# Uncertainty modelling in AI and engineering

## Lecture 1: Uncertainty, evidence theory

Thierry Denœux

Université de technologie de Compiègne, France
https://www.hds.utc.fr/~tdenoeux

ISCC, Fall 2026

# Part I

## Uncertainty

# Outline

# Uncertainty is everywhere

## Dennis Lindley's description

*There are things we know to be true, things we know to be false, and many things whose truth or falsity is not known to us.*

Uncertainty may concern:

- the past: who was the author of a text?
- the future: what will next year's economic growth be?
- the present: what is the value of an unknown physical constant?

Uncertainty is a consequence of lack of knowledge.

# Uncertainty is a matter of degree

We usually have some information, but not enough to identify the true answer with certainty.

## Climate sensitivity example (IPCC report)

*The likely range of equilibrium climate sensitivity has been narrowed (as compared to the previous report) to 2.5°C to 4.0°C (with a best estimate of 3.0°C) based on multiple lines of evidence, including improved understanding of cloud feedbacks.*

- This statement suggests two complementary notions:
  - what is sufficiently supported by evidence ("likely")
  - what is compatible with evidence ("possible")
- Uncertainty can decrease, and our beliefs can change, when new evidence is obtained and combined with a previous body of evidence.

# Aleatory vs. epistemic uncertainty

## Aleatory uncertainty

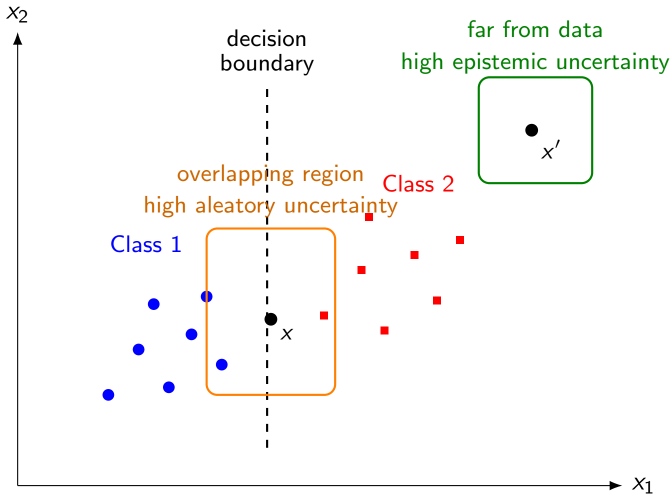
Uncertainty due to variability in the data-generating process, such as the result of tossing a coin.

## Epistemic uncertainty

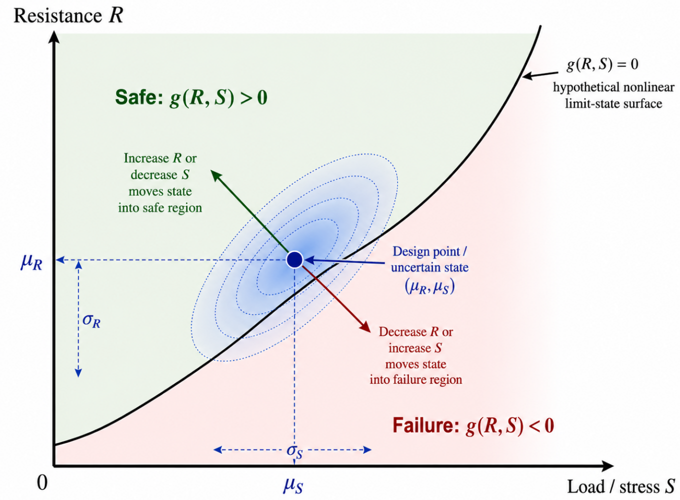
Uncertainty due to lack of knowledge, such as uncertainty about a unique future event or an unknown physical quantity.

In statistics, machine learning and engineering, both sources of uncertainty often coexist.

# Uncertainty in machine learning



# Uncertainty in reliability analysis



# A basic formal setting

We consider a question $\mathcal{Q}$ with one and only one correct answer.

## Unknown and frame

- $X$: unknown answer to the question;
- $\Theta$: set of all possible answers, called the frame of discernment;
- evidence provides only partial information about $X$.

The central problem of the course is:

How can partial information be represented, propagated and combined?

# Outline

# Set-based representations

The simplest representation of partial information is a subset $A \subseteq \Theta$.

## Interpretation

$$X \in A.$$

The true value is only known to belong to a set of possible values.

## Interval analysis

This idea underlies interval analysis: if

$$x \in [-1, 1], \qquad y \in [1, 2],$$

then for $z = \exp(x + y^2)$, interval arithmetic gives

$$y^2 \in [1, 4], \quad x + y^2 \in [0, 5], \quad z \in [1, e^5].$$

# Strengths and limits of sets

## Strength

If the inputs are guaranteed to lie in given sets, and if the model is correct, the output set is guaranteed to contain the true output.

## Limitation

Sets do not express degrees of doubt: a value is either possible or impossible.

Consequences:

- results may be too conservative to be useful;
- or too confident because some uncertainty has been ignored;
- high-dimensional boxes (Cartesian products of intervals) quickly become impractical.

# Probability theory

- Probability theory is the most widely used formalism for uncertainty.
- A probability measure requires assigning precise numerical probabilities to all relevant events.
- This is natural for repeatable random phenomena, i.e., aleatory uncertainty.
- It is questionable when information is scarce, qualitative, vague, or based on expert judgement, i.e., epistemic uncertainty.
- How can we represent evidence telling us that $X \in [a, b]$, and nothing more, using a probability distribution?

## Laplace's principle of indifference (PI)

If there is no reason to regard one possibility as more likely than another, then all possibilities should be assigned equal probability.

# Paradoxes

## Continuous case

- If we only know that $X \in [a, b]$, PI $\Rightarrow X \sim \text{Unif}([a, b])$.
- But if $Y = 1/X$, then $Y$ is not uniformly distributed on $[1/b, 1/a]$.
- The result depends on the chosen parametrisation, which is often arbitrary.

## Discrete case

- Assume that $A = \{\theta_1, \theta_2\}$. According to PI, $X \in A$ translates to

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

- If $\theta_2$ is split into two possible values $\theta_{21}$ and $\theta_{22}$, we get

$$p(\theta_1) = p(\theta_{21}) = p(\theta_{22}) = 1/3.$$

- Why should the probability of $\theta_1$ depend on the degree of detail with which $\theta_2$ is described?

# Outline

# Evidence theory

Evidence theory generalises both the set-based representations and probability theory by allowing probability mass to be assigned to subsets of $\Theta$.

## Basic idea

A source may support a proposition without specifying exactly which element of the proposition is true.

For an event $A$:

- $\mathrm{Bel}(A)$: degree to which evidence supports $A$;
- $\mathrm{Pl}(A)$: degree to which evidence is compatible $A$.

Typically,

$$\mathrm{Bel}(A) \leq \mathrm{Pl}(A).$$

# Fuzzy sets and possibility theory

- Possibility theory allows to account imprecise, vague information such as conveyed by natural language.
- Examples: "John is tall", "the temperature is high", etc.
- Such information can be represented by a fuzzy set, which captures the notion of a set with unsharp boundary. Mathematically, it is a function

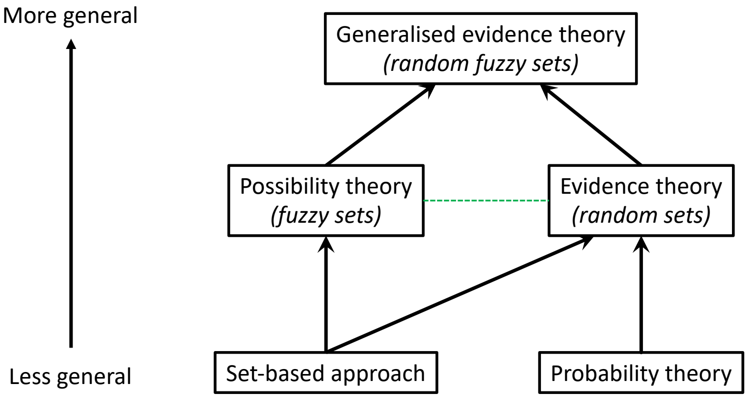
$$\widetilde{F} : \Theta \rightarrow [0,1].$$

- Evidence of the "$X$ is $\widetilde{F}$" can be represented by a possibility measure $\Pi : \mathcal{P}(\Theta) \rightarrow [0,1]$ such that

$$\Pi(A) = \sup_{\theta \in A} \widetilde{F}(\theta).$$

This is not a probability measure!

# A unified view




Random fuzzy sets provide a common framework for sets, probabilities, evidence theory and possibility theory.

# Generalised evidence theory

A unified framework based on random fuzzy sets.

## Guiding idea

Express different forms of uncertainty:

- randomness, as in random sets and probability theory;
- gradual membership, as in fuzzy sets and possibility theory.

This framework allows us to model, combine and propagate several forms of imperfect information within a single mathematical language.

# Themes covered in the course

Assuming from basic knowledge of classical uncertainty models: sets and probabilities, the course will introduce the main concepts progressively:

1. Evidence theory and belief functions
2. Random sets
3. Fuzzy sets and possibility theory
4. Random fuzzy sets as a unified framework
5. Application to uncertainty propagation
6. Application to statistical inference
7. Application to machine learning

The objective is understand the main concepts, and how to use them.

Part II

# Evidence Theory

# Roadmap

1. Representation of evidence
   - Mass functions
   - Belief, plausibility and commonality
   - Inversion and properties
2. Combination of evidence
   - Dempster's rule of combination
   - Discounting
3. Marginalisation and vacuous extension
   - Refinements and coarsenings
   - Product frames
   - Valuation-based systems

# Outline

1. Representation of evidence
   - Mass functions
   - Belief, plausibility and commonality
   - Inversion and properties
2. Combination of evidence
3. Marginalisation and vacuous extension

# Definitions

## Basic setting

- We consider a question $\mathcal{Q}$, to which there exists one and only one correct answer $X$, assumed to belong to some set $\Theta$.
- We call $X$ an unknown, and $\Theta$ its frame of discernment.

In this chapter, we assume that $\Theta$ is finite.

# Mass function

In evidence theory, we represent evidence about $X$ by assigning positive masses $m(A)$ to subsets $A \subseteq \Theta$, in such a way that the masses sum to 1.

## Mass function

A mass function on $\Theta$ is a mapping

$$m : \mathcal{P}(\Theta) \to [0,1]$$

satisfying

$$m(\varnothing) = 0, \qquad \sum_{A \subseteq \Theta} m(A) = 1.$$

- $m(A)$ is the degree to which the evidence supports $A$ as a whole. This mass is not split between elements of $A$.
- Sets $A$ such that $m(A) > 0$ are called focal sets.

# A small example

- Consider an intelligent vehicle equipped with a sensor and a machine learning algorithm recognising objects in the road scene.
- Let $X$ be the class of the object, in the $\Theta = \{\theta_1, \theta_2, \theta_3\}$, with

$$\theta_1 \equiv \text{pedestrian}, \quad \theta_2 \equiv \text{vehicle}, \quad \theta_3 \equiv \text{anything else}.$$

- Assume that the object has been classified as a pedestrian or a vehicle, but there is a 10% probability of sensor malfunction.
- This evidence can be represented by the following mass function:

$$m(\{\theta_1, \theta_2\}) = 0.9, \quad m(\Theta) = 0.1,$$

and $m(A) = 0$ for any $A \in \mathcal{P}(\Theta) \setminus \{\{\theta_1, \theta_2\}, \Theta\}$.

## Important distinction

0.90 assigned to $\{\theta_1, \theta_2\}$ is not the same as 0.45 on each singleton.

# Special mass functions

A mass function may have up to $2^{|\Theta|} - 1$ focal sets. However, in practice, we often define mass functions with a simple form (as in the previous example). A mass function is

- **Simple** if for some $A \subset \Theta$,
$$m(A) = s, \qquad m(\Theta) = 1 - s.$$
- **Categorical** if $m(A) = 1$ for one set $A \subset \Theta$; it represents certainty that the truth belongs to $A$.
- **Vacuous** if $m(\Theta) = 1$; it represents complete ignorance.
- **Consonant** if focal sets are nested: $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k$.
- **Bayesian** if all focal sets are singletons; then $m$ is equivalent to an ordinary probability mass function.

# Random-set interpretation

- Let $(\Omega, \mathcal{P}(\Omega), P)$ be a finite probability space and let $\overline{X} : \Omega \to \mathcal{P}(\Theta)$. The tuple

$$(\Omega, \mathcal{P}(\Omega), P, \Theta, \mathcal{P}(\Theta), \overline{X})$$

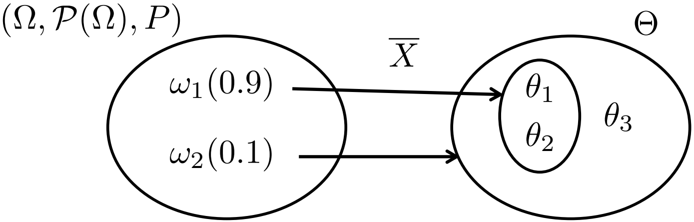
  is a (finite) random set.
- A random set induces a mass function

$$m(A) = P\{\omega : \overline{X}(\omega) = A\}.$$

- Meaning:
  - $\Omega$: set of interpretations of the evidence;
  - If $\omega \in \Omega$ holds, we know that $X \in \overline{X}(\omega)$, and nothing more.

# Back to the example

Here, $\Omega = \{\omega_1, \omega_2\}$, where $\omega_1$ stands for "the sensor is working", and $\omega_2$ stands for "the sensor is faulty".



# Three transforms of a mass function

For any $A \subseteq \Theta$,

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B), \qquad \mathrm{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B), \qquad Q(A) = \sum_{B \supseteq A} m(B).$$

**Belief** degree of support for $A$.

**Plausibility** degree of consistency with $A$ (lack of support for $A^c$).

**Commonality** degree of support for *all* elements of $A$ jointly.

$$\mathrm{Bel}(A) \leq \mathrm{Pl}(A), \qquad \mathrm{Pl}(A) = 1 - \mathrm{Bel}(A^c).$$

# Two-dimensional representation

The uncertainty about $A$ is represented by two numbers: $\text{Bel}(A)$ and $\text{Pl}(A)$, with $\text{Bel}(A) \leq \text{Pl}(A)$.

## Special cases

- $m$ vacuous: $\text{Bel}(A) = 0$ for all $A \neq \Theta$, and $\text{Pl}(A) = 1$ for all $A \neq \varnothing$.
- $m$ Bayesian: $\text{Bel}(A) = \text{Pl}(A)$ for all $A$, and Bel is a probability measure.

$$\text{Bel}(A) + \text{Bel}(A^c) \leq 1, \qquad \text{Pl}(A) + \text{Pl}(A^c) \geq 1.$$

# Computing the transforms in the example

| $A$ | $\emptyset$ | $\{\theta_1\}$ | $\{\theta_2\}$ | $\{\theta_1, \theta_2\}$ | $\{\theta_3\}$ | $\{\theta_1, \theta_3\}$ | $\{\theta_2, \theta_3\}$ | $\Theta$ |
|---|---|---|---|---|---|---|---|---|
| $m(A)$ | 0 | 0 | 0 | 0.90 | 0 | 0 | 0 | 0.10 |
| $\mathrm{Bel}(A)$ | 0 | 0 | 0 | 0.90 | 0 | 0 | 0 | 1 |
| $\mathrm{Pl}(A)$ | 0 | 1 | 1 | 1 | 0.10 | 1 | 1 | 1 |
| $Q(A)$ | 1 | 1 | 1 | 1 | 0.10 | 0.10 | 0.10 | 0.10 |

# Möbius inversion

The transformations are invertible.

## From belief to mass

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

## From commonality to mass

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

Therefore, $m$, Bel, Pl and $Q$ are equivalent representations of the same evidence.

# Key boundary conditions

We have

$$\text{Bel}(\emptyset) = \text{Pl}(\emptyset) = 0,$$

$$\text{Bel}(\Theta) = \text{Pl}(\Theta) = 1,$$

but

$$Q(\emptyset) = 1, \qquad Q(\Theta) = m(\Theta).$$

- Bel and Pl are monotone:

$$A \subseteq B \Rightarrow \begin{cases} \text{Bel}(A) \leq \text{Bel}(B) \\ \text{Pl}(A) \leq \text{Pl}(B) \end{cases}.$$

- $Q$ is antitone: $A \subseteq B \Rightarrow Q(A) \geq Q(B)$.

# Belief and plausibility are not additive

- For any subsets $A$ and $B$,

$$\text{Bel}(A \cup B) \geq \text{Bel}(A) + \text{Bel}(B) - \text{Bel}(A \cap B),$$

  where the inequality may be strict. We say that belief functions are super-additive.
- Conversely,

$$\text{Pl}(A \cup B) \leq \text{Pl}(A) + \text{Pl}(B) - \text{Pl}(A \cap B),$$

  We say that plausibility functions are subadditive.

## Reason

Masses assigned to sets intersecting both $A$ and $B$ support $A \cup B$ without supporting either $A$ or $B$ separately.

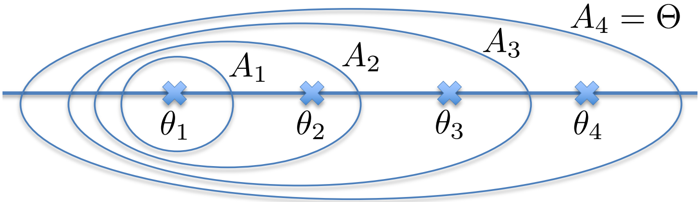
# Plausibility of singletons: contour function

- The contour function is
$$\mathrm{pl}(\theta) = \mathrm{Pl}(\{\theta\}) = \sum_{A \ni \theta} m(A).$$
- It measures how plausible each singleton is.
- It does not generally determine the whole mass function, except in special cases.

## Consonant case

If $m$ is consonant, then
$$\mathrm{Pl}(A) = \max_{\theta \in A} \mathrm{pl}(\theta).$$
For all $A, B$, we then have
$$\mathrm{Pl}(A \cup B) = \max\{\mathrm{Pl}(A), \mathrm{Pl}(B)\}, \quad \mathrm{Bel}(A \cap B) = \min\{\mathrm{Bel}(A), \mathrm{Bel}(B)\}$$

# From the contour function to the mass function

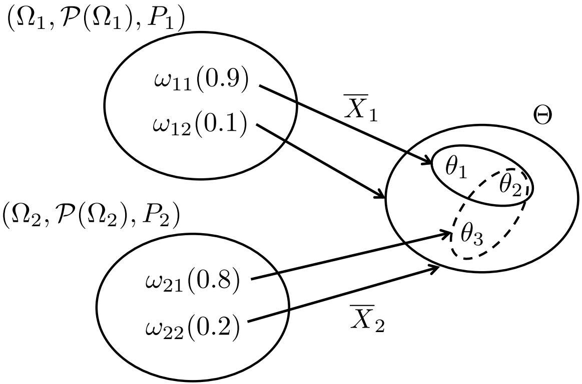



- Let pl be a contour function on the frame $\Theta = \{\theta_1, \ldots, \theta_n\}$, with elements arranged by decreasing order of plausibility, i.e.,
$$1 = \mathrm{pl}(\theta_1) \geq \mathrm{pl}(\theta_2) \geq \cdots \geq \mathrm{pl}(\theta_n),$$
and let $A_i$ denote the set $\{\theta_1, \ldots, \theta_i\}$, for $1 \leq i \leq n$.
- The corresponding mass function is
$$m(A_i) = \mathrm{pl}(\theta_i) - \mathrm{pl}(\theta_{i+1}), \quad 1 \leq i \leq n-1,$$
$$m(\Theta) = \mathrm{pl}(\theta_n).$$

# Outline

# Combination of evidence (no conflict)

A second sensor tells us that the object is not a pedestrian, and this sensor has 80% chance of functioning correctly.



# Combined random set and mass function

- Under this independence assumption, we have the joint probability measure:

$$(P_1 \otimes P_2)(\{\omega_{11}, \omega_{21}\}) = 0.72, \quad (P_1 \otimes P_2)(\{\omega_{11}, \omega_{22}\}) = 0.18,$$

$$(P_1 \otimes P_2)(\{\omega_{12}, \omega_{21}\}) = 0.08, \quad (P_1 \otimes P_2)(\{\omega_{12}, \omega_{22}\}) = 0.02,$$

- With the set-valued function

$$\overline{X}_1 \oplus \overline{X}_2 : (\omega_1, \omega_2) \mapsto \overline{X}_1(\omega_1) \cap \overline{X}_2(\omega_2),$$

we get yields the following combined mass function

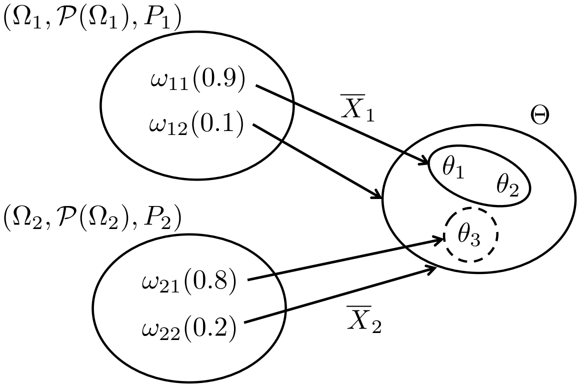
$$(m_1 \oplus m_2)(\{\theta_2\}) = 0.72, \quad (m_1 \oplus m_2)(\{\theta_1, \theta_2\}) = 0.18,$$

$$(m_1 \oplus m_2)(\{\theta_2, \theta_3\}) = 0.08 \quad (m_1 \oplus m_2)(\Theta) = 0.02.$$

- Mass function $m_1 \oplus m_2$ is called the orthogonal sum of $m_1$ and $m_2$.

# Conflict and normalisation

Assume now that the second piece of evidence tells us that the object is neither a pedestrian, nor a vehicle.



# Combined random set and mass function

- Assuming that both sensors are working leads to a contradiction. We must, therefore, conclude that at least one of the two sensors must be faulty, and condition the product probability measure $P_1 \otimes P_2$ on the set of possible pairs in $\Omega_1 \times \Omega_2$.
- Denoting the conditional probability measure as

$$P_{12}^* = (P_1 \otimes P_2)(\cdot \mid \{(\omega_{11}, \omega_{21})\}^c),$$

we get

$$P_{12}^*(\{\omega_{11}, \omega_{22}\}) = 0.18/0.28 = 9/14, \quad P_{12}^*(\{\omega_{12}, \omega_{21}\}) = 0.08/0.28 = 4/14,$$

$$P_{12}^*(\{\omega_{12}, \omega_{22}\}) = 1/14,$$

and the following combined mass function:

$$(m_1 \oplus m_2)(\{\theta_1, \theta_2\}) = 9/14, \quad (m_1 \oplus m_2)(\{\theta_3\}) = 4/14,$$

$$(m_1 \oplus m_2)(\Theta) = 1/14.$$

# Dempster's rule

Let $m_1$ and $m_2$ be two mass functions on the same frame $\Theta$.

## Degree of conflict

$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

## Orthogonal sum

If $\kappa < 1$, we define the orthogonal sum of $m_1$ and $m_2$ as

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

with $(m_1 \oplus m_2)(\emptyset) = 0$.

# Another example

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

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

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

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

# Commonality under combination

The commonality transform makes conjunctive combination simple:

$$Q_{m_1 \oplus m_2}(A) = \begin{cases} \frac{Q_{m_1}(A) Q_{m_2}(A)}{1-\kappa}, & A \neq \varnothing, \\ 1 & A = \varnothing. \end{cases}$$

## Practical message

Dempster's rule is multiplicative in the commonality domain.

# Algebraic properties

For normal mass functions on the same frame, when the rule is defined:

- Commutativity: $m_1 \oplus m_2 = m_2 \oplus m_1$.
- Associativity: $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$.
- Neutral element: the vacuous mass $m_\Theta(\Theta) = 1$.

## Consequence

Several independent pieces of evidence can be combined in any order, and ignorance changes nothing.

# What Dempster's rule assumes

Dempster's rule is appropriate when

- sources are independent and reliable;
- the frame is correctly specified.

## Warning

High conflict may indicate unreliable sources, dependence, or a wrong frame. In some cases, the initial model assumptions must be revised.

# Conditioning as a special case

- Conditioning a mass function $m$ on $B \subseteq \Theta$ is combination with the categorical mass $m_B(B) = 1$:

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

- Thus, for $A \neq \varnothing$,

$$m(A \mid B) = \frac{\sum_{C \cap B = A} m(C)}{Pl(B)}.$$

It is well defined iff $Pl(B) > 0$.

## Property

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

Generalises Bayesian conditioning.

# Computational cost

A direct combination over $\mathcal{P}(\Theta)$ can be expensive.

- Number of subsets: $2^{|\Theta|}$.
- Pairwise intersections of focal sets: $|\mathrm{Foc}(m_1)|\,|\mathrm{Foc}(m_2)|$.
- Repeated fusion may generate many focal sets.

## Key idea

Exploit structure: sparse focal sets, product frames, graphical decompositions.

# Meta-evidence about a source

A source may be unreliable.

- A sensor may fail.
- An expert may be only partially trustworthy.
- A classifier may be trained outside its domain of validity.

Let $\alpha \in [0, 1]$ denote the reliability of the source.

## Principle

Keep a fraction $\alpha$ of the source's committed evidence and transfer the remaining mass to total ignorance.

# Discounting operation

For a mass function $m$ and reliability $\alpha$,

$$\begin{aligned} {}^{\alpha}m(A) &= \alpha m(A), \qquad A \subset \Theta, \\ {}^{\alpha}m(\Theta) &= 1 - \alpha + \alpha m(\Theta). \end{aligned}$$

The quantity $1 - \alpha$ is called the discount rate.

## Special cases

- $\alpha = 1$: no discounting.
- $\alpha = 0$: vacuous mass function.
- Lower reliability means less commitment and wider belief–plausibility intervals.

# Effect on belief and plausibility

For $A \neq \Theta$,

$$\mathrm{Bel}_{^\alpha m}(A) = \alpha \, \mathrm{Bel}_m(A),$$

whereas

$$\mathrm{Pl}_{^\alpha m}(A) = 1 - \alpha + \alpha \, \mathrm{Pl}_m(A) \qquad (A \neq \emptyset).$$

## Interpretation

Discounting decreases belief in proper events and increases plausibility of their complements: it expresses loss of confidence.

# Discounting before combination

When combining several sources, reliability should normally be taken into account before fusion:

$$^{\alpha}m_1 \oplus {}^{\beta}m_2.$$

- Discounting reduces the influence of unreliable sources.
- It may also reduce conflict.
- Reliability is meta-evidence about the source, not evidence about the state itself.

# Outline

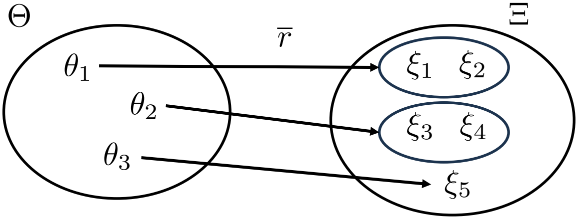
# Why change the frame?

- The definition of a frame is often a matter of convention.
- In the object recognition example, we could have defined a finer frame distinguishing, e.g.,
  - different kinds of vehicles (bicycles, cars, etc.);
  - different kinds of objects in the "anything else" class (such as "animal", "road sign", etc.).
- Different sources of information may provide evidence represented in frames of different granularities.

**Goal**

Express the same piece of evidence with different levels of detail without adding unwarranted information.

# Refinement and coarsening




- A frame $\Xi$ is a refinement of $\Theta$ iff there exists a mapping (called a refining)

$$\overline{r} : \mathcal{P}(\Theta) \rightarrow \mathcal{P}(\xi)$$

  such that
  - the subsets $\{\overline{r}(\{\theta\}) : \theta \in \Theta\}$ form a partition of $\Xi$;
  - for all $A \subseteq \Theta$, $\overline{r}(A) = \bigcup_{\theta \in A} \overline{r}(\{\theta\})$.
- We then say that $\Theta$ is a coarsening of $\Xi$.

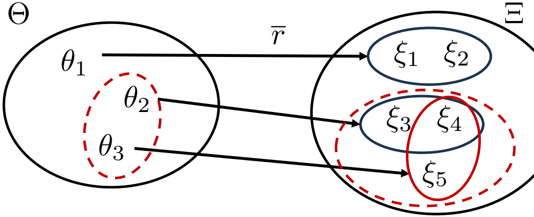
# Vacuous extension and restriction

- From $\Theta$ to $\Xi$: given a mass function $m^{\Theta}$ on $\Theta$, its vacuous extension $m^{\Theta\uparrow\Xi}$ in $\Xi$ is obtained by transferring each mass $m^{\Theta}(A)$ to $\overline{r}(A)$.
- From $\Xi$ to $\Theta$: given a mass function $m^{\Xi}$ on $\Xi$, its restriction $m^{\Xi\downarrow\Theta}$ in $\Theta$ is obtained by transferring each mass $m^{\Xi}(B)$ to the set

$$B^* := \{\theta \in \Theta : \overline{r}(\{\theta\}) \cap B \neq \varnothing\}.$$

- Example: $\{\xi_4, \xi_5\}^* = \{\theta_1, \theta_2\}$.

Going from a refined frame to a coarser frame loses detail.



# Product frames

- Let $X_1, \ldots, X_n$ be $n$ unknowns with frames $\Theta_1, \ldots, \Theta_n$.
- The product space is

$$\Theta = \Theta_1 \times \cdots \times \Theta_n.$$

- For $I \subseteq \{1, \ldots, n\}$, write

$$\Theta_I = \prod_{i \in I} \Theta_i,$$

and

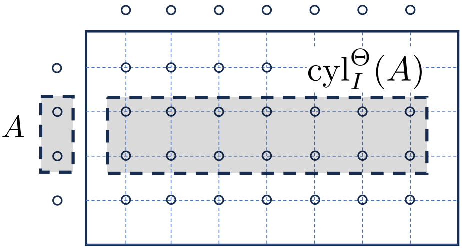
$$\Theta = \Theta_I \times \Theta_J, \quad J = I^c.$$

# Vacuous extension

The cylindrical extension of $A \subseteq \Theta_I$ in $\Theta$ is

$$\text{cyl}_I^\Theta(A) = A \times \Theta_J.$$

The mapping $\text{cyl}_I^\Theta : A \mapsto \text{cyl}_I^\Theta(A)$ is a refining.




## Vacuous extension

Let $m^{\Theta_I}$ be a mass function on $\Theta_I$. Its vacuous extension to $\Theta$ is

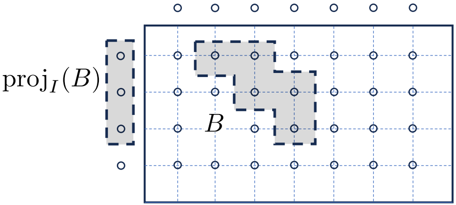
$$m^{\Theta_I \uparrow \Theta}(B) = \begin{cases} m^{\Theta_I}(A) & \text{if } B = \text{cyl}_I^\Theta(A) \text{ for some } A \subseteq \Theta_I, \\ 0, & \text{otherwise.} \end{cases}$$

Each mass $m^{\Theta_I}(A)$ is transferred to the cylindrical extension of $A$.

# Marginalisation

The restriction of $B \subseteq \Theta$ in $\Theta$ is its projection, defined as

$$\mathrm{proj}_I(B) = \{x_I \in \Theta_I : \exists x_J \in \Theta_J, (x_I, x_J) \in B\}.$$




## Marginal

Let $m$ be a mass function on $\Theta$. Its marginal on $\Theta_I$ is

$$m^{\Theta \downarrow \Theta_I}(A) = \sum_{B \subseteq \Theta:\ \mathrm{proj}_I(B) = A} m^{\Theta}(B).$$

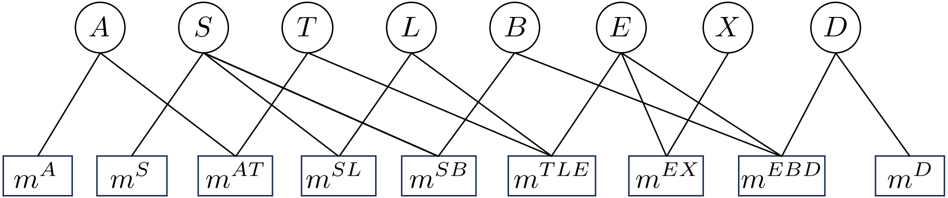
- Each joint focal set is projected onto the variables of interest.
- Marginalisation discards information about the remaining variables.
- It is the belief-function analogue of probabilistic marginalisation.

# Valuation-based systems

- Many problems can be formalised by defining
  - $n$ unknown $X_1, \ldots, X_n$ in frames $\Theta_i$, $i = 1, \ldots, n$, and
  - $q$ mass functions $m_1, \ldots, m_q$, each $m_i$ being defined in frame $\Theta_{I(i)} = \prod_{i=1}^{I(i)} \Theta_i$ for some subset $I(i) \subseteq \{1, \ldots, n\}$.
- Such a model is called an evidential valuation-based system (VBS).

# Chest clinic example

*Shortness-of-breath (**D**yspnoea) may be due to **T**uberculosis, **L**ung cancer or **B**ronchitis, or none of them, or more than one of them. A recent visit to **A**sia increases the chances of tuberculosis, while **S**moking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest **X**-ray detects that the patient has **E**ither lung cancer and tuberculosis; dyspnoea also does not discriminate between lung cancer and tuberculosis.*



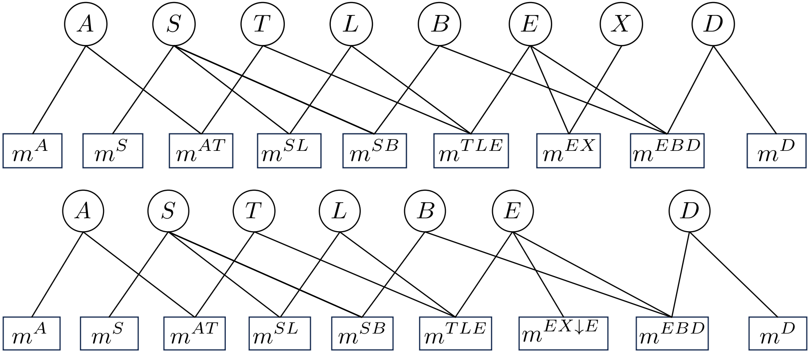
# Fusion algorithm: local combination and deletion

- Denote by $X_J$, $J \subset \{1, \ldots, n\}$ the subset of target variables.
- We want to compute

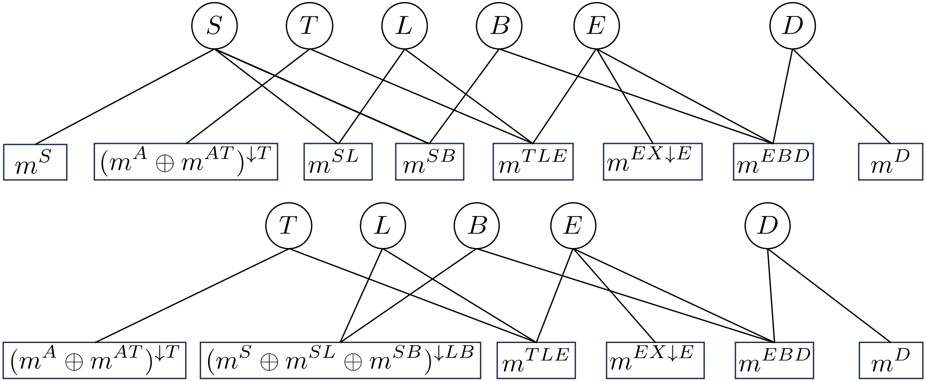
$$m^{\Theta_J} = (m_1 \oplus \ldots \oplus m_q)^{\downarrow \Theta_J}.$$

- The Fusion algorithm deletes variables in $J^c$ one at a time by iterating the following steps:
  1. combine all valuations involving a variable to be eliminated;
  2. marginalise the result to the remaining variables;
  3. continue until only target variables remain.
- Localisation reduces complexity when each piece of evidence concerns few variables.
- Different deletion sequences may induce different computational costs.
- One-step lookahead heuristic: select as the next unknown for deletion the one that leads to combination over the smallest state space, with ties broken randomly.

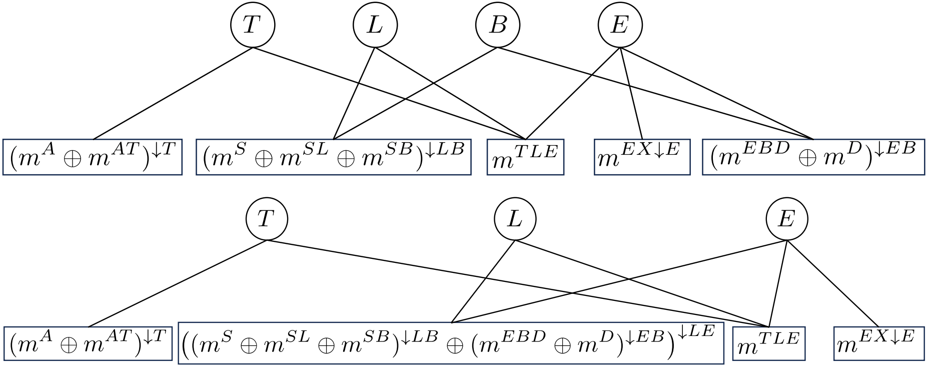
# Example (1/4)


A
S
T
L
B
E
X
D
m^A
m^S
m^AT
m^SL
m^SB
m^TLE
m^EX
m^EBD
m^D
A
S
T
L
B
E
D
m^A
m^S
m^AT
m^SL
m^SB
m^TLE
m^EX↓E
m^EBD
m^D

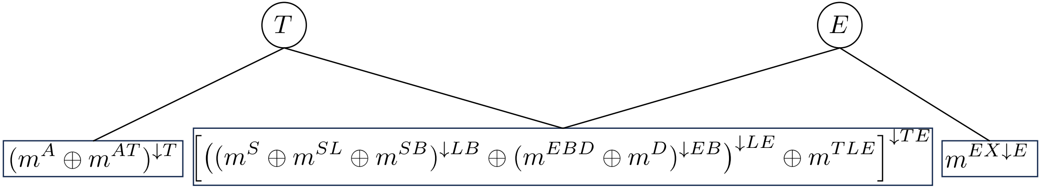
# Example (2/4)

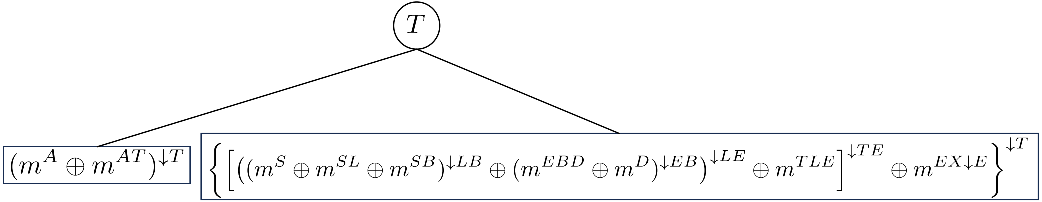

$S$
$T$
$L$
$B$
$E$
$D$
$m^S$
$(m^A \oplus m^{AT})^{\downarrow T}$
$m^{SL}$
$m^{SB}$
$m^{TLE}$
$m^{EX\downarrow E}$
$m^{EBD}$
$m^D$
$T$
$L$
$B$
$E$
$D$
$(m^A \oplus m^{AT})^{\downarrow T}$
$(m^S \oplus m^{SL} \oplus m^{SB})^{\downarrow LB}$
$m^{TLE}$
$m^{EX\downarrow E}$
$m^{EBD}$
$m^D$

# Example (3/4)


$T$
$L$
$B$
$E$
$(m^A \oplus m^{AT})^{\downarrow T}$
$(m^S \oplus m^{SL} \oplus m^{SB})^{\downarrow LB}$
$m^{TLE}$
$m^{EX\downarrow E}$
$(m^{EBD} \oplus m^D)^{\downarrow EB}$
$T$
$L$
$E$
$(m^A \oplus m^{AT})^{\downarrow T}$
$((m^S \oplus m^{SL} \oplus m^{SB})^{\downarrow LB} \oplus (m^{EBD} \oplus m^D)^{\downarrow EB})^{\downarrow LE}$
$m^{TLE}$
$m^{EX\downarrow E}$

# Example (4/4)


$T$
$E$
$(m^A \oplus m^{AT})^{\downarrow T}$
$\left[\left((m^S \oplus m^{SL} \oplus m^{SB})^{\downarrow LB} \oplus (m^{EBD} \oplus m^D)^{\downarrow EB}\right)^{\downarrow LE} \oplus m^{TLE}\right]^{\downarrow TE}$
$m^{EX \downarrow E}$


$T$
$(m^A \oplus m^{AT})^{\downarrow T}$
$\left\{\left[\left((m^S \oplus m^{SL} \oplus m^{SB})^{\downarrow LB} \oplus (m^{EBD} \oplus m^D)^{\downarrow EB}\right)^{\downarrow LE} \oplus m^{TLE}\right]^{\downarrow TE} \oplus m^{EX \downarrow E}\right\}^{\downarrow T}$

# Summary of the lecture

- A mass function allocates evidence to subsets of the frame.
- Belief, plausibility and commonality are equivalent transforms with different interpretations.
- Dempster's rule combines independent evidence by intersecting random sets and conditioning on non-empty intersections.
- Discounting expresses partial reliability of sources.
- Vacuous extension and marginalisation allow evidence to be handled in product frames.
- The Fusion algorithm uses the local structure of the evidence to combine and marginalise mass functions efficiently.

# Further reading

cf. http://www.hds.utc.fr/~tdenoeux


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

T. Denœux, 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", Volume 1, Chapter 4, Springer Verlag, pages 119-150, 2020.

T. Denœux.
*Modern Evidence Theory: A Unified Approach to Uncertainty in AI, Engineering and Statistics*.
Book manuscript (forthcoming).