Dempster-Shafer Theory

Jonathan Lawry

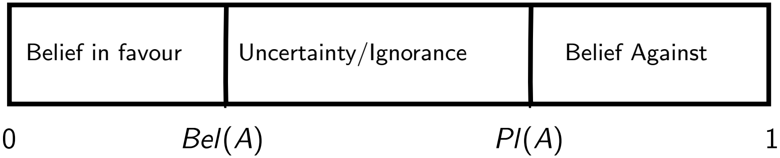
Department of Engineering Mathematics
University of Bristol

Introduction

- In DS theory uncertainty is quantified by using two uncertainty measures.
- Belief $Bel : 2^W \rightarrow [0,1]$ is a measure of the total amount of evidence in favour of a proposition.
- Plausibility $Pl : 2^W \rightarrow [0,1]$ is a measure of the total amount of evidence not against a proposition.
- $\forall A \subseteq W \;\; Bel(A) \leq Pl(A)$ and $Pl(A) - Bel(A)$ provides a direct measure of ignorance about the proposition A.

Pictorial Representation

- $Bel(A) + Bel(A^c) \leq 1$.
- $Bel(A)$ is the level of belief in favour of proposition A.
- $Pl(A) = 1 - Bel(A^c)$ is 1 minus the total belief in favour of A^c.
- $Pl(A) - Bel(A)$ is the level of ignorance concerning A.



Mass Assignments

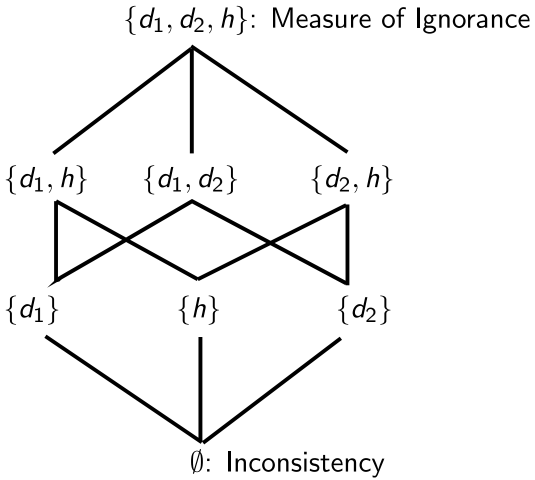
- **Mass Assignments (basic probability assignments):** Shafer-Dempster belief functions are defined by a probability distribution on the power set of W.

$$m : 2^W \to [0,1],\ m(\emptyset) = 0,\ \sum_{A \subseteq W} m(A) = 1$$

- $m(A)$ quantifies the total belief exactly for A i.e. For A but not for any $B \subset A$.
- **Belief:** $Bel(A) = \sum_{B \subseteq A} m(B)$
- **Plausibility:** $Pl(A) = \sum_{B : B \cap A \neq \emptyset} m(B)$

Example

- It is know that a patient is either healthy (h) or has one of two diseases d_1 or d_2. So $W = \{d_1, d_2, h\}$.



Example: 2

- What should be our belief that our patient is suffering from either d_1 or d_2 (i.e. that he is not healthy)?
- $Bel(\{d_1, d_2\})$ = sum of evidence that directly supports the patient not being healthy.
- $= m(\{d_1, d_2\}) + m(\{d_2\}) + m(\{d_1\})$
- What should be the plausibility of this proposition?
- $Pl(\{d_1, d_2\})$ = sum of the evidence which is consistent with the patient not being healthy.
- $= m(\{d_1, d_2\}) + m(\{d_2\}) + m(\{d_1\}) + m(\{d_1, h\}) + m(\{d_2, h\}) + m(\{d_1, d_2, h\})$

Generalising Probability

- If m is a mass assignment such that for $W = \{w_1, \ldots, w_n\}$

$$\sum_{i=1}^{n} m(\{w_i\}) = 1$$

- Then $Bel(A) = Pl(A) = P(A)$ where

$$P(A) = \sum_{w_i \in A} m(\{w_i\})$$

- In other words, $m(\{w_i\})$ is a probability distribution.

Axioms for DS-Belief

- $Bel(W) = 1$ and $Bel(\emptyset) = 0$
- $Bel(\bigcup_{i=1}^{n} A_i) \geq \sum_{\emptyset \neq S \subseteq \{A_1,\ldots,A_n\}} (-1)^{|S|-1} Bel(\bigcap_{A_i \in S} A_i)$
- This is a general version of super-additivity i.e. for $n = 2$ $Bel(A_1 \cup A_2) \geq Bel(A_1) + Bel(A_2) - Bel(A_1 \cap A_2)$
- Similarly, $Pl(W) = 1$ and $Pl(\emptyset) = 0$
- $Pl(\bigcup_{i=1}^{n} A_i) \leq \sum_{\emptyset \neq S \subseteq \{A_1,\ldots,A_n\}} (-1)^{|S|-1} Pl(\bigcap_{A_i \in S} A_i)$
- For $n = 2$ $Pl(A_1 \cup A_2) \leq Pl(A_1) + Pl(A_2) - Pl(A_1 \cap A_2)$

Belief Inversion

- We have seen how belief and plausibility is obtained from the mass assignment.
- However, it is also possible to determine the mass assignment from the belief function or from the plausibility function.
- From belief the inversion formula is:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B)$$

- From plausibiliy the inversion formula is:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} (1 - Pl(B^c))$$

Inversion Example

- Let $W = \{w_1, w_2, w_3\}$ and Bel be defined by: $Bel(\{w_1, w_2, w_3\}) = 1$, $Bel(\{w_1, w_2\}) = 0.5$, $Bel(\{w_1, w_3\}) = 0.2$, $Bel(\{w_2, w_3\}) = 0.3$, $Bel(\{w_1\}) = 0$, $Bel(\{w_2\}) = 0.3$, $Bel(\{w_3\}) = 0$
- Sets with 1 element: $m(\{w_1\}) = Bel(\{w_1\}) = 0$, $m(\{w_2\}) = Bel(\{w_2\}) = 0.3$, $m(\{w_3\}) = Bel(\{w_3\}) = 0$
- Sets with 2 elements: $m(\{w_2, w_3\}) = Bel(\{w_2, w_3\}) - m(\{w_2\}) - m(\{w_3\}) = 0.3 - 0.3 - 0 = 0$, similarly $m(\{w_1, w_3\}) = 0.2$ and $m(\{w_1, w_2\}) = 0.2$
- Sets with 3 elements: $m(\{w_1, w_2, w_3\}) = Bel(\{w_1, w_2, w_3\}) - m(\{w_1, w_2\}) - m(\{w_1, w_3\}) - m(\{w_2, w_3\}) - m(\{w_1\}) - m(\{w_2\}) - m(\{w_3\}) = 1 - 0 - 0.3 - 0 - 0 - 0.2 - 0.2 = 0.3$

Lower and Upper Probabilities

- Instead of defining a single belief measure the agent defines two measures, a lower probability $\underline{P} : 2^W \rightarrow [0, 1]$ and an upper probability $\overline{P} : 2^W \rightarrow [0, 1]$ where $\forall A \subseteq W$ $\underline{P}(A) \leq \overline{P}(A)$
- The measures $\underline{P}$ and $\overline{P}$ are *coherent* if $\forall A \subseteq W$ $\overline{P}(A^c) = 1 - \underline{P}(A)$
- One way of defining lower and upper probabilities is as *upper and lower envelopes*.
- Let $\mathcal{P}$ be a set of probability measures on 2^W. Then define:

$$\underline{P}(A) = \inf\{P(A) : P \in \mathcal{P}\} \text{ and } \underline{P}(A) = \sup\{P(A) : P \in \mathcal{P}\}$$

Example

- Let $W = \{w_1, w_2, w_3\}$ and let $\mathcal{P}$ be the following set of probability distributions: $\mathcal{P} = \{P_1, P_2, P_3\}$ defined according to the following probability distributions respectively:
- P_1: $p_1(w_1) = 0.2, p_1(w_2) = 0.3, p_1(w_3) = 0.5$,
  P_2: $p_2(w_1) = 0.8, p_2(w_2) = 0.1, p_2(w_3) = 0.1$,
  P_3: $p_3(w_1) = 0.15, p_3(w_2) = 0.6, p_3(w_3) = 0.25$
- $\underline{P}(\{w_1, w_3\}) =$
  $\min\{P_1(\{w_1, w_3\}), P_2(\{w_1, w_3\}), P_3(\{w_1, w_3\})\} = 0.4$
- $\overline{P}(\{w_1, w_3\}) =$
  $\max\{P_1(\{w_1, w_3\}), P_2(\{w_1, w_3\}), P_3(\{w_1, w_3\})\} = 0.9$

DS as Lower and Upper Probabilities

- *Bel* is a lower probability measure and *Pl* is an upper probability measure.
- For any belief and plausibility measures *Bel* and *Pl* there is a set of probability measures on 2^W $\mathcal{P}$ such that for all $A \subseteq W$

$$
\begin{aligned}
Bel(A) &= \inf\{P(A) : P \in \mathcal{P}\} \\
Pl(A) &= \sup\{P(A) : P \in \mathcal{P}\}
\end{aligned}
$$

- Where we define

$$\mathcal{P} = \{P : \forall A \subseteq W \;\; P(A) \geq Bel(A)\}$$

DS as Lower and Upper Probabilities

- We show that $\forall P \in \mathcal{P}$, $\forall A \subseteq W$ $Bel(A) \leq P(A) \leq Pl(A)$ and
- For all $A \subseteq W$ there exists $P \in \mathcal{P}$ such that $P(A) = Bel(A)$.
- Notice that since $\forall P \in \mathcal{P}$ $Bel(A^c) \leq P(A^c)$ we have that $1 - P(A^c) \leq 1 - Bel(A^c)$ and hence $P(A) \leq Pl(A)$.
- The second part is a bit harder...
- For every $B \subseteq W$ pick a single $w_B \in B$ such that if $B \not\subseteq A$ then $w_B \in A^c$.
- Define the probability distribution p such that $p(w) = \sum_{B : w_B = w} m(B)$

Updating Probabilities

- Suppose we have a prior probability measure P as characterised by probability distribution p.
- Further suppose that our evidence about the true state of the world is given by mass assignment m.
- In other words, its as if we are not certain what set we could be conditioning on but rather that we should give the answer $P(\bullet|B)$ with probability $m(B)$.
- To get a single probability measure we could then take the expected value so that:

$$P(\bullet|m) = \sum_{B \subseteq W} P(\bullet|B)m(B)$$

Updating Probabilities

- For any probability measure P on 2^W such that $P(B) > 0$ if $m(B) > 0$, $P(\bullet|m) \in \mathcal{P}$

$$\forall A \subseteq W \;\; P(A|m) = \sum_{B \subseteq W} P(A|B)m(B)$$

$$= \sum_{B \subseteq A} P(A|B)m(B) + \sum_{B \not\subseteq A} P(A|B)m(B)$$

$$= \sum_{B \subseteq A} m(B) + \sum_{B \not\subseteq A} P(A|B)m(B) \geq \sum_{B \subseteq A} m(B) = Bel(A)$$

- **Pignistic Distribution:** Take P to be the uniform measure so that

$$P_u(A|m) = \sum_{B \subseteq W} \frac{|A \cap B|}{|B|} m(B)$$

Pignistic Distribution

- Suppose we have the following mass assignment:

$$m(\{w_1, w_2, w_3\}) = 0.3, \;\; m(\{w_1, w_2\}) = 0.2,$$
$$m(\{w_1, w_3\}) = 0.2, \;\; m(\{w_2\}) = 0.3$$

- From this the Pignistic Distribution is given by:

$$\begin{aligned} p_u(w_1|m) &= \frac{1}{3}m(\{w_1, w_2, w_3\}) + \frac{1}{2}m(\{w_1, w_2\}) + \frac{1}{2}m(\{w_1, w_3\}) \\ &= \frac{0.3}{3} + \frac{0.2}{2} + \frac{0.2}{2} = 0.3 \\ p_u(w_2|m) &= \frac{0.3}{3} + \frac{0.2}{2} + 0.3 = 0.5 \\ p_u(w_3|m) &= \frac{0.3}{3} + \frac{0.2}{2} = 0.2 \end{aligned}$$

Lower Probabilities not Always Beliefs

- Consider the following lower probability measure:

$$\underline{P}(\{w_1\}) = 0,\ \underline{P}(\{w_2\}) = 0,\ \underline{P}(\{w_3\}) = 0$$
$$\underline{P}(\{w_1, w_2\}) = 0.4,\ \underline{P}(\{w_1, w_3\}) = 0.4,\ \underline{P}(\{w_2, w_3\}) = 0.4$$
$$\underline{P}(\{w_1, w_2, w_3\}) = 1$$

- Suppose that $\underline{P} = Bel$ for some belief function. Then we should be able to use inversion to find m. This gives:

$$m(\{w_1\}) = m(\{w_2\}) = m(\{w_3\}) = 0$$
$$m(\{w_1, w_2\}) = m(\{w_1, w_3\}) = m(\{w_2, w_3\}) = 0.4$$
$$m(\{w_1, w_2, w_3\}) = -0.2$$

Nested Evidence

- In some cases the underlying evidence for a DS belief function forms a nested hierarchy:
- In other words, there exists $F_1 \subseteq F_2 \subseteq \ldots \subseteq F_n \subseteq W$ such that $\sum_{i=1}^{n} m(F_i) = 1$.
- $F_1 \subseteq F_2 \subseteq \ldots \subseteq F_n \subseteq W$ are called the *focal sets* of m.
- Recall example $W = \{d_1, d_2, h\}$.
- Suppose all doctors agree that the patient may have disease d_2, a subset of doctors think that he may have disease d_1, of that subset some doctors also believe it is possible that he is healthy.
- All mass would then be allocated to the sets $\{d_2\} \subseteq \{d_1, d_2\} \subseteq \{d_1, d_2, h\}$

Possibility and Necessity Measures

- For a nested mass assignment the corresponding belief measure is called a *necessity measure* and the plausibility measure is called a *possibility measure*.
- Let F_j be the largest focal set for which $F_j \subseteq A$ then:

$$Bel(A) = Nec(A) = \sum_{B \subseteq A} m(B) = \sum_{i=1}^{j} m(F_i)$$

- Let F_k be the smallest focal set for which $A \cap F_k \neq \emptyset$ then:

$$Pl(A) = Pos(A) = \sum_{B : B \cap A \neq \emptyset} m(B) = \sum_{i=k}^{n} m(F_k)$$

Min-Max Rules

- Suppose F_k is the smallest focal set for which $A \cap F_k \neq \emptyset$.
- Suppose $F_{k'}$ is the smallest focal set for which $B \cap F_{k'} \neq \emptyset$.
- Suppose, w.l.o.g that $F_k \subseteq F_{k'}$ so that $Pos(A) \geq Pos(B)$.
- Clearly, F_k is the smallest focal set for which $(A \cup B) \cap F_k \neq \emptyset$. Therefore,

$$Pos(A \cup B) = \sum_{i=k}^{n} m(F_i) = Pos(A) = \max(Pos(A), Pos(B))$$

- Similarly, $Nec(A \cap B) = \min(Nec(A), Nec(B))$

Possibility and Necessity Axioms

- **Pos1:** $Pos(W) = 1,\ Pos(\emptyset) = 0.$
- **Pos2:** $Pos(A \cup B) = \max(Pos(A), Pos(B)).$
- **Nec1:** $Nec(W) = 1,\ Nec(\emptyset) = 0.$
- **Nec2:** $Nec(A \cap B) = \min(Nec(A), Nec(B)).$
- **Consequences:** $\forall A \subseteq W\ \ Pos(A) = 1$ or $Pos(A^c) = 1.$
- $\forall A \subseteq W\ \ Nec(A) = 0$ or $Nec(A^c) = 0$
- If $Nec(A) > 0$ then $Pos(A) = 1$
- If $Pos(A) < 1$ then $Nec(A) = 0$
- Either $[Nec(A), Pos(A)] = [x, 1]$ or $= [0, y]$ where $0 < x \leq 1$ and $0 \leq y < 1.$

Possibility Distributions

- For any $A \subseteq W$ $Pos(A) = \max\{Pos(\{w_i\}) : w_i \in A\}$
- Let $Pos(\{w_i\}) = \pi(w_i)$ then π is the possibility measure generating Pos.
- Analogy with probability distributions but with sum replaced by max.
- Also, notice that $max\{\pi(w_i) : w_i \in W\} = 1$
- Number of values required to specify a possibility measures is $n - 1$ compared to $2^n - 1$.
- Lotfi Zadeh orginally introduced possibility measures in terms of possibility distributions rather than seeing them as special cases of plausibility measures.

Hans Eats Eggs

- Consider the number of eggs that Hans could eat for breakfast. So $W = \{1, 2, 3, 4, 5, 6, 7, 8\}$.
- **Possibility:** $\pi(w)$ = the degree of ease with which Hans can eat w eggs.
- **Probability:** $p(w)$ = the probability that Hans will *actually* eat w eggs.

w	1	2	3	4	5	6	7	8
$\pi(w)$	1	1	1	1	0.8	0.6	0.4	0.2
$p(w)$	0.1	0.8	0.1	0	0	0	0	0

- Probability/possibility consistency: $p(w) \leq \pi(w)$

Possibilistic Inversion

- The mapping from *Pos* to the mass assignment m is much simpler than the general mapping from plausibility to mass assignment.
- In fact m can be determined from π as follows:
- Suppose $W = \{w_1, \ldots, w_n\}$ ordered such that $1 = \pi(w_1) \geq \pi(w_2) \geq \ldots \geq \pi(w_n)$ then:

$$
\begin{aligned}
m(\{w_1, \ldots, w_n\}) &= \pi(w_n) \\
m(\{w_1, \ldots, w_{n-1}\}) &= \pi(w_{n-1}) - \pi(w_n) \\
&\vdots \\
m(\{w_1, \ldots, w_i\}) &= \pi(w_i) - \pi(w_{i+1}) \\
&\vdots \\
m(\{w_1\}) &= 1 - \pi(w_2)
\end{aligned}
$$

Dempster's Rule of Combination

- Suppose we have two mass assignments (or belief functions) provided by two independent sources of evidence (perhaps two different agent's).
- Dempster's rule provides a method to combine these to give one single mass assignment.
- Suppose that m_1 and m_2 are two mass assignments then:

$$\forall A \subseteq W \;\; m_1 \oplus m_2(A) = \frac{\sum_{(B,C):B\cap C=A} m_1(B)m_2(C)}{1-\sum_{(B,C):B\cap C=\emptyset} m_1(B)m_2(C)}$$

Example: Dempster's Rule

- $m_1(\{w_1, w_2, w_3\}) = 0.3,\ m_1(\{w_1, w_2\}) = 0.2,$
  $m_1(\{w_1, w_3\}) = 0.2,\ m_1(\{w_2\}) = 0.3.$
- $m_2(\{w_1, w_2, w_3\}) = 0.4,\ m_2(\{w_1, w_2\}) = 0.1,$
  $m_2(\{w_2, w_3\}) = 0.25,\ m_2(\{w_3\}) = 0.25.$

$m_1 \oplus m_2$	$\{w_1, w_2, w_3\}$ 0.3	$\{w_1, w_2\}$ 0.2	$\{w_1, w_3\}$ 0.2	$\{w_2\}$ 0.3
$\{w_1, w_2, w_3\}$ 0.4	$\{w_1, w_2, w_3\}$ 0.12	$\{w_1, w_2\}$ 0.08	$\{w_1, w_3\}$ 0.08	$\{w_2\}$ 0.12
$\{w_1, w_2\}$ 0.1	$\{w_1, w_2\}$ 0.03	$\{w_1, w_2\}$ 0.02	$\{w_1\}$ 0.02	$\{w_2\}$ 0.03
$\{w_2, w_3\}$ 0.25	$\{w_2, w_3\}$ 0.075	$\{w_2\}$ 0.05	$\{w_3\}$ 0.05	$\{w_2\}$ 0.075
$\{w_3\}$ 0.25	$\{w_3\}$ 0.075	$\emptyset$ 0.05	$\{w_3\}$ 0.05	$\emptyset$ 0.075

Example: Dempster's Rule

- $1 - \sum_{(B,C):B \cap C = \emptyset} m_1(B) m_2(C) =$
  $1 - m_1(\{w_1, w_2\}) m_2(\{w_3\}) - m_1(\{w_2\}) m_2(\{w_3\}) =$
  $1 - 0.05 - 0.075 = 0.875.$

$$
\begin{gathered}
m_1 \oplus m_2(\{w_1, w_2\}) = \\
\frac{1}{0.875}(m_1(\{w_1, w_2\}) m_2(\{w_1, w_2, w_3\}) + \\
m_1(\{w_1, w_2, w_3\}) m_2(\{w_1, w_2\}) + m_1(\{w_1, w_2\}) m_2(\{w_1, w_2\})) \\
= \frac{1}{0.875}(0.08 + 0.03 + 0.02) = \frac{0.13}{0.875} = 0.1486
\end{gathered}
$$

Zadeh's Criticism

- Suppose a patient has one of three possible diseases: meningitis, concussion, tumor so that $W = \{m, c, t\}$
- Two independent doctors give the following diagnosis:
- **Doctor 1:** 'I am 99% sure it's meningitis, but there is a small 1% chance that it's concussion'.
- **Doctor 2:** 'I am 99% sure it's a tumor, but there is a small 1% chance that it's concussion'.
- $m_1(\{m\}) = 0.99,\ m_1(\{c\}) = 0.01$ and $m_2(\{t\}) = 0.99,\ m_2(\{c\}) = 0.01$
- $m_1 \oplus m_2(\{c\}) = 1$
- But both doctors agree that concussion is very unlikely!

Normalising and Inconsistency

- The problem seems to be a result of normalising out the combinations of focal sets which intersect to give $\emptyset$.
- If we don't normalise then we obtain the follow combined mass assignment:
- $m_1 \oplus' m_2(\emptyset) = 0.9999, \; m_1 \oplus' m_2(\{c\}) = 0.0001$
- This clearly indicates that the two sources of evidence are highly inconsistent.
- *Normalisation in Dempster's rule can lead to counter intuitive results when the sources of evidence are highly inconsistent.*

Conditional Belief Functions

- Let *Bel* be a prior belief function with mass assignment m_1.
- Let m_2 be the mass function defined such that $m_2(B) = 1$.
- Then define $Bel(A|B) = \sum_{S \subseteq A} m_1 \oplus m_2(S)$
- This is equivalent to:

$$Bel(A|B) = \frac{Bel(A \cup B^c) - Bel(B^c)}{1 - Bel(B^c)}$$

Example: Conditional Belief

- Suppose that for *Bel* the mass assignment is:
- $m_1(\{w_1, w_2, w_3\}) = 0.3,\ \ m_1(\{w_1, w_2\}) = 0.2,\ \ m_1(\{w_1, w_3\}) = 0.2,\ \ m_1(\{w_2\}) = 0.3$
- In this case what is $Bel(\{w_1, w_3\}|\{w_1, w_2\})$?
- Define $m_2(\{w_1, w_2\}) = 1$:

$m_1 \oplus m_2$	$\{w_1, w_2, w_3\}$ 0.3	$\{w_1, w_2\}$ 0.2	$\{w_1, w_3\}$ 0.2	$\{w_2\}$ 0.3
$\{w_1, w_2\}$ 1	$\{w_1, w_2\}$ 0.3	$\{w_1, w_2\}$ 0.2	$\{w_1\}$ 0.2	$\{w_2\}$ 0.3

- $m_1 \oplus m_2(\{w_1, w_2\}) = 0.5,\ \ m_1 \oplus m_2(\{w_1\}) = 0.2,\ \ m_1 \oplus m_2(\{w_2\}) = 0.3$
- $Bel(\{w_1, w_3\}|\{w_1, w_2\}) = m_1 \oplus m_2(\{w_1\}) = 0.2$