Representation and combination of evidence

Thierry Denœux

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

Beijing University of Technology,
Beijing, China
June-July 2016

Outline

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

Example

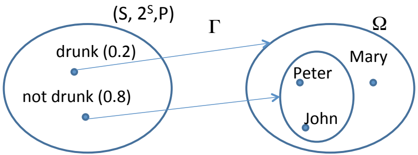
- When traveling by train, you find a page of a used newspaper, with an article announcing rain for tomorrow
- The date of the newspaper is missing. If is today's newspaper, you know that it will rain tomorrow (assuming the forecast is perfectly reliable). If not, you know nothing
- Assume your subjective probability that this is today's paper is 0.8
- The frame of discernment is $\Omega = \{\text{rain}, \neg\text{rain}\}$
- The evidence can be represented by the following mass function

$$m(\{\text{rain}\}) = 0.8, \quad m(\{\text{rain}, \neg\text{rain}\}) = 0.2$$

- The mass 0.2 is not committed to $\{\neg\text{rain}\}$, because there is no evidence that it will not rain

Mass function

Source



- A mass function m on Ω 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

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: $m_?$, 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

Outline

Belief function

- If interpretation s holds and $\Gamma(s) \subseteq A$ for some $A \subseteq \Omega$, we say that the evidence supports A.
- The probability that the evidence supports A is thus

$$\begin{aligned} Bel(A) &= P(\{s \in S | \Gamma(s) \subseteq A\}) \\ &= \sum_{B \subseteq A} m(B). \end{aligned}$$

- It can be interpreted as the total degree of support in A, or as a degree of belief that the truth is in A.
- The function $Bel : 2^{\Omega} \rightarrow [0, 1]$ is called a belief function

Plausibility function

- We can also consider the degree of support in $\overline{A}$,

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

- It is a measure of doubt in A (we doubt A if the complement of A is supported).
- The plausibility of A is defined as

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

  It is high when the complement $\overline{A}$ is not supported by the evidence.
- The function $Pl : 2^{\Omega} \rightarrow [0, 1]$ is called a plausibility function

Two-dimensional representation

- The uncertainty on 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 $m = m_?$ 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, $Bel(A) = Pl(A)$ for all A, and Bel is a probability measure.

Example

A	$\emptyset$	$\{\text{rain}\}$	$\{\neg\text{rain}\}$	$\{\text{rain}, \neg\text{rain}\}$
$m(A)$	0	0.8	0	0.2
$Bel(A)$	0	0.8	0	1
$pl(A)$	0	1	0.2	1

- We observe that

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

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

Characterization of belief and plausibility functions

Belief function

- Function Bel 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}

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

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

Characterization of belief and plausibility functions

Plausibility function

A function $Pl : 2^\Omega \rightarrow [0,1]$ is a plausibility function iff it is a completely alternating capacity, i.e., iff it satisfies the following conditions:

1. $Pl(\emptyset) = 0$;
2. $Pl(\Omega) = 1$;
3. For any $k \geq 2$ and any collection $A_1, \ldots, A_k$ of subsets of Ω,

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

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

Outline

Outline

1. Representation of evidence
   - Mass functions
   - Belief and plausibility functions
2. Relations with alternative theories
   - Possibility theory
   - Imprecise probabilities
3. Combination of evidence
   - Dempster's rule
   - Disjunctive rule
   - Dubois-Prade rule

Consonant belief function

- 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, for all $A, B \subseteq \Omega$,

$$Pl(A \cup B) = \max\left(Pl(A), Pl(B)\right)$$

$$Bel(A \cap B) = \min\left(Bel(A), Bel(B)\right)$$

- Pl is this a possibility measure, and Bel is the dual necessity measure

Contour function

- The contour function of a belief function *Bel* is defined by

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

- When *Bel* is consonant, it can be recovered from its contour function,

$$Pl(A) = \max_{\omega \in A} pl(\omega).$$

- The contour function is then a possibility distribution
- The theory of belief function can thus be considered as more expressive than possibility theory

From the contour function to the mass function

- Let pl be a contour on the frame $\Omega = \{\omega_1, \ldots, \omega_n\}$, 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 $1 \leq i \leq n$.
- Then, the corresponding mass function m is

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

Example

- Consider, for instance, the following contour distribution 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

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

Outline

Credal set

- A probability measure P on Ω is said to be compatible with Bel if
$$Bel(A) \leq P(A)$$
for all $A \subseteq \Omega$
- Equivalently, $P(A) \leq Pl(A)$ for all $A \subseteq \Omega$
- The set $\mathcal{P}(m)$ of probability measures compatible with m is called the credal set of m
$$\mathcal{P}(Bel) = \{P : \forall A \subseteq \Omega, Bel(A) \leq P(A))\}$$

Construction of $\mathcal{P}(Bel)$

- An arbitrary element of $\mathcal{P}(Bel)$ can be obtained by distributing each mass $m(A)$ among the elements of A.
- More precisely, let $\alpha(\omega, A)$ be the fraction of $m(A)$ allocated to the element ω. We have
$$\sum_{\omega \in A} \alpha(\omega, A) = m(A).$$
- By summing up the numbers $\alpha(\omega, A)$ for each ω, we get a probability mass function on Ω,
$$p_\alpha(\omega) = \sum_{A \ni \omega} \alpha(\omega, A).$$
- It can be verified that
$$P_\alpha(A) = \sum_{\omega \in A} p_\alpha(\omega) \geq Bel(A),$$
for all $A \subseteq \Omega$.

Belief functions are coherent lower probabilities

- It can be shown (Dempster, 1967) that any element of the credal set $\mathcal{P}(Bel)$ can be obtained in that way.
- Furthermore, the bounds in the inequalities $Bel(A) \leq P(A)$ and $P(A) \leq Pl(A)$ are attained. We thus have, for all $A \subseteq \Omega$,

$$Bel(A) = \min_{P \in \mathcal{P}(Bel)} P(A)$$

$$Pl(A) = \max_{P \in \mathcal{P}(Bel)} P(A)$$

- We say that *Bel* is a coherent lower probability.
- Not all lower envelopes of sets of probability measures are belief functions!

A counterexample

- Suppose a fair coin is tossed twice, in such a way that the outcome of the second toss may depend on the outcome of the first toss.
- The outcome of the experiment can be denoted by $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$.
- Let $H_1 = \{(H,H),(H,T)\}$ and $H_2 = \{(H,H),(T,H)\}$ the events that we get Heads in the first and second toss, respectively.
- Let $\mathcal{P}$ be the set of probability measures on Ω which assign $P(H_1) = P(H_2) = 1/2$ and have an arbitrary degree of dependence between tosses.
- Let P_* be the lower envelope of $\mathcal{P}$.

A counterexample – continued

- It is clear that $P_*(H_1) = 1/2$, $P_*(H_2) = 1/2$ and $P_*(H_1 \cap H_2) = 0$ (as the occurrence Heads in the first toss may never lead to getting Heads in the second toss).
- Now, in the case of complete positive dependence, $P(H_1 \cup H_2) = P(H_1) = 1/2$, hence $P_*(H_1 \cup H_2) \leq 1/2$.
- We thus have

$$P_*(H_1 \cup H_2) < P_*(H_1) + P_*(H_2) - P_*(H_1 \cap H_2),$$

which violates the complete monotonicity condition for $k = 2$.

Two different theories

- Mathematically, the notion of coherent lower probability is thus more general than that of belief function.
- However, the definition of the credal set associated with a belief function is purely formal, as these probabilities have no particular interpretation in our framework.
- The theory of belief functions is not a theory of imprecise probabilities.

Outline

1. Representation of evidence
   - Mass functions
   - Belief and plausibility functions
2. Relations with alternative theories
   - Possibility theory
   - Imprecise probabilities
3. Combination of evidence
   - Dempster's rule
   - Disjunctive rule
   - Dubois-Prade rule

Outline

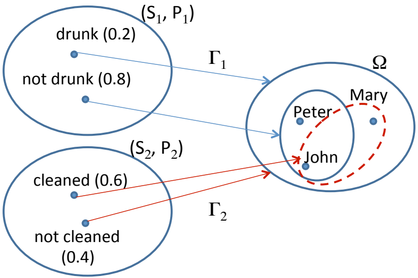
Dempster's rule

Rain example continued

- The first item of evidence gave us: $m_1(\{\text{rain}\}) = 0.8$, $m_1(\Omega) = 0.2$
- New piece of evidence: upon arriving in the train station, you watch the weather bulletin on TV, saying that it will not rain tomorrow, and the forecast has 60 % reliability.
- This second piece if evidence can be represented by the mass function: $m_2(\{\neg\text{rain}\}) = 0.6$, $m_2(\Omega) = 0.4$
- How to combine these two pieces of evidence?

Dempster's rule

Justification



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

	reliable (0.6)	not reliable (0.4)
today (0.8)	$\emptyset$, 0.48	$\{\text{rain}\}$, 0.32
not today (0.2)	$\{\neg\text{rain}\}$, 0.12	Ω, 0.08

We then get the following combined mass function,

$$\begin{aligned} m(\{\text{rain}\}) &= 0.32/0.52 \approx 0.62 \\ m(\{\neg\text{rain}\}) &= 0.12/0.52 \approx 0.23 \\ m(\Omega) &= 0.08/0.52 \approx 0.15 \end{aligned}$$

Dempster's rule

Definition

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

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

$$(m_1 \oplus m_2)(\{a\}) = (0.02 + 0.03 + 0.12)/0.6 = 0.17/0.6$$
$$(m_1 \oplus m_2)(\{b\}) = 0.2/0.6$$
$$(m_1 \oplus m_2)(\{a,b\}) = 0.08/0.6$$
$$(m_1 \oplus m_2)(\{c\}) = 0.15/0.6.$$

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

Recovering Jeffrey's conditioning

- Jeffrey's conditioning is also recovered as a special case where:
  - We start with an additive probability measure P;
  - Given a partition $E_1, \ldots, E_n$, we receive new evidence, represented by a belief function Bel whose focal sets are all unions of E_i's.
- The combined belief function $P \oplus Bel$ is Bayesian. Let $q_i = (P \oplus Bel)(E_i)$.
- Then, for all $A \subseteq \Omega$,

$$(P \oplus Bel)(A) = \sum_{i=1}^{n} q_i P(A|E_i)$$

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.

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

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

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

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

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

Outline

Disjunctive rule

Definition and justification

- Let (S_1, P_1, Γ_1) and (S_2, P_2, Γ_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:

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

Disjunctive rule

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

The resulting mass function is

$$
\begin{aligned}
(m_1 \cup m_2)(\{a,b\}) &= 0.05 + 0.2 + 0.02 + 0.08 = 0.35 \\
(m_1 \cup m_2)(\{b,c\}) &= 0.25 \\
(m_1 \cup m_2)(\{a,c\}) &= 0.03 + 0.15 = 0.18 \\
(m_1 \cup m_2)(\Omega) &= 0.1 + 0.12 = 0.22.
\end{aligned}
$$

Disjunctive rule

Properties

- Commutativity, associativity.
- No neutral element.
- $m_?$ is an absorbing element.
- Expression using belief functions:

$$Bel_1 \cup Bel_2 = Bel_1 \cdot Bel_2$$

Outline

Definition

- In general, the disjunctive rule may be preferred in case of heavy conflict between the different pieces of evidence.
- An alternative rule, which is somehow intermediate between the disjunctive and conjunctive rules, has been proposed by Dubois and Prade (1988). It is defined as follows:

$$(m_1 \uplus m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) + \sum_{\{B \cap C = \emptyset, B \cup C = A\}} m_1(B) m_2(C),$$

for all $A \subseteq \Omega$, $A \neq \emptyset$, and $(m_1 \uplus m_2)(\emptyset) = 0$.

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

$$
\begin{aligned}
(m_1 \uplus m_2)(\{a,b\}) &= 0.05 + 0.08 = 0.13 \\
(m_1 \uplus m_2)(\{b\}) &= 0.2 \\
(m_1 \uplus m_2)(\{b,c\}) &= 0.25 \\
(m_1 \uplus m_2)(\{a\}) &= 0.02 + 0.03 + 0.12 = 0.17 \\
(m_1 \uplus m_2)(\{c\}) &= 0.15 \\
(m_1 \uplus m_2)(\Omega) &= 0.1.
\end{aligned}
$$

Properties

- The DP rule boils down to the conjunctive and disjunctive rules when, respectively, the degree of conflict is equal to zero and one.
- In other cases, it has some intermediate behavior.
- It is not associative. If several pieces of evidence are available, they should be combined at once using an obvious n-ary extension of the above formula.