IPML

IMPRECISE PROBABILISTIC MACHINE LEARNING

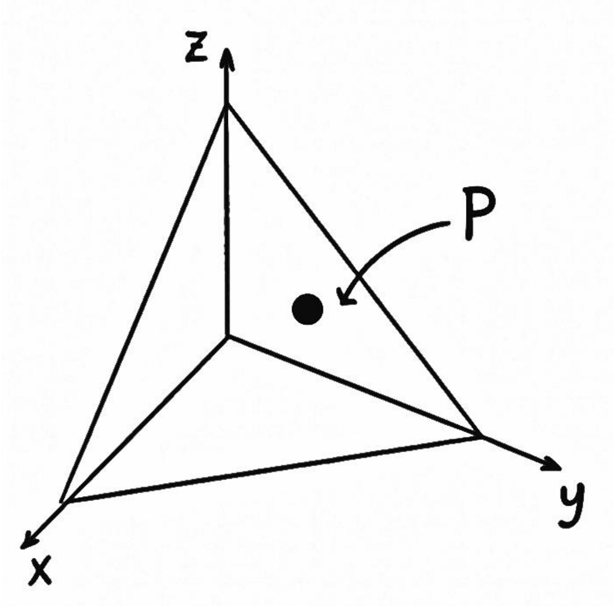
# Lecture 5: Convex Sets of Probabilities

Krikamol Muandet

21 November 2025

# Outline

1. Decision Making under Uncertainty
2. Theory of Credal Sets
3. Connections to IP Models
4. **Bonus**: Learning and Reasoning

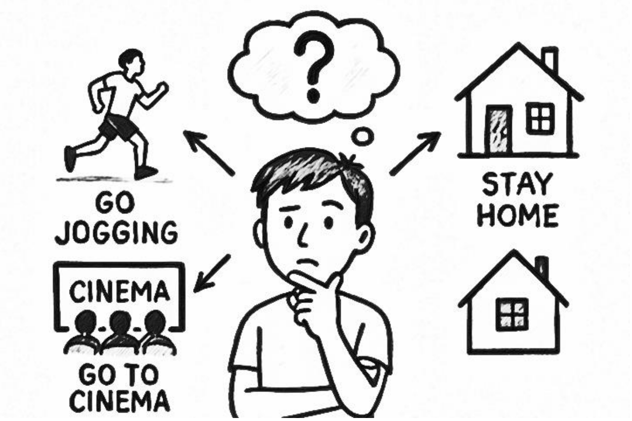


# Decision Making under Uncertainty

# Decision Making under Uncertainty

- Consider a simple scenario: *You need to make decision on which action to take given the state of the world.*

| Act \ State | Sunny | Cloudy |
|---|---|---|
| Jogging | 10 | 5 |
| Cinema | -5 | 12 |
| Stay Home | 0 | 0 |




- An **act** is a *random* function of the state of the world.
- Foundation of **Reinforcement Learning (RL)**

# Expected Utility Theory

- Given mutually exclusive outcomes, a **lottery** is a scenario where each outcome will happen with a given probability, e.g.,

$$L = p_{\mathrm{TH}}\mathrm{Thailand} + p_{\mathrm{US}}\mathrm{USA} + p_{\mathrm{DE}}\mathrm{Germany}, \qquad p_{\mathrm{TH}} + p_{\mathrm{US}} + p_{\mathrm{DE}} = 1$$

- A Von-Neumann Morgenstern (VNM) **rational agent** has a preference that fulfills these four axioms:
  1. **Completeness**: For any lotteries $L$ and $M$, either $L \succcurlyeq M$ or $M \succcurlyeq L$.
  2. **Transitivity**: If $L \succcurlyeq M$ and $M \succcurlyeq N$, then $L \succcurlyeq N$.
  3. **Continuity**: If $L \preccurlyeq M \preccurlyeq N$, then there exists a probability $p \in [0,1]$ such that $pL + (1-p)N \sim M$.
  4. **Independence**: For any $M$ and $p \in [0,1)$: $L \preccurlyeq N$ if and only if $(1-p)L + pM \preccurlyeq (1-p)N + pM$.

Neumann, John von and Morgenstern, Oskar, *Theory of Games and Economic Behavior*. Princeton, NJ. Princeton University Press, 1953.

# Expected Utility Theory

- For a (VNM-)rational agent, there exists a **utility function** $u$ which assigns to each outcome $S$ a real number $u(S)$ such that for any two lotteries:

$$L \preccurlyeq M \text{ if and only if } \mathbb{E}[u(L)] < \mathbb{E}[u(M)]$$

where $\mathbb{E}[u(L)]$ is given by

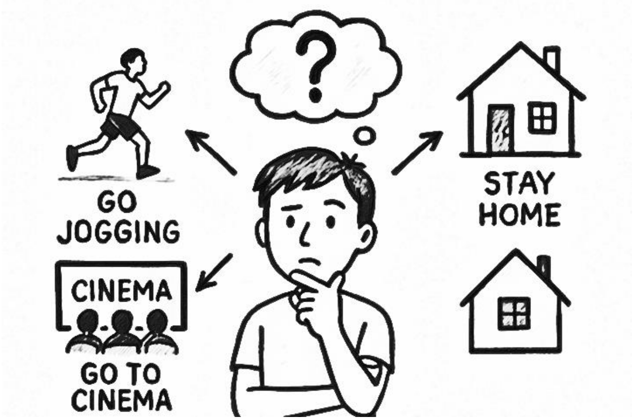
$$\mathbb{E}[u(p_1 S_1 + \cdots + p_\mathrm{n} S_\mathrm{n})] = p_1 u(S_1) + \cdots + p_\mathrm{n} u(S_\mathrm{n})$$

- The utility function $u$ is unique up to **affine transformations**, i.e., adding a constant and multiplying by a positive scalar.

# Decision Making under Uncertainty

- Consider a simple scenario: *You need to make decision on which action to take given the state of the world.*

| Act \ State | Sunny | Cloudy |
|---|---|---|
| Jogging | [10, 5, 2] | [5, 2, 0] |
| Cinema | [-5, 2, 10] | [12, 5, 10] |
| Stay Home | [0, -5, 5] | [0, 5, -5] |

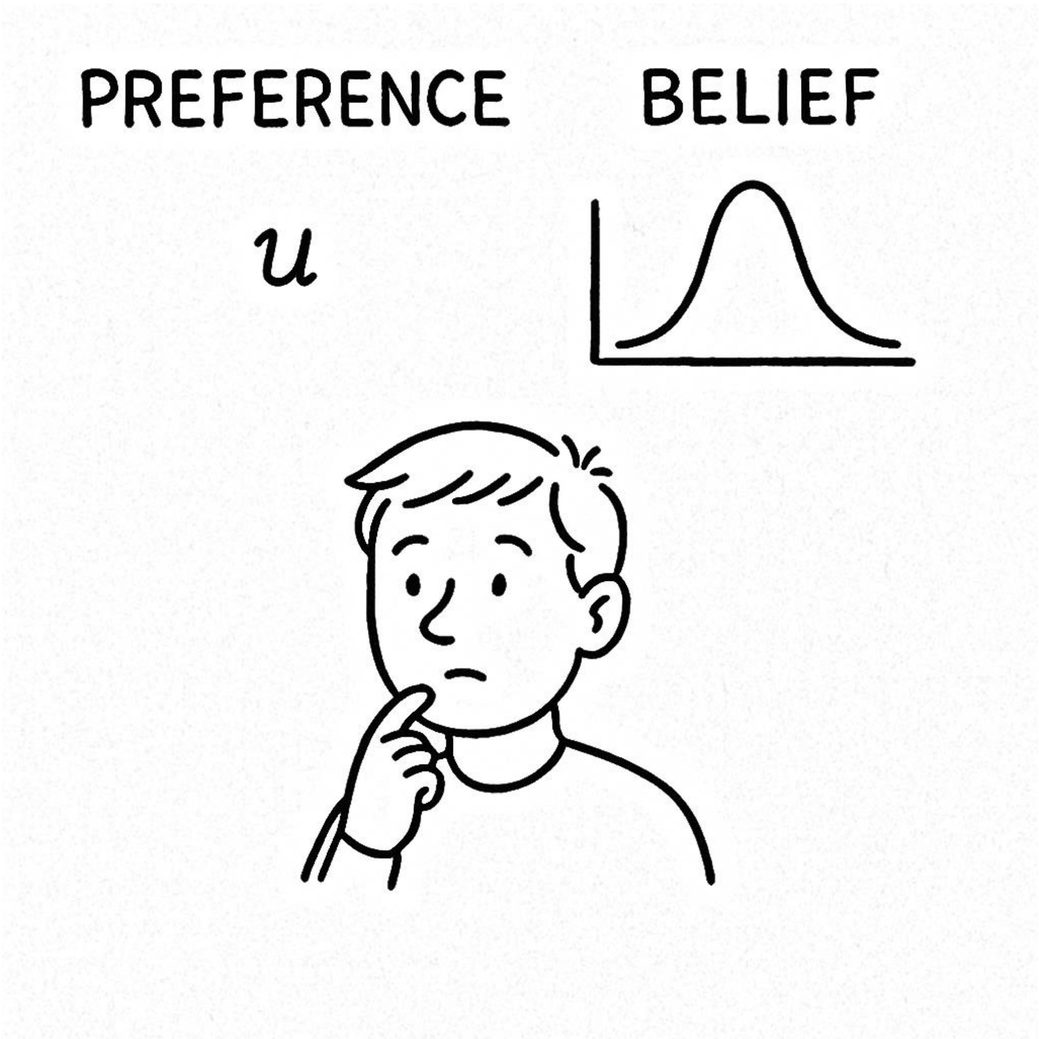



- Consider (VNM-) agents whose beliefs over states are

$$\boldsymbol{p}_{\mathrm{a}} = (0.1, 0.9), \ \ \boldsymbol{p}_{\mathrm{b}} = (0.5, 0.5), \ \ \boldsymbol{p}_{\mathrm{c}} = (0.9, 0.1)$$

- Which act would each agent pick?

# Bayesian Agent



# Bayesian Framework

- For a state space $\Omega = \{\omega_1, \ldots, \omega_n\}$, the **belief** of an agent about which $\omega \in \Omega$ obtains is represented by **a single probability distribution $p(\boldsymbol{\omega})$.**
- The agent always prefers an act $a \in \mathcal{A}$ with **higher expected utility**, i.e., for $a_1, a_2 \in \mathcal{A}$:

$$a_1 \succ a_2 \quad \text{if and only if} \quad \mathbb{E}_{p(\omega)}[u(\omega, a_1)] > \mathbb{E}_{p(\omega)}[u(\omega, a_2)]$$

- What if the agent's belief is represented as a **set of distributions**?
  - This can handle imprecision, ignorance, and collective belief.
  - What does it mean for an agent to hold a set of probability distributions as its belief state?
  - How to choose the optimal action under such imprecise beliefs?

# Beyond Single Probability Distribution

- Consider $\mathcal{P} = \{p, q\}$ and $a_1, a_2 \in \mathcal{A}$ such that

$$\mathbb{E}_p[u(\omega, a_1)] > \mathbb{E}_p[u(\omega, a_2)] \quad \text{but} \quad \mathbb{E}_q[u(\omega, a_1)] < \mathbb{E}_q[u(\omega, a_2)]$$

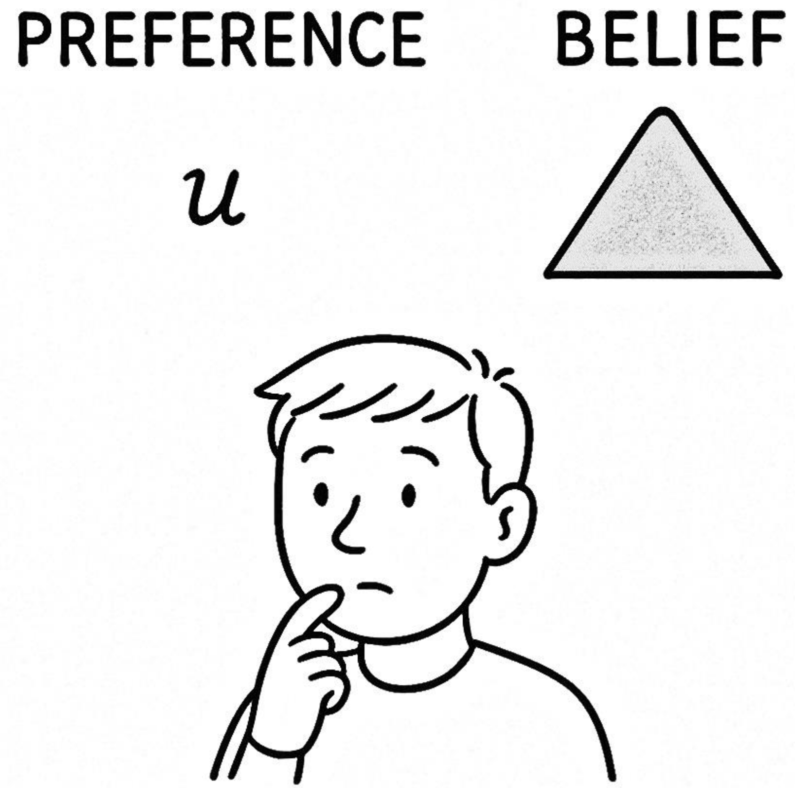
- Hence, $a_1$ and $a_2$ are **incomparable** with respect to the expected utility.
- Unlike Bayesian setting, the agent has a **partial order** of preferences:
  - **Incomplete belief**: The agent is unsure that any single distribution is "true" (as in sensitivity analysis or robust Bayesian) or lacks the time or resources to specify one precisely.
  - **Exhaustive belief**: Even after careful consideration, the agent cannot fully rank all acts. Some remain genuinely incomparable.

# Quasi-Bayesian Framework

- Let $\prec$ be a partial order of preferences on $\mathcal{A}$ which fulfills the **quasi-Bayesian rationality axioms**.
- Then, there exists a unique nonempty convex set $\mathcal{K}$ of finitely additive probability measures such that

$$a_1 \prec a_2 \Leftrightarrow \mathbb{E}_p[u(\omega, a_1)] < \mathbb{E}_p[u(\omega, a_2)], \forall p \in \mathcal{K}$$

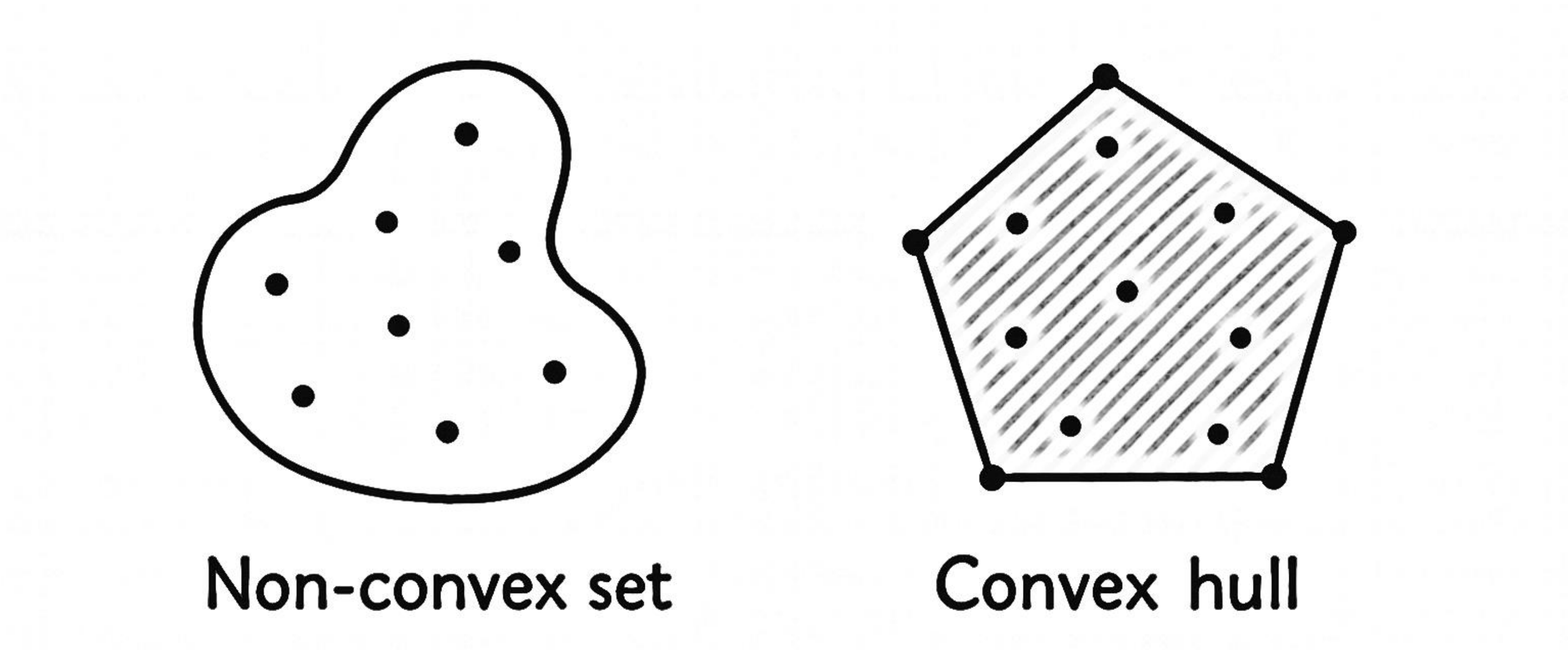
- The set $\mathcal{K}$ is the **credal set** representing $\prec$.
- Some acts are better than others, and some acts cannot be compared.






F. J. Giron and S. Rios, "*Quasi-Bayesian behaviour: A more realistic approach to decision making?*", In: Bayesian Statistics, University Press, 1980, pp. 17–38.

# Why Convex Credal Sets?




A convexification of a credal set preserves the partial order of preference.

# Theory of Credal Sets

# Credal Set

- A credal set $\mathcal{K}$ on a finite outcome space $\Omega$ is a **closed**, **convex** subset of the probability simplex:

$$\mathcal{K} \subseteq \Delta^{|\Omega|-1} = \left\{ \boldsymbol{p} \in R^{|\Omega|} : \ p_i \geq 0, \sum_{i=1}^{|\Omega|} p_i = 1 \right\}$$

- **Convexity** ensures that any mixture of admissible probability measures is itself admissible.
- By the Krein–Milman theorem, the credal set $\mathcal{K}(\Omega)$ can be equivalently described by its **extreme points** $\mathrm{ext}[\mathcal{K}(\Omega)]$.

# Special Cases

- Finitely generated credal set (FGCS)
- The vacuous credal set $\mathcal{P}^S \coloneqq \{\, p \in \mathcal{P} : \ p(S) = 1\}$ for some $S \in \mathcal{E}$.
- The vacuous credal set $\mathcal{C} \coloneqq \mathcal{P}$ (**complete ignorance**)
- Singleton credal sets $\mathcal{C} \coloneqq \{p\}$ (**precise**)
- Linear-vacuous or $\epsilon$-contamination credal sets:

$$\mathcal{C}_{p,\epsilon} \coloneqq \{(1-\epsilon)p + \epsilon q : \ q \in \mathcal{P}\}$$

# Construction of Credal Sets

- Interval probabilities
- Coherent lower previsions
- Elicitation from expert opinion
- Relative likelihood $\mathrm{C}_{\alpha} = \{h : \gamma(h) \geq \alpha\}$
- Ensemble and data-driven methods
- Conformal prediction

# Lower and Upper Expectations

- The **expectation** of a function $f$ on $\Omega$ with respect to the credal set $\mathcal{K}$ forms a closed interval $[\underline{\mathbb{E}}[f], \overline{\mathbb{E}}[f]]$:

$$\underline{\mathbb{E}}[f] = \min_{p \in \mathcal{K}} \mathbb{E}_p[f] = \min_{p \in \mathrm{ext}[\mathcal{K}]} \mathbb{E}_p[f]$$

$$\overline{\mathbb{E}}[f] = \max_{p \in \mathcal{K}} \mathbb{E}_p[f] = \max_{p \in \mathrm{ext}[\mathcal{K}]} \mathbb{E}_p[f]$$

- We have already shown in L3 that $\underline{\mathbb{E}}[f] = -\overline{\mathbb{E}}[-f]$.
- A **lower prevision** is an **affinely superadditive** lower expectation:

$$\underline{\mathbb{E}}[f+g] \geq \underline{\mathbb{E}}[f] + \underline{\mathbb{E}}[g] \qquad \underline{\mathbb{E}}[\alpha f + \beta] = \alpha \underline{\mathbb{E}}[f] + \beta$$

# Lower and Upper Envelopes

- A credal set induces an interval of probabilities for a random variable: $[\underline{p}(\omega), \overline{p}(\omega)]$ called **lower and upper envelopes (coherent lower/upper probabilities)**:

$$\underline{p}(\omega) = \inf_{p \in \mathcal{K}} p(\omega) = \inf_{p \in \mathrm{ext}[\mathcal{K}]} p(\omega)$$

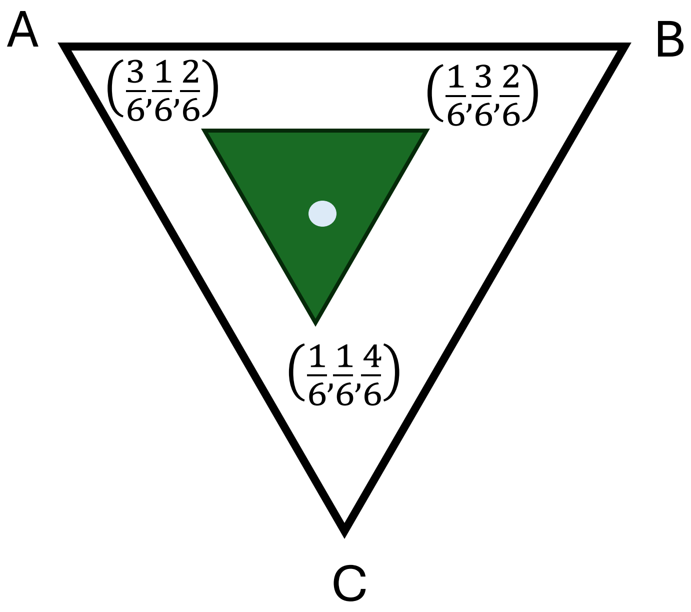
$$\overline{p}(\omega) = \sup_{p \in \mathcal{K}} p(\omega) = \sup_{p \in \mathrm{ext}[\mathcal{K}]} p(\omega)$$

- The conjugate relation:

$$\underline{p}(\omega) = 1 - \overline{p}(\omega^{\mathrm{c}}).$$

# Example

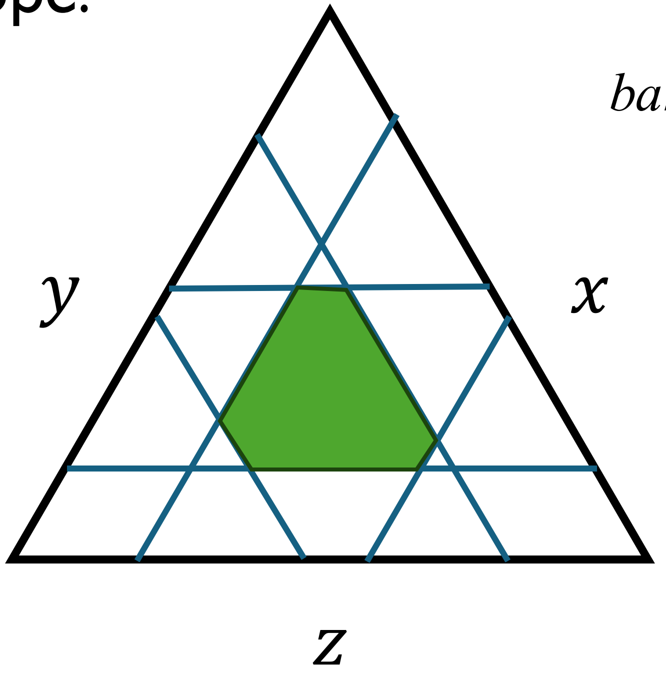
- $\mathcal{C}_{p,\epsilon} \coloneqq \{(1-\epsilon)p + \epsilon q : q \in \mathcal{P}\}$ with $\epsilon = \frac{1}{3}$ and $p = (p_{\mathrm{A}}, p_{\mathrm{B}}, p_{\mathrm{C}}) = \left(\frac{1}{4}, \frac{1}{4}, \frac{2}{4}\right)$.
- Find $\underline{P}(\{\mathrm{B}, \mathrm{C}\})$ and $\underline{\mathbb{E}}(f)$ with $f = (1, 0, -1)$



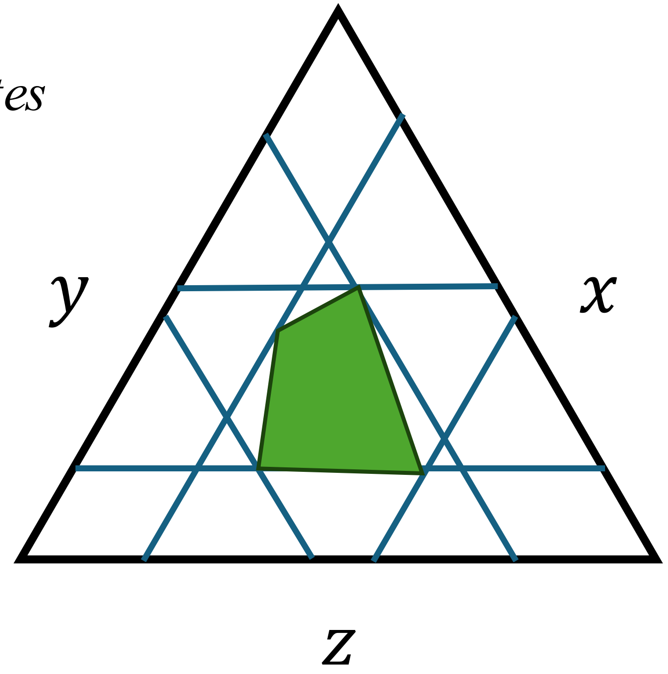
# Connections to IP Models

# Interval of Probabilities

- There can be several set of distributions that generate the same lower envelope.




*baricentric coordinates*




$$\mathcal{K} = \{p : \; p(x) \geq \underline{p}(x), \forall x\}$$

# Lower Expectations

- From a lower expectation model $\underline{\mathbb{E}}[\cdot]$, we can create a set of probability distributions:

$$\mathcal{K} = \{p : \mathbb{E}_p[f] \geq \underline{\mathbb{E}}[f], \forall f\}$$

- If $\underline{\mathbb{E}}[\cdot]$ is **superadditive** and **affinely homogeneous**:

$$\underline{\mathbb{E}}[f + g] \geq \underline{\mathbb{E}}[f] + \underline{\mathbb{E}}[g], \qquad \underline{\mathbb{E}}[\alpha f + \beta] = \alpha \underline{\mathbb{E}}[f] + \beta$$

then there is a one-to-one correspondance between $\mathcal{K}$ and $\underline{\mathbb{E}}[\cdot]$.

# Lower Probabilities

- A **lower/upper probability** pair is a pair of non-negative functions $(\underline{p}, \overline{p})$:
    1. $\overline{p}(\omega) \;=\; 1 - \underline{p}(\omega^c)$
    2. $\underline{p}(\emptyset) = 0,\;\; \underline{p}(\Omega) = 1$
    3. $\underline{p}(\omega_1 \cup \omega_2) \geq \underline{p}(\omega_1) + \underline{p}(\omega_2)$ for any disjoint $\omega_1, \omega_2$
    4. $\overline{p}(\omega_1 \cup \omega_2) \leq \overline{p}(\omega_1) + \overline{p}(\omega_2)$ for any disjoint $\omega_1, \omega_2$

- A function $\underline{p}$ is **2-monotone Choquet capacity** if it is positive and
    1. $\underline{p}(\emptyset) = 0,\;\; \underline{p}(\Omega) = 1$
    2. $\underline{p}(\omega_1 \cup \omega_2) \geq \underline{p}(\omega_1) + \underline{p}(\omega_2) - \overline{p}(\omega_1 \cap \omega_2)$ for any $\omega_1, \omega_2$
- There is a correspondence between the 2-monotone lower probability and the set of all probability distributions that **dominate** it.

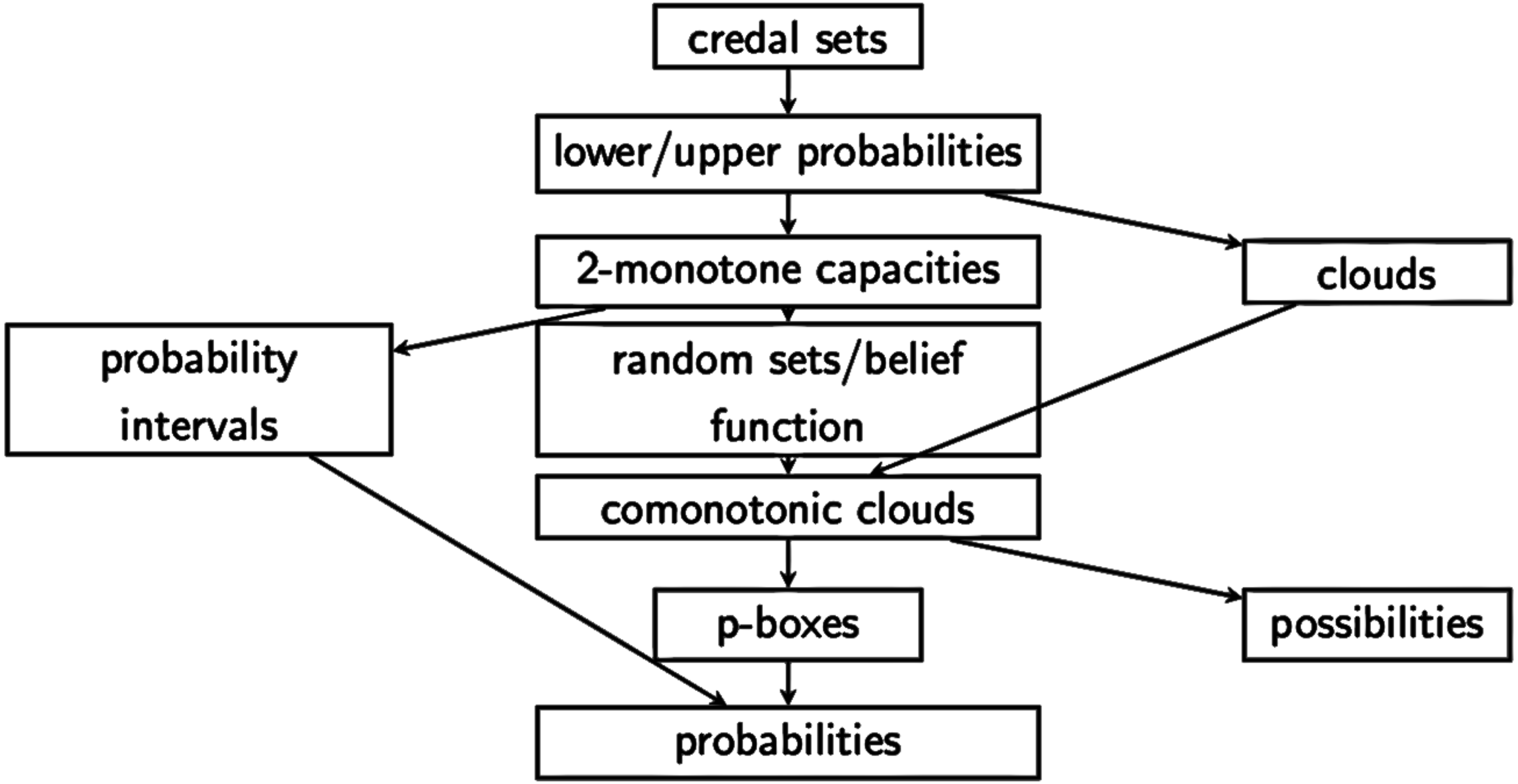
# Choquet Capacities

- A positive function $\nu(\cdot)$ is an **$n$-monotone Choquet capacity** if
  1. $\nu(\emptyset) = 0, \;\; \nu(\Omega) = 1$
  2. For every integer $k$ with $1 \leq k \leq n$, and for every collection of sets $\omega_1, \dots, \omega_k$, the following inequality holds:

$$\sum_{\emptyset \neq I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} \, \nu\left(\bigcup_{i \in I} \omega_i\right) \geq \nu\left(\bigcap_{i=1}^{k} \omega_i\right)$$

- Any lower probability is 1-monotone and 2-monotone lower probabilities are 2-monotone capacities.
- If a lower probability is $n$-monotone for all $n$, then it is called **infinite monotone** or **belief function**.

# Summary


credal sets
lower/upper probabilities
2-monotone capacities
clouds
probability intervals
random sets/belief function
comonotonic clouds
p-boxes
possibilities
probabilities

# Recommended Reading

- Introduction to the Theory of Sets of Probabilities by Fabio Cozman

# Exercises