

Imprecise Probabilistic Machine Learning

Winter Semester 2026/27

Saarland University / University of Hamburg / Public Enrolment

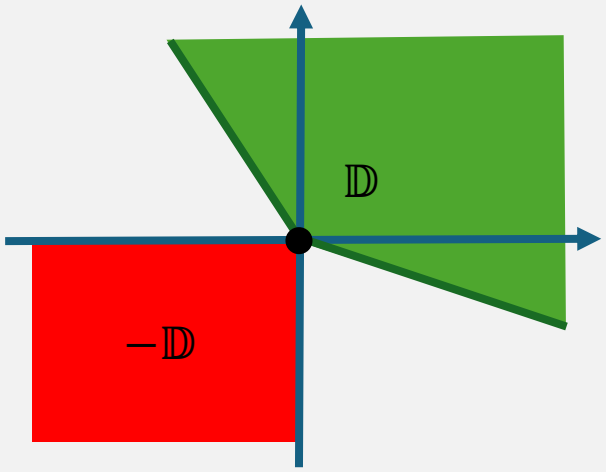
Lecture 4: Possibility Theory

Krikamol Muandet

06 November 2025

Recap

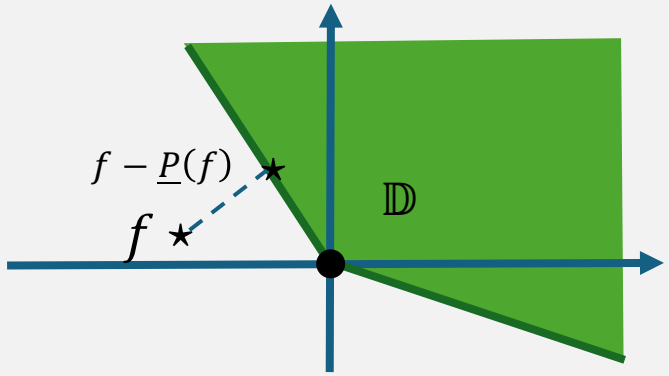
Acceptable Gambles



(A1) If $f < 0$, then $f \notin \mathbb{D}$
(A2) If $f \geq 0$, then $f \in \mathbb{D}$
(A3) If $f \in \mathbb{D}$, then $\lambda f \in \mathbb{D}$ $(\forall \lambda \in \mathbb{R}, \lambda > 0)$
(A4) If $f \in \mathbb{D}$ and $g \in \mathbb{D}$, then $f + g \in \mathbb{D}$

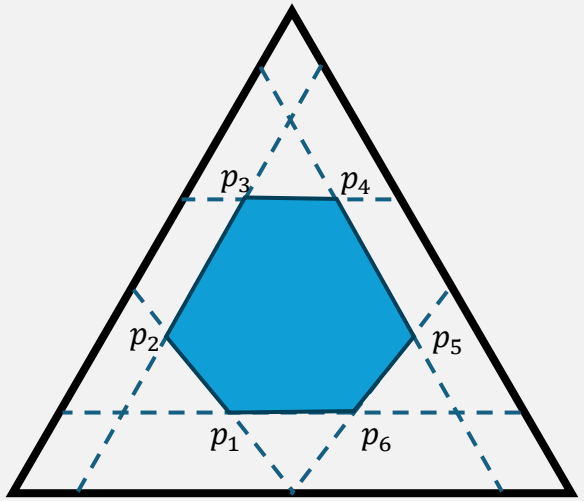
Lower Prevision

$$\underline{P}(f) := \sup\{\mu \in \mathbb{R} : f - \mu \in \mathbb{D}\}$$



(LP1). $\underline{P}(f) \geq \inf_{\omega \in \Omega} f(\omega)$
(LP2). $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$
(LP3). $\underline{P}(\lambda f) = \lambda \underline{P}(f)$

Credal Set



$$\underline{P}(f) = \inf_{P \in \mathcal{M}[\underline{P}]} P(f),$$
$$\mathcal{M}[\underline{P}] = \{P \in \mathcal{P} : P(f) \geq \underline{P}(f), \forall f \in \mathbb{D}\}$$

Possibility Theory

Will it rain tomorrow?

It is **probable** that it will rain tomorrow



10% 20% 70%

“Rain is more likely than the alternatives.”

It is **possible** that it will rain tomorrow



Rain cannot be ruled out

“Possible does not necessarily mean likely.”

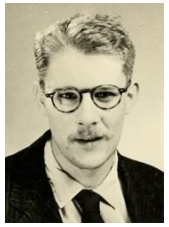
Probability distributes *likelihood*. **Possibility** expresses *compatibility*

Historical Background



G. L. S. Shackle (1940-1970's) called *degree of potential surprise* of an event its *degree of impossibility* (retrospectively, the degree of necessity of the opposite event)

- **Epistemic:** Impossibility is understood as disbelief.
- Potential surprise is valued on a disbelief scale $[0, u^*]$.



D. Lewis (1973) considered a relation between possible worlds that he call *comparative possibility* (counterfactuals).

"A world j is at least as similar to world i as world k is."



L. J. Cohen (1977) considered the problem of legal reasoning and introduced so-called *Baconian probabilities (degree of provability)*. A hypothesis and its negation cannot both be provable together to any extent.



L. A. Zadeh (1978) proposed an interpretation of membership functions of fuzzy sets as possibility distributions.

Possibility Distributions

☀️ Sunny	⛅ Cloudy	🌧️ Rainy
1.0	0.7	0.3
Fully compatible	Less compatible	Still possible, less compatible

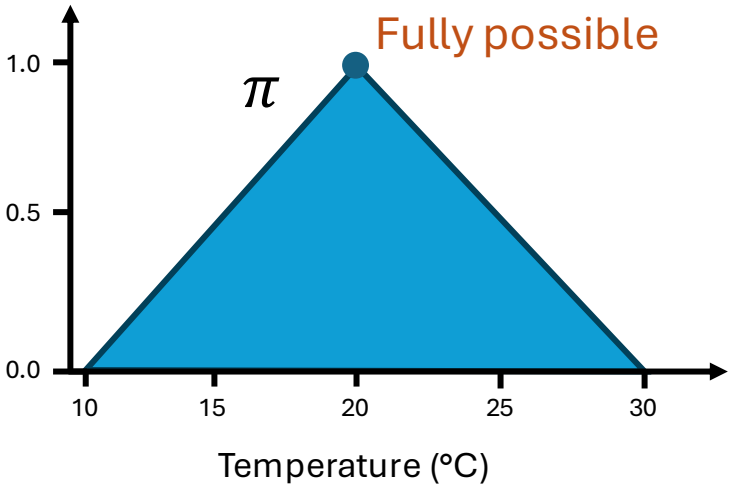
$$\pi: \Omega \to [0,1], \qquad \max_{\omega \in \Omega} \pi(\omega) = 1 \qquad (\text{finite } \Omega)$$



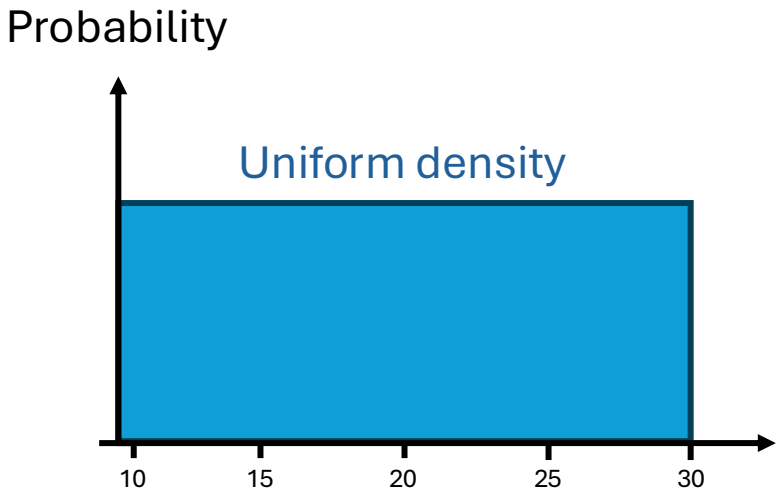
- Possibility 1 does not mean certainty. Possibility values need not sum to 1.
- The function $\pi(\omega)$ represents **an agent's knowledge about which state in Ω is actual.**
  - $\pi(\omega) = 0 \Longrightarrow$ state ω is rejected as **impossible**.
  - $\pi(\omega) = 1 \Longrightarrow$ state ω is **totally possible (plausible)**.
- **Normalization**: At least one state in Ω remains fully possible:

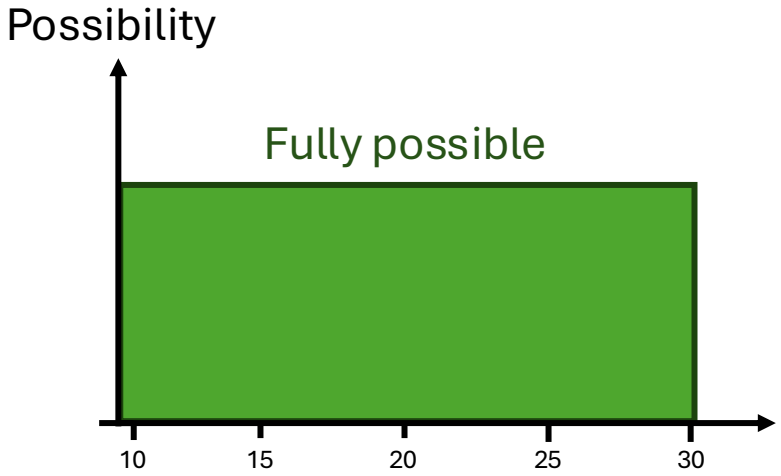
$$\exists \omega \in \Omega, \qquad \pi(\omega) = 1$$

Possibility Distributions



The area under the curve need not equal one







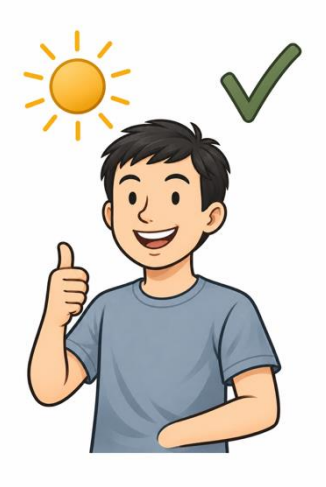
Principle of Minimal Specificity

"Any hypothesis not known to be impossible cannot be ruled out."

- A possibility distribution π is at least as **specific** as another π' if and only if for each state of affairs ω: $\pi(\omega) \leq \pi'(\omega)$.

$$(1, 0.3, 0.5) \leq (1, 0.5, 0.7) \leq (1, 0.8, 0.9)$$

- Two extreme forms of knowledge in possibilistic framework:



Complete knowledge

$\pi(\omega_0) = 1$ and $\pi(\omega) = 0$ for all $\omega \neq \omega_0$

Example: $\pi = (1, 0, 0)$



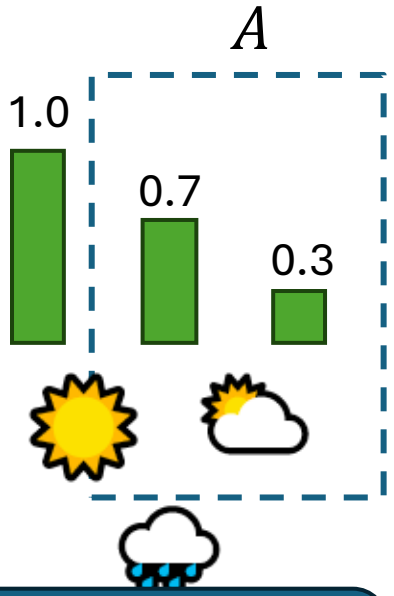
Complete ignorance

$\pi(\omega) = 1$ for all $\omega \in \Omega$

Example: $\pi = (1, 1, 1)$

From States to Events

$\Omega = \{\text{sunny, cloudy, rainy}\}, \qquad A = \{\text{cloudy, rainy}\} = \text{“not sunny”}$



An event is **possible** to the degree of its most possible member:

$$\Pi(A) = \max_{\omega \in A} \pi(\omega)$$

$$\Pi(A) = \Pi(\{\text{cloudy, rainy}\}) = \max(0.7, 0.3) = 0.7$$

To keep “cloudy or rainy” possible, at least one of these states must remain possible. The most compatible state determines the event’s possibility.

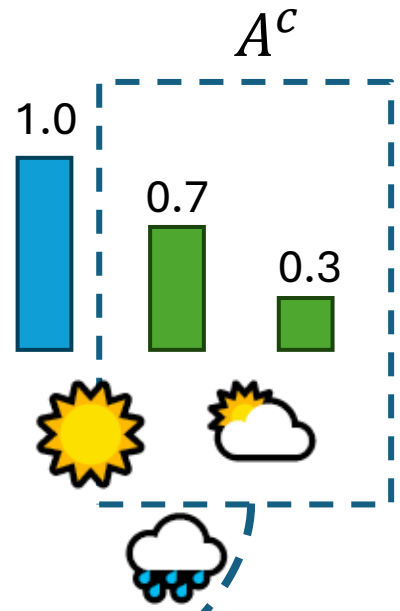
Combining alternatives uses the maximum; their possibilities do not accumulate.

$$\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$$

Exercise: Find the possibility values of the events $\emptyset$, Ω, and $\{\omega\}$:

$$\Pi(\emptyset), \Pi(\Omega), \Pi(\{\omega\})$$

Possibility versus Necessity



We found that “cloudy or rainy” has possibility 0.7. What does this tell us about “sunny”?

Possibility: look inside the event	**Necessity:** look outside the event
How compatible is **sunny** with our knowledge?	How strongly is **not sunny** excluded?
$\Pi(A) = \max_{\omega \in A} \pi(\omega)$	$\mathrm{N}(A) = 1 - \Pi(A^c)$
$\Pi(\text{sunny}) = 1$	$\mathrm{N}(\text{sunny}) = 1 - 0.7 = 0.3$

“Sunny is fully possible, but alternatives remain possible – so sunny is not certain.”

Exercise: What is the interpretation of this?

$$\Pi(A) = \Pi(A^c) = 1 \Longrightarrow \mathrm{N}(A) = \mathrm{N}(A^c) = 0$$

Possibility and Necessity Measures

From a possibility distribution π, we can derive the **possibility** and **necessity** measures:

Possibility (upper probability)

$$\Pi(A) \coloneqq \max_{\omega \in A} \pi(\omega)$$

Necessity (lower probability)

$$N(A) \coloneqq \min_{\omega \notin A} (1 - \pi(\omega))$$

"Does event A occur?"

- $\Pi(A)$ evaluates to what extent A is **consistent with π**
- $\mathrm{N}(A)$ evaluates to what extent **alternatives are excluded**

Possibility and Necessity Measures

A possibility measure satisfies:

1. **Non-negativity**: $\pi(\omega) = \Pi(\{\omega\}) \geq 0$ for all outcomes $\omega \in \Omega$
2. **Maxitivity**: $\Pi(A) = \max_{\omega \in A} \pi(\omega)$ for all events $A \subseteq \Omega$
3. **Normed**: $\Pi(\Omega) = 1$

- The ***possibility-necessity duality***:

$$N(A) = 1 - \Pi(A^c).$$

- Generally, $\Pi(\Omega) = \mathrm{N}(\Omega) = 1$ and $\Pi(\emptyset) = \mathrm{N}(\emptyset) = 0$.

Connections to Imprecise Probability

- A possibility measure induces a **credal set**:

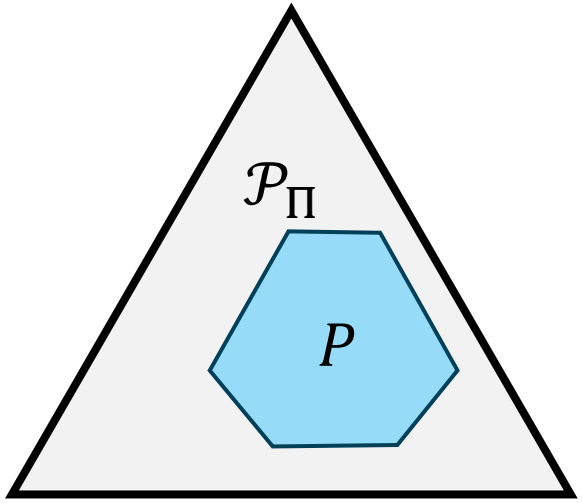
$$\mathcal{P}_\Pi = \{P \in \Delta(\Omega) : P(A) \leq \Pi(A), \forall A \subseteq \Omega\}$$

Lower Envelope

$$N(A) = \inf_{P \in \mathcal{P}_\Pi} P(A)$$

Upper Envelope

$$\Pi(A) = \sup_{P \in \mathcal{P}_\Pi} P(A)$$



Important: A credal set induced by a possibility measure has a specific structure, known as a *consonant* credal model (to be discussed further in the lecture on belief function theory).

- For every compatible probability distribution $P \in \mathcal{P}_\Pi$:

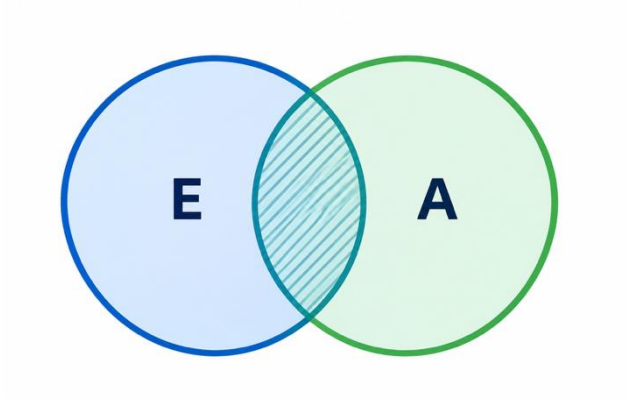
$$N(A) \leq P(A) \leq \Pi(A)$$

- A possibility distribution specifies a family of compatible precise probability distributions, rather than a unique distribution.

Boolean Case

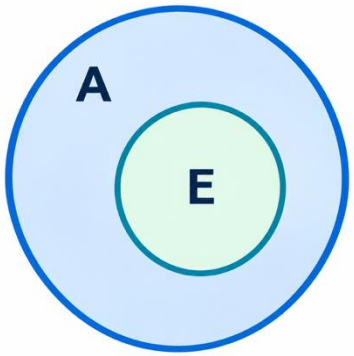
The possibility distribution comes down to the disjunctive (epistemic) set $\mathrm{E} \subseteq S$

$\Pi(A) = 1$ if $A \cap E \neq \emptyset$, and 0 otherwise.



The function Π checks whether proposition A is logically **consistent with** the available information or not.

$N(A) = 1$ if $E \subseteq A$, and 0 otherwise.



The function N checks whether proposition A is logically **entailed by** the available information or not.

Basic Properties

- $\Pi(A \cup \mathrm{B}) = \max\big(\Pi(A), \Pi(B)\big)$
  If at least one of A or B is possible enough, then their union is as possible as the more plausible one.

- $N(A \cap \mathrm{B}) = \min\big(N(A), N(B)\big)$
  The certainty of a conjunction can only be as strong as the weakest link.

- $\Pi(A) = 1$ whenever $N(A) > 0$

Certainty Qualification

- "**A is certain to degree α**" can be modeled by the constraint $N(A) \geq \alpha$.
- It represents a family of possible epistemic states π that obey this constraint.
- The *least specific* possibilistic distribution:

$$\pi_{(A,\alpha)}(s) = \begin{cases} 1 & \text{if } s \in A \\ 1 - \alpha & \text{otherwise} \end{cases}$$

- If $\alpha = 1$, we get the **characteristic function** of A.
- If $\alpha = 0$, we get **total ignorance**.

Joint Possibility Distributions

- Possibility distributions over Cartesian products of domains $S_1 \times S_2 \times \cdots \times S_m$ are called **joint possibility distributions**:

$$\pi(s_1, s_2, \ldots, s_m)$$

- The projection of $\pi_k^{\downarrow}$ of π onto S_k is defined as

$$\pi_k^{\downarrow}(s_k) = \Pi(S_1 \times \cdots S_{k-1} \times \{s_k\} \times \cdots S_{k+1} \times S_m) = \sup_{s_i \in S_i, i \neq k} \pi(s_1, \ldots, s_n)$$

- It's clear that $\pi(s_1, \ldots, s_n) \leq \min_{k=1}^{m} \pi_k^{\downarrow}(s_k)$: *A joint possibility distribution is at least as specific as the Cartesian product of its projections*.

Conditioning

- Conditioning can be defined from a Bayesian-like equation:

$$\Pi(B \cap A) = \Pi(B|A) \star \Pi(A)$$

where $\Pi(A) > 0$ and $\star$ is a t-norm. Moreover, $N(B|A) = 1 - \Pi(B^c|A)$.

- If operation $\star$ is the minimum, the **minimal specificity principle** gives

$$\Pi(B|A) = \begin{cases} 1 & \text{if } \Pi(B \cap A) = \Pi(A) > 0 \\ \Pi(B \cap A) & \text{otherwise} \end{cases}$$

- We have $N(B|A) > 0$ if and only if $\Pi(B \cap A) > \Pi(\mathrm{B}^c \cap A)$.
- Also, if $\Pi(B|A) > \Pi(B)$, then $\Pi(B|A) = 1$.
- In numerical setting, $\star$ is a product so that $\Pi(B|A) = \Pi(B \cap A)/\Pi(A)$

Possibilistic Independence

- **Unrelatedness:** $\Pi(A \cap B) = \min\big(\Pi(A), \Pi(B)\big)$

  When it doesn't hold, it indicate an epistemic form of mutual exclusion between A and B.

- **Causal independence:** $\Pi(B \mid A) = \Pi(B)$

  A form of directed epistemic independence whereby learning A doesn't affect the plausibility of B.

Intepretations of Possibility

1. **Qualitative** possibility theory
   - The possibility space S is the set of interpretations of a formal propositional language based on a finite set of Boolean attributes.
   - Connections with modal logic, comparative possibility, non-monotonic inference, and possibilistic logic, etc

2. **Quantitative** possibility theory
   - The possibility degrees range in the unit interval, and are considered in connection with belief function and imprecise probability theory.
   - A degree of possibility can be viewed as an **upper probability bound**.

Applications in Machine Learning

Robust Clustering and Classification

- **Possibilistic C-Means (PCM)**
  - Relax normalisation: membership degrees need not sum to one.
  - Objective: minimise intra-cluster distance and penalise low membership.
  - Handles overlapping clusters and noisy samples.
  - Krishnapuram & Keller, 1993; Lesot & Bouchon-Meunier, 2003.

- **Possibilistic SVMs and k-NN**
  - Replace probabilistic margin with possibility degrees.
  - Allows independent plausibility for each class.
  - Robust to label noise and outliers.
  - Lee et al., 2005; Denoeux, 1995.

Learning from Imprecise, Incomplete, or Ambiguous Data

- Uncertain labels or features (intervals, linguistic descriptors, partial labels)
- Represent these uncertainties as **possibility distributions:** $\pi(x_i)$ reflecting compatibility with each class or value.
- Learning algorithms can treat $\pi(x_i)$ as upper bounds on probabilities and optimise with respect to **necessity-weighted** or **possibility-weighted** losses.
- **Possibilistic regression**: Instead of minimising log-likelihood, minimise possibility-based loss:

$$L(\theta) = \sum_{\mathrm{i}=1}^{\mathrm{n}} -\log \Pi_\theta(y_i|x_i)$$

which penalises **inconsistency** more than **improbability**.

Information Fusion and Ensemble Learning

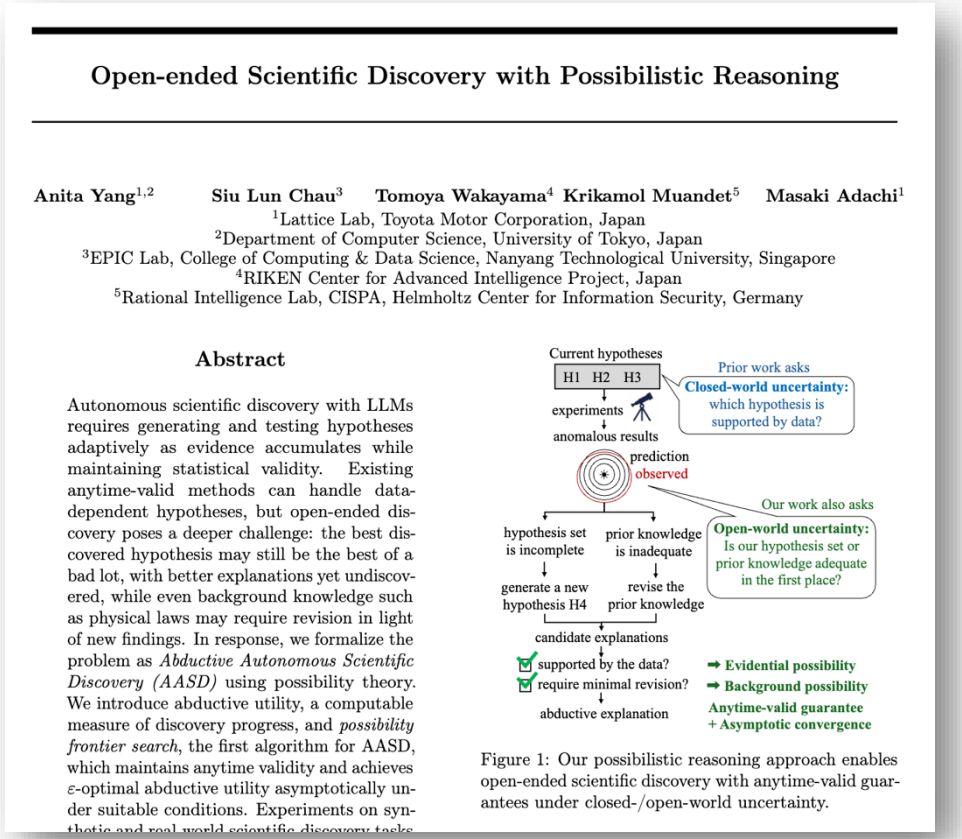
- In ML systems, information comes from *multiple, uncertain sources*: different sensors, data modalities (text, image, audio), ensemble models, human experts, or even different "**belief models**" (causal, probabilistic, symbolic).
- Aggregate multiple uncertain sources via max-min combination:

$$\Pi_{\text{fused}}(A) = \max_i \min(\Pi_{\text{i}}(A), w_i)$$

- Handles conflicting evidence and avoids overconfidence.
- **Applications**: Sensor and multi-modal fusion, expert opinion aggregation, ensemble model integration under epistemic disagreement

Open-ended Scientific Discovery

Possibilistic reasoning is particularly suitable for scientific discovery because **it allows us to reason about what is possible** without prematurely committing to what is probable.



Open-ended Scientific Discovery with Possibilistic Reasoning

Anita Yang[1,2] Siu Lun Chau[3] Tomoya Wakayama[4] Krikamol Muandet[5] Masaki Adachi[1]
[1]Lattice Lab, Toyota Motor Corporation, Japan
[2]Department of Computer Science, University of Tokyo, Japan
[3]EPIC Lab, College of Computing & Data Science, Nanyang Technological University, Singapore
[4]RIKEN Center for Advanced Intelligence Project, Japan
[5]Rational Intelligence Lab, CISPA, Helmholtz Center for Information Security, Germany

Abstract

Autonomous scientific discovery with LLMs requires generating and testing hypotheses adaptively as evidence accumulates while maintaining statistical validity. Existing anytime-valid methods can handle data-dependent hypotheses, but open-ended discovery poses a deeper challenge: the best discovered hypothesis may still be the best of a bad lot, with better explanations yet undiscovered, while even background knowledge such as physical laws may require revision in light of new findings. In response, we formalize the problem as *Abductive Autonomous Scientific Discovery (AASD)* using possibility theory. We introduce abductive utility, a computable measure of discovery progress, and *possibility frontier search*, the first algorithm for AASD, which maintains anytime validity and achieves ε-optimal abductive utility asymptotically under suitable conditions. Experiments on synthetic and real-world scientific discovery tasks

Figure 1: Our possibilistic reasoning approach enables open-ended scientific discovery with anytime-valid guarantees under closed-/open-world uncertainty.

- **Preserve alternatives**: Do not prematurely exclude plausible scientific explanations.
- **Refine through evidence**: Progressively constrain the space of possible hypotheses.
- **Avoid unjustified precision**: Represent what the evidence warrants, rather than imposing precise probabilities.

Paper: https://arxiv.org/pdf/2610.11289

Recommended Reading

- **Possibility Theory and Statistical Reasoning**
  By Didier Dubois (2006)
- **Possibility Theory and its Applications: Where Do we Stand ?**
  By Didier Dubois and Henri Prade (2014)
- **Possibilistic inferential models: a review**
  By Ryan Martin (2025)