Introduction to belief functions

Thierry Denœux

Université de Technologie de Compiègne, France
HEUDIASYC (UMR CNRS 7253)
https://www.hds.utc.fr/~tdenoeux

Fourth School on Belief Functions and their Applications
Xi’An, China, July 5, 2017
Contents of this lecture

1. Historical perspective, motivations
2. Fundamental concepts: belief, plausibility, commonality, conditioning, basic combination rules
3. Some more advanced concepts: cautious rule, multidimensional belief functions, belief functions in infinite spaces
Uncertain reasoning

- In science and engineering we always need to reason with partial knowledge and uncertain information (from sensors, experts, models, etc.)

- Different sources of uncertainty
  - Variability of entities in populations and outcomes of random (repeatable) experiments $\rightarrow$ Aleatory uncertainty. Example: drawing a ball from an urn. Cannot be reduced
  - Lack of knowledge $\rightarrow$ Epistemic uncertainty. Example: inability to distinguish the color of a ball because of color blindness. Can be reduced

- Classical ways of representing uncertainty
  1. Using probabilities
  2. Using set (e.g., interval analysis), or propositional logic
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)

1. Inability to represent ignorance
2. Not a plausibility model of how people make decisions based on weak information
The wine/water paradox

- **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 liquids. 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

$$P(X \leq 2) = (2 - 1/3)/(3 - 1/3) = 5/8$$
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. All we know is that $Y \in [1/3, 3]$. According to the PI, $Y \sim U_{[1/3,3]}$. Consequently

$$P(Y \geq 1/2) = (3 - 1/2)/(3 - 1/3) = 15/16$$
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. All we know is that $Y \in [1/3, 3]$. According to the PI, $Y \sim U_{[1/3,3]}$. Consequently

$$P(Y \geq 1/2) = (3 - 1/2)/(3 - 1/3) = 15/16$$

However, $P(X \leq 2) = P(Y \geq 1/2)!$
Ellsberg’s paradox

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

Most people strictly prefer A to B, hence \( P(\text{red}) > P(\text{black}) \), but they strictly prefer D to C, hence \( P(\text{black}) > P(\text{red}) \).
Ellsberg’s paradox

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 $P(\text{black}) > P(\text{red})$. 
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 \( P(red) > P(black) \), but they strictly prefer **D** to **C**, hence \( P(black) > P(red) \)
Partial knowledge about some variable $X$ is described by a set $E$ of possible values for $X$ (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 $Y$ surely containing $y$.

Advantage: computationally simpler than the probabilistic approach in many cases (interval analysis).

Drawback: no way to express doubt, conservative approach.
Theory of belief functions

History

- A formal framework for representing and reasoning with uncertain information
- Also known as Dempster-Shafer theory or Evidence theory
- Originates from the work of Dempster (1968) in the context of statistical inference.
- Formalized by Shafer (1976) as a theory of evidence
- Popularized and developed by Smets in the 1980’s and 1990’s under the name Transferable Belief Model
- Starting from the 1990’s, growing number of applications in information fusion, classification, reliability and risk analysis, etc.
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.).

Dempster-Shafer reasoning produces the same results as probabilistic reasoning or interval analysis when provided with the same information.

However, the greater expressive power of the theory of belief functions allows us to represent what we know in a more faithful way.
Relationships with other theories

- Fuzzy sets & Possibility theory
- Imprecise probability
- Rough sets
- DS theory
- Probability theory
Basic notions
- Mass functions
- Belief and plausibility functions
- Dempster’s rule

Selected advanced topics
- Informational orderings
- Cautious rule
- Belief functions on product spaces
- Belief functions on infinite spaces
Outline

1. Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
Mass function

Definition

- Let $X$ be a variable taking values in a finite set $\Omega$ (frame of discernment).
- Evidence about $X$ may be represented by a 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 property will be assumed hereafter, unless otherwise specified.
Example: the broken sensor

- Let $X$ be some physical quantity (e.g., a temperature), talking values in $\Omega$.
- A sensor returns a set of values $A \subset \Omega$, for instance, $A = [20, 22]$.
- However, the sensor may be broken, in which case the value it returns is completely arbitrary.
- There is a probability $p = 0.1$ that the sensor is broken.
- What can we say about $X$? How to represent the available information (evidence)?
Here, the probability $p$ is not about $X$, but about the state of a sensor.

Let $S = \{\text{working, broken}\}$ the set of possible sensor states.

- If the state is “working”, we know that $X \in A$.
- If the state is “broken”, we just know that $X \in \Omega$, and nothing more.

This uncertain evidence can be represented by a mass function $m$ on $\Omega$, such that

$$m(A) = 0.9, \quad m(\Omega) = 0.1$$
A mass function $m$ on $\Omega$ may be viewed as arising from
- A set $S = \{s_1, \ldots, s_r\}$ of states (interpretations)
- A probability measure $P$ on $S$
- A multi-valued mapping $\Gamma : S \rightarrow 2^\Omega$

The four-tuple $(S, 2^S, P, \Gamma)$ is called a source for $m$

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

$m(A)$ is the probability of knowing that $X \in A$, and nothing more, given the available evidence
Special cases

- If the evidence tells us that \( X \in A \) for sure and nothing more, for some \( A \subseteq \Omega \), then we have a **logical** mass function \( m[A] \) such that \( m[A](A) = 1 \)
  - \( m[A] \) is equivalent to \( A \)
  - Special case: \( m? \), the **vacuous** mass function, represents total ignorance

- If each interpretation \( s_i \) of the evidence points to a single value of \( X \), then all focal sets are singletons and \( m \) is said to be **Bayesian**. It is equivalent to a probability distribution

- A Dempster-Shafer mass function can thus be seen as
  - a generalized set
  - a generalized probability distribution

- Total ignorance is represented by the vacuous mass function \( m? \) such that \( m?(\Omega) = 1 \)
Outline

1 Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
Belief function

- If the evidence tells us that the truth is in $A$, and $A \subseteq B$, we say that the evidence supports $B$.

- Given a normalized mass function $m$, the probability that the evidence supports $B$ is thus

$$Bel(B) = \sum_{A \subseteq B} m(A)$$

- The number $Bel(B)$ is called the degree of belief in $B$, and the function $B \rightarrow Bel(B)$ is called a belief function.
If the evidence does not support \( \overline{B} \), it is consistent with \( B \).

The probability that the evidence is consistent with \( B \) is thus

\[
Pl(B) = \sum_{A \cap B \neq \emptyset} m(A) = 1 - Bel(\overline{B}).
\]

The number \( Pl(B) \) is called the plausibility of \( B \), and the function \( B \rightarrow Pl(B) \) is called a plausibility function.
The uncertainty on a proposition $B$ is represented by two numbers: $Bel(B)$ and $Pl(B)$, with $Bel(B) \leq Pl(B)$.

The intervals $[Bel(B), Pl(B)]$ have maximum length when $m$ is the vacuous mass function. Then,

$$[Bel(B), Pl(B)] = [0, 1]$$

for all subset $B$ of $\Omega$, except $\emptyset$ and $\Omega$.

The intervals $[Bel(B), Pl(B)]$ are reduced to points when the focal sets of $m$ are singletons (m is then said to be Bayesian); then,

$$Bel(B) = Pl(B)$$

for all $B$, and $Bel$ is a probability measure.
Basic notions

Belief and plausibility functions

Broken sensor example

From

\[ m(A) = 0.9, \quad m(\Omega) = 0.1 \]

we get

\[ Bel(A) = m(A) = 0.9, \quad Pl(A) = m(A) + m(\Omega) = 1 \]

\[ Bel(\overline{A}) = 0, \quad Pl(\overline{A}) = m(\Omega) = 0.1 \]

\[ Bel(\Omega) = Pl(\Omega) = 1 \]

We observe that

\[ Bel(A \cup \overline{A}) \geq Bel(A) + Bel(\overline{A}) \]

\[ Pl(A \cup \overline{A}) \leq Pl(A) + Pl(\overline{A}) \]

\textit{Bel} and \textit{Pl} are \textbf{non additive measures}.
Characterization of belief functions

- Function $Bel : 2^\Omega \to [0, 1]$ is a **completely monotone capacity**: it verifies $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$ and

$$Bel \left( \bigcup_{i=1}^{k} A_i \right) \geq \sum_{\emptyset \neq I \subseteq \{1, \ldots, k\}} (-1)^{|I|+1} Bel \left( \bigcap_{i \in I} A_i \right).$$

for any $k \geq 2$ and for any family $A_1, \ldots, A_k$ in $2^\Omega$.

- Conversely, to any completely monotone capacity $Bel$ corresponds a unique mass function $m$ such that:

$$m(A) = \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B), \quad \forall A \subseteq \Omega.$$
Let $m$ be a mass function, $Bel$ and $Pl$ the corresponding belief and plausibility functions.

For all $A \subseteq \Omega$,

\[
Bel(A) = 1 - Pl(\overline{A})
\]

\[
m(A) = \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B)
\]

\[
m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|+1} Pl(\overline{B})
\]

$m$, $Bel$ and $Pl$ are thus three equivalent representations of
- a piece of evidence or, equivalently
- a state of belief induced by this evidence
When the focal sets of \( m \) are nested: \( A_1 \subset A_2 \subset \ldots \subset A_r \), \( m \) is said to be consonant

The following relations then hold

\[
P_l(A \cup B) = \max \left( P_l(A), P_l(B) \right), \quad \forall A, B \subseteq \Omega
\]

\( P_l \) is this a possibility measure, and \( B_e l \) is the dual necessity measure

The possibility distribution is the contour function

\[
pl(x) = P_l(\{x\}), \quad \forall x \in \Omega
\]

The theory of belief function can thus be considered as more expressive than possibility theory (but the combination operations are different, see later).
Credal set

- A probability measure $P$ on $\Omega$ is said to be compatible with $m$ if
  \[ \forall A \subseteq \Omega, \quad Bel(A) \leq P(A) \leq Pl(A) \]

- The set $\mathcal{P}(m)$ of probability measures compatible with $m$ is called the credal set of $m$

  \[ \mathcal{P}(m) = \{ P : \forall A \subseteq \Omega, Bel(A) \leq P(A) \} \]

- $Bel$ is the lower envelope of $\mathcal{P}(m)$

  \[ \forall A \subseteq \Omega, \quad Bel(A) = \min_{P \in \mathcal{P}(m)} P(A) \]

- Not all lower envelopes of sets of probability measures are belief functions!
Outline

1. Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
The first item of evidence gave us: $m_1(A) = 0.9$, $m_1(\Omega) = 0.1$.

Another sensor returns another set of values $B$, and it is in working condition with probability 0.8.

This second piece of evidence can be represented by the mass function: $m_2(B) = 0.8$, $m_2(\Omega) = 0.2$

How to combine these two pieces of evidence?
If interpretations $s_1 \in S_1$ and $s_2 \in S_2$ both hold, then $X \in \Gamma_1(s_1) \cap \Gamma_2(s_2)$

If the two pieces of evidence are independent, then the probability that $s_1$ and $s_2$ both hold is $P_1(\{s_1\})P_2(\{s_2\})$
Computation

We then get the following combined mass function,

\[
\begin{align*}
m(A \cap B) &= 0.72 \\
m(A) &= 0.18 \\
m(B) &= 0.08 \\
m(\Omega) &= 0.02
\end{align*}
\]
Case of conflicting pieces of evidence

- If $\Gamma_1(s_1) \cap \Gamma_2(s_2) = \emptyset$, we know that $s_1$ and $s_2$ cannot hold simultaneously.
- The joint probability distribution on $S_1 \times S_2$ must be conditioned to eliminate impossible pairs of interpretation.
Computation

We then get the following combined mass function,

\[
m(\emptyset) = 0 \\
m(A) = 0.18/0.28 \approx 0.64 \\
m(B) = 0.08/0.28 \approx 0.29 \\
m(\Omega) = 0.02/0.28 \approx 0.07
\]
Dempster’s rule

- Let $m_1$ and $m_2$ be two mass functions and
  
  $$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

  their degree of conflict

- If $\kappa < 1$, then $m_1$ and $m_2$ can be combined as
  
  $$(m_1 \oplus m_2)(A) = \frac{1}{1 - \kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \neq \emptyset$$

  and $(m_1 \oplus m_2)(\emptyset) = 0$
Dempster’s rule

Properties

- Commutativity, associativity. Neutral element: $m_?$
- Generalization of intersection: if $m[A]$ and $m[B]$ are logical mass functions and $A \cap B \neq \emptyset$, then
  
  $$m[A] \oplus m[B] = m[A \cap B]$$

- If either $m_1$ or $m_2$ is Bayesian, then so is $m_1 \oplus m_2$ (as the intersection of a singleton with another subset is either a singleton, or the empty set).
Dempster’s conditioning

- Conditioning is a special case, where a mass function $m$ is combined with a logical mass function $m[A]$. Notation:
  $$m \oplus m[A] = m(\cdot | A)$$

- It can be shown that
  $$Pl(B|A) = \frac{Pl(A \cap B)}{Pl(A)}.$$

- Generalization of Bayes’ conditioning: if $m$ is a Bayesian mass function and $m[A]$ is a logical mass function, then $m \oplus m[A]$ is a Bayesian mass function corresponding to the conditioning of $m$ by $A$. 
Commonality function

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

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

- Conversely,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} Q(B)$$

- $Q$ is another equivalent representation of a belief function.
Let $Q_1$ and $Q_2$ be the commonality functions associated to $m_1$ and $m_2$. Let $Q_1 \oplus Q_2$ be the commonality function associated to $m_1 \oplus m_2$. We have

$$(Q_1 \oplus Q_2)(A) = \frac{1}{1 - \kappa} Q_1(A) \cdot Q_2(A), \quad \forall A \subseteq \Omega, A \neq \emptyset$$

$$(Q_1 \oplus Q_2)(\emptyset) = 1$$

In particular, $pl(\omega) = Q(\{\omega\})$. Consequently,

$$pl_1 \oplus pl_2 = (1 - \kappa)^{-1} pl_1 pl_2.$$
Mass functions expressing pieces of evidence are always normalized.

Smets introduced the unnormalized Dempster’s rule (TBM conjunctive rule $\cap$), which may yield an unnormalized mass function.

He proposed to interpret $m(\emptyset)$ as the mass committed to the hypothesis that $X$ might not take its value in $\Omega$ (open-world assumption).

I now think that this interpretation is problematic, as $m(\emptyset)$ increases mechanically when combining more and more items of evidence.

Claim: unnormalized mass functions (and $\cap$) are convenient mathematically, but only normalized mass functions make sense.

In particular, $Bel$ and $Pl$ should always be computed from normalized mass functions.
TBM disjunctive rule

Let \((S_1, P_1, \Gamma_1)\) and \((S_2, P_2, \Gamma_2)\) be sources associated to two pieces of evidence.

If interpretation \(s_k \in S_k\) holds and piece of evidence \(k\) is reliable, then we can conclude that \(X \in \Gamma_k(s_k)\).

If interpretation \(s \in S_1\) and \(s_2 \in S_2\) both hold and we assume that at least one of the two pieces of evidence is reliable, then we can conclude that \(X \in \Gamma_1(s_1) \cup \Gamma_2(s_2)\).

This leads to the TBM disjunctive rule:

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

\(Bel_1 \cup Bel_2 = Bel_1 \cdot Bel_2\)
Outline

1. Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
Let $m_1$ and $m_2$ 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 logical mass functions
- $m_{[A]}$ is more committed than $m_{[B]}$ iff $A \subseteq B$

Extension to arbitrary mass functions?
Plausibility ordering

- $m_1$ is pl-more committed than $m_2$ (noted $m_1 \sqsubseteq_{pl} m_2$) if
  
  $$Pl_1(A) \leq Pl_2(A), \quad \forall A \subseteq \Omega$$

  or, equivalently,
  
  $$Bel_1(A) \geq Bel_2(A), \quad \forall A \subseteq \Omega$$

- Imprecise probability interpretation:

  $$m_1 \sqsubseteq_{pl} m_2 \iff \mathcal{P}(m_1) \subseteq \mathcal{P}(m_2)$$

- Properties:
  - Extension of set inclusion:
    
    $$m_{[A]} \sqsubseteq_{pl} m_{[B]} \iff A \subseteq B$$
  - Greatest element: vacuous mass function $m_?$
Commonality ordering

- If $m_1 = m \oplus m_2$ for some $m$, and if there is no conflict between $m$ and $m_2$, then $Q_1(A) = Q(A)Q_2(A) \leq Q_2(A)$ for all $A \subseteq \Omega$.
- This property suggests that smaller values of the commonality function are associated with richer information content of the mass function.
- $m_1$ 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:
    \[ m_{[A]} \sqsubseteq_q m_{[B]} \iff A \subseteq B \]
  - Greatest element: vacuous mass function $m_?$
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 $S(A, B) =$ proportion of $m_2(B)$ transferred to $A \subseteq B$

- $S$: specialization matrix

Properties:
- Extension of set inclusion
- Greatest element: $m_?$
- $m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2 \end{cases}$
Least Commitment Principle

Definition

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. Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
Cautious rule

Motivations

- The basic rules $\oplus$ 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.
Cautious rule

Problem

- 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 $Q_{12} = \min(Q_1, Q_2)$.
- 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.
Simple and separable mass functions

- **Definition**: $m$ is **simple mass function** if it has the following form

  $m(A) = 1 - w(A)$
  
  $m(\Omega) = w(A)$

  for some $A \subset \Omega$, $A \neq \emptyset$ and $w(A) \in [0, 1]$. It is denoted by $A^{w(A)}$.

- **Property**: $A^{w_1(A)} \oplus A^{w_2(A)} = A^{w_1(A)w_2(A)}$

- A (normalized) mass function is **separable** if it can be written as the $\oplus$ combination of simple mass functions

  $m = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w(A)}$

  with $0 \leq w(A) \leq 1$ for all $A \subset \Omega$, $A \neq \emptyset$
The $w$-ordering

- Let $m_1$ and $m_2$ be two mass functions
- We say that $m_1$ is $w$-less committed than $m_2$ (denoted by $m_1 \subseteq_w m_2$) if
  \[ m_1 = m_2 \oplus m \]
  for some separable mass function $m$
- How to check this condition?
Weight function

Definition

- Let $m$ be a non dogmatic mass function, i.e., $m(\Omega) > 0$
- The weight function $w : 2^\Omega \to (0, +\infty)$ is defined by $w(\Omega) = 1$ and
  \[
  \ln w(A) = - \sum_{B \supseteq A} (-1)^{|B|-|A|} \ln Q(B), \quad \forall A \subset \Omega
  \]
- It can be shown that $Q$ can be recovered from $w$ as follows
  \[
  \ln Q(A) = - \sum_{\Omega \supset B \supseteq A} \ln w(B), \quad \forall A \subseteq \Omega
  \]
- $m$ can also be recovered from $w$ by
  \[
  m = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w(A)}
  \]
  although $A^{w(A)}$ is not a proper mass function when $w(A) > 1$
Weight function

Properties

- $m$ is separable iff
  \[ w(A) \leq 1, \quad \forall A \subset \Omega, A \neq \emptyset \]

- Dempster’s rule can be computed using the $w$-function by
  \[
m_1 \oplus m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w_1(A)w_2(A)}
  \]

- Characterization of the $w$-ordering
  \[ m_1 \sqsubseteq_w m_2 \iff w_1(A) \leq w_2(A), \quad \forall A \subset \Omega, A \neq \emptyset \]
Cautious rule

Definition

- Let $m_1$ and $m_2$ be two non dogmatic mass functions with weight functions $w_1$ and $w_2$.
- The $w$-least committed element in $S_w(m_1) \cap S_w(m_2)$ exists and is unique. It is defined by:

$$ m_1 \ominus m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A_{\min(w_1(A),w_2(A))} $$

- Operator $\ominus$ is called the (normalized) cautious rule.
**Computation**

**Cautious rule computation**

<table>
<thead>
<tr>
<th>$m$-space</th>
<th>$w$-space</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>$w_1$</td>
</tr>
<tr>
<td>$m_2$</td>
<td>$w_2$</td>
</tr>
<tr>
<td>$m_1 \land m_2$</td>
<td>$\min(w_1, w_2)$</td>
</tr>
</tbody>
</table>

Remark: we often have simple mass functions in the first place, so that the $w$ function is readily available.
Cautious rule

Properties

- Commutative, associative
- Idempotent: \( \forall m, m \otimes m = m \)
- Distributivity of \( \oplus \) with respect to \( \otimes \)

\[
(m_1 \oplus m_2) \otimes (m_1 \oplus m_3) = m_1 \oplus (m_2 \otimes m_3), \forall m_1, m_2, m_3
\]

The common item of evidence \( m_1 \) is not counted twice!

- No neutral element, but \( m \otimes m = m \) iff \( m \) is separable
## Basic rules

<table>
<thead>
<tr>
<th>Sources</th>
<th>independent</th>
<th>dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>All reliable</td>
<td>$\oplus$</td>
<td>$\wedge$</td>
</tr>
<tr>
<td>At least one reliable</td>
<td>$\cup$</td>
<td>$\vee$</td>
</tr>
</tbody>
</table>

$\bigvee$ is the bold disjunctive rule
Outline

1. Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
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)
Multidimensional belief functions
Marginalization, vacuous extension

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 $m_{XY}$ on $\Omega_{XY}$ can be seen as a generalized relation between variables $X$ and $Y$
Two basic operations on product frames
1. Express a joint mass function $m_{XY}$ in the coarser frame $\Omega_X$ or $\Omega_Y$ (marginalization)
2. Express a marginal mass function $m_X$ on $\Omega_X$ in the finer frame $\Omega_{XY}$ (vacuous extension)
Marginalization

Problem: express \( m_{XY} \) in \( \Omega_X \)

Solution: transfer each mass \( m_{XY}(A) \) to the projection of \( A \) on \( \Omega_X \)

Marginal mass function

\[
m_{XY \downarrow X}(B) = \sum_{\{A \subseteq \Omega_{XY}, A \downarrow \Omega_X = B\}} m_{XY}(A) \quad \forall B \subseteq \Omega_X
\]

Generalizes both set projection and probabilistic marginalization
Vacuous extension

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

Vacuous extension:

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

What can we say about $Y$?

Solution:

$$m_Y = (m_X \uparrow_{XY} \oplus m_{XY}) \downarrow_Y$$

Simpler notation:

$$m_Y = (m_X \oplus m_{XY}) \downarrow_Y$$

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

1 Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster’s rule

2 Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Belief functions on product spaces
   - Belief functions on infinite spaces
Belief function: general definition

- Let $\Omega$ be a set (finite or not) and $\mathcal{B}$ be an algebra of subsets of $\Omega$ (a nonempty family of subsets of $\Omega$, closed under complementation and finite intersection).
- A belief function (BF) on $\mathcal{B}$ is a mapping $\text{Bel} : \mathcal{B} \rightarrow [0, 1]$ verifying $\text{Bel}(\emptyset) = 0$, $\text{Bel}(\Omega) = 1$ and the complete monotonicity property: for any $k \geq 2$ and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

  \[ \text{Bel}\left(\bigcup_{i=1}^{k} B_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1, \ldots, k\}} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} B_i\right) \]

- A function $\text{Pl} : \mathcal{B} \rightarrow [0, 1]$ is a plausibility function iff $B \rightarrow 1 - \text{Pl}(\overline{B})$ is a belief function.
Let $S$ be a state space, $\mathcal{A}$ an algebra of subsets of $S$, $\mathbb{P}$ a finitely additive probability on $(S, \mathcal{A})$

Let $\Omega$ be a set and $\mathcal{B}$ an algebra of subsets of $\Omega$

$\Gamma$ a multivalued mapping from $S$ to $2^{\Omega} \setminus \{\emptyset\}$

The four-tuple $(S, \mathcal{A}, \mathbb{P}, \Gamma)$ is called a source

Under some conditions, it induces a belief function on $(\Omega, \mathcal{B})$
Strong measurability

Lower and upper inverses: for all $B \in \mathcal{B}$,

$$\Gamma_*(B) = B_* = \{s \in S | \Gamma(s) \neq \emptyset, \Gamma(s) \subseteq B\}$$

$$\Gamma^*(B) = B^* = \{s \in S | \Gamma(s) \cap B \neq \emptyset\}$$

$\Gamma$ is strongly measurable wrt $\mathcal{A}$ and $\mathcal{B}$ if, for all $B \in \mathcal{B}$, $B^* \in \mathcal{A}$

$(\forall B \in \mathcal{B}, B^* \in \mathcal{A}) \iff (\forall B \in \mathcal{B}, B_* \in \mathcal{A})$
Belief function induced by a source

Lower and upper probabilities

\[ \forall B \in \mathcal{B}, \quad P^*_*(B) = \frac{P(B^*_*)}{P(\Omega^*)}, \quad P^*(B) = \frac{P(B^*)}{P(\Omega^*)} = 1 - Bel(\overline{B}) \]

- \( P_* \) is a BF, and \( P^* \) is the dual plausibility function
- Conversely, for any belief function, there is a source that induces it (Shafer’s thesis, 1973)
Interpretation

- Typically, $\Omega$ is the domain of an unknown quantity $\omega$, and $S$ is a set of interpretations of a given piece of evidence about $\omega$.
- If $s \in S$ holds, then the evidence tells us that $\omega \in \Gamma(s)$, and nothing more.
- Then
  - $Bel(B)$ is the probability that the evidence supports $B$.
  - $Pl(B)$ is the probability that the evidence is consistent with $B$. 
Let $\pi$ be a mapping from $\Omega$ to $S = [0, 1]$ s.t. $\sup \pi = 1$
Let $\Gamma$ be the multi-valued mapping from $S$ to $2^\Omega$ defined by
\[ \forall s \in [0, 1], \quad \Gamma(s) = \{ \omega \in \Omega | \pi(\omega) \geq s \} \]
The source $(S, \mathcal{B}(S), \lambda, \Gamma)$ defines a consonant BF on $\Omega$, such that $pl(\omega) = \pi(\omega)$ (contour function)
The corresponding plausibility function is a possibility measure
\[ \forall B \subseteq \Omega, \quad Pl(B) = \sup_{\omega \in B} pl(\omega) \]
Let \((U, V)\) be a bi-dimensional random vector from a probability space \((S, \mathcal{A}, P)\) to \(\mathbb{R}^2\) such that \(U \leq V\) a.s.

Multi-valued mapping:

\[
\Gamma : s \rightarrow \Gamma(s) = [U(s), V(s)]
\]

The source \((S, \mathcal{A}, P, \Gamma)\) is a random closed interval. It defines a BF on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\)
Let \((S_i, \mathcal{A}_i, \mathbb{P}_i, \Gamma_i), i = 1, 2\) be two sources representing independent items of evidence, inducing BF \(Bel_1\) and \(Bel_2\).

The combined BF \(Bel = Bel_1 \oplus Bel_2\) is induced by the source \((S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2, \Gamma)\) with

\[
\Gamma(s_1, s_2) = \Gamma_1(s_1) \cap \Gamma_2(s_2)
\]
Approximate computation

**Monte Carlo simulation**

**Require:** Desired number of focal sets $N$

$i \leftarrow 0$

**while** $i < N$ **do**

Draw $s_1$ in $S_1$ from $\mathbb{P}_1$

Draw $s_2$ in $S_2$ from $\mathbb{P}_2$

$\Gamma \cap (s_1, s_2) \leftarrow \Gamma_1(s_1) \cap \Gamma_2(s_2)$

**if** $\Gamma \cap (s_1, s_2) \neq \emptyset$ **then**

$i \leftarrow i + 1$

$B_i \leftarrow \Gamma \cap (s_1, s_2)$

**end if**

**end while**

$\hat{Bel}(B) \leftarrow \frac{1}{N} \# \{ i \in \{1, \ldots, N \} | B_i \subseteq B \}$

$\hat{Pl}(B) \leftarrow \frac{1}{N} \# \{ i \in \{1, \ldots, N \} | B_i \cap B \neq \emptyset \}$
The theory of belief functions: 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)

The mathematical theory of belief functions in infinite spaces exists. We need practical models
References

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

cf. [http://www.hds.utc.fr/~tdenoeux](http://www.hds.utc.fr/~tdenoeux)