IPML

IMPRECISE PROBABILISTIC MACHINE LEARNING

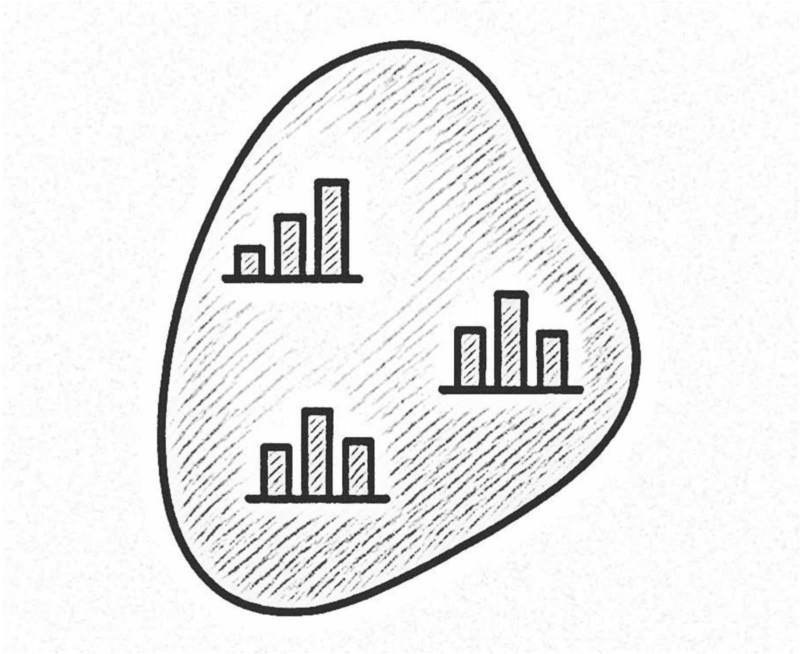
Lecture 3: Possibility Theory

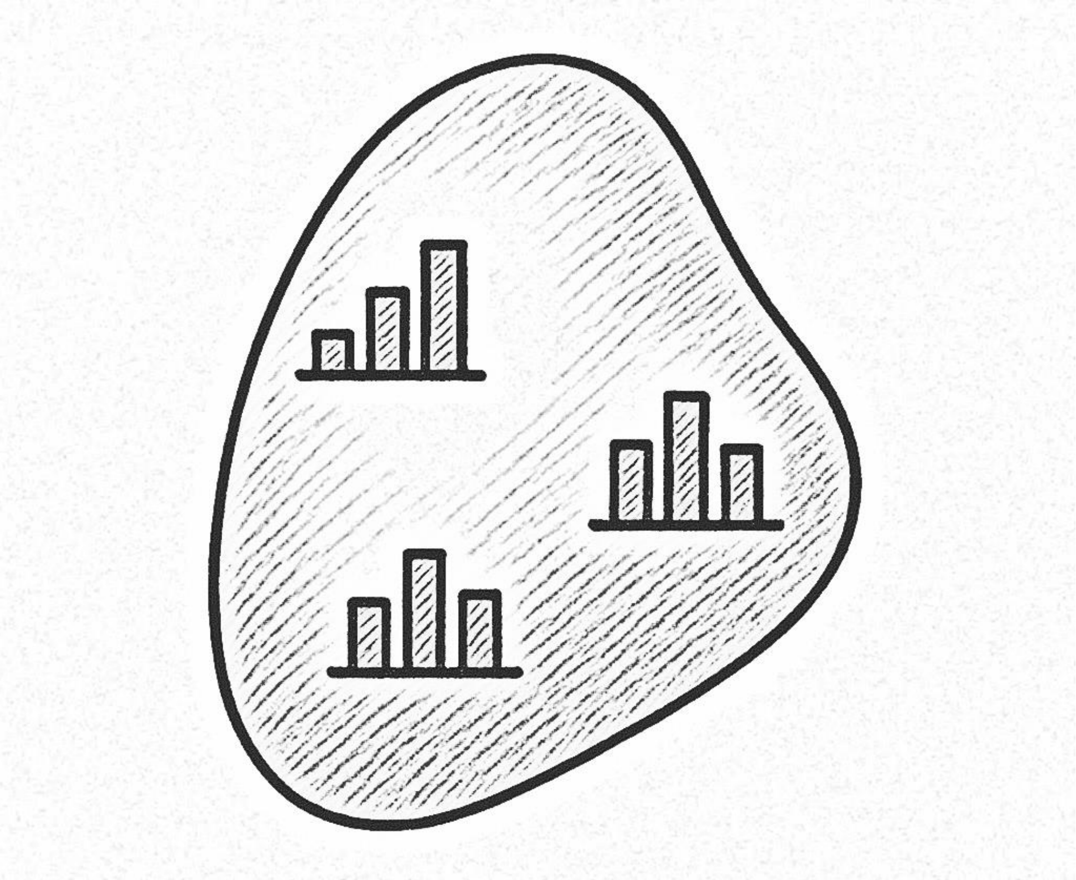
Krikamol Muandet

07 November 2025

Outline

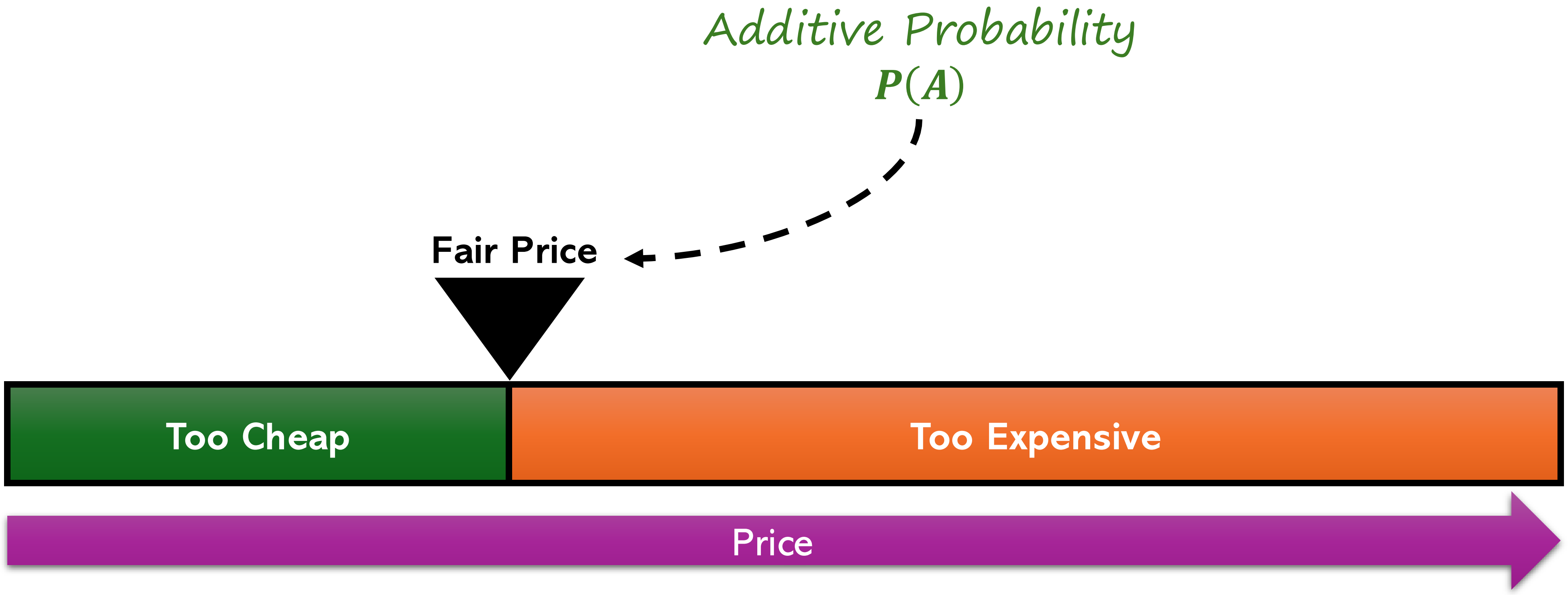
1. Imprecise Probability
2. Possibility Theory
3. Applications in Machine Learning



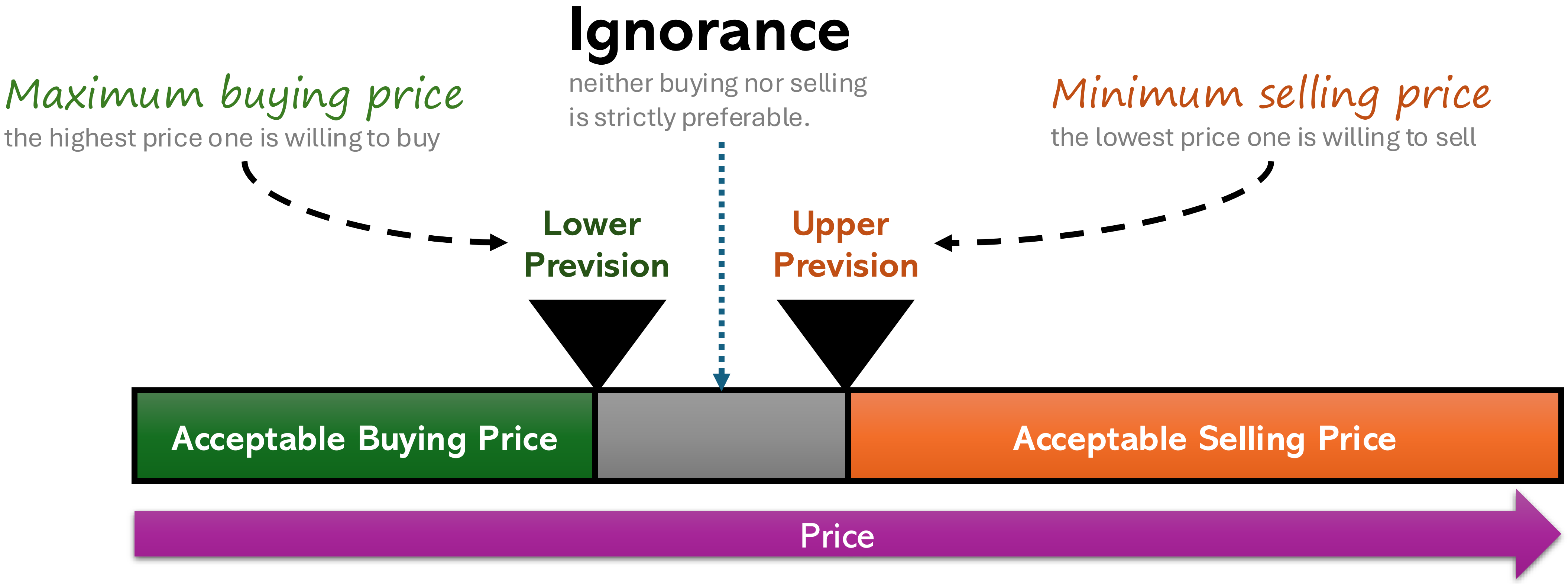


Recap on Imprecise Probability

Betting Perspective



Betting Perspective

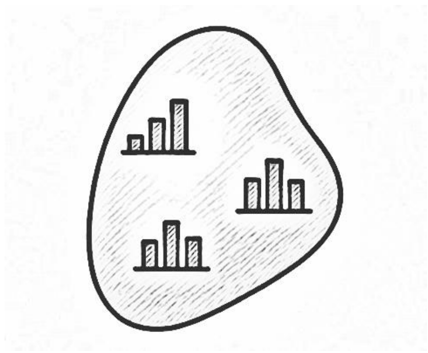


Probability Intervals

- Let X be a random variable taking values in a *finite set* $\mathcal{X}$
- A **probability interval** $(\underline{p}, \overline{p})$ is a pair of lower and upper probability mass functions satisfying:
  1. $0 \leq \underline{p}_x \leq \overline{p}_x \leq 1$ for all outcome $x \in \mathcal{X}$ (**bounded**)
  2. $\sum_{x \in \mathcal{X}} \underline{p}_x \leq 1 \leq \sum_{x \in \mathcal{X}} \overline{p}_x$ (**proper**)
  3. $\underline{p}_x \geq 1 - \sum_{z \neq x} \overline{p}_z$ and $\overline{p}_x \geq 1 - \sum_{z \neq x} \underline{p}_z$ for all $x \in \mathcal{X}$ (**reachable**)

- A set of compatible pmfs is called the **credal set**:

$$\mathcal{C} = \{p : \underline{p}_x \leq p_x \leq \overline{p}_x\}$$



Lower and Upper Probabilities

The lower and upper probabilities can be defined as

$$\underline{P}(S) := \max\left\{\sum_{x\in S} \underline{p}_x, 1 - \sum_{x\in S^c} \overline{p}_x\right\}, \qquad \overline{P}(S) := \min\left\{\sum_{x\in S} \overline{p}_x, 1 - \sum_{x\in S^c} \underline{p}_x\right\}$$

Exercise: Consider the following lower and upper pmf:

	X	Y	XY	YY
$\overline{p}$	3/8	3/8	5/8	5/8
$\underline{p}$	1/8	1/8	3/8	0/8

Then, calculate:

- $\overline{P}(\{Y, XY\}) =?$
- $\underline{P}(\{X, Y\}) =?$

Lower and Upper Previsions

- Let $\mathcal{C}$ be a non-empty subset of a set of all probabilities, i.e., $\mathcal{C} \subseteq \mathcal{P}$. The **lower and upper previsions** can be defined as

$$\underline{P}(S) \coloneqq \inf_{p\in\mathcal{C}} P_p(S), \qquad \overline{P}(S) \coloneqq \sup_{p\in\mathcal{C}} P_p(S)$$

$$\underline{E}(f) \coloneqq \inf_{p\in\mathcal{C}} E_p(f), \qquad \overline{E}(f) \coloneqq \sup_{p\in\mathcal{C}} E_p(f)$$

- Examples of credal sets:
  - The vacuous credal set $\mathcal{P}^{\mathrm{S}} \coloneqq \{\, p \in \mathcal{P} : \ p(S) = 1\}$ for some $S \in \mathcal{E}$.
  - The vacuous credal set $\mathcal{C} \coloneqq \mathcal{P}$ **(complete ignorance)**
  - The singleton credal set $\mathcal{C} \coloneqq \{p\}$ **(precise)**
  - and many more.

Coherent Lower Prevision

A **gamble** f on $\mathcal{X}$ is a bounded real-valued map on $\mathcal{X}$ representing an *uncertain reward*. Two types of transaction involving a gamble f :

1. Accepting to **buy** f for a price $\mu \Leftrightarrow$ *accepting the gamble* $f - \mu$
2. Accepting to **sell** f for a price $\lambda \Leftrightarrow$ *accepting the gamble* $\lambda - f$

Lower prevision $\underline{P}(f)$ represents *supremum* acceptable buying price for f:

$$\underline{P}(f) \coloneqq \sup\{\, \boldsymbol{\mu} \in R : f - \boldsymbol{\mu} \in \mathcal{D} \,\}$$

Upper prevision $\overline{P}(f)$ represents *infimum* acceptable selling price for f:

$$\overline{P}(f) \coloneqq \inf\{\, \boldsymbol{\lambda} \in R : \boldsymbol{\lambda} - f \in \mathcal{D} \,\}$$

Remark: When $f = I_A$, $\underline{P}(I_A) \coloneqq \underline{P}(A)$ is called a *coherent lower probability.*

Coherent Lower Prevision

A lower prevision $\underline{P}$ should ***avoid sure loss***, meaning that

$$\sup_{x\in\mathcal{X}} \sum_{\mathrm{i=1}}^{\mathrm{n}}[f_i(x) - \underline{P}(f_i)\ ] \geq 0 \quad \text{for all } n \geq 0 \text{ and all } f_1, \dots, f_n$$

If this is not satisfied, then there are $n > 0, f_1, \dots, f_n$ and $\delta > 0$ such that

$$\sup_{x\in\mathcal{X}} \sum_{\mathrm{i=1}}^{\mathrm{n}}[f_i(x) - (\underline{P}(f_i) - \delta)] \leq -\delta$$

Conclusion: the sum of desirable transactions $f_i(x) - (\underline{P}(f_i) - \delta)$ results in a loss of at least δ, no matter the outcome of the experiment.

Coherent Lower Prevision

A stronger rationality condition, ***coherence***, requires that supremum acceptable buying price for a gamble f cannot be raised by considering a positive linear combination of a finite number of other acceptable gambles:

$$\sup_{x\in\mathcal{X}} \sum_{i=1}^{n} \left[f_i(x) - \underline{P}(f_i)\right] - m\left[f_0(x) - \underline{P}(f_0)\right] \geq 0$$

for all natural $n, m \geq 0$ and all $f_0, f_1, \ldots, f_n$.

Exercise: What if the lower prevision is incoherent?

Possibility Theory

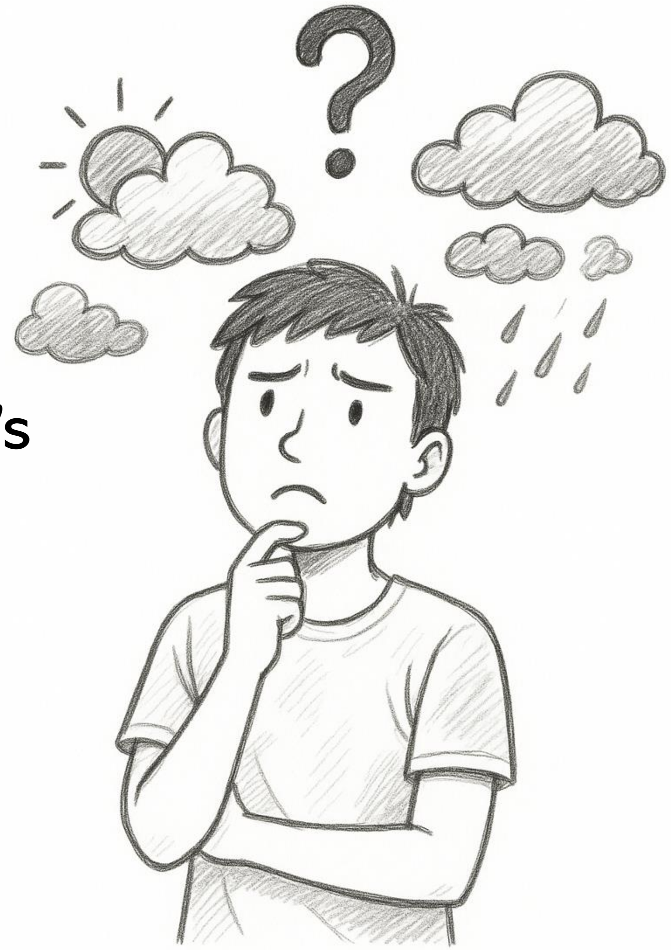
Historical Background

- **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



- A set of **states of affairs** S (e.g., sunny, cloudy, and rainy)
- A **possibility distribution** $\pi: S \rightarrow [0,1]$ represents an agent's *epistemic state* of the actual state in S.
  - $\pi(s) = 0 \Longrightarrow$ state s is rejected as **impossible**.
  - $\pi(s) = 1 \Longrightarrow$ state s is **totally possible (plausible)**.
- The larger $\pi(s)$, the more plausible the state s is.
- If S is exhaustive, at least one of the elements of S should be the actual state: $\exists s, \pi(s) = 1$ (normalisation)

$\pi(\text{sunny}) = 1.0,$ quite possible that it's sunny

$\pi(\text{cloudy}) = 0.7,$ somewhat possible that it's cloudy

$\pi(\text{rainy}) = 0.3$ rather unlikely (but not impossible) that it's rainy

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 s: $\pi(s) \leq \pi'(s)$.
- Two extreme forms of partial knowledge in possibilistic framework:

Complete knowledge
$\pi(s_0) = 1$ and $\pi(s) = 0$ for all $s \neq s_0$

Complete ignorance
$\pi(s) = 1$ for all $s \in S$

Possibility and Necessity Measures

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

Possibility (upper probability)

$$\Pi(A) \coloneqq \bar{P}(A) \coloneqq \max_{x \in A} \pi_x$$

Necessity (lower probability)

$$N(A) \coloneqq \underline{P}(A) \coloneqq \min_{x \notin A} (1 - \pi_x)$$

"Does event A occur?"

- $\Pi(A)$ evaluates to what extent A is **consistent with π**
- $\mathrm{N}(A)$ evaluates to what extent A is **certainly implied by π**

Possibility and Necessity Measures

Formally, for a random variable X taking values in a set of outcomes S, a possibility measure satisfies:

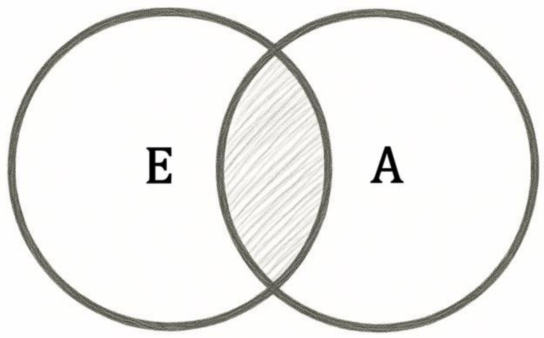
1. **Non-negativity**: $\pi_x \coloneqq \Pi(X = x) = \Pi(\{x\}) \geq 0$ for all outcomes $x \in S$
2. **Maxitivity**: $\Pi(A) = \max_{x \in A} \pi_x$ for all events $A \subseteq \mathcal{X}$
3. **Normed**: $\Pi(S) = 1$

- The ***possibility-necessity duality***: $N(A) = 1 - \Pi(A^c)$.
- Generally, $\Pi(S) = \mathrm{N}(S) = 1$ and $\Pi(\emptyset) = \mathrm{N}(\emptyset) = 0$.

Boolean Case

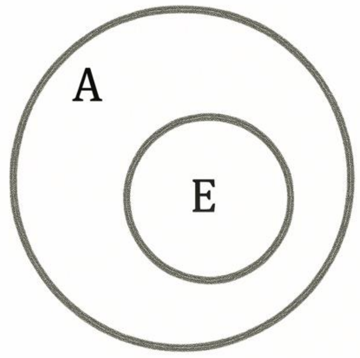
The possibility distribution comes down to the disjunctive (epistemic) set $\mathrm{E} \subseteq S$

$\Pi(A) = 1$ 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**:

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

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

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

- It's clear that $\pi(s_1, \dots, s_n) \leq \min_{k=1}^{m} \pi_k^{\downarrow}(s_k)$: *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:** $\Pi(B\,|A) = \Pi(B)$
  A form of directed epistemic independence whereby learning A doesn't affect the plausibility of B.

Intepretations of Possibility

- **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

- **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:

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

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

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)

Exercises

Exercises

Consider a random variable X with a finite outcome space

$$S = \{s_1, s_2, s_3, s_4\}.$$

You are told that the probabilities of each elementary event are not known precisely, but lie within the following **interval constraints**:

Outcome	Lower bound l_i	Upper bound u_i
s_1	0.1	0.3
s_2	0.2	0.4
s_3	0.1	0.5
s_4	0.0	0.3

The only global constraint is the probabilities must sum to one: $\sum_{i=1}^{4} P(s_i) = 1$.

Exercises

Construct the Credal Set

- Write the set $\mathcal{P} = \{P : \; l_i \leq P(s_i) \leq u_i, \sum_{\mathrm{i}=1}^{4} P(s_i) = 1\}$.
  Sol: $\mathcal{P} = \{P : \; 0.1 \leq P(s_1) \leq 0.3, 0.2 \leq P(s_2) \leq 0.4, 0.1 \leq P(s_3) \leq 0.5, 0.0 \leq P(s_4) \leq 0.3, \sum_{\mathrm{i}=1}^{4} P(s_i) = 1\}$
- How many free variables does this set have?
  Sol: 4 variables − 1 linear equality = 3.
- What is its geometric shape in R^4?
  Sol: convex polytope = intersection of the probability 3-simplex with the axis-aligned box $[l_1, u_1] \times \cdots \times [l_4, u_4]$

Exercises

Compute the **lower** and **upper** probabilities of the following events:

- $A_1 = \{s_1, s_2\}$
- $A_3 = \{s_2, s_3, s_4\}$
- $A_2 = \{s_3\}$

Sol: Using the formula, we have

- $\underline{P}(A_1) = \max(0.3,0.2) = 0.3,\ \overline{P}\,(A_1) = \min(0.7,0.9) = 0.7$
- $\underline{P}(A_2) = \max(0.1,0.0) = 0.1,\ \overline{P}\,(A_2) = \min(0.5,0.7) = 0.5$
- $\underline{P}(A_3) = \max(0.3,0.7) = 0.7,\ \overline{P}\,(A_3) = \min(1.2,0.9) = 0.9$

using the formulas:

$$\underline{P}(S) \coloneqq \max\left\{\sum_{x\in S}\underline{p}_x, 1-\sum_{x\in S^c}\overline{p}_x\right\}, \qquad \overline{P}(S) \coloneqq \min\left\{\sum_{x\in S}\overline{p}_x, 1-\sum_{x\in S^c}\underline{p}_x\right\}$$

Exercises

Verify coherence: Show that your results satisfy:

- $\underline{P}(\emptyset) = 0$, $\overline{P}(S) = 1$ (by formular)
- $\underline{P}(A) + \overline{P}(\bar{A}) = 1$ (by duality: $A_1: 0.3 + 0.7 = 1, A_2: 0.1 + 0.9 = 1, A_3: 0.7 + 0.3 = 1$)
- $\underline{P}(A) \leq \overline{P}(A)$ for all A ($A_1: 0.3 \leq 0.7, A_2: 0.1 \leq 0.5, A_3: 0.7 \leq 0.9$)

Interpret the Results

- Explain the meaning of the gap $\overline{P}(A) - \underline{P}(A)$. (Imprecision width $w(A) = \overline{P}(A) - \underline{P}(A)$. $w(A_1) = 0.7 - 0.3 = 0.4, w(A_2) = 0.5 - 0.1 = 0.4, w(A_3) = 0.9 - 0.7 = 0.2$)
- What kind of uncertainty does it reflect? (It reflects **epistemic uncertainty** due to interval bounds.)
- For which event do you observe the highest imprecision, and why? (A_1 and A_2 are most imprecise. Largest freedom in allocating mass to their complements)

Exercises

Assume that these intervals originate from a possibility distribution:

$$\pi(s_1) = 0.8, \qquad \pi(s_2) = 1.0, \qquad \pi(s_3) = 0.7, \qquad \pi(s_4) = 0.5.$$

- Compute: $\Pi(A) = \max_{x \in A} \pi(x)$, $N(A) = 1 - \Pi(\bar{A})$.
- Compare (Π, N) with your $(\overline{P}, \underline{P})$ from the previous exercises.

Sol: For A_1, A_2, A_3, we have

- $A_1 = \{s_1, s_2\}$: $\Pi(A_1) = \max(0.8,1.0) = \mathbf{1.0}, \Pi(A_1^c) = \max(0.7,0.5) = 0.7 \Longrightarrow \mathrm{N}(A_1) = \mathbf{0.3}$
- $A_2 = \{s_3\}$: $\Pi(A_2) = 0.7, \Pi(A_2^c) = \max(0.8,1.0,0.5) = 1.0 \Longrightarrow \mathrm{N}(A_1) = \mathbf{0.0}$
- $A_3 = \{s_2, s_3, s_4\}$: $\Pi(A_3) = \max(1.0,0.7,0.5) = \mathbf{1.0}, \Pi(A_3^c) = \pi(s_1) = 0.8 \Longrightarrow \mathrm{N}(A_3) = \mathbf{0.2}$

Exercises

- **Group 1**: Show that for any gamble f, we have $\overline{P}(f) = -\underline{P}(-f)$.

$$\underline{P}(f) \coloneqq \sup\{\, \boldsymbol{\mu} \in R : f - \boldsymbol{\mu} \in \mathcal{D} \,\}$$

$$\overline{P}(f) \coloneqq \inf\{\, \boldsymbol{\lambda} \in R : \boldsymbol{\lambda} - f \in \mathcal{D} \,\}$$

- **Group 2**: Given the maxitivity $\Pi(A \cup \mathrm{B}) = \max\big(\Pi(A), \Pi(B)\big)$, show the minitivity property of the necessity measure:

$$N(A \cap \mathrm{B}) = \min\big(N(A), N(B)\big)$$

Exercises

- **Group 1**: Show that for any gamble f, we have $\overline{P}(f) = -\underline{P}(-f)$.

  Sol: By the definition of the upper prevision,

$$\begin{aligned}\overline{P}(f) &= \inf\{\, \lambda \in R : \lambda - f \in \mathcal{D} \,\} \\ &= \inf\{\, \lambda \in R : -f - (-\lambda) \in \mathcal{D} \,\} \\ &= \inf\{\, -\mu \in R : -f - \mu \in \mathcal{D} \,\} \\ &= -\sup\{\, \mu \in R : -f - \mu \in \mathcal{D} \,\} \\ &= -\underline{P}(-f).\end{aligned}$$

Exercises

- **Group 2**: Given the maxitivity $\Pi(A \cup \mathrm{B}) = \max\big(\Pi(A), \Pi(B)\big)$, show the minitivity property of the necessity measure:

$$N(A \cap \mathrm{B}) = \min\big(N(A), N(B)\big)$$

Sol: It follows from the possibility-necessity duality that

$$\begin{aligned} N(A \cap \mathrm{B}) &= 1 - \Pi\big((A \cap B)^{\mathrm{c}}\big) \\ &= 1 - \Pi(A^{\mathrm{c}} \cup B^{\mathrm{c}}) \\ &= 1 - \max(\Pi(A^{\mathrm{c}}), \Pi(B^{\mathrm{c}})) \\ &= \min(1 - \Pi(A^{\mathrm{c}}), 1 - \Pi(B^{\mathrm{c}})) \\ &= \min(N(A), N(B)). \end{aligned}$$

Exercises

- **Group 1&2**: Show the lower prevision must be **coherent**.

$$\sup_{x\in\mathcal{X}} \sum_{i=1}^{n} [f_i(x) - \underline{P}(f_i)] - m[f_0(x) - \underline{P}(f_0)] \geq 0$$

Sol: Suppose that the above inequality doesn't hold for some non-negative n, m and some $f_0, f_1, \dots, f_n$. If $m = 0$, this means that $\underline{P}$ incurs sure loss. Now, assume that $m > 0$. Then, there is some $\delta > 0$ such that

$$\sum_{i=1}^{n} [f_i - (\underline{P}(f_i) - \delta)] \leq m[f_0 - (\underline{P}(f_0) + \delta)]$$

The terms on both sides of the inequality should be desirable, but this means that the our agent should accept to buy the gamble f_0 for the price $\underline{P}(f_0) + \delta$, which is strictly higher than the supremum buying price he has specified for it $\Longrightarrow$ **incoherence**.