_A = m^*(\cdot | A).$$
Dempster’s rule
Expression using commonalities

- **Commonality function**: let $q : 2^\Omega \to [0, 1]$ be defined as

$$q(A) = \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \Omega.$$ 

- Conversely,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} q(B), \quad \forall A \subseteq \Omega.$$ 

- Interpretation: $q(A) = m(A|A)$, for any $A \subseteq \Omega$.

- Expression of the unnormalized Dempster’s rule using commonalities:

$$(q_1 \ominus q_2)(A) = q_1(A) \cdot q_2(A), \quad \forall A \subseteq \Omega.$$
TBM disjunctive rule
Definition and justification

- Let \((\Theta_1, P_1, \Gamma_1)\) and \((\Theta_2, P_2, \Gamma_2)\) be the multi-valued mapping frameworks associated to two pieces of evidence.
- If interpretation \(\theta_k \in \Theta_k\) holds and piece of evidence \(k\) is reliable, then we can conclude that \(X \in \Gamma_k(\theta_k)\).
- If interpretation \(\theta_1 \in \Theta_1\) and \(\theta_2 \in \Theta_2\) both hold and we assume that at least one of the two pieces of evidence is reliable, then we can conclude that \(X \in \Gamma_1(\theta_1) \cup \Gamma_2(\theta_2)\).
- This leads to the TBM disjunctive rule:

\[
(m_1 \cup m_2)(A) = \sum_{B \cup C = A} m_1(B)m_2(C), \quad \forall A \subseteq \Omega
\]
TBM disjunctive rule

Properties

- Commutativity, associativity.
- Neutral element: $m_{\emptyset}$
- Let $b = bel + m(\emptyset)$ (implicability function). We have:

\[(b_1 \cup b_2) = b_1 \cdot b_2\]

- De Morgan laws for $\cap$ and $\cup$:

\[
\begin{align*}
\overline{m_1 \cup m_2} & = \overline{m_1} \cap \overline{m_2}, \\
\overline{m_1 \cap m_2} & = \overline{m_1} \cup \overline{m_2},
\end{align*}
\]

where $\overline{m}$ denotes the complement of $m$ defined by $\overline{m}(A) = m(\overline{A})$ for all $A \subseteq \Omega$. 
Selecting a combination rule

- All three rules $\cap$, $\oplus$ and $\cup$ assume the pieces of evidence to be **independent**.
- The conjunctive rules $\cap$ and $\oplus$ further assume that the pieces of evidence are **both reliable**;
- The TBM disjunctive rule $\cup$ only assumes that at least one of the items of evidence combined is reliable (weaker assumption).
- $\cap$ vs. $\oplus$:
  - $\cap$ keeps track of the **conflict** between items of evidence: very useful in some applications.
  - $\cap$ also makes sense under the **open-world assumption**.
  - The conflict increases with the number of combined mass functions: normalization is often necessary at some point.
- What to do with dependent items of evidence? → **Cautious** rule
Outline

1 Basics
   - Belief representation
   - Information fusion
   - Decision making

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions
A decision problem can be formalized by defining:

- A set of **acts** $A = \{a_1, \ldots, a_s\}$;
- A set of **states of the world** $\Omega$;
- A **loss function** $L : A \times \Omega \rightarrow \mathbb{R}$, such that $L(a, \omega)$ is the loss incurred if we select act $a$ and the true state is $\omega$.

**Bayesian framework**

- Uncertainty on $\Omega$ is described by a **probability measure** $P$;
- Define the **risk** of each act $a$ as the expected loss if $a$ is selected:

$$R(a) = \mathbb{E}_P[L(a, \cdot)] = \sum_{\omega \in \Omega} L(a, \omega)P(\{\omega\}).$$

- Select an act with **minimal risk**.

**Extension to the belief function framework?**

Thierry Denœux
Let \( m \) be a normalized mass function, and \( \mathcal{P}(m) \) the set of compatible probability measures on \( \Omega \), i.e., the set of \( P \) verifying

\[
\text{bel}(A) \leq P(A) \leq \text{pl}(A), \quad \forall A \subseteq \Omega.
\]

The lower and upper expected risk of each act \( a \) are defined, respectively, as:

\[
\underline{R}(a) = \mathbb{E}_m[L(a, \cdot)] = \inf_{P \in \mathcal{P}(m)} R_P(a) = \sum_{A \subseteq \Omega} m(A) \min_{\omega \in A} L(a, \omega)
\]

\[
\overline{R}(a) = \mathbb{E}_m[L(a, \cdot)] = \sup_{P \in \mathcal{P}(m)} R_P(a) = \sum_{A \subseteq \Omega} m(A) \max_{\omega \in A} L(a, \omega)
\]
Decision making

Strategies

- For each act $a$ we have a risk interval $[R(a), \overline{R}(a)]$. How to compare these intervals?
- Three strategies:
  1. $a$ is preferred to $a'$ iff $\overline{R}(a) \leq R(a')$;
  2. $a$ is preferred to $a'$ iff $R(a) \leq R(a')$ (optimistic strategy);
  3. $a$ is preferred to $a'$ iff $\overline{R}(a) \leq \overline{R}(a')$ (pessimistic strategy).
- Strategy 1 yields only a partial preorder: $a$ and $a'$ are not comparable if $\overline{R}(a) > R(a')$ and $\overline{R}(a') > \overline{R}(a)$. 
Decision making

Special case

- Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, $A = \{a_1, \ldots, a_K\}$, where $a_i$ is the act of selecting $\omega_i$.
- Let

$$L(a_i, \omega_j) = \begin{cases} 0 & \text{if } i = j \text{ (the true state has been selected)}, \\ 1 & \text{otherwise} \end{cases}.$$ 

- Then $\underline{R}(a_i) = 1 - pl(\omega_i)$ and $\overline{R}(a_i) = 1 - bel(\omega_i)$.
- The lower (resp., upper) risk is minimized by selecting the hypothesis with the largest plausibility (resp., degree of belief).
We have \( m(\{r\}) = 1/3, m(\{b, y\}) = 2/3. \)

<table>
<thead>
<tr>
<th></th>
<th>( r )</th>
<th>( b )</th>
<th>( y )</th>
<th>( R )</th>
<th>( \overline{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>-100</td>
<td>0</td>
<td>0</td>
<td>-100/3</td>
<td>-100/3</td>
</tr>
<tr>
<td>( B )</td>
<td>0</td>
<td>-100</td>
<td>0</td>
<td>-200/3</td>
<td>0</td>
</tr>
<tr>
<td>( C )</td>
<td>-100</td>
<td>0</td>
<td>-100</td>
<td>-100</td>
<td>-100/3</td>
</tr>
<tr>
<td>( D )</td>
<td>0</td>
<td>-100</td>
<td>-100</td>
<td>-200/3</td>
<td>-200/3</td>
</tr>
</tbody>
</table>

The observed behavior (preferring \( A \) to \( B \) and \( D \) to \( C \)) is explained by the pessimistic strategy.
How to find a **compromise** between the pessimistic strategy (minimizing the upper expected risk) and the optimistic one (minimizing the lower expected risk)?

**Two approaches:**

- **Hurwicz criterion:** \(a\) is preferred to \(a'\) iff \(R_\rho(a) \leq R_\rho(a')\) with

  \[
  R_\rho(a) = (1 - \rho)R(a) + \rho \overline{R}(a).
  \]

  and \(\rho \in [0, 1]\) is a **pessimism index** describing the attitude of the decision maker in the face of ambiguity.

- **Pignistic transformation** (Transferable Belief Model).
The “Dutch book” argument: in order to avoid Dutch books (sequences of bets resulting in sure loss), we have to base our decisions on a probability distribution on $\Omega$.

The TBM postulates that uncertain reasoning and decision making are two fundamentally different operations occurring at two different levels:

- **Uncertain reasoning** is performed at the credal level using the formalism of belief functions.
- **Decision making** is performed at the pignistic level, after the $m$ on $\Omega$ has been transformed into a probability measure.
The pignistic transformation $Bet$ transforms a normalized mass function $m$ into a probability measure $P_m = Bet(m)$ as follows:

$$P_m(A) = \sum_{\emptyset \neq B \subseteq \Omega} m(B) \frac{|A \cap B|}{|B|}, \quad \forall A \subseteq \Omega.$$ 

It can be shown that $bel(A) \leq P_m(A) \leq pl(A)$, hence $P_m \in \mathcal{P}(m)$. Consequently,

$$R(a) \leq R_{P_m}(a) \leq \overline{R}(a), \quad \forall a \in A.$$
Decision making

Example

Let $m(\{John\}) = 0.48$, $m(\{John, Mary\}) = 0.12$, 
$m(\{Peter, John\}) = 0.32$, $m(\Omega) = 0.08$.

We have

$$P_m(\{John\}) = 0.48 + \frac{0.12}{2} + \frac{0.32}{2} + \frac{0.08}{3} \approx 0.73,$$

$$P_m(\{Peter\}) = \frac{0.32}{2} + \frac{0.08}{3} \approx 0.19$$

$$P_m(\{Mary\}) = \frac{0.12}{2} + \frac{0.08}{3} \approx 0.09$$
Outline

1 Basics
   • Belief representation
   • Information fusion
   • Decision making

2 Selected advanced topics
   • Informational orderings
   • Cautious rule
   • Multidimensional belief functions
Let $m_1$ et $m_2$ be two mass functions on $\Omega$.

In what sense can we say that $m_1$ is more informative (committed) than $m_2$?

Special case:
- Let $m_A$ and $m_B$ be two categorical mass functions.
- $m_A$ is more committed than $m_B$ iff $A \subseteq B$.

Extension to arbitrary mass functions?
Plausibility and commonality orderings

- \( m_1 \) is pl-more committed than \( m_2 \) (noted \( m_1 \sqsubseteq_{pl} m_2 \)) if
  \[ pl_1(A) \leq pl_2(A), \quad \forall A \subseteq \Omega. \]

- \( m_1 \) is q-more committed than \( m_2 \) (noted \( m_1 \sqsubseteq_{q} m_2 \)) if
  \[ q_1(A) \leq q_2(A), \quad \forall A \subseteq \Omega. \]

Properties:
- Extension of set inclusion:
  \[ m_A \sqsubseteq_{pl} m_B \iff m_A \sqsubseteq_{q} m_B \iff A \subseteq B. \]
- Greatest element: vacuous mass function \( m_\Omega \).
Strong (specialization) ordering

- $m_1$ is a **specialization** of $m_2$ (noted $m_1 \sqsubseteq_s m_2$) if $m_1$ can be obtained from $m_2$ by distributing each mass $m_2(B)$ to subsets of $B$:

$$m_1(A) = \sum_{B \subseteq \Omega} S(A, B) m_2(B), \quad \forall A \subseteq \Omega,$$

where $S(A, B) = \text{proportion of } m_2(B) \text{ transferred to } A \subseteq B$.

- $S$: specialization matrix.

- Properties:
  - Extension of set inclusion;
  - Greatest element: $m_\Omega$;
  - $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{pl} m_2$ and $m_1 \sqsubseteq_{q} m_2$. 

Where $S(A, B) = \text{proportion of } m_2(B)$ transferred to $A \subseteq B$. 
Definition (Least Commitment Principle)

*When several belief functions are compatible with a set of constraints, the least informative according to some informational ordering (if it exists) should be selected.*

A very powerful method for constructing belief functions!
Outline

1. Basics
   - Belief representation
   - Information fusion
   - Decision making

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions
Cautious rule
Motivations

The standard rules $\cap$, $\oplus$ and $\cup$ assume the sources of information to be **independent**, e.g.
- experts with non overlapping experience/knowledge;
- non overlapping datasets.

What to do in case of **non independent evidence**?
- Describe the nature of the interaction between sources (difficult, requires a lot of information);
- Use a combination rule that **tolerates redundancy** in the combined information.

Such rules can be derived from the LCP using **suitable informational orderings**.
Cautious rule

Principle

- Two sources provide mass functions $m_1$ and $m_2$, and the sources are both considered to be reliable.
- After receiving these $m_1$ and $m_2$, the agent’s state of belief should be represented by a mass function $m_{12}$ more committed than $m_1$, and more committed than $m_2$.
- Let $S_x(m)$ be the set of mass functions $m'$ such that $m' \sqsubseteq_x m$, for some $x \in \{pl, q, s, \cdots \}$. We thus impose that $m_{12} \in S_x(m_1) \cap S_x(m_2)$.
- According to the LCP, we should select the $x$-least committed element in $S_x(m_1) \cap S_x(m_2)$, if it exists.
The above approach works for special cases.

Example (Dubois, Prade, Smets 2001): if \( m_1 \) and \( m_2 \) are consonant, then the \( q \)-least committed element in \( S_q(m_1) \cap S_q(m_2) \) exists and it is unique: it is the consonant mass function with commonality function \( q_{12} = q_1 \land q_2 \).

In general, neither existence nor uniqueness of a solution can be guaranteed with any of the \( x \)-orderings, \( x \in \{pl, q, s\} \).

We need to define a new ordering relation.

This ordering will be based on the (conjunctive) canonical decomposition of belief functions.
Definition: $m$ is simple mass function if it has the following form

\[
m(A) = 1 - w_A
\]
\[
m(\Omega) = w_A,
\]

with $A \subset \Omega$ and $w_A \in [0, 1]$.

Notation: $A^{w_A}$.

Property: $A^{w_1} \cap A^{w_2} = A^{w_1 w_2}$.

A mass function is separable if it can be written as the combination of simple mass functions:

\[
m = \bigcap_{A \subset \Omega} A^{w(A)}
\]

with $0 \leq w(A) \leq 1$ for all $A \subset \Omega$. 
Let $m_{12} = m_1 \ominus m_2$. We have $q_{12} = q_1 \cdot q_2$.

Assume we no longer trust $m_2$ and we wish to subtract it from $m_{12}$.

If $m_2$ is non dogmatic (i.e. $m_2(\Omega) > 0$ or, equivalently, $q_2(A) > 0, \forall A$), $m_1$ can be retrieved as

$$q_1 = q_{12} / q_2.$$

We note $m_1 = m_{12} \ominus m_2$.

Remark: $m_1 \ominus m_2$ may not be a valid mass function!
**Theorem (Smets, 1995)**

Any non dogmatic mass function \( m(Ω) > 0 \) can be canonically decomposed as:

\[
m = \left( \bigcap_{A \subset Ω} A^{w_C(A)} \right) \otimes \left( \bigcap_{A \subset Ω} A^{w_D(A)} \right)
\]

with \( w_C(A) \in (0, 1] \), \( w_D(A) \in (0, 1] \) and \( \max(w_C(A), w_D(A)) = 1 \) for all \( A \subset Ω \).

- Let \( w = w_C/w_D \).
- Function \( w : 2^Ω \setminus Ω \rightarrow \mathbb{R}^*_+ \) is called the (conjunctive) weight function.
- It is a new equivalent representation of a non dogmatic mass function (together with \( bel, pl, q, b \)).
Properties of $w$

- Function $w$ is directly available when $m$ is built by accumulating simple mass functions (common situation).
- Calculation of $w$ from $q$:

$$\ln w(A) = - \sum_{B \supseteq A} (-1)^{|B| - |A|} \ln q(B), \quad \forall A \subset \Omega.$$ 

- Conversely,

$$\ln q(A) = - \sum_{\Omega \supseteq B \supseteq A} \ln w(B), \quad \forall A \subseteq \Omega.$$ 

- TBM conjunctive rule:

$$w_1 \ominus w_2 = w_1 \cdot w_2.$$
Let $m_1$ and $m_2$ be two non dogmatic mass functions. We say that $m_1$ is \textit{w-more committed} than $m_2$ (denoted as $m_1 \sqsubseteq_w m_2$) if $w_1 \leq w_2$.

Interpretation: $m_1 = m_2 \ominus m$ with $m$ separable.

Properties:

- $m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2, \end{cases}$

- $m_\Omega$ is the \textit{only maximal element} of $\sqsubseteq_w$:

  $$m_\Omega \sqsubseteq_w m \Rightarrow m = m_\Omega.$$
Theorem

Let \( m_1 \) and \( m_2 \) be two nondogmatic BBAs. The \( w \)-least committed element in \( S_w(m_1) \cap S_w(m_2) \) exists and is unique. It is defined by the following weight function:

\[
w_{1 \wedge 2}(A) = w_1(A) \wedge w_2(A), \quad \forall A \subset \Omega.
\]

Definition (cautious conjunctive rule)

\[
m_1 \wedge m_2 = \bigcap_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)}.
\]

Thierry Denœux
Cautious rule

Definition

Theorem

Let $m_1$ and $m_2$ be two nondogmatic BBAs. The $w$-least committed element in $S_w(m_1) \cap S_w(m_2)$ exists and is unique. It is defined by the following weight function:

$$w_{1\wedge 2}(A) = w_1(A) \land w_2(A), \quad \forall A \subset \Omega.$$ 

Definition (cautious conjunctive rule)

$$m_1 \ominus m_2 = \bigodot_{A \subset \Omega} A^{w_1(A) \land w_2(A)}.$$
Cautious rule computation

<table>
<thead>
<tr>
<th>$m$-space</th>
<th>$w$-space</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>$m_2$</td>
<td>$\rightarrow$</td>
</tr>
<tr>
<td>$m_1 \land m_2$</td>
<td>$\leftarrow$</td>
</tr>
</tbody>
</table>
Cautious rule

Properties

- Commutative, associative
- Idempotent: $\forall m, m \sqcap m = m$
- Distributivity of $\sqcup$ with respect to $\sqcap$:

$$(m_1 \sqcup m_2) \sqcap (m_1 \sqcup m_3) = m_1 \sqcup (m_2 \sqcap m_3), \forall m_1, m_2, m_3.$$ 

The same item of evidence $m_1$ is not counted twice!

- No neutral element, but $m_\Omega \sqcap m = m$ iff $m$ is separable.
Related rules

- **Normalized cautious rule:**
  \[
  (m_1 \odot^* m_2)(A) = \begin{cases} 
  \frac{(m_1 \odot m_2)(A)}{1 - (m_1 \odot m_2)(\emptyset)} & \text{if } A \neq \emptyset \\
  0 & \text{if } A = \emptyset.
  \end{cases}
  \]

- **Bold disjunctive rule:**
  \[
  m_1 \uplus m_2 = \overline{m_1 \odot m_2}.
  \]

- Both \( \odot^* \) and \( \uplus \) are commutative, associative and idempotent.
Global picture

Six basic rules:

<table>
<thead>
<tr>
<th>Sources</th>
<th>independent</th>
<th>dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>All reliable</td>
<td>open world</td>
<td>□</td>
</tr>
<tr>
<td>closed world</td>
<td>⊕</td>
<td>△</td>
</tr>
<tr>
<td>At least one reliable</td>
<td>∪</td>
<td>∨</td>
</tr>
</tbody>
</table>
Outline

1 Basics
   - Belief representation
   - Information fusion
   - Decision making

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions
Multidimensional belief functions

Motivations

In many applications, we need to express uncertain information about several variables taking values in different domains.

Example: fault tree (logical relations between Boolean variables and probabilistic or evidential information about elementary events).
Fault tree example
(Dempster & Kong, 1988)
Let $X$ and $Y$ be two variables defined on frames $\Omega_X$ and $\Omega_Y$.

Let $\Omega_{XY} = \Omega_X \times \Omega_Y$ be the product frame.

A mass function $m^{\Omega_{XY}}$ on $\Omega_{XY}$ can be seen as an uncertain relation between variables $X$ and $Y$.

Two basic operations on product frames:

1. Express a joint mass function $m^{\Omega_{XY}}$ in the coarser frame $\Omega_X$ or $\Omega_Y$ (marginalization);
2. Express a marginal mass function $m^{\Omega_X}$ on $\Omega_X$ in the finer frame $\Omega_{XY}$ (vacuous extension).
Marginalization

- Problem: express $m_{\Omega^{XY}}$ in $\Omega_X$.
- Solution: transfer each mass $m_{\Omega^{XY}}(A)$ to the projection of $A$ on $\Omega_X$:

$$m_{\Omega^{XY}\downarrow\Omega_X}(B) = \sum_{\{A \subseteq \Omega^{XY}, A\downarrow\Omega_X = B\}} m_{\Omega^{XY}}(A), \forall B \subseteq \Omega_X.$$

- Marginal mass function

- Generalizes both set projection and probabilistic marginalization.
Vacuous extension:

- Problem: express $m^{\Omega_X}$ in $\Omega_{XY}$.
- Solution: transfer each mass $m^{\Omega_X}(B)$ to the cylindrical extension of $B$: $B \times \Omega_Y$.

Vacuous extension:

$$m^{\Omega_X \uparrow \Omega_{XY}}(A) = \begin{cases} m^{\Omega_X}(B) & \text{if } A = B \times \Omega_Y \\ 0 & \text{otherwise.} \end{cases}$$
Assume that we have:
- Partial knowledge of $X$ formalized as a mass function $m_{\Omega X}$;
- A joint mass function $m_{\Omega XY}$ representing an uncertain relation between $X$ and $Y$.

What can we say about $Y$?

Solution:

$$m_{\Omega Y} = \left( m_{\Omega X} \uparrow_{\Omega XY} \ominus m_{\Omega XY} \right) \downarrow_{\Omega Y}.$$ 

Infeasible with many variables and large frames of discernment, but efficient algorithms exist to carry out the operations in frames of minimal dimensions.
Fault tree example

<table>
<thead>
<tr>
<th>Cause</th>
<th>$m({1})$</th>
<th>$m({0})$</th>
<th>$m({0, 1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0.05</td>
<td>0.90</td>
<td>0.05</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.05</td>
<td>0.90</td>
<td>0.05</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.005</td>
<td>0.99</td>
<td>0.005</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.01</td>
<td>0.985</td>
<td>0.005</td>
</tr>
<tr>
<td>$X_5$</td>
<td>0.002</td>
<td>0.995</td>
<td>0.003</td>
</tr>
<tr>
<td>$G$</td>
<td>0.001</td>
<td>0.99</td>
<td>0.009</td>
</tr>
<tr>
<td>$M$</td>
<td>0.02</td>
<td>0.951</td>
<td>0.029</td>
</tr>
<tr>
<td>$F$</td>
<td>0.019</td>
<td>0.961</td>
<td>0.02</td>
</tr>
</tbody>
</table>
The theory of belief function: a very general formalism for representing imprecision and uncertainty that extends both probabilistic and set-theoretic frameworks:

- Belief functions can be seen both as generalized sets and as generalized probability measures;
- Reasoning mechanisms extend both set-theoretic notions (intersection, union, cylindrical extension, inclusion relations, etc.) and probabilistic notions (conditioning, marginalization, Bayes theorem, stochastic ordering, etc.).

The theory of belief function can also be seen as more general than Possibility theory (possibility measures are particular plausibility functions).
G. Shafer.


Ph. Smets and R. Kennes.

The Transferable Belief Model.


D. Dubois and H. Prade.

A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets.

T. Denœux.
Analysis of evidence-theoretic decision rules for pattern classification.

T. Denœux.
Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence.