Theory of Belief Functions: Application to machine learning and statistical inference

Lecture 1: basic concepts

Thierry Denœux

Summer 2023

Outline of the course I

- Course homepage:
  `https://www.hds.utc.fr/~tdenoeux/dokuwiki/en/bf`
- Roadmap:
  1. Basic notions:
     - Belief functions on finite sets. Dempster's rule (lecture + exercises)
     - Decision making (lecture + exercises)
  2. First applications: classification and statistical inference
     - *"A k-nearest neighbor classification rule based on Dempster-Shafer theory"* (paper reading + exercises in R)
     - *"A neural network classifier based on Dempster-Shafer theory"* (paper reading + exercises in R)
     - *"Constructing belief functions from sample data using multinomial confidence regions"* (paper reading + exercises in R)
  3. Advanced concepts:
     - Random sets and belief functions in a general framework (lecture + exercises)
     - Possibility and Random fuzzy sets
     - *"Reasoning with fuzzy and uncertain evidence using epistemic random fuzzy sets: general framework and practical models"* (paper reading + exercises)

Outline of the course II

4. Statistical and ML applications:
   - Statistical prediction using belief functions: application to linear and logistic regression
   - *"Prediction of future observations using belief functions: a likelihood-based approach"* (paper reading + exercises)
   - *"Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets: the ENNreg model"* (paper reading + exercises)
5. Statistical inference and learning from uncertain data:
   - *"Maximum likelihood estimation from Uncertain Data in the Belief Function Framework"* (paper reading + exercises)
   - *"Parametric Classification with Soft Labels using the Evidential EM Algorithm"* (paper reading + exercises)
6. Project
7. Project presentation

What we will study in this course

- A mathematical formalism called
  - Dempster-Shafer (DS) theory
  - Evidence theory
  - Theory of belief functions
- This formalism was introduced by A. P. Dempster in the 1960's for statistical inference, and developed by G. Shafer in the late 1970's into a general theory for reasoning under uncertainty.
- DS encompasses probability theory and set-membership approaches such as interval analysis as special cases: it is very general.
- Many applications in AI (expert systems, machine learning), engineering (information fusion, uncertainty quantification, risk analysis), statistics (statistical estimation and prediction), etc.
- Some applications to econometrics. A new research avenue to explore!

Outline

Reminder: probability mass functions and measures

- Let Ω be a finite set (sample space, universe or discourse,...)
- A probability mass function is mapping $p : \Omega \to [0, 1]$ such that

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

- The corresponding probability measure is the mapping $P : 2^\Omega \to [0, 1]$ defined by

$$P(A) = \sum_{\omega \in A} p(\omega) \quad \text{for all } A \subseteq \Omega$$

Properties

- $P(\emptyset) = 0$, $P(\Omega) = 1$
- Additivity:

$$\forall A, B \subseteq \Omega, \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- More generally, for any $k \geq 2$ and for any family $A_1, \ldots, A_k$ of subsets of Ω,

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

Interpretations

1. Objective:
   - Ω is the set of possible outcomes of a random experiment
   - $p(\omega)$ is the limit frequency of outcome ω in a series of repetitions of the random experiment
   - $P(A)$ is the limit frequency of the event "$\omega \in A$"
2. Subjective:
   - Ω is the set of possible answers to some question
   - $P(A)$ is a agent's degree of belief that the true answer belongs to A
   - Probability theory is the mainstream formalism for representing uncertainty in AI

Should degrees of belief be additive?

The case of complete ignorance

- To highlight the implications of the additivity assumption, it is useful to consider the extreme (but frequent) situation of complete ignorance.
- How to define a probability measure on Ω in that case?
- The only sensible solution is provided by Laplace's principle of indifference (PI): "In the absence of any relevant evidence, agents should distribute their credence (or 'degrees of belief') equally among all the possible outcomes under consideration".
- As shown by the following example, this principle leads to paradoxes.

Is there life around Sirius?

- Consider the question: "Are there or are there not living beings in orbit around the star Sirius"?
- The set of possibilities can be denoted by $\Theta = \{\theta_1, \theta_2\}$, where
  - θ_1 is the possibility that there is life
  - θ_2 is the possibility that there is not
- As we are completely ignorant about this question, the probabilities should be, according to the PI:

$$p(\theta_1) = p(\theta_2) = 1/2$$

Is there life around Sirius? (continued)

- We could also have considered a refined set of possibilities, such as $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where
  - ω_1 corresponds to the possibility that there is life around Sirius
  - ω_2 corresponds to the possibility that there are planets but no life, and
  - ω_3 corresponds to the possibility that there are not even planets
- With this new set of probabilities, complete ignorance is represented by

$$p(\omega_1) = p(\omega_2) = p(\omega_3) = 1/3$$

- But θ_1 has the same meaning as ω_1 and θ_2 has the same meaning as $\{\omega_2, \omega_3\}$, so the probability distributions on Θ and Ω are inconsistent.

The solution: relax the additivity property

- The Dempster-Shafer theory was introduced to solve such paradoxes.
- It replaces the probability measure P by two nonadditive measures: a belief function and a plausibility function.
- The probabilistic formalism is recovered as a special case (known frequencies or proportions in a population).

Outline

Mass function

Definition

- Let Ω be the finite set of possible answers to some question X.
- To emphasize the fact that the granularity of Ω is a matter of choice, Ω is sometimes called the frame of discernment
- A mass function is a mapping $m : 2^{\Omega} \rightarrow [0, 1]$ such that

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

and

$$m(\emptyset) = 0$$

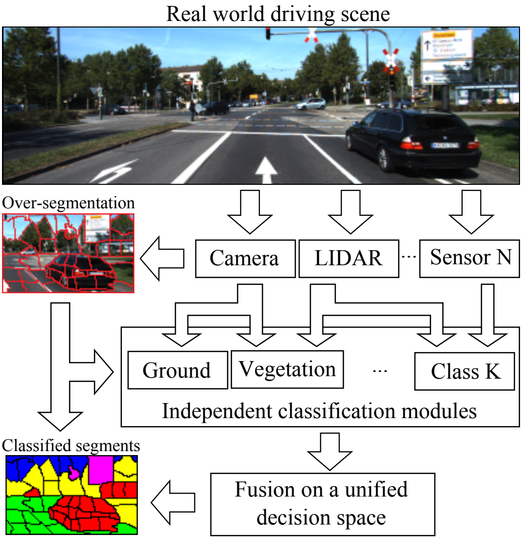
- Every subset A of Ω such that $m(A) > 0$ is a focal set of m

Mass function
Interpretation

- In DS theory, a mass function m on Ω is used as a representation of evidence, i.e., partial information about the question of interest.
- It is usually induced by
  - A mapping Γ from a set S of interpretations (or possible meanings) of the evidence
  - Known probabilities on S
  - Each probability $p(s)$ for $s \in S$ is then transferred to subset $\Gamma(s) \subseteq \Omega$, and

$$m(A) = \sum_{\{s \in S : \Gamma(s) = A\}} p(s)$$

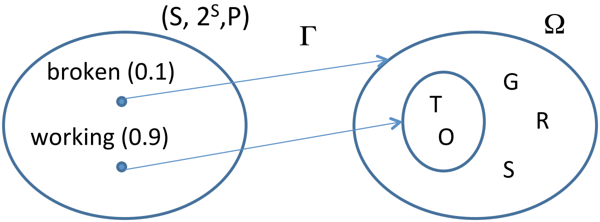
Example: road scene analysis



Example: road scene analysis (continued)

- Let X be the contents of some region in the image, and $\Omega = \{G, R, T, O, S\}$, corresponding to the possibilities **G**rass, **R**oad, **T**ree/Bush, **O**bstacle, **S**ky.
- Assume that a lidar sensor (laser telemeter) returns the information $X \in \{T, O\}$, but we there is a probability $p = 0.1$ that the information is not reliable (because, e.g., the sensor is out of order).
- How to represent this information by a mass function?

Formalization



- Here, the probability p is not about X, but about the state of a sensor.
- Let $S = \{\text{working}, \text{broken}\}$ the set of possible sensor states.
  - If the state is "working", we know that $X \in \{T, O\}$.
  - 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 Ω, such that

$$m(\{T, O\}) = 0.9, \quad m(\Omega) = 0.1$$

Special cases

Logical mass function: 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 $m_A(A) = 1$. Example: $m_{\{T,O\}}$ denotes the mass function such that $m_{\{T,O\}}(\{T,O\}) = 1$.

Vacuous mass function: In particular, m_Ω represents total ignorance; it is called the vacuous mass function

Bayesian mass function: If all focal sets of m are singletons, m is said to be Bayesian. It is equivalent to a probability mass function. Example: $m(\{T\}) = 0.5$, $m(\{O\}) = 0.5$.

A Dempster-Shafer mass function can thus be seen as

- A generalized set
- A generalized probability distribution

Outline

Defiinitions

- Given a mass function m on Ω, the corresponding belief and plausibility functions are mappings from 2^Ω to [0,1] defined as follows:

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

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

- Interpretation:
  - $Bel(A)$ is a measure of the strength with which A is supported by the available evidence (taking into accounts all subsets $B \subseteq A$); it is a degree of belief in A
  - $Pl(A)$ is a measure of the lack of support given to the complement of A, it is a degree of lack of belief in $\overline{A}$

Elementary properties

- $Bel(\emptyset) = Pl(\emptyset) = 0$
- $Bel(\Omega) = Pl(\Omega) = 1$
- For all $A \subseteq \Omega$,

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

- Superadditivity of *Bel*:

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

- Subadditivity of *Pl*:

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

Two-dimensional representation

- The uncertainty about a proposition A is represented by two numbers: $Bel(A)$ and $Pl(A)$, with $Bel(A) \leq Pl(A)$
- The intervals $[Bel(A), Pl(A)]$ have maximum length when m is vacuous: then, $Bel(A) = 0$ for all $A \neq \Omega$, and $Pl(A) = 1$ for all $A \neq \emptyset$.
- The intervals $[Bel(A), Pl(A)]$ have minimum length when m is Bayesian. Then,

$$Bel(A) = Pl(A) = \sum_{\omega \in A} m(\{\omega\})$$

for all A, and Bel is a probability measure.

Road scene analysis example

- We had $\Omega = \{G, R, T, O, S\}$ and

$$m(\{T, O\}) = 0.9, \quad m(\Omega) = 0.1$$

- What are the credibility and the plausibility that the region corresponds or does not correspond to a tree?

$$Bel(\{T\}) = 0, \quad Pl(\{T\}) = 0.9 + 0.1 = 1$$

$$Bel(\overline{\{T\}}) = 0, \quad Pl(\overline{\{T\}}) = 1$$

But

$$Bel(\{T\} \cup \overline{\{T\}}) = Bel(\Omega) = 1$$

and

$$Pl(\{T\} \cup \overline{\{T\}}) = Pl(\Omega) = 1.$$

Characterization of belief functions

Theorem

Let $F : 2^\Omega \to [0,1]$. The following two statements are equivalent:

Statement 1 *There exists a mass function $m : 2^\Omega \to [0,1]$ such that $F(A) = \sum_{B \subseteq A} m(B)$ for all $A \subseteq \Omega$ (i.e., F is a belief function).*

Statement 2 *Function F has the following 3 properties:*

1. $F(\emptyset) = 0$
2. $F(\Omega) = 1$
3. *For any $k \geq 2$ and for any family $A_1, \ldots, A_k$ in 2^Ω,*

$$F\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} F\left(\bigcap_{i \in I} A_i\right) \qquad (2)$$

(Property (2) is called complete monotonicity).

Relations between *m*, *Bel* and *Pl*

- Let m be a mass function, Bel and Pl the corresponding belief and plausibility functions
- Thanks to the following equations, given any one of these functions, we can recover the other two: for all $A \subseteq \Omega$,

$$
\begin{aligned}
Bel(A) &= \sum_{B \subseteq A} m(B) \\
Pl(A) &= 1 - Bel(\overline{A}) \\
Bel(A) &= 1 - Pl(\overline{A}) \\
m(A) &= \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B)
\end{aligned}
$$

- m, Bel et Pl are thus three equivalent representations of a piece of evidence.

Outline

Definition and theorem

Consonant mass functions are an important special case.

Definition (Consonant mass function)

A mass function m is consonant iff its focal sets are nested, i.e., for any two focal set A_i and A_j, $A_i \subseteq A_j$ or $A_j \subseteq A_i$

Theorem

Let m be a mass function, and let Bel and Pl be the corresponding belief and plausibility functions. The following statements are equivalent:

1. *m is consonant*
2. *For any $A, B \subseteq \Omega$, $Bel(A \cap B) = \min\left[Bel(A), Bel(B)\right]$*
3. *For any $A, B \subseteq \Omega$, $Pl(A \cup B) = \max\left[Pl(A), Pl(B)\right]$*
4. *For any $A \subseteq \Omega$, $Pl(A) = \max_{\omega \in A} pl(\omega)$, where $pl(\omega) = Pl(\{\omega\})$*

(Function $pl : \Omega \rightarrow [0, 1]$ is called the contour function).

Proof of $1 \Rightarrow 2$

- Let m be a consonant mass function with focal sets $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_r$.
- For any $A, B \subseteq \Omega$, let i_1 and i_2 be the largest indices such that $A_i \subseteq A$ and $A_i \subseteq B$, respectively.
- Then, $A_i \subseteq A \cap B$ iff $i \leq \min(i_1, i_2)$ and

$$
\begin{aligned}
Bel(A \cap B) &= \sum_{i=1}^{\min(i_1,i_2)} m(A_i) \\
&= \min\left(\sum_{i=1}^{i_1} m(A_i), \sum_{i=1}^{i_2} m(A_i)\right) \\
&= \min(Bel(A), Bel(B)).
\end{aligned}
$$

Proof of $2 \Rightarrow 3$

Now, from the equality $\overline{A \cup B} = \overline{A} \cap \overline{B}$, we have

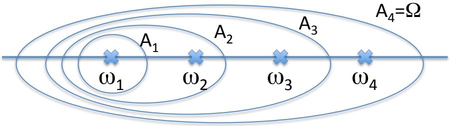
$$\begin{aligned} Pl(A \cup B) &= 1 - Bel(\overline{A \cup B}) \\ &= 1 - Bel(\overline{A} \cap \overline{B}) \\ &= 1 - \min(Bel(\overline{A}), Bel(\overline{B})) \\ &= \max(1 - Bel(\overline{A}), 1 - Bel(\overline{B})) \\ &= \max(Pl(A), Pl(B)). \end{aligned}$$

Proof of $3 \Rightarrow 4$

- Assume that $Pl(A \cup B) = \max(Pl(A), Pl(B))$ for all $A, B \subseteq \Omega$.
- Let Π_n be the following property: $Pl(A) = \max_{\omega \in A} pl(\omega)$ for all $A \subseteq \Omega$ such that $|A| \leq n$.
- We prove Π_n for all $n \geq 0$ by induction:
  - Π_1 and Π_2 trivially true.
  - Assume Π_n is true and let $A \subseteq \Omega$ such that $|A| = n + 1$. We can write $A = B \cup \{\omega_0\}$ with $|B| = n$. Consequently,

$$\begin{aligned} Pl(A) &= \max(Pl(B), pl(\omega_0)) \\ &= \max(\max_{\omega \in B} pl(\omega), pl(\omega_0)) \\ &= \max_{\omega \in A} pl(\omega) \end{aligned}$$

Proof of $4 \Rightarrow 1$ I



- Let Pl be a plausibility function verifying $Pl(A) = \max_{\omega \in A} pl(\omega)$ for all A.
- Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be the frame of discernment with elements arranged by decreasing order of plausibility, i.e.,

$$1 = pl(\omega_1) \geq pl(\omega_2) \geq \ldots \geq pl(\omega_n)$$

  and let A_i denote the set $\{\omega_1, \ldots, \omega_i\}$, for $1 \leq i \leq n$.
- Let m denote the following consonant mass function:

$$m(A_i) = pl(\omega_i) - pl(\omega_{i+1}), \quad 1 \leq i \leq n-1$$
$$m(\Omega) = pl(\omega_n).$$

Example

For instance, for the following contour function defined on the frame $\Omega = \{a, b, c, d\}$:

ω	a	b	c	d
$pl(\omega)$	0.3	0.5	1	0.7

the corresponding mass function is

$$\begin{aligned} m(\{c\}) &= 1 - 0.7 = 0.3 \\ m(\{c, d\}) &= 0.7 - 0.5 = 0.2 \\ m(\{c, d, b\}) &= 0.5 - 0.3 = 0.2 \\ m(\{c, d, b, a\}) &= 0.3. \end{aligned}$$

Proof of 4 $\Rightarrow$ 1 (continued)

- Let Pl_m be the plausibility function induced by m.
- For any subset A of Ω, let $i_A = \min\{1 \leq i \leq n : \omega_i \in A\}$.
- $A_i \cap A \neq \emptyset$ iff $i \geq i_A$.
- Consequently,

$$\begin{aligned}
Pl_m(A) &= \sum_{i=i_A}^{n} m(A_i) \\
&= pl(\omega_{i_A}) - pl(\omega_{i_A+1}) + pl(\omega_{i_A+1}) - pl(\omega_{i_A+2}) + \ldots - pl(\omega_n) + pl(\omega_n) \\
&= pl(\omega_{i_A}) \\
&= \max_{\omega \in A} pl(\omega) = Pl(A),
\end{aligned}$$

i.e., $Pl_m = Pl$.

Outline

Outline

Road scene example continued

- Variable X was defined as the type of object in some region of the image, and the frame was $\Omega = \{G, R, T, O, S\}$, corresponding to the possibilities **G**rass, **R**oad, **T**ree/Bush, **O**bstacle, **S**ky
- A lidar sensor gave us the following mass function:

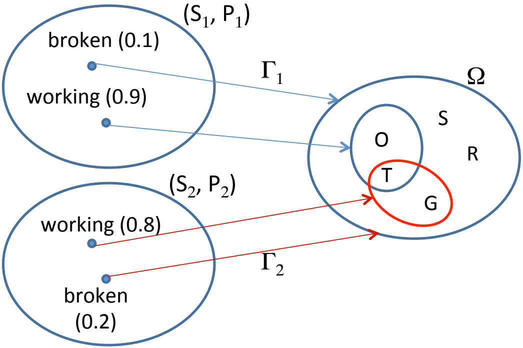
$$m_1(\{T, O\}) = 0.9, \quad m_1(\Omega) = 0.1$$

- Now, assume that a camera returns the mass function:

$$m_2(\{G, T\}) = 0.8, \quad m_2(\Omega) = 0.2$$

- How to combine these two pieces of evidence?

Analysis



- 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 $P_1(\{s_1\})P_2(\{s_2\})$

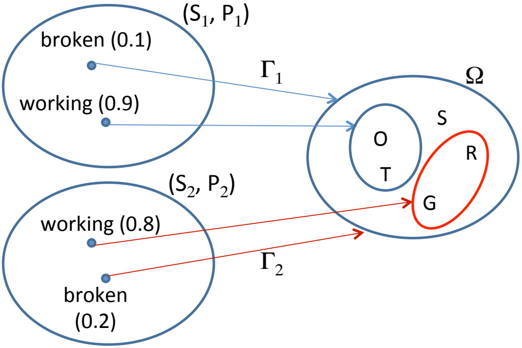
Computation

$m_1 \backslash m_2$	$\{T, G\}$ (0.8)	Ω (0.2)
$\{O, T\}$ (0.9)	$\{T\}$ (0.72)	$\{O, T\}$ (0.18)
Ω (0.1)	$\{T, G\}$ (0.08)	Ω (0.02)

We then get the following combined mass function,

$$\begin{aligned} m(\{T\}) &= 0.72 \\ m(\{O, T\}) &= 0.18 \\ m(\{T, G\}) &= 0.08 \\ m(\Omega) &= 0.02 \end{aligned}$$

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 such pairs

Computation

$m_1 \backslash m_2$	$\{G, R\}$ (0.8)	Ω (0.2)
$\{O, T\}$ (0.9)	$\emptyset$ (0.72)	$\{O, T\}$ (0.18)
Ω (0.1)	$\{G, R\}$ (0.08)	Ω (0.02)

We then get the following combined mass function,

$$
\begin{aligned}
m(\emptyset) &= 0 \\
m(\{O, T\}) &= 0.18/0.28 = 9/14 \\
m(\{G, R\}) &= 0.08/0.28 = 4/14 \\
m(\Omega) &= 0.02/0.28 = 1/14
\end{aligned}
$$

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

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \neq \emptyset \tag{3}$$

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

- $m_1 \oplus m_2$ is called the orthogonal sum of m_1 and m_2
- This rule can be used to combine mass functions induced by independent pieces of evidence

Another example

A	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	$\{a,b,c\}$
$m_1(A)$	0	0	0.5	0.2	0	0.3	0	0
$m_2(A)$	0	0.1	0	0.4	0.5	0	0	0

		m_2		
		$\{a\}, 0.1$	$\{a,b\}, 0.4$	$\{c\}, 0.5$
	$\{b\}, 0.5$	$\emptyset, 0.05$	$\{b\}, 0.2$	$\emptyset, 0.25$
m_1	$\{a,b\}, 0.2$	$\{a\}, 0.02$	$\{a,b\}, 0.08$	$\emptyset, 0.1$
	$\{a,c\}, 0.3$	$\{a\}, 0.03$	$\{a\}, 0.12$	$\{c\}, 0.15$

The degree of conflict is $\kappa = 0.05 + 0.25 + 0.1 = 0.4$. The combined mass function is

$$
\begin{aligned}
(m_1 \oplus m_2)(\{a\}) &= (0.02 + 0.03 + 0.12)/0.6 = 0.17/0.6 \approx 0.2833 \\
(m_1 \oplus m_2)(\{b\}) &= 0.2/0.6 = 1/3 \\
(m_1 \oplus m_2)(\{a,b\}) &= 0.08/0.6 \approx 0.1333 \\
(m_1 \oplus m_2)(\{c\}) &= 0.15/0.6 = 0.25.
\end{aligned}
$$

Properties

1. Commutativity, associativity. Neutral element: $m_?$
2. Generalization of intersection: if m_A and m_B are logical mass functions and $A \cap B \neq \emptyset$, then
$$m_A \oplus m_B = m_{A \cap B}$$
3. 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).

Outline

Dempster's rule conditioning

- Conditioning is a special case, where a mass function m is combined with a logical mass function m_B. Notation:

$$m \oplus m_B = m(\cdot \mid B)$$

- We thus have $m(A \mid B) = 0$ for any A not included in B and, for any $A \subseteq B$,

$$m(A \mid B) = (1-\kappa)^{-1} \sum_{C \cap B = A} m(C), \tag{4}$$

where the degree of conflict κ is

$$\kappa = \sum_{C \cap B = \emptyset} m(C) = 1 - \sum_{C \cap B \neq \emptyset} m(C) = 1 - Pl(B).$$

Conditional plausibility function

Proposition

The plausibility function $Pl(\cdot|B)$ induced by $m(\cdot|B)$ is given by

$$Pl(A \mid B) = \frac{Pl(A \cap B)}{Pl(B)}$$

Proof: We have

$$\begin{aligned} Pl(A \mid B) &= \sum_{\{C: C\cap A \neq \emptyset\}} m(C|B) \\ &= Pl(B)^{-1} \sum_{\{C: C\cap A \neq \emptyset\}} \sum_{\{D: D\cap B = C\}} m(D) \\ &= Pl(B)^{-1} \sum_{\{D: D\cap B\cap A \neq \emptyset\}} m(D) = \frac{Pl(A\cap B)}{Pl(B)} \end{aligned}$$

If Pl is a probability measure, $Pl(\cdot \mid B)$ is, thus, the conditional probability measure given B: Dempster's rule of combination thus extends Bayesian conditioning.

Outline

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,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} Q(B) \tag{5}$$

- Q is another equivalent representation of a belief function.
- Properties: $Q(\emptyset) = 1$ and $Q(\Omega) = m(\Omega)$

Commonality function and Dempster's rule

- 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 $(Q_1 \oplus Q_2)(\emptyset) = 1$ and, for all non empty subset A of Ω,

$$(Q_1 \oplus Q_2)(A) = (1-\kappa)^{-1} Q_1(A) \cdot Q_2(A).$$

Proof

$$
\begin{aligned}
(Q_1 \oplus Q_2)(A) &= \sum_{B \supseteq A} (m_1 \oplus m_2)(B) \\
&= (1-\kappa)^{-1} \sum_{B \supseteq A} \sum_{C \cap D = B} m_1(C) m_2(D) \\
&= (1-\kappa)^{-1} \sum_{C \cap D \supseteq A} m_1(C) m_2(D) \\
&= (1-\kappa)^{-1} \sum_{C \supseteq A, D \supseteq A} m_1(C) m_2(D) \\
&= (1-\kappa)^{-1} \left(\sum_{C \supseteq A} m_1(C) \right) \left(\sum_{D \supseteq A} m_2(D) \right) \\
&= (1-\kappa)^{-1} Q_1(A) \cdot Q_2(A).
\end{aligned}
$$

Product rule for commonality and contour functions

- Using (5) with $A = \emptyset$, we get

$$\sum_{\emptyset \neq B \subseteq \Omega} (-1)^{|B|} Q(B) = -Q(\emptyset) = -1, \tag{6}$$

  which makes it possible to compute the commonality function once commonality numbers are determined up to some multiplicative constant. (See following example)
- Given two mass functions m_1 and m_2, we can thus combine them either using (3), or by converting them to commonality functions, multiplying them pointwise, and computing the corresponding mass function using (5).
- In particular, $pl(\omega) = Q(\{\omega\})$. Consequently,

$$pl_1 \oplus pl_2 = (1-\kappa)^{-1} pl_1 pl_2.$$

Example (cf. Slide 43)

A	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	$\{a,b,c\}$
$Q_1(A)$	1	0.5	0.7	0.2	0.3	0.3	0	0
$Q_2(A)$	1	0.5	0.4	0.4	0.5	0	0	0
$Q_1(A)Q_2(A)$	1	0.25	0.28	0.08	0.15	0	0	0

$$(1-\kappa)^{-1} = -1/(-0.25-0.28-0.15+0.08) = 1.6667$$

$$\Rightarrow \kappa = 1-(1/1.6667) = 0.4$$

A	$\emptyset$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{c\}$	$\{a,c\}$	$\{b,c\}$	$\{a,b,c\}$
$(Q_1 \oplus Q_2)(A)$	1	0.4167	0.4667	0.1333	0.25	0	0	0
$(m_1 \oplus m_2)(A)$	0	0.2833	0.3333	0.1333	0.25	0	0	0