Introduction to belief functions

Thierry Denœux[1]

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

Spring School BFTA 2013
Carthage, Tunisia, May 20, 2013

Contents of this lecture

1. Historical perspective, motivations.
2. Fundamental concepts: belief, plausibility, commonality, Conditioning, basic combination rules.
3. Some more advanced concepts: least commitment principle, cautious rule, multidimensional belief functions.

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 frameworks for reasoning with uncertainty:
  1. Probability theory;
  2. Set-membership approach (e.g., interval analysis).

Probability theory
Interpretations

- 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 plausibility model of how people make decisions based on weak information.

Inability to represent ignorance

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?

Inability to represent ignorance

The wine/water paradox (continued)

- Let X denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. According to the PI, $X \sim \mathcal{U}_{[1/3,3]}$. 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, $Y \sim \mathcal{U}_{[1/3,3]}$. Consequently:

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

- However, $P(X \leq 2) = P(Y \geq 1/3)$!

Decision making
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 $P(red) > P(black)$, but they strictly prefer D to C, hence $P(black) > P(red)$.

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 θ 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 $\mathbb{Y}$ 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.

Theory of belief functions
Main idea

- 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.)
- Dempter-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.

Outline

1. Basics
   - Representation of evidence
   - Combination of evidence
   - Decision making
2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions

Outline

Mass function
Definition

- Let X be a variable taking values in a finite set Ω (frame of discernment).
- Evidence about X may be represented by a mass function $m : 2^\Omega \rightarrow [0, 1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

- Every A of Ω 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..

Murder example

- A murder has been committed. There are three suspects: $\Omega = \{Peter, John, Mary\}$.
- 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

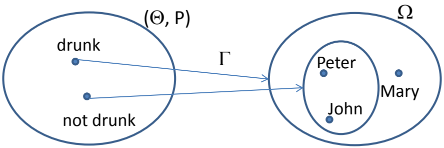
$$m(\{Peter, John\}) = 0.8,$$

$$m(\Omega) = 0.2$$

- The mass 0.2 is not committed to $\{Mary\}$, because the testimony does not accuse Mary at all!

Mass function

Random set interpretation



- A mass function m on Ω may be viewed as arising from
  - A set $\Theta = \{\theta_1, \ldots, \theta_r\}$ of interpretations;
  - A probability measure P on Θ;
  - A multi-valued mapping $\Gamma : \Theta \rightarrow 2^{\Omega}$.
- Meaning: under interpretation θ_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

- 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_Ω, the vacuous mass function, represents total ignorance.
- If each interpretation θ_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.

Belief function

- The total degree of support for A can be defined as the probability that the evidence implies A:

$$Bel(A) = P(\{\theta \in \Theta | \Gamma(\theta) \subseteq A\}) = \sum_{B \subseteq A} m(B).$$

- Function $Bel : 2^\Omega \to [0,1]$ is called a belief function.
- It 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^Ω.

Plausibility function

- The plausibility of A is the probability that the evidence does not contradict A. It is defined as

$$Pl(A) = P(\{\theta \in \Theta | \Gamma(\theta) \cap A \neq \emptyset\}) = \sum_{B \cap A \neq \emptyset} m(B)$$

- Properties:
  - $Pl(\emptyset) = 0$, $Pl(\Omega) = 1$;
  - $Bel(A) \leq Pl(A)$, $\forall A \subseteq \Omega$;
  - $Pl(A) = 1 - Bel(\overline{A})$, $\forall A \subseteq \Omega$.
- If m is Bayesian, $Bel = Pl$ (probability measure).

Murder example

A	$\emptyset$	$\{P\}$	$\{J\}$	$\{P,J\}$	$\{M\}$	$\{P,M\}$	$\{J,M\}$	Ω
$m(A)$	0	0	0	0.8	0	0	0	0.2
$Bel(A)$	0	0	0	0.8	0	0	0	1
$pl(A)$	0	1	1	1	0.2	1	1	1

- We observe that

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$$

$$Pl(A \cup B) \leq Pl(A) + Pl(B) - Pl(A \cap B)$$

Wine/water paradox revisited

- Let X denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. This is modeled by the logical mass function m_X such that $m_X([1/3, 3]) = 1$. Consequently:

$$Bel_X([2, 3]) = 0, \quad Pl_X([2, 3]) = 1.$$

- 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 logical mass function m_Y such that $m_Y([1/3, 3]) = 1$. Consequently:

$$Bel_Y([1/3, 1/2]) = 0, \quad Pl_Y([1/3, 1/2]) = 1.$$

Relations between *m*, *Bel* et *Pl*

- Let m be a normalized mass function, *Bel* and *Pl* the corresponding belief and plausibility functions.
- Relations:

$$Bel(A) = 1 - Pl(\overline{A}), \quad \forall A \subseteq \Omega$$

$$m(A) = \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B), \forall A \subseteq \Omega$$

- *m*, *Bel* et *Pl* are thus three equivalent representations of
  - a piece of evidence or, equivalently,
  - a state of belief induced by this evidence.

Relationship with Possibility theory

- 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:

$$Pl(A \cup B) = \max\left(Pl(A), Pl(B)\right), \quad \forall A, B \subseteq \Omega.$$

- Pl is this a possibility measure, and Bel is the dual necessity measure.
- The possibility distribution is the contour function:

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

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

Outline

1. Basics
   - Representation of evidence
   - Combination of evidence
   - Decision making

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

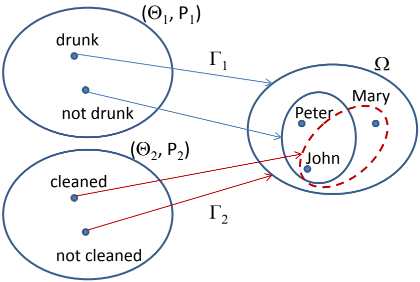
Dempster's rule

Murder example continued

- The first item of evidence gave us: $m_1(\{Peter, John\}) = 0.8$, $m_1(\Omega) = 0.2$.
- 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: $m_2(\{John, Mary\}) = 0.6$, $m_2(\Omega) = 0.4$.
- How to combine these two pieces of evidence?

Dempster's rule

Justification



- If interpretations $\theta_1 \in \Theta_1$ 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 the probability that θ_1 and θ_2 both hold is $P_1(\{\theta_1\})P_2(\{\theta_2\})$.
- If $\Gamma_1(\theta_1) \cap \Gamma_2(\theta_2) = \emptyset$, we know that θ_1 and θ_2 cannot hold simultaneously.
- The joint probability distribution on $\Theta_1 \times \Theta_2$ must be conditioned to eliminate such pairs.

Dempster's rule
Definition

- Let m_1 and m_2 be two mass functions and

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

their degree of conflict.
- If $K < 1$, then m_1 and m_2 can be combined as

$$(m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \neq \emptyset,$$

and $(m_1 \oplus m_2)(\emptyset) = 0$.

Dempster's rule
Properties

- Commutativity, associativity. Neutral element: m_Ω.
- Generalization of intersection: if m_A and m_B are categorical mass functions and $A \cap B \neq \emptyset$, then

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

- Generalization of probabilistic 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.
- Notation for conditioning (special case):

$$m \oplus m_A = m(\cdot|A).$$

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,

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

- Expression of $\oplus$ using commonalities:

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

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

Remarks on normalization

- Mass functions expressing pieces of evidence are always normalized.
- Smets introduced the unnormalized Dempster's rule (TBM conjunctive rule $\textcircled{\cap}$), 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 Ω (open-world assumption).
- We think that this interpretation is problematic, as $m(\emptyset)$ increases mechanically when combining more and more items of evidence.
- Our claim: unnormalized mass functions (and $\textcircled{\cap}$) 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 $(\Theta_1, P_1, \Gamma_1)$ and $(\Theta_2, P_2, \Gamma_2)$ be the random sets 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:

$$(m_1 \bigcirc\!\!\!\!\cup\, m_2)(A) = \sum_{B \cup C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega$$

- $Bel_1 \bigcirc\!\!\!\!\cup\, Bel_2 = Bel_1 \cdot Bel_2$.

Selecting a combination rule

- All three rules $\textcircled{\cap}$, $\oplus$ and $\textcircled{\cup}$ assume the pieces of evidence to be independent.
- The conjunctive rules $\textcircled{\cap}$ and $\oplus$ further assume that the pieces of evidence are both reliable;
- The TBM disjunctive rule $\textcircled{\cup}$ only assumes that at least one of the items of evidence combined is reliable (weaker assumption).
- $\textcircled{\cap}$ vs. $\oplus$:
  - $\textcircled{\cap}$ is simpler to compute (no normalization).
  - $\textcircled{\cap}$ keeps track of the conflict between items of evidence: very useful in some applications.
  - 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? $\rightarrow$ Cautious rule

Outline

Decision making
Problem formulation

- A decision problem can be formalized by defining:
  - A set of acts $\mathcal{A} = \{a_1, \ldots, a_s\}$;
  - A set of states of the world Ω;
  - A loss function $L : \mathcal{A} \times \Omega \to \mathbb{R}$, such that $L(a, \omega)$ is the loss incurred if we select act a and the true state is ω.
- Bayesian framework
  - Uncertainty on Ω is described by a probability measure P;
  - Define the risk of each act a as the expected loss if a is selected:

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

  - Select an act with minimal risk.
- Extension when uncertainty on Ω is described by a belief function?

Decision making

Compatible probabilities

- Let m be a normalized mass function, and $\mathcal{P}(m)$ the set of probability measures on Ω such that

$$Bel(A) \leq P(A) \leq Pl(A), \quad \forall A \subseteq \Omega.$$

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

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

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

Decision making

Strategies

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

Decision making

Special case

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

Decision making

Coming back to Ellsberg's paradox

We have $m(\{r\}) = 1/3$, $m(\{b, y\}) = 2/3$.

	r	b	y	$\underline{R}$	$\overline{R}$
A	-100	0	0	-100/3	-100/3
B	0	-100	0	-200/3	0
C	-100	0	-100	-100	-100/3
D	0	-100	-100	-200/3	-200/3

The observed behavior (preferring A to B and D to C) is explained by the pessimistic strategy.

Decision making
Other decision strategies

- 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

$$R_\rho(a) = (1-\rho)\underline{R}(a) + \rho\overline{R}(a).$$

    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).

Decision making

TBM approach

- 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 Ω.
- 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 Ω has been transformed into a probability measure.

Decision making

Pignistic transformation

- 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:

$$Bel(A) \leq P_m(A) \leq Pl(A), \quad \forall A \subseteq \Omega.$$

- Consequently,

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

Decision making
Example

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

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

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

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

Which decision rule to use?

- The two most widely used decision rules are: the maximum plausibility (optimistic strategy) and the maximum pignistic probability.
- Smets argued strongly in favor of the latter based (1) the avoidance of Dutch books and (2) the following requirement:

$$Bet(\alpha m_1 + (1-\alpha)m_2) = \alpha Bet(m_1) + (1-\alpha)Bet(m_2).$$

  It is not clear, however, why this property should be required.
- A practical argument in favor of the maximum plausibility rule is as follows: if $m_{12} = m_1 \oplus m_2$, then

$$pl_{12}(\omega) \propto pl_1(\omega)pl_2(\omega).$$

  When combining several mass functions, we do not need to compute the complete mass function to make a decision.

Outline

Informational comparison of belief functions

- Let m_1 et m_2 be two normalized mass functions on Ω.
- 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 and commonality orderings

- m_1 is pl-more committed than m_2 (noted $m_1 \sqsubseteq_{pl} m_2$) if $\mathcal{P}(m_1) \subseteq \mathcal{P}(m_2)$, which is equivalent to

$$Pl_1(A) \leq Pl_2(A), \quad \forall A \subseteq \Omega.$$

- 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:

$$m_A \sqsubseteq_{pl} m_B \Leftrightarrow m_A \sqsubseteq_q m_B \Leftrightarrow A \subseteq B.$$

  - Greatest element: vacuous mass function m_Ω.

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_Ω;
  - $m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2. \end{cases}$

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

Cautious rule
Motivations

- The basic rules $\oplus$ and $\textcircled{\cup}$ 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 $\mathcal{S}_x(m)$ 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 $m_{12} \in \mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2)$.
- According to the LCP, we should select the x-least committed element in $\mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2)$, 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 $\mathcal{S}_q(m_1) \cap \mathcal{S}_q(m_2)$ 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

$$
\begin{aligned}
m(A) &= 1 - w(A)\\
m(\Omega) &= w(A),
\end{aligned}
$$

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 normalized mass functions.
- We say that m_1 is w-less committed than m_2 (denoted by $m_1 \sqsubseteq_w m_2$) 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:

$$\ln Q(A) = -\sum_{\Omega \supset B \not\supseteq A} \ln w(B), \quad \forall A \subseteq \Omega$$

- 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:

$$m_1 \sqsubseteq_w m_2 \Leftrightarrow w_1(A) \leq w_2(A), \quad \forall A \subset \Omega, A \neq \emptyset.$$

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 $\mathcal{S}_w(m_1) \cap \mathcal{S}_w(m_2)$ exists and is unique. It is defined by:

$$m_1 \owedge m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{\min(w_1(A), w_2(A))}.$$

- Operator $\owedge$ is called the (normalized) cautious rule.

Cautious rule
Computation

Cautious rule computation

m-space		w-space
m_1	$\longrightarrow$	w_1
m_2	$\longrightarrow$	w_2
$m_1 \owedge m_2$	$\longleftarrow$	$\min(w_1, w_2)$

Cautious rule
Properties

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

$$(m_1 \oplus m_2) \owedge (m_1 \oplus m_3) = m_1 \oplus (m_2 \owedge m_3), \forall m_1, m_2, m_3.$$

  The same item of evidence m_1 is not counted twice!
- No neutral element, but $m_\Omega \owedge m = m$ iff m is separable.

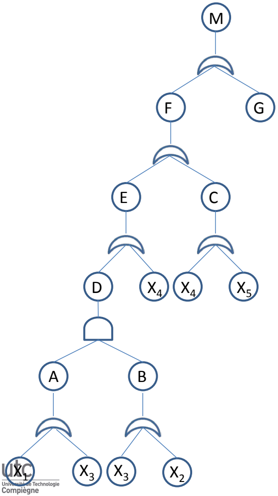
Basic rules

Sources	independent	dependent
All reliable	$\oplus$	$\owedge$
At least one reliable	$\cup\!\!\!\bigcirc$	$\ovee$

$\ovee$ is the bold disjunctive rule.

Outline

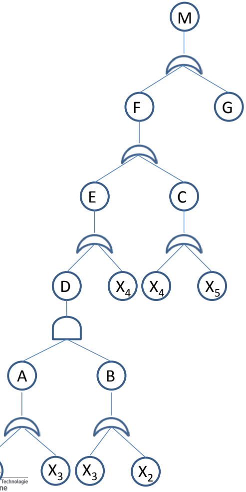
Multidimensional belief functions
Motivations

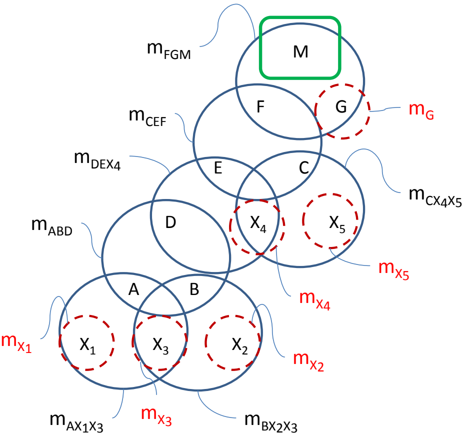


- In many applications, we need to express uncertain information about several variables taking values in different domains.
- Example: fault tree (logical relations between Boolean variables and probabilistic or evidential information about elementary events).

Fault tree example

(Dempster & Kong, 1988)





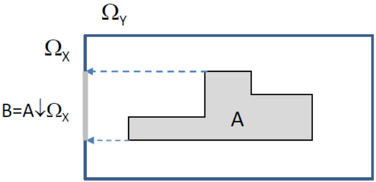
Hypergraph

Multidimensional belief functions

Marginalization, vacuous extension

- Let X and Y be two variables defined on frames Ω_X and Ω_Y.
- Let $\Omega_{XY} = \Omega_X \times \Omega_Y$ be the product frame.
- A mass function m^{XY} on Ω_{XY} can be seen as an 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 Ω_X or Ω_Y (marginalization);
  2. Express a marginal mass function m^X on Ω_X in the finer frame Ω_{XY} (vacuous extension).

Marginalization



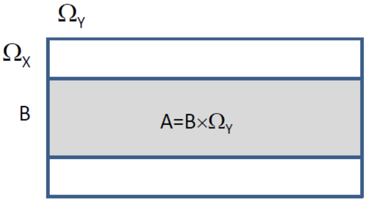
- Problem: express m^{XY} in Ω_X.
- Solution: transfer each mass $m^{XY}(A)$ to the projection of A on Ω_X.

- 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 Ω_{XY}.
- Solution: transfer each mass $m^X(B)$ to the cylindrical extension of B: $B \times \Omega_Y$.

- Vacuous extension:

$$m^{X\uparrow XY}(A) = \begin{cases} m^X(B) & \text{if } A = B \times \Omega_Y \\ 0 & \text{otherwise.} \end{cases}$$

Operations in product frames

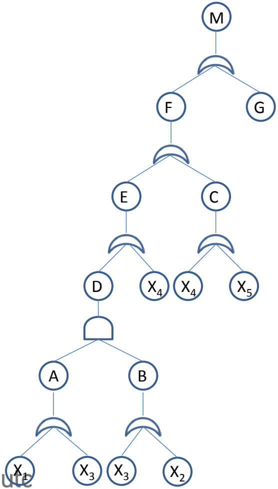
Application to approximate reasoning

- 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:

$$m^Y = \left(m^{X\uparrow XY} \oplus m^{XY}\right)^{\downarrow Y}.$$

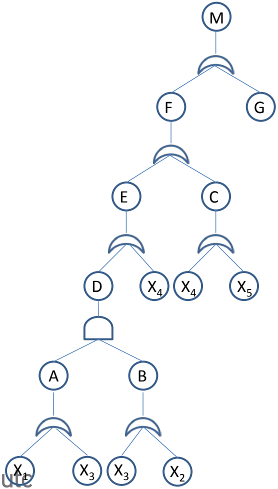
- Infeasible with many variables and large frames of discernment, but efficient algorithms exist to carry out the operations in frames of minimal dimensions.

Fault tree example



Cause	$m(\{1\})$	$m(\{0\})$	$m(\{0,1\})$
X_1	0.05	0.90	0.05
X_2	0.05	0.90	0.05
X_3	0.005	0.99	0.005
X_4	0.01	0.985	0.005
X_5	0.002	0.995	0.003
G	0.001	0.99	0.009
M	0.02	0.951	0.029
F	0.019	0.961	0.02

Fault tree example (continued)



Cause	$m(\{1\})$	$m(\{0\})$	$m(\{0,1\})$
M	1	0	0
G	0.197	0.796	0.007
F	0.800	0.196	0.004
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_1	0.236	0.724	0.040
X_2	0.236	0.724	0.040
X_3	0.200	0.796	0.004
X_4	0.302	0.694	0.004
X_5	0.099	0.898	0.003

Summary

- 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).

References I
cf. http://www.hds.utc.fr/˜tdenoeux

G. Shafer.
A mathematical theory of evidence. Princeton University Press, Princeton, N.J., 1976.

Ph. Smets and R. Kennes.
The Transferable Belief Model.
Artificial Intelligence, 66:191-243, 1994.

D. Dubois and H. Prade.
A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets.
International Journal of General Systems, 12(3):193-226, 1986.

T. Denœux.
Analysis of evidence-theoretic decision rules for pattern classification.
Pattern Recognition, 30(7):1095-1107, 1997.

References II

cf. http://www.hds.utc.fr/~tdenoeux

T. Denœux.

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

Artificial Intelligence, Vol. 172, pages 234-264, 2008.