Theory of Belief Functions

Chapter 1: Basic notions

Thierry Denœux

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

August 2019

Uncertainty

- Topic of this course: modeling and propagation of uncertainty.

  "There are some things that you know to be true, and others that you know to be false; yet, despite this extensive knowledge that you have, there remain many things whose truth or falsity is not known to you. We say that you are uncertain about them. You are uncertain, to varying degrees, about everything in the future; much of the past is hidden from you; and there is a lot of the present about which you do not have full information. Uncertainty is everywhere and you cannot escape from it."

  Dennis Lindley, "Understanding uncertainty".

- Classical formalisms:
  1. Probabilities
  2. Sets

Inability of probabilities to represent ignorance

- 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

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

Inability of probabilities to represent ignorance (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.

Inability of sets to express doubt

- Sometimes, uncertain information about a variable X can be expressed by a set A.
- For instance: what is the temperature in this room?
- In particular, interval analysis uses intervals, which can be propagated in equations.
- For instance, if $X \in [\underline{x}, \overline{x}]$ and $Y \in [\underline{y}, \overline{y}]$, then

$$X + Y \in [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$$

- The meaning of $X \in [\underline{x}, \overline{x}]$ is that X is guaranteed to belong to this interval.
- As we cannot express doubt, we need to be very conservative and select wide intervals. These intervals grow even bigger when they are propagated in equations, making the final results sometimes useless.

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), etc.
- Recently, there has been a revived interested in its application to Statistical Inference and Computational Statistics (classification, clustering).

Outline

Outline

Mass function

Definition

- Let X be a variable taking one and only one value in a finite set Ω, called the frame of discernment
- Evidence (uncertain information) about X can be represented by a mass function $m : 2^{\Omega} \rightarrow [0, 1]$ such that

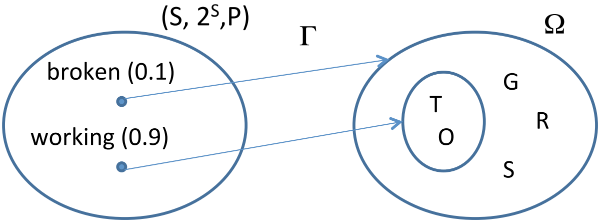
$$\sum_{A \subseteq \Omega} m(A) = 1$$

- Every subset 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 throughout this course, unless otherwise specified.

Example

- Consider the road scene analysis example. Let X be the type of object in some region of 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?

Analysis



- 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

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

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

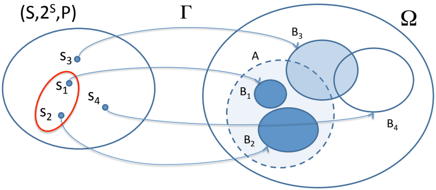
Special cases

- 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 all focal sets of m 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 the evidence tells us that the truth is in B, and $B \subseteq A$, we say that the evidence supports A.



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

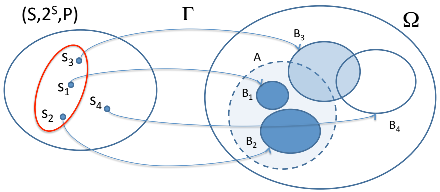
$$Bel(A) = \sum_{B \subseteq A} m(B)$$

- The number $Bel(A)$ is called the credibility of A, or the degree of belief in A, and the function $A \rightarrow Bel(A)$ is called a belief function.

- Property: $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$.

Plausibility function

- If the evidence does not support $\overline{A}$, it is consistent with A.



- The probability that the evidence is consistent with B is thus

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

- The number $Pl(A)$ is called the plausibility of A, and the function $A \rightarrow Pl(A)$ is called a plausibility function.

- Property: $Pl(\emptyset) = 0$, $Pl(\Omega) = 1$.

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 $m = m_?$ 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 / 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 $Pl(\{T\} \cup \overline{\{T\}}) = Pl(\Omega) = 1$.

- 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 $Pl : 2^\Omega \to [0,1]$ 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,

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

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

Consonant mass 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 $A, B \subseteq \Omega$,

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

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

- Pl is said to be a possibility measure, and Bel is the dual necessity measure

Proof

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

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

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

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

Contour function

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

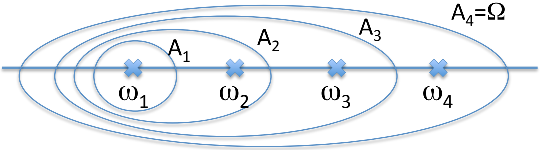
$$pl(\omega) = Pl(\{\omega\}), \quad \forall \omega \in \Omega$$

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

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

$$Bel(A) = 1 - Pl(\overline{A}) = 1 - \max_{\omega \notin A} pl(\omega) = \min_{\omega \notin A}(1 - pl(\omega)).$$

From the contour function to the mass function



- Let pl be a contour function 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

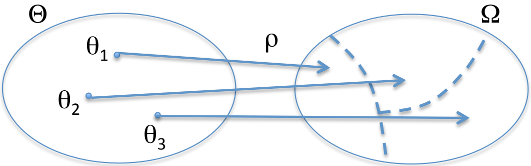
Example

- Let us come back to the road scene analysis example, with $\Omega = \{G, R, T, O, S\}$.
- Assume that we have a vegetation detector, which can determine if a region of the image contains vegetation or not. For this detector, the frame of discernment is $\Theta = \{V, \neg V\}$, where V means that there is vegetation, and $\neg V$ means that there is no vegetation.
- We have the correspondence

$$\begin{array}{rcl} V & \rightarrow & \{G, T\} \\ \neg V & \rightarrow & \{R, O, S\} \end{array}$$

- The elements of Ω can be obtained by splitting some or all of the elements of Θ. We say that Ω is a refinement of Θ, and Θ is a coarsening of Ω

General definition



Definition

A frame Ω is a refinement of a frame Θ iff there is a mapping $\rho : 2^\Theta \to 2^\Omega$ (called a refining) such that:

- $\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega$ is a partition of Ω, and
- For all $A \subseteq \Omega$, $\rho(A) = \bigcup_{\theta \in A} \rho(\{\theta\})$.

Vacuous extension

- In the road scene example, assume that the vegetation detector provide the following mass function on Θ:

$$m^{\Theta}(\{V\}) = 0.6, \quad m^{\Theta}(\{\neg V\}) = 0.3, \quad m^{\Theta}(\Theta) = 0.1$$

- How to express m^{Θ} in Ω?
- Solution: for all $A \subseteq \Theta$, we transfer the mass $m^{\Theta}(A)$ to $\rho(A)$. Here,

$$\begin{aligned} m^{\Theta}(\{V\}) = 0.6 \quad &\rightarrow \quad \rho(\{V\}) = \{G, T\} \\ m^{\Theta}(\{\neg V\}) = 0.3 \quad &\rightarrow \quad \rho(\{\neg V\}) = \{R, O, S\} \\ m^{\Theta}(\Theta) = 0.1 \quad &\rightarrow \quad \rho(\Theta) = \Omega \end{aligned}$$

- We finally the following mass function on Ω,

$$m^{\Theta\uparrow\Omega}(\{G, T\}) = 0.6, \quad m^{\Theta\uparrow\Omega}(\{R, O, S\}) = 0.3, \quad m^{\Theta\uparrow\Omega}(\Omega) = 0.1.$$

- $m^{\Theta\uparrow\Omega}$ is called the vacuous extension of m^{Θ} in Ω.

Expression of information in a coarser frame

- Let us now assume that we have the following mass function on Ω,

$$m^\Omega(\{T\}) = 0.4, \quad m^\Omega(\{T, O\}) = 0.3, \quad m^\Omega(\{R, S\}) = 0.3.$$

- How to express m^Ω in Θ?
- We cannot do it without loss of information, because, for instance, there is no $A \subseteq \Theta$ such that $\rho(A) = \{T\}$: the mapping ρ does not have an inverse.

Inner and outer reductions

- We can define two generalized inverses of ρ:

$$\underline{\rho}^{-1}(B) = \{\theta \in \Theta | \rho(\{\theta\}) \subseteq B\},$$
$$\overline{\rho}^{-1}(B) = \{\theta \in \Theta | \rho(\{\theta\}) \cap B \neq \emptyset\},$$

  for any subset B of Ω. The subsets $\underline{\rho}^{-1}(B)$ and $\overline{\rho}^{-1}(B)$ are called, respectively, the inner reduction and the outer reduction of B.
- Here, for instance,

$$\underline{\rho}^{-1}(\{T\}) = \underline{\rho}^{-1}(\{T, O\}) = \underline{\rho}^{-1}(\{R, S\}) = \emptyset$$

$$\overline{\rho}^{-1}(\{T\}) = \{V\}, \quad \overline{\rho}^{-1}(\{T, O\}) = \{V, \neg V\}, \quad \overline{\rho}^{-1}(\{R, S\}) = \{\neg V\}$$

Restriction

- The restriction of m^Ω in Θ is obtained by transferring each mass $m^\Omega(B)$ to the outer reduction of B: for all subset A of Θ,

$$m^{\Omega\downarrow\Theta}(A) = \sum_{\overline{\rho}^{-1}(B)=A} m^\Omega(B).$$

- In the example, we thus have

$$m^{\Omega\downarrow\Theta}(\{V\}) = 0.4, \quad m^{\Omega\downarrow\Theta}(\Theta) = 0.3, \quad m^{\Omega\downarrow\Theta}(\{\neg V\}) = 0.3.$$

- Remark: the vacuous extension of $m^{\Omega\downarrow\Theta}$ is

$$m^{(\Omega\downarrow\Theta)\uparrow\Omega}(\{G,T\}) = 0.4, \quad m^{(\Omega\downarrow\Theta)\uparrow\Omega}(\Omega) = 0.3,$$

$$m^{(\Omega\downarrow\Theta)\uparrow\Omega}(\{R,S,O\}) = 0.3.$$

It is less precise that m^Ω: we have lost information when expressing m^Ω in a coarser frame.

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:

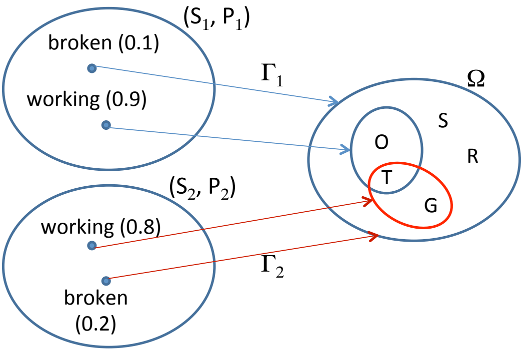
$$m_1(\{T, O\}) = 0.9, \quad m_1(\Omega) = 0.1$$

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

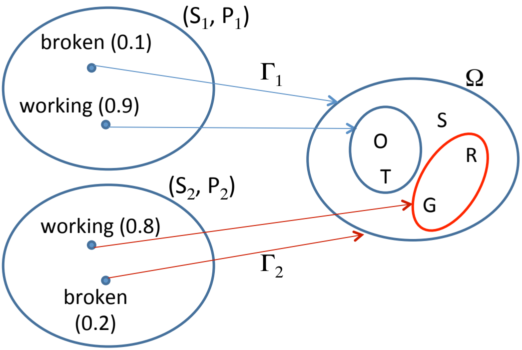
Computation

$m_1 \backslash m_2$	$\{T, G\}$ (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

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

  and $(m_1 \oplus m_2)(\emptyset) = 0$
- $m_1 \oplus m_2$ is called the orthogonal sum of m_1 and m_2

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).

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:

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

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

$$m(A \mid B) = (1-\kappa)^{-1} \sum_{C \cap B = A} m(C), \tag{2}$$

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

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

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

- 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 $Q : 2^\Omega \to [0,1]$ be defined as

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

- Conversely,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} Q(B) \tag{3}$$

- 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,

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

Product rule for commonality and contour functions

- Using (3) with $A = \emptyset$, we get

$$\sum_{B \subseteq \Omega} (-1)^{|B|} Q(B) = 0, \tag{4}$$

  which makes it possible to compute the commonality function once commonality numbers are determined up to some multiplicative constant.
- Given two mass functions m_1 and m_2, we can thus combine them either using (1), or by converting them to commonality functions, multiplying them pointwise, and computing the corresponding mass function using (3).
- In particular, $pl(\omega) = Q(\{\omega\})$. Consequently,

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

Computational complexity

- The orthogonal sum of two mass functions m_1 and m_2 can be performed in two ways: either directly using (1) (mass-based approach), or by computing the product of commonalities.
- Using the mass-based approach, the time needed to compute the combination is proportional to $|\mathcal{F}(m_1)||\mathcal{F}(m_2)||\Omega|$, where $\mathcal{F}(m_i) \subseteq 2^\Omega$ is the collection of focal sets of m_i. In the worst case where both mass functions have $2^{|\Omega|} - 1$ focal sets, the computing time thus becomes proportional to $2^{2|\Omega|}|\Omega|$.
- The other approach implies converting the mass functions into commonalities, multiplying the commonalities pointwise, normalizing, and computing the combined mass function using (3). The conversion from one of the equivalent functions m, Bel, Pl and Q to another can be performed in time proportional to $|\Omega|^2 2^{|\Omega|}$.
- The mass-based approach is thus more efficient when the number of focal sets is much smaller than the cardinality of 2^Ω.

Summarization

- If computing time is limited, we may use approximations. Useful strategy: approximate each mass function by a simpler mass function with fewer focal sets.
- The simplest method is the Summarization algorithm, which works as follows.
- Let k be the maximum allowed number k of focal sets. Let $F_1, \ldots, F_n$ be the focal sets of m ranked by decreasing mass, i.e., $m(F_1) \geq m(F_2) \geq \ldots \geq m(F_n)$. If $n > k$, the $n - k$ focal sets $F_{k+1}, \ldots, F_n$ are replaced by their union, and m is approximated by

$$m'(F_i) = m(F_i), \quad i = 1, \ldots, k,$$

$$m'\left(\bigcup_{i=k+1}^{n} F_i\right) = \sum_{i=k+1}^{n} m(F_i).$$

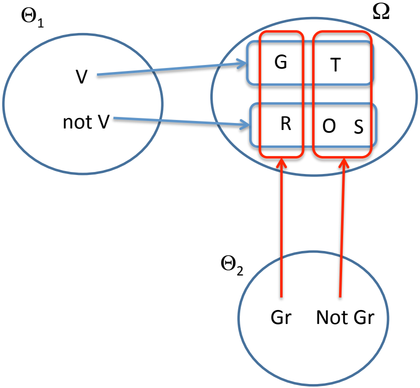
Outline

Compatible frames of discernment

Definition

Two frames are compatible if they have a common refinement.

Example:



Combination of mass functions on compatible frames

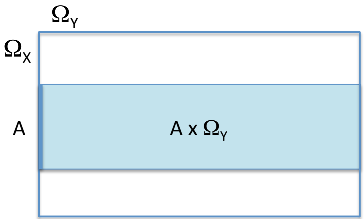
- Let m^{Θ_1} and m^{Θ_2} be two mass functions defined on compatible frames Θ_1 and Θ_2 with common refinement Ω.
- The orthogonal sum of m^{Θ_1} and m^{Θ_2} in Ω is

$$m^{\Theta_1} \oplus m^{\Theta_2} = m^{\Theta_1 \uparrow \Omega} \oplus m^{\Theta_2 \uparrow \Omega}$$

- Example: assume that $m^{\Theta_1}(\{V\}) = 0.3$, $m^{\Theta_1}(\{\neg V\}) = 0.5$, $m^{\Theta_1}(\{V, \neg V\}) = 0.2$, and $m^{\Theta_2}(\{Gr\}) = 0.4$, $m^{\Theta_2}(\{\neg Gr\}) = 0.5$, $m^{\Theta_2}(\{Gr, \neg Gr\}) = 0.1$. Compute $m^{\Theta_1} \oplus m^{\Theta_2}$.

Cylindrical extension

- Let us now assume that we have two frames Ω_X and Ω_Y related to two different questions about, e.g., the values of two unknown variables X and Y.
- Let $\Omega_{XY} = \Omega_X \times \Omega_Y$ be the product space. It is a refinement of both Ω_X and Ω_Y.

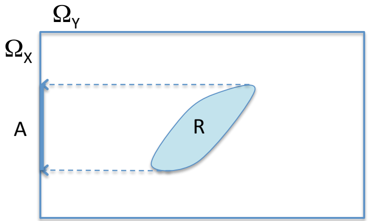


- We can define the following refining ρ from 2^{Ω_X} to $2^{\Omega_X \times \Omega_Y}$:

$$\rho(A) = A \times \Omega_Y,$$

for all $A \subseteq \Omega_X$. The set $\rho(A)$ is called the cylindrical extension of A in Ω_{XY} and is denoted by $A \uparrow \Omega_{XY}$.

Projection



- Conversely, let R be a subset of Ω_{XY}.
- Its outer reduction is

$$\begin{aligned}\overline{\rho}^{-1}(R) &= \{x \in \Omega_X | \rho(\{x\}) \cap R \neq \emptyset\} \\ &= \{x \in \Omega_X | \exists y \in \Omega_Y, (x,y) \in R\}.\end{aligned}$$

- This set is denoted by $R \downarrow \Omega_X$ and is called the projection of R on Ω_X

Vacuous extension and marginalization

- The vacuous extension of a mass function m^X from Ω_X to Ω_{XY} is obtained by transferring each mass $m^{\Omega_X}(B)$ for any subset B of Ω_X to the cylindrical extension of B:

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

- Conversely, the restriction of a joint mass function m^{XY} on Ω_X is

$$m^{XY\downarrow X}(A) = \sum_{B\downarrow\Omega_X = A} m^{XY}(B),$$

for all $A \subseteq \Omega_X$. The mass functions $m^{XY\downarrow X}$ and $m^{XY\downarrow Y}$ are called the marginals of m^{XY} and the operation that computes the marginals from a joint mass function is called marginalization. This operation extends both set projection and probabilistic marginalization.

Application to approximate reasoning

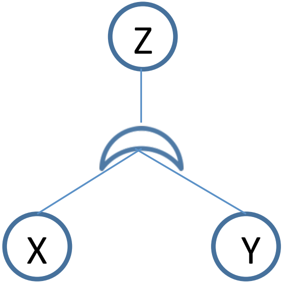
- Assume that we have:
  - Partial knowledge of X formalized as a mass function m^X
  - A joint mass function m^{XY} representing an uncertain relation between X and Y
- What can we say about Y?
- Solution:

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

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

Example

- A machine fails if any one of two components fails.
- Let Z, X and Y be the binary variables describing the states of the two components, and the machine.



- We have the following prior knowledge about the states of the components:

$$m^X(\{1\}) = 0.1, m^X(\{0\}) = 0.3,$$

$$m^X(\{0,1\}) = 0.6$$

$$m^Y(\{0,1\}) = 1$$

- We observe that the machine fails. What are our beliefs about the states of the two components?

Solution

- Pieces of evidence:

$$m_0^{XYZ}(\{(1,1,1),(1,0,1),(0,1,1),(0,0,0)\}) = 1$$

$$m^{X\uparrow XYZ}(\{1\}\times\Omega_{YZ}) = 0.1,\ m^{X\uparrow XYZ}(\{0\}\times\Omega_{YZ}) = 0.3,\ m^{X\uparrow XYZ}(\Omega_{XYZ}) = 0.6$$
$$m^{Y\uparrow XYZ}(\Omega_{XYZ}) = 1, \quad m^{Z\uparrow XYZ}(\Omega_{XY} \times \{1\}) = 1$$

- Let $m_1^{XYZ} = m_0^{XYZ} \oplus m^{X\uparrow XYZ} \oplus m^{Z\uparrow XYZ}$. We have

$$m_1^{XYZ}(\{(1,1,1),(1,0,1)\}) = 0.1,\ m_1^{XYZ}(\{(0,1,1)\}) = 0.3,$$

$$m_1^{XYZ}(\{(1,1,1),(1,0,1),(0,1,1)\}) = 0.6$$

- Marginalizing on X and Y, we get

$$m_1^{XYZ\downarrow X}(\{1\}) = 0.1, m_1^{XYZ\downarrow X}(\{0\}) = 0.3, m_1^{XYZ\downarrow X}(\{0,1\}) = 0.6$$

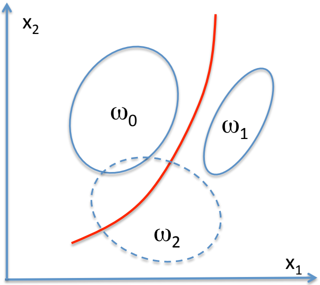
$$m_1^{XYZ\downarrow Y}(\{1\}) = 0.3, m_1^{XYZ\downarrow Y}(\{0,1\}) = 0.7$$

Outline

Outline

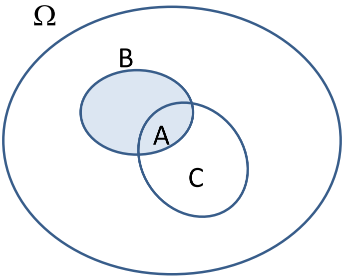
Example

- Assume a machine diagnosis system was built from a classifier trained with examples from two states: ω_0 (normal state), ω_1 (fault).
- In operating condition, we discover a new fault ω_2.



- The mass function m_0 computed by the classifier is conditioned on $B = \{\omega_0, \omega_1\}$.
- After the frame has been extended to include ω_2, m must be deconditioned to the frame $\Omega = \{\omega_0, \omega_1, \omega_2\}$.

Deconditioning



- Let m_0 be a mass function on Ω expressing our beliefs about X in a context where we know that $X \in B$
- We want to build a mass function m verifying the constraint

$$m(\cdot \mid B) = m \oplus m_B = m_0$$

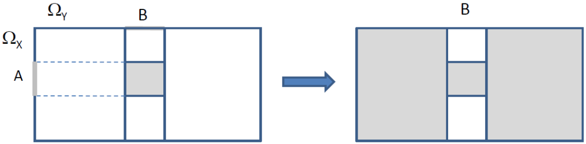
- Any m built from m_0 by transferring each mass $m_0(A)$ to $A \cup C$ for some $C \subseteq \overline{B}$ satisfies the constraint

- Conservative attitude: transfer $m_0(A)$ to the largest such set, which is $A \cup \overline{B}$

$$m(D) = \begin{cases} m_0(A) & \text{if } D = A \cup \overline{B} \text{ for some } A \subseteq B \\ 0 & \text{otherwise} \end{cases}$$

Deconditioning

Conditional embedding



- More complex situation: two frames Ω_X and Ω_Y
- Let m_0^X be a mass function on Ω_X expressing our beliefs about X in a context where we know that $Y \in B$ for some $B \subseteq \Omega_Y$
- We want to find m^{XY} such that $\left(m^{XY} \oplus m_B^Y\right)^{\downarrow X} = m_0^X$
- Least committed solution: transfer $m_0^X(A)$ to

$$(A \times B) \cup (\Omega_X \times \overline{B}) = (A \times \Omega_Y) \cup (\Omega_X \times \overline{B})$$

- Consistent with logical implication: $(B \rightarrow A) \equiv (\neg B \vee A)$.
- Notation $m^{XY} = (m_0^X)^{\Uparrow XY}$.

Outline

Discounting

Problem statement

- A source of information provides:
  - a value
  - a set of values
  - a probability distribution, etc.
- The information is:
  - not fully reliable or
  - not fully relevant.
- Examples:
  - Possibly faulty sensor
  - Measurement performed in unfavorable experimental conditions
  - Information is related to a situation or an object that only has some similarity with the situation or the object considered (case-based reasoning).

Discounting

Formalization

- A source S provides a mass function m_0^Ω.
- S may be reliable or not. Let $\mathcal{R} = \{R, NR\}$.
- Assumptions:
  1. If S is reliable, we accept m_0^Ω as a representation of our beliefs:
  $$m^\Omega(\cdot \mid R) = m_0^\Omega$$
  2. If S is not reliable, we know nothing:
  $$m^\Omega(\cdot \mid NR) = m_?^\Omega$$
  3. The source has a probability α of not being reliable:
  $$m^{\mathcal{R}}(\{NR\}) = \alpha, \quad m^{\mathcal{R}}(\{R\}) = 1 - \alpha$$

Discounting

Solution

- We compute:

$$
{}^{\alpha}m^{\Omega} = \left(\underbrace{m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}}_{m_1} \oplus \underbrace{m^{\Omega}(\cdot \mid R)^{\Uparrow\Omega\times\mathcal{R}}}_{m_2} \right)^{\downarrow\Omega}
$$

- We have

$$
m_1(\Omega \times \{NR\}) = \alpha, \quad m_1(\Omega \times \{R\}) = 1 - \alpha
$$

$$
m_2\left((A \times \{R\}) \cup (\Omega \times \{NR\})\right) = m_0(A), \quad \forall A \subseteq \Omega
$$

- Let $m_{12} = m_1 \oplus m_2$. We have

$$
m_{12}(A \times \{R\}) = (1 - \alpha)m_0(A), \quad \forall A \subseteq \Omega
$$

$$
m_{12}(\Omega \times \{NR\}) = \alpha
$$

Discounting

Solution (continued)

- Marginalizing on Ω, we get

$$ {}^{\alpha}m^{\Omega}(A) = \begin{cases} (1-\alpha)m_0^{\Omega}(A) & \text{if } A \subset \Omega \\ (1-\alpha)m_0^{\Omega}(\Omega) + \alpha & \text{if } A = \Omega. \end{cases} $$

- Equivalent expression

$$ {}^{\alpha}m^{\Omega} = (1-\alpha)m_0^{\Omega} + \alpha m_?^{\Omega} $$

- α is called the discount rate. It is the probability that the source is not reliable.