# Introduction to the theory of belief functions

Frédéric Pichon

Laboratory of Computer Engineering and Automation of Artois (LGI2A)
Université d'Artois, Béthune, France
http://www.lgi2a.univ-artois.fr/~pichon

7th School on Belief Functions and their Applications
Granada, Spain, October 19, 2025

# History

- The theory originates from the work of Dempster (1967)[1] in the context of statistical inference.
- It was formalized by Shafer (1976)[2] as a theory of evidence.

$\rightarrow$ Also known as Dempster-Shafer theory or Evidence theory.

- Smets contributed to its development around the 1990's through his Transferable Belief Model (TBM) interpretation of the theory.
- Since then, it has found applications in a wide range of areas, including information fusion, machine learning, reliability, risk analysis, optimization, and preference modeling, as well as in various application domains such as medicine, defense, finance, and climate change.

---

[1] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics, 38:325–339, 1967.

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

# Key features

- A formal framework for reasoning with uncertain information.
- It extends both probabilistic and logical/set-based (such as interval analysis) reasonings:
  - It includes extensions of set-theoretic/logical notions (intersection, union, inclusion, inconsistency) and probabilistic notions (conditioning, marginalization, entropy)
  - Any reasoning done with sets or with probabilities alone, is recovered.
- However, its greatest expressive power allows also reasonings involving both sets and probabilities, which is often the case.
- Moreover, it is easily put in practice thanks to breaking down, using a set of tools, the intricate available evidence into simpler judgements.

$\rightarrow$ General and operational framework for uncertain reasoning.

# Contents of this lecture

1. Representation of evidence: mass, belief and plausibility functions.
2. Combination of evidence: Dempster's rule, compatible frames.
3. Comparison of evidence: uncertainty principles, informational orderings.

# Outline

1. Representation of evidence
   - Mass function
   - Belief and plausibility functions
2. Combination of evidence
   - Dempster's rule
   - Compatible frames
3. Comparison of evidence
   - Uncertainty principles
   - Inclusion relation
   - Uncertainty measures

# Outline

# Outline

1. Representation of evidence
   - Mass function
   - Belief and plausibility functions
2. Combination of evidence
   - Dempster's rule
   - Compatible frames
3. Comparison of evidence
   - Uncertainty principles
   - Inclusion relation
   - Uncertainty measures

# Mass function

## Definition

Let $\Theta$ be a finite set called the frame of discernment. A mass function on $\Theta$ is a mapping $m : 2^{\Theta} \to [0, 1]$ such that

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

Any $A \subseteq \Theta$ such that $m(A) > 0$ is a focal set of $m$.

A mass function is used to represent a state of knowledge about an uncertain variable $X$ taking values in $\Theta$, induced by a body of evidence.

# Example

- An attack from the red country will soon be launched on one of five targets (2 cities, 3 military bases) in the blue country.


- The target $X \in \Theta = \{c1, c2, b1, b2, b3\}$ is of interest.
- A piece of information that $X \in \{b1, c1\}$ has been received via a communication link with an ally.
- However, there is a 0.1 probability that this link is out of order.

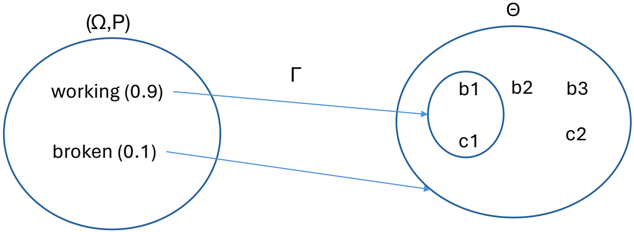
# Formalization

- Let $\Omega = \{\text{working, broken}\}$ be the set of possible states of the link.
  - If it is working, we know that $X \in \{b1, c1\}$.
  - It it is broken, we know nothing, i.e., we just know that $X \in \Theta$.
- There is a probability $P(\text{working}) = 0.9$ that it is working, hence of knowing that $X \in \{b1, c1\}$.
- There is a probability $P(\text{broken}) = 0.1$ that it is broken, hence of knowing that $X \in \Theta$.
- This state of knowledge can be represented by the mass function $m$ such that:

$$m(\{b1, c1\}) = 0.9, \quad m(\Theta) = 0.1$$

- $m(A)$: probability of knowing only that $X \in A$.

# Formalization (continued)




- In addition to the set $\Omega$ of states of the link, equipped with the probability measure $P$, let $\Gamma : \Omega \to 2^{\Theta}$ be the mapping from the states to what is known about $X$, e.g., $\Gamma(\text{working}) = \{b1, c1\}$.
- The triple $(\Omega, P, \Gamma)$ represents the piece of evidence, and mass function $m$ models the state of knowledge it generates about $X$.
- This meaning (semantics) of a mass function is in line with the random code semantics proposed by Shafer for mass functions.

# Random code semantics

## Shafer (1981)[3]

- Suppose we receive a coded message about $X$.
- The actual code used is unknown, but we know :
  - it was one in a finite set $\Omega$;
  - the chance $P(\omega)$ of each code $\omega \in \Omega$ being selected.
- Furthermore, we know that the meaning of the message is $X \in \Gamma(\omega)$, with $\Gamma(\omega)$ a nonempty subset of $\Theta$, if code $\omega$ was used.
- The probability that the message means $X \in A$ is then:

$$m(A) := P(\{\omega \in \Omega : \Gamma(\omega) = A\}), \quad \forall A \in 2^{\Theta}.$$

$\rightarrow$ A mass function is obtained by fitting a piece of evidence to such message $(\Omega, P, \Gamma)$.

[3] G. Shafer. Constructive probability. Synthese, 48(1):1–60, 1981.

# Example

## Laskey (1987)[4]

- A spy is sent to discover whether the enemy intends to attack at dawn.
- She observes a nonempty subset of $\Theta = \{yes, no\}$ (i.e. she may observe A={yes}, the enemy will attack; A={no}, the enemy will not attack; or $A = \Theta$, she is unable to determine whether the enemy will attack).
- There are two possible codes $\omega_1$ and $\omega_2$, with probabilities $P(\omega_1) = 1/3$ and $P(\omega_2) = 2/3$, and with coding schemes

$$\omega_1(\{yes\}) = APPLE, \quad \omega_1(\{no\}) = CHERRY, \quad \omega_1(\Theta) = BANANA,$$
$$\omega_2(\{yes\}) = APPLE, \quad \omega_2(\{no\}) = BANANA, \quad \omega_2(\Theta) = CHERRY.$$

[4] K. B. Laskey. Belief in belief functions: an examination of Shafer's canonical examples, Proc. of the Third Conference on UAI, pages 39 - 46, 1987.

# Example (continued)

- The spy sends the coded message BANANA.
- The meaning of this message depends on the code used:

$$\begin{aligned}\Gamma(\omega_1) &= \omega_1^{-1}(BANANA) = \Theta, \\ \Gamma(\omega_2) &= \omega_2^{-1}(BANANA) = \{no\}.\end{aligned}$$

- Hence, the state of knowledge about the attack is represented by

$$m(\Theta) = 1/3, \quad m(\{no\}) = 2/3.$$

# Random set

- The triple/message $(\Omega, P, \Gamma)$ representing a piece of evidence is a random set.
- A random set $(\Omega, P, \Gamma)$ always induces a mass function $m : 2^\Theta \to [0,1]$ from

$$m(A) := P(\{\omega \in \Omega : \Gamma(\omega) = A\}), \quad \forall A \in 2^\Theta.$$

- Conversely, any mass function $m : 2^\Theta \to [0,1]$ can be seen as generated by the random set $(\Omega, P, \Gamma)$ with

$$\Omega = 2^\Theta,$$

$$P(\{A\}) = m(A), \quad A \subseteq \Theta,$$

  and

$$\Gamma(A) = A, \quad A \subseteq \Theta.$$

- More on random (fuzzy) sets in Lectures 7 & 9 (continuous variables).

# Special cases

- Logical mass function if it has only one focal set $A \subseteq \Theta$, i.e., the evidence tells us that $X \in A$ for sure and nothing more. It is equivalent to the set $A$ and is denoted $m_{[A]}$ ($m_{[A]}(A) = 1$).
  - Vacuous mass function if $\Theta$ is the only focal set. It represents total ignorance.
- Bayesian mass function if its focal sets are singletons, i.e., each possible code/state $\omega$ of the evidence points to a single value of $\Theta$. It is equivalent to a probability distribution.

# The case of inconsistent knowledge

- In some cases, it may happen that the modeling of the evidence yields a random set $(\Omega, P, \Gamma)$ such that $\Gamma(\omega) = \emptyset$ for some $\omega$ and

$$m(\emptyset) = P(\{\omega \in \Omega : \Gamma(\omega) = \emptyset\}) > 0,$$

  i.e., there is a nonzero probability that the induced knowledge is inconsistent.
- $1 - m(\emptyset) = P(\{\omega \in \Omega : \Gamma(\omega) \neq \emptyset\})$ is the probability that the induced knowledge is consistent and, since $m$ models this knowledge, it is by extension a measure of the consistency of $m$.
- It allows us to assess the validity of the assumptions that lead to the modeling of the evidence by this random set.

# Unnnormalized mass function

- What to do if such a case of unnormalized mass function, i.e., $m(\emptyset) > 0$, occur?
- Try and resolve the inconsistency, so that $m(\emptyset) = 0$, by revising the modeling assumptions (next two slides, Shafer's solution in the context of the random code metaphor)
- However, if it is not clear which modeling assumptions should be revised or because it can sometimes be convenient, an option (as in Smets's TBM) is to allow $m(\emptyset) > 0$ and thus leave $m$ as is.

# Random code semantics extended

Shafer (1990)[5] and (1992)[6]

- In the random code metaphor, when the message is decoded using the different codes, we may sometimes get nonsense.
- The codes with which we get nonsense cannot be the one actually used.
- $\Gamma(\omega) = \emptyset$ for some $\omega$ is used to indicate that $\omega$ could not be the code actually used.
- The observation that some codes cannot be the one actually used is taken into account by revising the chance distribution $P$ for the codes, so as to eliminate them, i.e., $P$ is conditioned on the event

$$\Theta^* = \{\omega \in \Omega : \Gamma(\omega) \neq \emptyset\}.$$

[5] G. Shafer. Perspectives on the theory and practice of belief functions. Int. J. Approx. Reason., 4(5-6):323-362, 1990.

[6] G. Shafer. Rejoinders to comments on "Perspectives on the theory and practice of belief functions". Int. J. Approx. Reason., 6(3):445-480, 1992.

# Normalization

- In this extended version of the random code metaphor, the probability that the message $(\Omega, P, \Gamma)$ means $X \in A$ is:

$$m^*(A) := P^*(\{\omega \in \Omega : \Gamma(\omega) = A\}), \quad \forall A \in 2^\Theta,$$

  with $P^*$ the probability measure resulting from the conditioning of $P$ on the event $\Theta^*$.
- Mass function $m^*$ is normalized, i.e., $m^*(\emptyset) = 0$.
- It can be obtained directly from the mass function $m$ induced from the random set $(\Omega, P, \Gamma)$ as:

$$m^*(A) = \frac{m(A)}{1 - m(\emptyset)}, \quad \forall A \in 2^\Theta \setminus \{\emptyset\}.$$

- This latter operation is called (Dempster's) normalization.
- Remark: a mass function is (totally) consistent if and only if it is normalized.

# Outline

# Belief and plausibility functions

## Definition

Given a mass function $m$ on $\Theta$, the corresponding belief and plausibility functions are mappings $bel : 2^\Theta \to [0,1]$ and $pl : 2^\Theta \to [0,1]$ such that

$$\begin{aligned} bel(A) &= \sum_{B \subseteq A, B \neq \emptyset} m(B), \\ pl(A) &= \sum_{B \cap A \neq \emptyset} m(B). \end{aligned}$$

- $bel(A)$ is the probability that the proposition $X \in A$ is implied by the noncontradictory evidence.
- $pl(A)$ is the probability that the proposition $X \in A$ is consistent with the evidence.

# Interpretation

- Assume the evidence is that $X \in B$ for some nonempty $B \subseteq \Theta$.
- Consider the proposition $X \in A$ for some $A \subseteq \Theta$.
- If $B \subseteq A$, then we know that $X \in A$, i.e., the proposition is implied by the evidence.
- The interpretation of $bel(A)$ follows then from that of $m$.
- The interpretation of $pl$ is obtained similarly by remarking that, given evidence $X \in B$ and proposition $X \in A$, if $B \cap A \neq \emptyset$, then we cannot exclude that $X \in A$, i.e., the proposition is consistent with the evidence.

# Attack location example

- We had $\Theta = \{c1, c2, b1, b2, b3\}$ and

$$m(\{b1, c1\}) = 0.9, \quad m(\Theta) = 0.1$$

- Degrees of belief and plausibility of some subsets of $\Theta$:

| $A$ | $\emptyset$ | $\{b1\}$ | $\{c1\}$ | $\{b1, c1\}$ | $\{b1, c1, c2\}$ | $\{b1, b2\}$ | $\{b2, c2\}$ | $\Theta$ |
|---|---|---|---|---|---|---|---|---|
| $bel(A)$ | 0 | 0 | 0 | 0.9 | 0.9 | 0 | 0 | 1 |
| $pl(A)$ | 0 | 1 | 1 | 1 | 1 | 1 | 0.1 | 1 |

# Relations between $m$, $bel$ and $pl$

- Let $m$ be a mass function, $bel$ and $pl$ the corresponding belief and plausibility functions.
- We have

$$bel(A) = pl(\Theta) - pl(\overline{A}), \quad A \subseteq \Theta,$$

$$\begin{aligned} m(A) &= \sum_{B \subseteq A} (-1)^{|A|-|B|} bel(B), \quad A \subseteq \Theta, A \neq \emptyset, \\ m(\emptyset) &= 1 - bel(\Theta), \end{aligned}$$

$$\begin{aligned} m(A) &= \sum_{B \subseteq A} (-1)^{|A|-|B|+1} pl(\overline{B}), \quad A \subseteq \Theta, A \neq \emptyset, \\ m(\emptyset) &= 1 - pl(\Theta). \end{aligned}$$

- $m$, $bel$ and $pl$ are thus three equivalent representations of a state of knowledge.

# Elementary properties

- For all $A \subseteq \Theta$, $bel(A) \leq pl(A)$.
- $bel(\emptyset) = pl(\emptyset) = 0$.
- $bel(\Theta) = pl(\Theta) = 1 - m(\emptyset)$.
- In the attack location example, we get, for $A = \{b1, c1\}$:

| | $A$ | $\overline{A}$ | $\Theta$ |
|---|---|---|---|
| $bel$ | 0.9 | 0 | 1 |
| $pl$ | 1 | 0.1 | 1 |

- We observe that

$$\begin{aligned} bel(A) + bel(\overline{A}) &\leq bel(A \cup \overline{A}) = bel(\Theta), \\ pl(A) + pl(\overline{A}) &\geq pl(A \cup \overline{A}) = pl(\Theta). \end{aligned}$$

- *bel* and *pl* are nonadditive measures.

# Contour function

## Definition

Given a mass function $m$ on $\Theta$, the corresponding contour function is the mapping $c : \Theta \to [0, 1]$

$$\begin{aligned} c(\theta) &= \sum_{\theta \in B} m(B) \\ &= pl(\{\theta\}). \end{aligned}$$

In some particular cases with respect to the focal sets of $m$, it is possible to recover $m$ from $c$.

# Particular cases

## Relationship with probability theory

- When a mass function $m$ is Bayesian, i.e., its focal sets are singletons, we have, for all $A \subseteq \Theta$,

$$bel(A) = pl(A) = \sum_{\theta \in A} m(\{\theta\}).$$

- $bel = pl$ is then a probability measure.
- If $pl$ is a probability measure, $c$ is its probability distribution, i.e.,

$$pl(A) = \sum_{\theta \in A} c(\theta).$$

- $m$ is recovered directly from $c$ as they coincide.

# Particular cases

## Relationship with possibility theory

- When a mass function $m$ is consonant, which means that its focal sets $A_1, \ldots, A_r$ are nested, i.e., $A_1 \subseteq \cdots \subseteq A_r$, we have, for all $A, B \subseteq \Omega$,
$$pl(A \cup B) = \max(pl(A), pl(B)),$$
$$bel(A \cup B) = \min(bel(A), bel(B)).$$
- $pl$ is then a possibility measure and $bel$ is the dual necessity measure.
- If $pl$ is a possibility measure, $c$ is its possibility distribution, i.e.,
$$pl(A) = \max_{\theta \in A} c(\theta).$$

# Particular cases

## Relationship with possibility theory (continued)

- $m$ is recovered from $c$ as follows.
- Let $c_1 > \ldots > c_r$ be the distinct values taken by $c$, arranged in decreasing order, and $c_{r+1} = 0$.
- Let $A_i = \{\theta | c(\theta) \geq c_i\}$, $i = 1, \ldots, r$.
- Then, we have, for any $A \subseteq \Theta$,

$$m(A) = \begin{cases} c_i - c_{i+1} & \text{if } A = A_i,\ i = 1, \ldots, r, \\ 1 - c_1 & \text{if } A = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

# Mass function from contour function

- Let $m_c$ denote the mass function computed from some contour function $c$ according to the formula on the previous slide.
- Approximation : Let $m$ be a mass function, with contour function $c$ such that $c_1 = 1$.Then, among the mass functions that are consonant and lower approximations of $m$ (their plausibility functions are dominated by $pl$), $m_c$ is a reasonable [7] choice.
- Elicitation: Let $c$ be a contour function such that $c_1 = 1$. Then, among the mass functions that have this contour function, $m_c$ is a reasonable[8] choice.
- (In both cases, "reasonable" means respecting the so-called maximum uncertainty principle, see later.)
- (Approximation of belief functions will be covered in Lecture 10.)

[7] See D. Dubois, H. Prade. A set-theoretic view of belief functions–logical operations and approximation by fuzzy sets. Int. J. Gen. Syst., 12:193-226, 1986.

[8] See T. Denœux. Methods for building belief functions. Fifth BFAS School on Belief Functions and Their Applications, Sienna, Italy, October 27–31, 2019.

# Consistency of a mass function

- Let $m$ be a mass function, with focal sets $A_1, \ldots, A_r$ and contour function $c$.
- The condition $c_1 = 1$, i.e., $\max_{\theta \in \Theta} c(\theta) = 1$, is equivalent to
$$\cap_{i=1}^{r} A_i \neq \emptyset.$$
- Such a mass function is said to be fully consistent.
- Full consistency is a stronger form of consistency than that exhibited by normalized mass functions ($A_i \neq \emptyset, i = 1 \ldots, r$), since
$$\cap_{i=1}^{r} A_i \neq \emptyset \Rightarrow A_i \neq \emptyset, i = 1 \ldots, r.$$
- The quantity $\max_{\theta \in \Theta} c(\theta)$ has been proposed[9] as an alternative (stronger) measure of consistency for $m$, to $1 - m(\emptyset)$.
- Remark: $c_1 = 1 - m_c(\emptyset)$, hence the full consistency of $m$ is nothing but the consistency of its approximation $m_c$.

[9] S. Destercke, T. Burger. Toward an axiomatic definition of conflict between belief functions. IEEE Trans. Cybern., 43(2):585–596, 2013.

# Outline

1. Representation of evidence
   - Mass function
   - Belief and plausibility functions
2. Combination of evidence
   - Dempster's rule
   - Compatible frames
3. Comparison of evidence
   - Uncertainty principles
   - Inclusion relation
   - Uncertainty measures

# Outline

1. Representation of evidence
   - Mass function
   - Belief and plausibility functions
2. Combination of evidence
   - Dempster's rule
   - Compatible frames
3. Comparison of evidence
   - Uncertainty principles
   - Inclusion relation
   - Uncertainty measures

# Independent and reliable messages

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$, with $\Gamma_i : \Omega_i \to 2^\Theta \setminus \{\emptyset\}$, $i = 1, 2$, be two messages representing two pieces of evidence about $X$ and inducing mass functions $m_1$ and $m_2$, respectively.
- Assume that these messages are independent, i.e., the chance $P_{12}(\omega_1, \omega_2)$ that the pair of codes $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ was chosen is equal to $P_1(\omega_1) \cdot P_2(\omega_2)$.
- Assume further that they are reliable: if the actual codes were $\omega_1$ and $\omega_2$, we know for sure that $X \in \Gamma_\cap(\omega_1, \omega_2) := \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$.

# Conjunctive rule

- Under the preceding assumptions of independence and reliability, our body of evidence is represented by the random set $(\Omega_1 \times \Omega_2, P_{12}, \Gamma_\cap)$, which induces the state of knowledge about $X$ modeled by the mass function denoted $m_1 \circledcap m_2$, called the conjunctive sum of $m_1$ and $m_2$, and defined as

$$(m_1 \circledcap m_2)(A) := P_{12}(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = A\}).$$

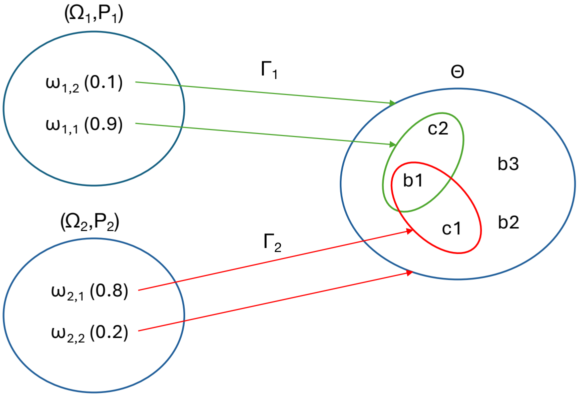
- It is easy to show that

$$(m_1 \circledcap m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C).$$

- The binary operation $\circledcap$ is called the conjunctive rule (unnormalized Dempster's rule).

# Example

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ be the following coded messages sent by two spies about the target $X \in \Theta = \{c1, c2, b1, b2, b3\}$.



# Example (continued)

- Message $(\Omega_1, P_1, \Gamma_1)$ induces mass function

$$m_1(\{b1, c2\}) = 0.9, \quad m_1(\Theta) = 0.1.$$

- Message $(\Omega_2, P_2, \Gamma_2)$ induces mass function

$$m_2(\{b1, c1\}) = 0.8, \quad m_2(\Theta) = 0.2.$$

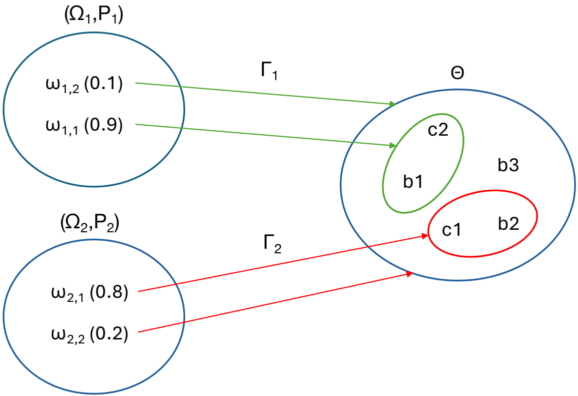
- Assuming they are independent and reliable, we obtain

| $m_2 \setminus m_1$ | $\{b1, c2\}$ 0.9 | $\Theta$ 0.1 |
|---|---|---|
| $\{b1, c1\}$ 0.8 | $\{b1, c1\} \cap \{b1, c2\} = \{b1\}$ 0.8*0.9=0.72 | $\{b1, c1\}$ 0.08 |
| $\Theta$ 0.2 | $\{b1, c2\}$ 0.18 | $\Theta$ 0.02 |

$$\begin{aligned} m_{1 \cap 2}(\{b1\}) &= 0.72 \\ m_{1 \cap 2}(\{b1, c1\}) &= 0.08 \\ m_{1 \cap 2}(\{b1, c2\}) &= 0.18 \\ m_{1 \cap 2}(\Theta) &= 0.02 \end{aligned}$$

# Case of conflicting evidence

- Suppose $(\Omega_2, P_2, \Gamma_2)$ is rather the following:



# Case of conflicting evidence (continued)

- Hence, $m_2$ is now

$$m_2(\{b2, c1\}) = 0.8, \quad m_2(\Theta) = 0.2.$$

- We obtain

| $m_2 \setminus m_1$ | $\{b1, c2\}$ 0.9 | $\Theta$ 0.1 |
|---|---|---|
| $\{b2, c1\}$ 0.8 | $\{b2, c1\} \cap \{b1, c2\} = \emptyset$ 0.8*0.9=0.72 | $\{b2, c1\}$ 0.08 |
| $\Theta$ 0.2 | $\{b1, c2\}$ 0.18 | $\Theta$ 0.02 |

$$\begin{aligned} m_{1 \textcircled{\cap} 2}(\emptyset) &= 0.72 \\ m_{1 \textcircled{\cap} 2}(\{b2, c1\}) &= 0.08 \\ m_{1 \textcircled{\cap} 2}(\{b1, c2\}) &= 0.18 \\ m_{1 \textcircled{\cap} 2}(\Theta) &= 0.02 \end{aligned}$$

# Analysis

- We are in a case where the body of evidence is modeled by a random set $(\Omega_1 \times \Omega_2, P_{12}, \Gamma_\cap)$ such that $\Gamma_\cap(\omega_1, \omega_2) = \emptyset$ for some $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ and

$$m_{1 \bigcirc\!\!\!\!\cap\, 2}(\emptyset) = P_{12}(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = \emptyset\}) > 0$$

- This latter probability is known as the degree of conflict between $m_1$ and $m_2$. As we have seen, it is a measure of the invalidity of the modeling assumptions that lead to this random set.
- Their most common revision, in order to resolve the inconsistency, is basically Shafer's solution for the case of inconsistent knowledge seen earlier:
  - if $\Gamma_\cap(\omega_1, \omega_2) = \emptyset$, then it is an observation that $(\omega_1, \omega_2)$ could not be the pair of codes actually used.

$\rightarrow$ The distribution $P_{12}$ must be revised to eliminate such pairs, i.e., conditioned on $\Theta_\cap = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) \neq \emptyset\}$.

# Dempster's rule

- Following this reasoning, the probability of knowing that $X \in A$ from two independent and reliable messages $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ inducing mass functions $m_1$ and $m_2$, is

$$(m_1 \oplus m_2)(A) := P_\cap(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = A\}),$$

  with $P_\cap$ the probability measure resulting from the conditioning of $P_{12}$ on the event $\Theta_\cap$.
- $m_1 \oplus m_2$ is called the orthogonal sum of $m_1$ and $m_2$.
- The orthogonal sum is well defined if $P_{12}(\Theta_\cap) > 0$.
- It is easy to show that

$$(m_1 \oplus m_2)(A) = \frac{(m_1 \textcircled{\cap} m_2)(A)}{1 - (m_1 \textcircled{\cap} m_2)(\emptyset)}, \quad \forall A \in 2^\Theta \setminus \{\emptyset\}.$$

- The binary operation $\oplus$ is called Dempster's rule.

# Assumptions underlying Dempster's rule

- In general, $P_\cap \neq P_{12}$, hence the reasoning leading to Dempster's rule has induced a revision of the modeling assumptions (in the sense that the distribution on the codes is no more $P_{12}$) and, specifically, some dependence between the messages.
- This may seem contradictory with associating Dempster's rule, still, with the assumptions that the messages are "independent and reliable".
- This apparent contradiction is resolved if we consider that the codes have indeed been drawn according to $P_{12} = P_1 \times P_2$. And then, after this experiment has taken place, we receive an observation about the pair of codes actually drawn (specifically, that some pairs could not have been the one drawn) and we must condition $P_{12}$ on this observation.

# Example

## Dempster's rule

- We had

$$m_1(\{b1, c2\}) = 0.9, \quad m_1(\Theta) = 0.1$$

and

$$m_2(\{b2, c1\}) = 0.8, \quad m_2(\Theta) = 0.2.$$

- We obtain

| $m_2 \setminus m_1$ | $\{b1, c2\}$ 0.9 | $\Theta$ 0.1 |
|---|---|---|
| $\{b2, c1\}$ 0.8 | $\emptyset$ 0.72 | $\{b2, c1\}$ 0.08 |
| $\Theta$ 0.2 | $\{b1, c2\}$ 0.18 | $\Theta$ 0.02 |

$$\begin{aligned}
m_{1\oplus 2}(\emptyset) &= 0 \\
m_{1\oplus 2}(\{b2, c1\}) &= 0.08/0.28 = 9/14 \\
m_{1\oplus 2}(\{b1, c2\}) &= 0.18/0.28 = 4/14 \\
m_{1\oplus 2}(\Theta) &= 0.02/0.28 = 1/14
\end{aligned}$$

# Properties of Dempster's rule

- Commutativity: $m_1 \oplus m_2 = m_2 \oplus m_1$
- Associativity: $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$
- Insensitivity to vacuous information (vacuous mass function as neutral element): $m \oplus m_{[\Theta]} = m$
- Generalization of set intersection: if $A \cap B \neq \emptyset$ then

$$m_{[A]} \oplus m_{[B]} = m_{[A \cap B]}$$

- Generalization of probabilistic conditioning: if $m$ is a Bayesian mass function and $m_{[A]}$ is a logical mass function, then

$$m(\cdot|A) := m \oplus m_{[A]}$$

  is a Bayesian mass function corresponding to the conditioning of $m$ by $A$.

# Applicability of Dempster's rule

## Zadeh's example

- Let $X \in \Theta = \{a, b, c\}$ and two experts providing mass functions $m_1$ and $m_2$ about $X$ such that

$$m_1(\{a\}) = 0.99, m_1(\{b\}) = 0.01, m_2(\{b\}) = 0.01, m_2(\{c\}) = 0.99$$

- We have $m_{1\oplus 2}(\{b\}) = 1$.
- As both experts considered $b$ to be very unlikely, some authors claim this result to be counterintuitive, and use it to question Dempster's rule.
- However, if you accept the assumptions underlying Dempster's rule, then this is the only reasonable conclusion: expert 1 tells that $c$ is impossible, and expert 2 tells that $a$ is impossible, hence $b$ is the only remaining possibility.
- The question is not whether Dempster's rule produces sound results or not, but rather whether its underlying assumptions hold.

# Applicability of Dempster's rule

## Zadeh's example (continued) and complexity

- As we have seen, the degree of conflict $(m_1 \bigcirc\!\!\!\!\cap\, m_2)(\emptyset)$ is a way to assess to validity of the assumptions that the pieces of evidence are reliable and independent.
- In Zadeh's example, we have a conflict of 0.9999, which suggests that these assumptions may not be valid.

$\rightarrow$ Lecture 3 will cover alternative rules, corresponding to other assumptions.

- Another issue with Dempster's rule is its computational complexity: in the worst case, exponential with respect to $|\Theta|$.

$\rightarrow$ Lecture 3 will show that its complexity can be managed in practical applications.

# Outline

# Granularity of the frame of discernement

- The granularity of the frame of discernement is always, to some extent, a matter of convention, as any element of the frame can always be split into several possibilities.
- Example:
  - Assume an analyst, who can determine whether the target is a military base.
  - His frame of discernment is

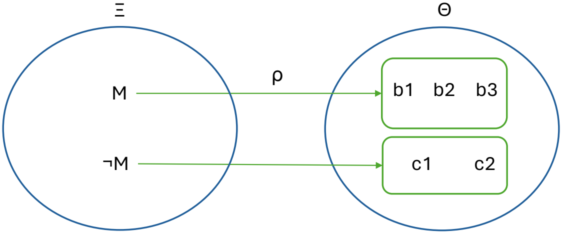
$$\Xi = \{M, \neg M\},$$

    where $M$ means that the target is a military base.
  - Recall the frame $\Theta = \{c1, c2, b1, b2, b3\}$. We have

$$M \rightarrow \{b1, b2, b3\}, \qquad \neg M \rightarrow \{c1, c2\}.$$

  - By splitting the elements of $\Xi$, we can obtain the elements of $\Theta$.

# Refinement and coarsening




## Definition

A frame $\Theta$ is a refinement of a frame $\Xi$, and $\Xi$ a coarsening of $\Theta$, if there exists a mapping $\rho : 2^{\Xi} \to 2^{\Theta}$, called a refining, such that

- $\{\rho(\{\xi\}) \mid \xi \in \Xi\} \subseteq 2^{\Theta}$ is a partition of $\Theta$,
- and, for all $A \subseteq \Xi$,
$$\rho(A) = \bigcup_{\xi \in A} \rho(\{\xi\}).$$

# Vacuous extension

## Expression of knowledge in a finer frame

- Assume the analyst provides the following mass function on $\Xi$:

$$m^{\Xi}(\{M\}) = 0.3, \qquad m^{\Xi}(\{\neg M\}) = 0.6, \qquad m^{\Xi}(\Xi) = 0.1.$$

- How to express the state of knowledge $m^{\Xi}$ in the finer frame $\Theta$?

$\rightarrow$ Transfer $m^{\Xi}(A)$ to $\rho(A)$, for all $A \subseteq \Xi$:

$$\begin{aligned} m^{\Xi}(\{M\}) = 0.3 &\rightarrow \rho(\{M\}) = \{b1, b2, b3\} \\ m^{\Xi}(\{\neg M\}) = 0.6 &\rightarrow \rho(\{\neg M\}) = \{c1, c2\} \\ m^{\Xi}(\Xi) = 0.1 &\rightarrow \rho(\Xi) = \Theta \end{aligned}$$

- We get the following mass function on $\Theta$:

$$m^{\Xi\uparrow\Theta}(\{c1, c2\}) = 0.6, m^{\Xi\uparrow\Theta}(\{b1, b2, b3\}) = 0.3, m^{\Xi\uparrow\Theta}(\Theta) = 0.1.$$

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

# Outer reduction

- Suppose now that we have the following mass function on $\Theta$:

$$m^{\Theta}(\{c2\}) = 0.4, \qquad m^{\Theta}(\{c2, b2\}) = 0.3, \qquad m^{\Theta}(\{b1, b3\}) = 0.3.$$

- How to express $m^{\Theta}$ in the coarser frame $\Xi$?
- It is not so obvious as the mapping $\rho$ is not invertible, e.g., there is no $A \subseteq \Xi$ such that $\rho(A) = \{c2\}$.
- A solution is to rely on a generalized inverse of $\rho$, called outer reduction and defined as, for any $B \subseteq \Theta$,

$$\rho^{\dagger}(B) = \{\xi \in \Xi : \rho(\{\xi\}) \cap B \neq \emptyset\}.$$

- We have for our example:

$$\rho^{\dagger}(\{c2\}) = \{\neg M\}, \quad \rho^{\dagger}(\{c2, b2\}) = \{M, \neg M\}, \quad \rho^{\dagger}(\{b1, b3\}) = \{M\}.$$

# Restriction

## Expression of knowledge in a coarser frame

### Definition

The restriction of $m^{\Theta}$ in $\Xi$ transfers each mass $m^{\Theta}(B)$ to the outer reduction of $B$: for all $A \subseteq \Xi$,

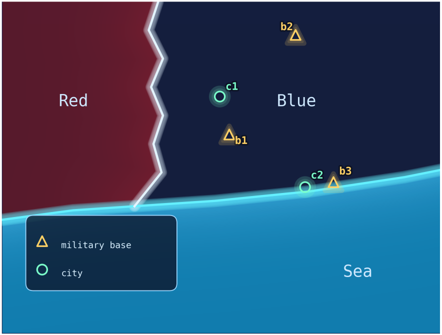
$$m^{\Theta\downarrow\Xi}(A) = \sum_{\rho^{\dagger}(B)=A} m^{\Theta}(B).$$

- In the example, the restriction of $m^{\Theta}$ in $\Xi$ is

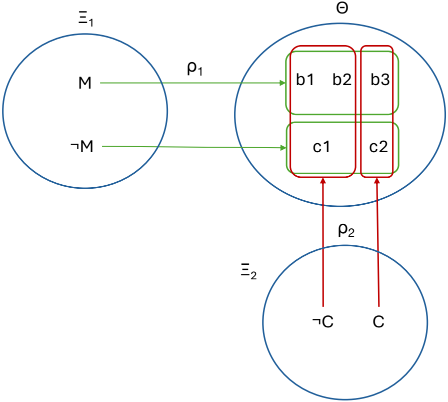
$$m^{\Theta\downarrow\Xi}(\{\neg M\}) = 0.4, \qquad m^{\Theta\downarrow\Xi}(\Xi) = 0.3, \qquad m^{\Theta\downarrow\Xi}(\{M\}) = 0.3.$$

- Remark: in general, $m^{(\Theta\downarrow\Xi)\uparrow\Theta} \neq m^{\Theta}$, i.e., information is lost when expressing $m^{\Theta}$ in a coarser frame.

# Compatible frames




$$\Xi_2 = \{\text{Coast, Not Coast}\}$$




## Definition

Two frames are compatible if they have a common refinement.

# Combination on compatible frames

## Definition

Let $m^{\Xi_1}$ and $m^{\Xi_2}$ be mass functions on compatible frames $\Xi_1$ and $\Xi_2$ with common refinement $\Theta$. Their orthogonal sum in $\Theta$ is

$$m^{\Xi_1} \oplus m^{\Xi_2} = m^{\Xi_1 \uparrow \Theta} \oplus m^{\Xi_2 \uparrow \Theta}$$

- Example:

$$m^{\Xi_1}(\{M\}) = 0.5, \quad m^{\Xi_1}(\{\neg M\}) = 0.3, \quad m^{\Xi_1}(\Xi_1) = 0.2,$$
$$m^{\Xi_2}(\{C\}) = 0.4, \quad m^{\Xi_2}(\{\neg C\}) = 0.5, \quad m^{\Xi_2}(\Xi_2) = 0.1.$$

- Their orthogonal sum in $\Theta$ is

| | | $m^{\Xi_2 \uparrow \Theta}$: $\{c2, b3\}$, 0.4 | $m^{\Xi_2 \uparrow \Theta}$: $\{c1, b1, b2\}$, 0.5 | $m^{\Xi_2 \uparrow \Theta}$: $\Theta$, 0.1 |
|---|---|---|---|---|
| $m^{\Xi_1 \uparrow \Theta}$ | $\{c1, c2\}$, 0.3 | $\{c2\}$, 0.12 | $\{c1\}$, 0.15 | $\{c1, c2\}$, 0.03 |
| $m^{\Xi_1 \uparrow \Theta}$ | $\{b1, b2, b3\}$, 0.5 | $\{b3\}$, 0.20 | $\{b1, b2\}$, 0.25 | $\{b1, b2, b3\}$, 0.05 |
| $m^{\Xi_1 \uparrow \Theta}$ | $\Theta$, 0.2 | $\{c2, b3\}$, 0.08 | $\{c1, b1, b2\}$, 0.10 | $\Theta$, 0.02 |

# Case of product frames

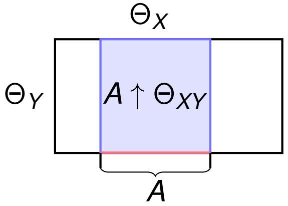
- Assume now that we have two uncertain variables $X$ and $Y$ with frames $\Theta_X$ and $\Theta_Y$.
- Example:
  - $X$ is whether the target is a military base, with $\Theta_X = \{M, \neg M\}$
  - $Y$ is whether the enemy will attack at dawn, with $\Theta_Y = \{D, \neg D\}$.
- Let $\Theta_{XY} = \Theta_X \times \Theta_Y$ be the product space.

# Cylindrical extension and projection

- $\Theta_{XY}$ is a refinement of $\Theta_X$ (and $\Theta_Y$) with refining $\rho : 2^{\Theta_X} \to 2^{\Theta_{XY}}$ defined as, for all $A \subseteq \Theta_X$,

$$\rho(A) = A \times \Theta_Y.$$

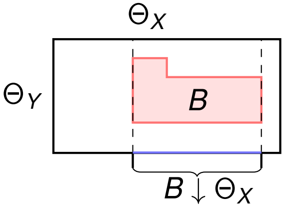
- $\rho(A)$ is called the cylindrical extension of $A$ in $\Theta_{XY}$ and is denoted by $A \uparrow \Theta_{XY}$.




- The outer reduction of a subset $B$ of $\Theta_{XY}$ is

$$\begin{aligned}\rho^{\dagger}(B) &= \{x \in \Theta_X \mid \rho(\{x\}) \cap B \neq \emptyset\} \\ &= \{x \in \Theta_X \mid \exists\, y \in \Theta_Y,\ (x, y) \in B\}.\end{aligned}$$

- $\rho^{\dagger}(B)$ is called the projection of $B$ on $\Theta_X$ and is denoted by $B \downarrow \Theta_X$.



# Vacuous extension and marginalization

- The vacuous extension of $m^{\Theta_X}$ in $\Theta_{XY}$ transfers each mass $m^{\Theta_X}(B)$, for any $B \subseteq \Theta_X$, to the cylindrical extension of $B$:

$$m^{\Theta_X \uparrow \Theta_{XY}}(A) = \begin{cases} m^{\Theta_X}(B) & \text{if } A = B \times \Theta_Y, \\ 0 & \text{otherwise.} \end{cases}$$

- Conversely, the restriction, called marginalization, of a joint mass function $m^{\Theta_{XY}}$ in $\Theta_X$ transfers each mass $m^{\Theta_{XY}}(B)$ to the outer reduction (projection) of $B$ on $\Theta_X$: for all $A \subseteq \Theta_X$,

$$m^{\Theta_{XY} \downarrow \Theta_X}(A) = \sum_{B \downarrow \Theta_X = A} m^{\Theta_{XY}}(B).$$

- The marginalization operation extends both set projection and probabilistic marginalization.

# Application to evidential reasoning

- Many applied problems can be modeled by defining variables and relations between them. Given partial information about some of them, the problem is then to infer the values of variables of interest.
- Such problem can be cast in the belief function framework.
- For simplicity, assume that we have only two variables $X$ and $Y$. Furthermore, suppose we have
  - Partial knowledge of $Y$ in the form of a mass function $m^{\Theta_Y}$;
  - A joint mass function $m^{\Theta_{XY}}$ representing an uncertain relation between $X$ and $Y$.
- What can we say about $X$?

$\rightarrow$ Compute

$$\left(m^{\Theta_Y \uparrow \Theta_{XY}} \oplus m^{\Theta_{XY}}\right)^{\downarrow \Theta_X}.$$

- Remark: such operations are intractable with many variables and large frames of discernment. However, efficient algorithms exist to carry them out in frames of minimal dimensions $\rightarrow$ Lecture 4.

# Example

- $X \in \Theta_X = \{M, \neg M\}$ is whether the target is a military base and $Y \in \Theta_Y = \{D, \neg D\}$ is whether the enemy will attack at dawn.
- Assume mass function $m^{\Theta_Y}$ about $Y$ is such that

$$m^{\Theta_Y}(\{D\}) = 0.3, \quad m^{\Theta_Y}(\{\neg D\}) = 0.5, \quad m^{\Theta_Y}(\Theta_Y) = 0.2$$

- Furthermore, we have a rule "If the enemy attacks at dawn, then the target is a military base" (for short "If $D$ then $M$"), which is reliable with 0.7 probability.
- This piece of evidence can be modeled by the joint mass function $m^{\Theta_{XY}}$ such that

$$m^{\Theta_{XY}}(\{(D, M), (\neg D, M), (\neg D, \neg M)\}) = 0.7, \quad m^{\Theta_{XY}}(\Theta_{XY}) = 0.3$$

# Example (continued)

- Computation of $m^{\Theta_Y \uparrow \Theta_{XY}} \oplus m^{\Theta_{XY}}$:

| | | $m^{\Theta_{XY}}$: $\{(D,M),(\neg D,M),(\neg D,\neg M)\}$, 0.7 | $m^{\Theta_{XY}}$: $\Theta_{XY}$, 0.3 |
|---|---|---|---|
| $m^{\Theta_Y \uparrow \Theta_{XY}}$ | $\{D \times \Theta_X\}$, 0.3 | $\{(D,M)\}$, 0.21 | $\{D \times \Theta_X\}$, 0.09 |
| | $\{\neg D \times \Theta_X\}$, 0.5 | $\{(\neg D,M),(\neg D,\neg M)\}$, 0.35 | $\{\neg D \times \Theta_X\}$, 0.15 |
| | $\Theta_{XY}$, 0.2 | $\{(D,M),(\neg D,M),(\neg D,\neg M)\}$, 0.14 | $\Theta_{XY}$, 0.06 |

- Marginalizing on $\Theta_X$, we get

$$
\begin{aligned}
\left(\left(m^{\Theta_Y \uparrow \Theta_{XY}} \oplus m^{\Theta_{XY}}\right)^{\downarrow \Theta_X}\right)(\{M\}) &= 0.21,\\
\left(\left(m^{\Theta_Y \uparrow \Theta_{XY}} \oplus m^{\Theta_{XY}}\right)^{\downarrow \Theta_X}\right)(\Theta_X) &= 0.79.
\end{aligned}
$$

# Outline

# Outline

# Maximum and minimum uncertainty principles

- In various problems, we need to substitute a knowledge state by another one, which has to be selected among a set of candidate knowledge states.
- If the candidate knowledge states are more (less) informative than the original one, then the substitute should be the least (most) informative one among them.
- This is known as the maximum (minimum) uncertainty, or the minimum information gain (loss), principle[10].
- Principle of maximum (minimum) entropy / nonspecificity in probability / possibility theory, least (most) commitment principle in belief function theory.

[10] G. J. Klir. Uncertainty and Information: Foundations of Generalized Information Theory. John Wiley & Sons, 2005

# Example

## Maximum uncertainty principle

- Suppose only marginal probability distributions $P^{\Theta_X}$ and $P^{\Theta_Y}$ for the preceding variables $X$ and $Y$ are available, and their joint probability distribution $P^{\Theta_{XY}}$ is needed.
- The marginals amount to constraints for the joint: we only known that the joint belongs to the set

$$\mathcal{P} = \{P^{\Theta_{XY}} | P^{\Theta_{XY}\downarrow\Theta_X} = P^{\Theta_X}, P^{\Theta_{XY}\downarrow\Theta_Y} = P^{\Theta_Y}\}.$$

- Going from the initial knowledge state $\mathcal{P}$ to any $P^{\Theta_{XY}} \in \mathcal{P}$ represents a gain of information.
- We should thus apply the maximum uncertainty principle, i.e., select the most uncertain (least informative) distribution in $\mathcal{P}$.
- Uncertainty $U(P^{\Theta})$ of a probability distribution $P^{\Theta}$ is typically evaluated using Shannon entropy: $U(P^{\Theta}) = -\sum_{\theta\in\Theta} P^{\Theta}(\theta) \log P^{\Theta}(\theta)$.
- The least informative distribution in $\mathcal{P}$ is $P^{\Theta_{XY}} = P^{\Theta_X} \times P^{\Theta_Y}$.

# Example

## Minimum uncertainty principle

- Assume an uncertain variable $Z$ about the number of days before the attack, with $\Theta_Z = \{1, \ldots, 30\}$.
- Suppose it is originally know that $Z \in \{3, 5, 7\}$ and this piece of information must be approximated by a conservative interval, i.e., $I = [\ell, u]$, with $\ell, u \in \Theta_Z$, $\ell \leq u$, such that $\{3, 5, 7\} \subseteq I$.
- Going from the initial knowledge state $\{3, 5, 7\}$ to any such $I$ represents a loss of information.
- We should thus apply the minimum uncertainty principle, i.e., select the least uncertain (most informative) such $I$.
- Uncertainty $U(A)$ of a subset $A \subseteq \Theta$ is typically evaluated by its nonspecificity, which is related to its cardinality: $U(A) = f(|A|)$, with $f$ a nondecreasing function ($f = \log_2$ is the Hartley measure).
- The most informative interval that includes $\{3, 5, 7\}$ is $I = [3, 7]$.

# Uncertainty principles in belief function theory

- In order to be able to apply the uncertainty principles with mass functions, we need a way to compare them with respect to their uncertainty (information content).
- This can be approached in two ways:
  - qualitatively, by establishing inclusion relations (partial orders) between mass functions;
  - quantitatively, through the use of uncertainty measures.

# Outline

1. Representation of evidence
   - Mass function
   - Belief and plausibility functions
2. Combination of evidence
   - Dempster's rule
   - Compatible frames
3. Comparison of evidence
   - Uncertainty principles
   - Inclusion relation
   - Uncertainty measures

# Specialization

- Given two pieces of information $X \in A$ and $X \in B$, with $A, B \subseteq \Theta$, $X \in A$ is at least as informative as $X \in B$ if $A \subseteq B$.
- An extension of this ordering between sets to mass functions is the specialization ordering.
- Given two mass functions $m_1$ and $m_2$ on $\Theta$, $m_1$ is at least as informative as $m_2$, noted $m_1 \sqsubseteq m_2$, if $m_1$ can be obtained from $m_2$ by distributing each mass $m_2(B)$ to subsets of $B$, i.e.,

$$m_1(A) = \sum_B S(A, B) m_2(B), \quad \forall A,$$

  where $S(A, B) =$ proportion of $m_2(B)$ transferred to $A \subseteq B$.
- Properties
  - Extension of set inclusion: $m_{[A]} \sqsubseteq m_{[B]} \Leftrightarrow A \subseteq B$
  - Greatest element: vacuous mass function $m_{[\Theta]}$
  - $m_1 \sqsubseteq m_2 \Rightarrow pl_1 \leq pl_2$
- Remark: other inclusion relations exist.

# Application: Ballooning extension

## Problem

- Suppose only a mass function $m(\cdot|B)$ about a variable $X$ with frame $\Theta$ is available, i.e., we have evidence about $X$ assuming that some proposition $X \in B$ holds, and the mass function $m$ such that $m(\cdot|B) = m \oplus m_{[B]}$ is needed.
- The conditional $m(\cdot|B)$ amounts to constraints for $m$: we only know that $m$ belongs to the set

$$\mathcal{M} = \{m | m(\cdot|B) = m \oplus m_{[B]}\}$$

- For a similar reason as in the previous probabilistic example, the maximum uncertainty principle should be applied, i.e., we should select the most uncertain (least informative) $m$ in $\mathcal{M}$.

# Application: Ballooning extension

## Solution

### Proposition

*The $\sqsubseteq$-least informative element $m' \in \mathcal{M}$ is obtained by transferring each mass $m(A|B)$ to $A \cup \overline{B}$:*

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

- $m'$ is known as the ballooning extension[11] of $m(\cdot|B)$.
- Recall that the vacuous extension allows us to express a state of knowledge $m^{\Xi}$ in the finer frame $\Theta$, where every element of $\Xi$ is split into several elements of $\Theta$.
- The ballooning extension allows us to express a state of knowledge $m^{\Theta}$ in an extended frame $\Theta'$, where $\Theta'$ contains all the elements of $\Theta$ and some new elements ($\Theta$ is here to $\Theta'$ what $B$ is to $\Theta$ in the proposition).

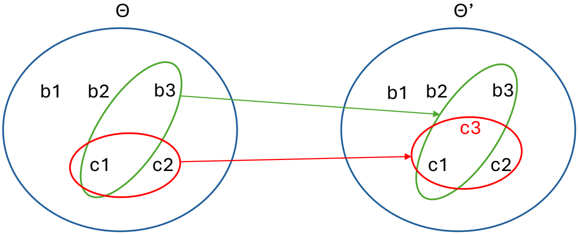
[11] aka deconditioning or conditional embedding.

# Application: Ballooning extension

Example

- Suppose a mass function $m$ about $X \in \Theta = \{c1, c2, b1, b2, b3\}$ such that $m(\{c1, b3\}) = 0.6$ and $m(\{c1, c2\}) = 0.4$.
- We learn that actually there is one more city, called $c3$, that could be the target, hence the frame is extended to $\Theta' = \{c1, c2, c3, b1, b2, b3\}$
- The ballooning extension of $m$ from $\Theta$ to $\Theta'$ is the mass function $m'$ such that

$$m'(\{c1, b3, c3\}) = 0.6, \quad m'(\{c1, c2, c3\}) = 0.4$$



# Application: Ballooning extension

## Case of product frames

- Suppose now that we have two uncertain variables $X$ and $Y$ with frames $\Theta_X$ and $\Theta_Y$.
- We have a piece of evidence about the value of $X$ assuming that some proposition $Y \in B$ holds, for some $B \in \Theta_Y$.
- This piece of evidence is modeled by a mass function on $\Theta_X$ denoted by $m_B^{\Theta_X}$.
- The mass function $m^{\Theta_{XY}}$ such that $m_B^{\Theta_X} = (m^{\Theta_{XY}} \oplus m_{[B]}^{\Theta_Y \uparrow \Theta_{XY}})^{\downarrow \Theta_X}$ is needed.
- The $\sqsubseteq$-least informative solution is the mass function on $\Theta_{XY}$ denoted $(m_B^{\Theta_X})^{\Uparrow \Theta_{XY}}$ and obtained by transferring $m_B^{\Theta_X}(A)$ to $(A \times B) \cup (\Theta_X \times \overline{B})$.

# Application: Ballooning extension

## Case of product frames: Example

- $X \in \Theta_X = \{M, \neg M\}$ is whether the target is a military base and $Y \in \Theta_Y = \{D, \neg D\}$ is whether the enemy will attack at dawn.
- We have the following piece of evidence: if the enemy attacks at dawn, then the probability to know that the target is a military base is 0.7 and the probability to know nothing is 0.3.
- This piece of evidence can be modeled by the mass function

$$m^{\Theta_X}_{\{D\}}(\{M\}) = 0.7, \quad m^{\Theta_X}_{\{D\}}(\Theta_X) = 0.3$$

- Its ballooning extension on $\Theta_{XY}$ is

$$\begin{aligned} (m^{\Theta_X}_{\{D\}})^{\Uparrow \Theta_{XY}}(\{(D, M), (\neg D, M), (\neg D, \neg M)\}) &= 0.7, \\ (m^{\Theta_X}_{\{D\}})^{\Uparrow \Theta_{XY}}(\Theta_{XY}) &= 0.3 \end{aligned}$$

- Remark: same mass function as in the previous example and corresponding to the rule “If $D$ then $M$” reliable with 0.7 probability.

# Outline

1. Representation of evidence
   - Mass function
   - Belief and plausibility functions
2. Combination of evidence
   - Dempster's rule
   - Compatible frames
3. Comparison of evidence
   - Uncertainty principles
   - Inclusion relation
   - Uncertainty measures

# Different categories of measures

- Instead of seeking the least (most) informative specialization, a measure of uncertainty can be maximized (minimized).
- As mass functions extend both the probabilistic and logical/set formalisms, there have been various proposals inspired by the measurement of uncertainty in these frameworks.
- Uncertainty measures for mass functions can be categorized into:
  - imprecision (or nonspecificity) measures (tied to the extension of sets);
  - inconsistency (or conflict) mesures (origins in both the probabilistic and logical approaches [12] );
  - total uncertainty measures, capturing both imprecision and inconsistency.
- Measures often defined for normalized mass functions, at least originally, and satisfying diverse lists of properties → Lecture 2.

[12] A.-L. Jousselme, F. Pichon, N. Ben Abdallah, S. Destercke. A note about entropy and inconsistency in evidence theory. Proc. of BELIEF 2021, pages 215-223.

# Imprecision measures

- Idea: uncertainty (in the sense of imprecision) is higher when masses are allocated to larger focal sets

$$I(m) = f\left(\sum_{\emptyset \neq A \subseteq \Theta} m(A) g(|A|)\right)$$

with $f = g = Id$ (cardinality), $f = Id$ and $g = \log_2$ (nonspecificity), $f = \log_2$ and $g = Id$ ("additive cardinality")
- Interpretation: mean imprecision
- Maximum for ignorance $m_{[\Theta]}$, minimum for contradiction $m_{[\emptyset]}$.
- $|A| \leq |B| \Leftrightarrow I(m_{[A]}) \leq I(m_{[B]})$ (only $\Rightarrow$ for nonspecificity, requires $B \neq \emptyset$ for $\Leftrightarrow$).
- Nonspecificity and additive cardinality extend the Hartley measure.
- $I(m) = f(\sum_{\theta \in \Theta} c(\theta))$ for $g = Id$.
- $I(m)$ is constant ($= f(g(1))$) for $m$ Bayesian (and it is the minimum for normalized mass functions), hence the need to measure another dimension of uncertainty.

# Inconsistency measures (1/2)

- Idea: uncertainty (in the sense of inconsistency) is higher when masses are allocated to more inconsistent focal sets

$$E(m) = -\sum_{A \subseteq \Theta} m(A) \log_2(g(A))$$

with $g = pl$ (dissonance), $g = bel$ (confusion), and $m$ normalized.

- Interpretation: $-\log_2 g(A)$ can be interpreted as a degree to which the evidence is inconsistent with focal set $X \in A$, hence mean value of the inconsistency among focal sets.
- $E(m)$ extends Shannon entropy, in particular dissonance is also maximized when $m$ is the uniform Bayesian mass function ($m(\{\theta\}) = 1/|\Theta|$ for all $\theta \in \Theta$).
- Minimized when, respectively, $m$ is fully consistent (nonempty intersection of the focal sets) and logical.

# Inconsistency measures (2/2)

- Alternative measures using the consistency measures seen earlier and a decreasing function $h$:

$$C(m) = h(1 - m_0(\emptyset))$$

  with $h(x) = 1 - x$ and $m_0 = m$ (conflict) or $m_0 = m_c$ (strong conflict), or with $h = -\log_2$ and $m_0 = m_c$ ("additive strong conflict")
- For $m_0 = m$, minimum when $m$ is consistent (i.e., normalized).
- For $m_0 = m_c$, minimum when $m$ is fully consistent.
- Maximum for contradiction $m_{[\emptyset]}$
- Normalized mass functions (only $m_0 = m_c$ useful): $C(m)$ maximum for uniform Bayesian mass function and, for $h = -\log_2$ and $m$ Bayesian, $C(m) = -\log_2 \max_{\theta \in \Theta} m(\{\theta\})$, i.e., is the min-entropy (*MinE*) of probability distribution $m$.
- Inconsistency measure $\phi$ for sets: $\phi(A) = 1$ if $A = \emptyset$, $\phi(A) = 0$ if $A \neq \emptyset$. We have $\phi(A) \leq \phi(B) \Leftrightarrow C(m_{[A]}) \leq C(m_{[B]})$.

# Total uncertainty measures (1/2)

- Idea: capture both imprecision and inconsistency in a single uncertainty measure.
- Compound measures, which add up a measure of imprecision and a measure of inconsistency.
- Entropy-based measures, which evaluate the uncertainty in a probability transformation of $m$.
- Total uncertainty of Denœux (2025)[13] is an example of both kinds

$$T(m) = I(m) + C(m) = MinE(p_m)$$

with $I$ and $C$ the additive measures, and $p_m(\theta) = \frac{c(\theta)}{\sum_{\theta' \in \Theta} c(\theta')}$.

[13] T. Denœux. Uncertainty measures in a generalized theory of evidence. Fuzzy Sets and Systems, 520:109546, 2025.

# Total uncertainty measures (2/2)

- Aggregate uncertainty

$$AU(m) = \max_{P \in \mathcal{P}(m)} S(P)$$

  where $\mathcal{P}(m)$ is the set of probability measures that dominate *bel* (credal set of $m \rightarrow$ Lecture 6) and $S$ is the Shannon entropy.
- Both $T$ and $AU$ are defined for normalized mass functions.
- Properties of $AU$ and $T$:
  - Maximum for ignorance and the uniform Bayesian mass function.
  - Minimum for certain mass functions (i.e., $m$ such that $m(\{\theta\}) = 1$ for some $\theta \in \Theta$).
  - Additive: an uncertainty measure $U$ for mass functions is additive if

$$U(m^{\Theta_X} \oplus m^{\Theta_Y}) = U(m^{\Theta_X}) + U(m^{\Theta_Y}).$$

  - $AU$ reduces to the Shannon entropy when $m$ is Bayesian and to the Hartley measure when $m$ is logical.

# Example of application of uncertainty measures

- Suppose we are given the degrees of belief of $r$ subsets, i.e., $bel(A_i) = \alpha_i$, $i = 1, \ldots, r$, for an unknown mass function $m$.
- For a similar reason as in previous examples, we should then select the most uncertain $m$ that satisfies these constraints.
- Comparing the uncertainty of mass functions using, e.g., the cardinality measure, we have then the following linear optimization problem to solve:

$$\max_m \sum_{\emptyset \neq A \subseteq \Theta} m(A)|A|$$

under the constraints:

$$\begin{aligned} \sum_{\emptyset \neq B \subseteq A_i} m(B) &= \alpha_i, \quad i = 1, \ldots, r, \\ \sum_{A \subseteq \Theta} m(A) &= 1. \end{aligned}$$

# Relation between inclusion, imprecision and inconsistency (1/3)

- For all $A, B \subseteq \Theta$, we have

$$A \subseteq B \Rightarrow |A| \leq |B| \text{ and } \phi(A) \geq \phi(B)$$

- This monotonicity wrt inclusion property extends to mass functions: for any mass functions $m$ and $m'$, and any measures $I$ and $C$,

$$m \sqsubseteq m' \Rightarrow I(m) \leq I(m') \text{ and } C(m) \geq C(m')$$

- Consequence: Consider a set $\mathcal{M}$ of mass functions that are more specialized than a mass function $m'$, i.e., for all $m \in \mathcal{M}, m \sqsubseteq m'$.
- Suppose a mass function has to be chosen from $\mathcal{M}$, as a substitute to $m'$, using some uncertainty measure $U$.
- If $U = I$, then the maximum uncertainty principle is at play: all $m \in \mathcal{M}$ are more informative, according to I, than $m'$, and to minimize information gain, the least informative, i.e., the least precise, should be chosen.

# Relation between inclusion, imprecision and inconsistency (2/3)

- Conversely, if $U = C$, then the minimum uncertainty principle is at play: all $m \in \mathcal{M}$ are less informative, according to C, than $m'$, and to minimize information loss, the most informative, i.e., the least inconsistent (conflicting), should be chosen.
- To sum up, if $U = I$, choose $\operatorname{argmax}_{m\in\mathcal{M}} I(m)$, and if $U = C$, choose $\operatorname{argmin}_{m\in\mathcal{M}} C(m)$.
- Special case: $\mathcal{M}$ is a chain wrt $\sqsubseteq$
  - Instead of using a measure $U$, using $\sqsubseteq$ to chose the substitute of $m'$ yields a unique solution: the $\sqsubseteq$-greatest element in $\mathcal{M}$
  - This element is also a solution to the above optimization problems
- Remark: if $\mathcal{M}$ is a set of less specialized mass functions than $m'$, then argmax and argmin are swapped and $\sqsubseteq$-greatest becomes $\sqsubseteq$-least.

# Relation between inclusion, imprecision and inconsistency (3/3)

- If one wants to take into account both imprecision (maximization) and inconsistency (minimization), then one will typically be facing a bi-objective optimization problem, for which various approaches may then be used.
- An alternative to account for both dimensions might be to use a measure of total uncertainty. However, some caution should be exercised, at least for $U = T$: if $m \sqsubseteq m'$, we can have either $T(m) < T(m')$ or $T(m) > T(m')$, hence neither the max nor the min uncertainty principle seem appropriate.
- One could then instead search for the mass function that preserves the most (in terms of gain or loss) the information content of $m'$, i.e., $\mathrm{argmin}_{m \in \mathcal{M}} |T(m) - T(m')|$.

# Recommended readings

## The foundation

G. Shafer.
A Mathematical Theory of Evidence.
Princeton University Press, Princeton, NJ, 1976.

## The classics

R.R. Yager and L. Liu (Eds.)
Classic Works of the Dempster–Shafer Theory of Belief Functions.
Springer, Berlin, 2008.

## The recent summary

T. Denoeux, D. Dubois and H. Prade.
Representations of Uncertainty in Artificial Intelligence: Beyond Probability and Possibility.
In P. Marquis, O. Papini and H. Prade (Eds), “A Guided Tour of Artificial Intelligence Research”, Vol.1, Chap. 4, pages 119-150, Springer Verlag, 2020.

## The intellectual autobiography

G. Shafer.
A Mathematical Theory of Evidence turns 40.
International Journal of Approximate Reasoning, 79:7–25, 2016.

Thank you for your attention.