IPML

IMPRECISE PROBABILISTIC MACHINE LEARNING

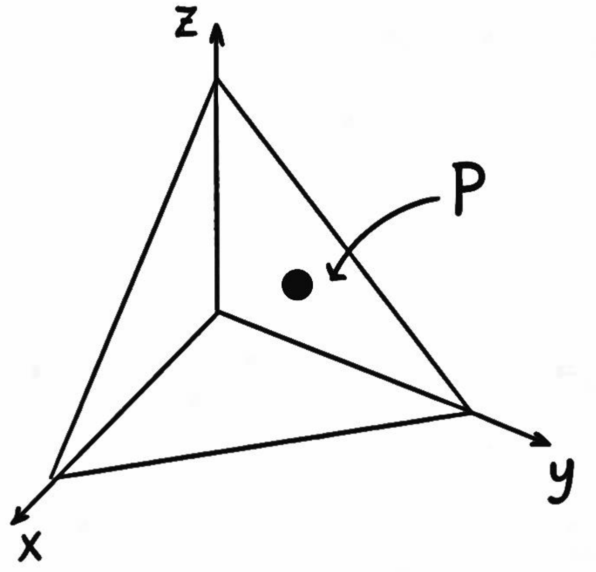
Lecture 5: Convex Sets of Probabilities

Krikamol Muandet

21 November 2025

Outline

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

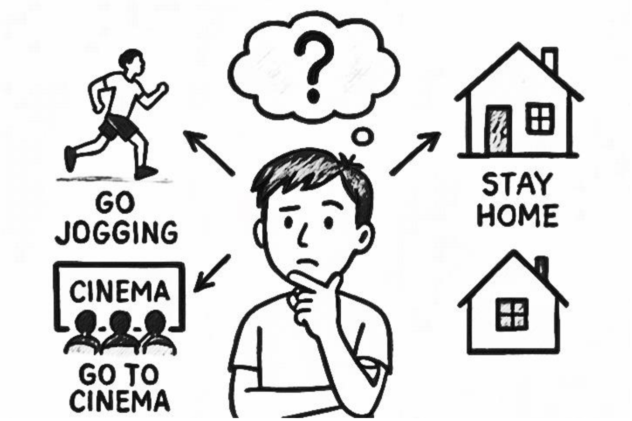


Decision Making under Uncertainty

Decision Making under Uncertainty

- A simple scenario: *You must select an action whose outcome is contingent on the realised 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.,

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

- A Von-Neumann Morgenstern (VNM) **rational agent**'s preference fulfills :
  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:

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

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

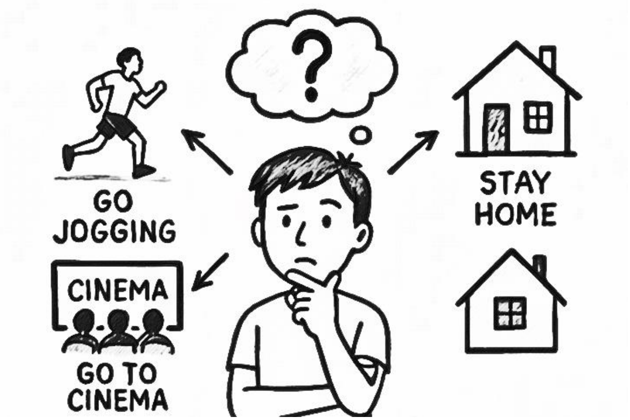
$$\mathbb{E}[u(p_1 S_1 + \cdots + p_\mathrm{n} S_\mathrm{n})] = p_1 u(S_1) + \cdots + p_\mathrm{n} u(S_\mathrm{n})$$

- 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

- A simple scenario: *You must select an action whose outcome is contingent on the realised 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?

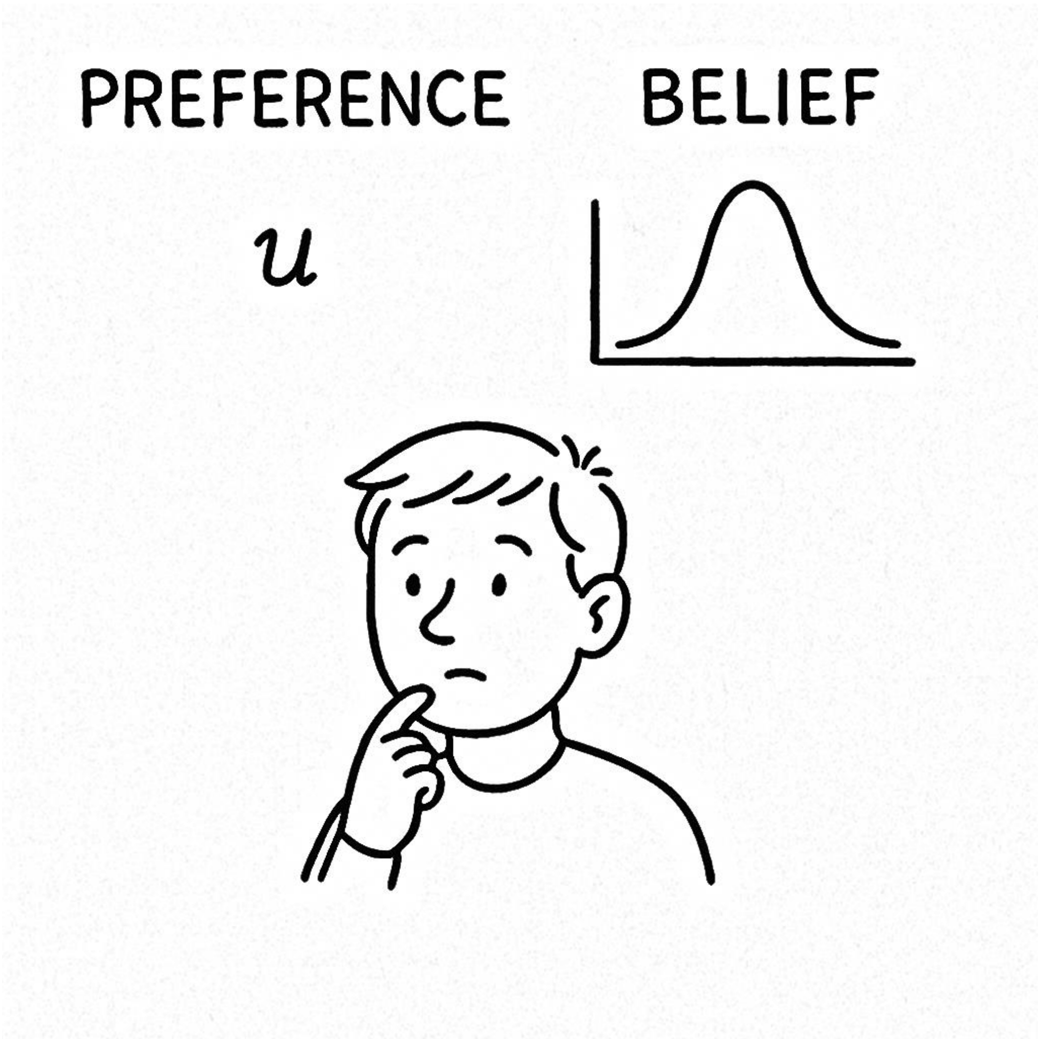
For an agent with $\boldsymbol{p}_{\mathrm{a}} = (0.1, 0.9)$

- $\mathbb{E}[u(\text{Jogging})] = 0.1 \times 10 + 0.9 \times 5$
  $= 1.0 + 4.5 = \mathbf{5.5}$
- $\mathbb{E}[u(\text{Cinema})] = 0.1 \times -5 + 0.9 \times 12$
  $= -0.5 + \ 10.8 = \mathbf{10.3}$
- $\mathbb{E}[u(\text{Home})] = 0.1 \times 0 + 0.9 \times 0$
  $= \ 0 + 0 = \mathbf{0}$

Hence, the agent will go to the cinema.

In Bayesianism, a belief state (or uncertainty) of an agent is represented by a **single probability distribution** over the states of the world.

Bayesian Agent



Bayesian Framework

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

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

- Here, $a_1 \succcurlyeq a_2$ means that a_1 is **at least as good as** a_2.

Bayesian Framework

- What if the agent's belief is now represented as a **set of distributions**?
  - This can handle imprecision, ignorance, and collective belief.
  - But 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 a set $\mathcal{K} = \{p, q\}$ and any two actions $a_1, a_2 \in \mathcal{A}$ such that

$$\mathbb{E}_{p(\omega)}[u(\omega, a_1)] > \mathbb{E}_{p(\omega)}[u(\omega, a_2)], \text{ but } \mathbb{E}_{q(\omega)}[u(\omega, a_1)] < \mathbb{E}_{q(\omega)}[u(\omega, a_2)]$$

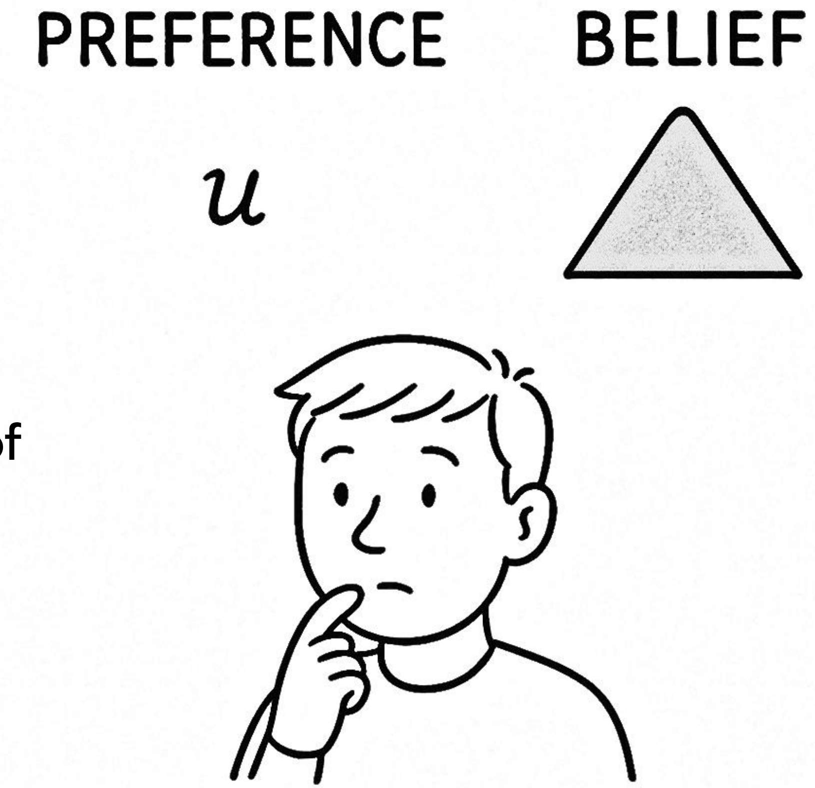
- Clearly, a_1 and a_2 are **incomparable** with respect to the expected utility.
- Unlike Bayesian setting, the agent has a **partial order** of preferences:
  1. **Incomplete belief**: The agent is unsure that any single distribution is "true" (as in sensitivity analysis or robust Bayesian) or lacks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

Quasi-Bayesian Framework

- Let $\preccurlyeq$ 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

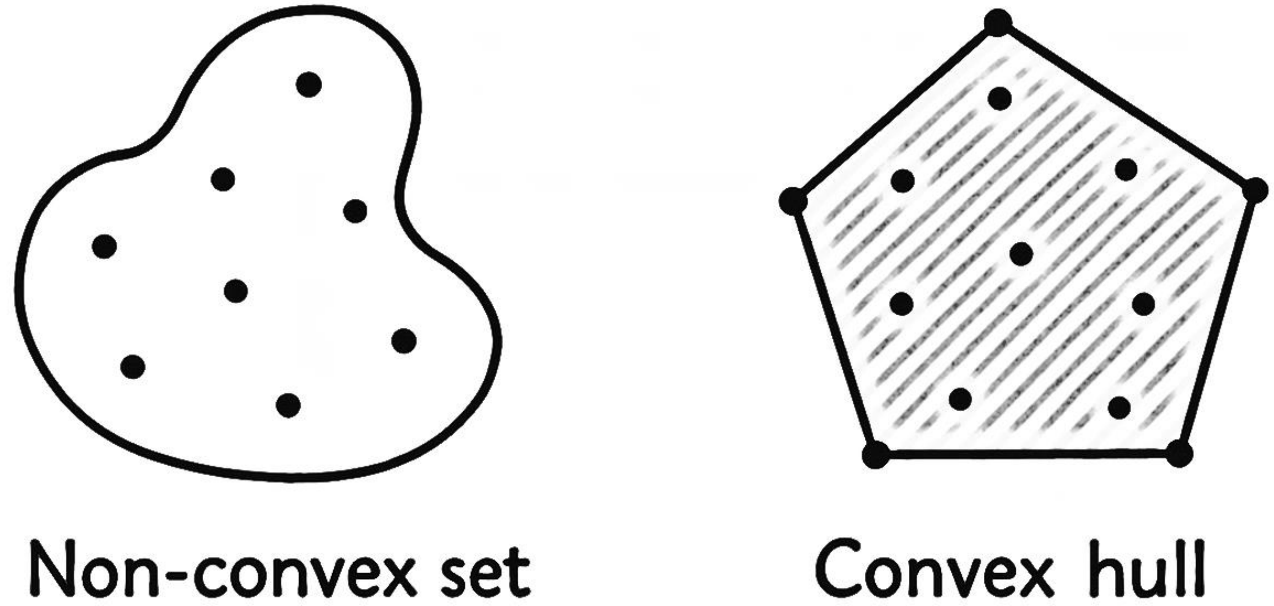
$$a_1 \preccurlyeq a_2 \Leftrightarrow \mathbb{E}_p[u(\omega, a_1)] \leq \mathbb{E}_p[u(\omega, a_2)], \forall p \in \mathcal{K}$$

- The set $\mathcal{K}$ is the **credal set** representing $\preccurlyeq$.
- **Implication**: 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 Ω is a **closed**, **convex** subset of the probability simplex:

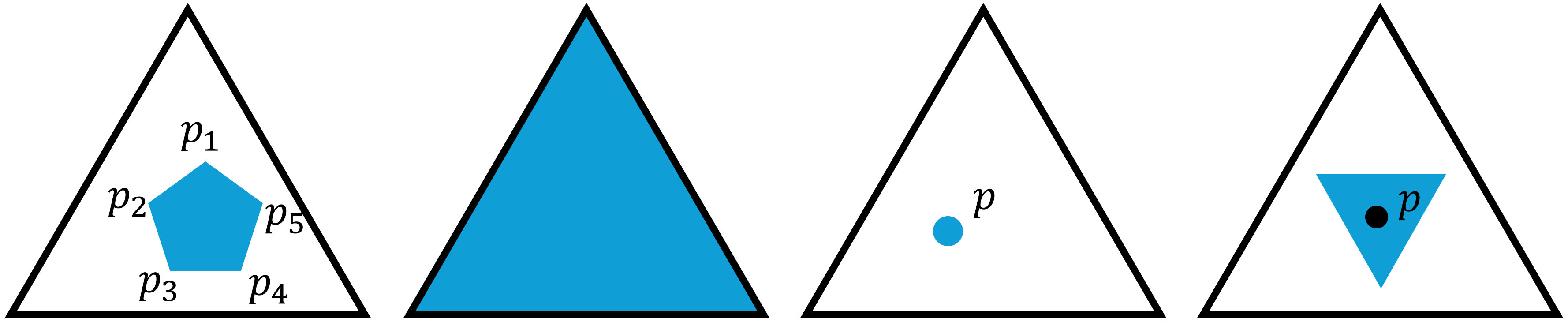
$$\mathcal{K} \subseteq \Delta^{|\Omega|-1} = \left\{ \boldsymbol{p} \in R^{|\Omega|} : \; p_i \geq 0, \sum_{i=1}^{|\Omega|} p_i = 1 \right\}$$

- **Convexity** ensures that any mixture of admissible probability measures is itself admissible, i.e., $p, q \in \mathcal{K} \Longrightarrow \alpha p + (1 - \alpha) q \in \mathcal{K}$ for any $\alpha \in [0,1]$.
- By the **Krein–Milman theorem**, the credal set $\mathcal{K}(\Omega)$ can be equivalently described by its **extreme points** $\text{ext}[\mathcal{K}(\Omega)]$.

Special Cases

- A finitely generated credal set (FGCS) $\mathcal{K} \coloneqq \mathrm{ConvexHull}(\{p_1, \dots, p_n\})$
- A vacuous credal set $\mathcal{K}^S \coloneqq \{\, p \in \Delta^{|\Omega|-1} \colon p(S) = 1\}$ for some $S \in \mathcal{E}$.
- A vacuous credal set $\mathcal{K} \coloneqq \Delta^{|\Omega|-1}$ (complete ignorance)
- A singleton credal set $\mathcal{K} \coloneqq \{p\}$ (precise belief)
- A linear-vacuous or ϵ-contamination credal set:

$$\mathcal{K}_{p,\epsilon} \coloneqq \left\{(1-\epsilon)p + \epsilon q \,:\, q \in \Delta^{|\Omega|-1}\right\}$$



Construction of Credal Sets

- **Interval probabilities:** The credal set arises as the *intersection of probability constraints* $L(\omega) \leq p(\omega) \leq U(\omega)$ with the simplex.
- **Coherent lower previsions:** The credal set is the set of distributions dominating all lower bounds associated with the lower prevision:

$$\mathcal{K} := \{p : p(A) \geq \underline{P}(A), \forall A\}$$

- **Elicitation from expert opinion:** $\mathcal{K} := \mathrm{ConvexHull}(\{p_1, \ldots, p_n\})$
- **Relative likelihood**: The set of models whose likelihood exceeds a given fraction α of the maximum likelihood, i.e., $\mathcal{K}_\alpha = \{h : \gamma(h) \geq \alpha\}$
- **Ensemble and data-driven methods**
- **Conformal prediction**

Lower and Upper Expectations

- The **expectation** of a function f on Ω 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:

$$\underline{\mathbb{E}}[f+g] \geq \underline{\mathbb{E}}[f] + \underline{\mathbb{E}}[g], \qquad \underline{\mathbb{E}}[\alpha f + \beta] = \alpha \underline{\mathbb{E}}[f] + \beta,\ \alpha > 0, \beta \in R$$

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)**:

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

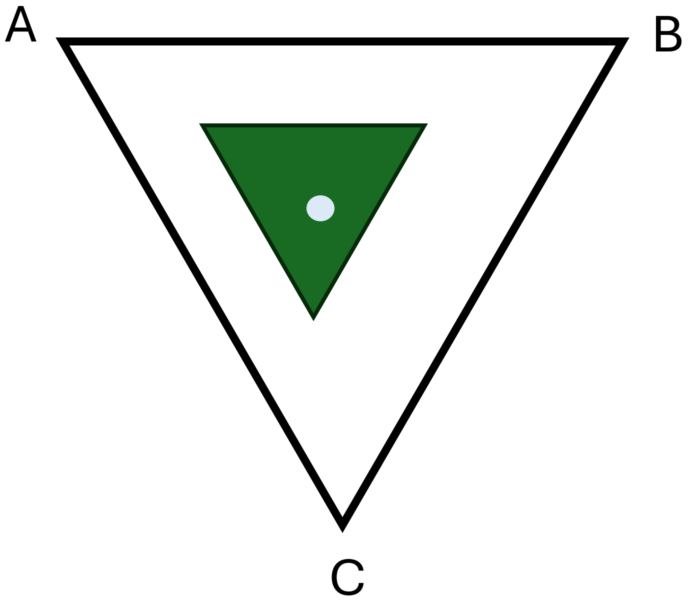
$$\overline{p}(\omega) = \sup_{p \in \mathcal{K}} p(\omega) = \sup_{p \in \mathrm{ext}[\mathcal{K}]} p(\omega)$$

- The conjugate relation:

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

Example

- $\mathcal{K}_{p,\epsilon} \coloneqq \{(1-\epsilon)p + \epsilon q : q \in \Delta^3\}$ 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 $f = (1, 0, -1)$



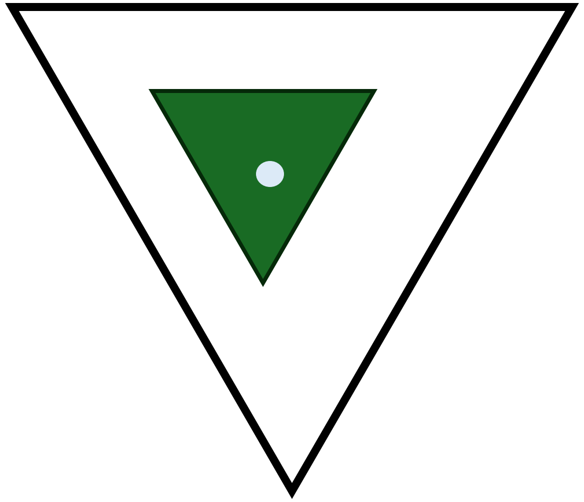
First, we can write the credal set as $\mathcal{K}_{p,\frac{1}{3}} \coloneqq \left\{\frac{2}{3}p + \frac{1}{3}q : q \in \Delta^3\right\}$:

$$\begin{aligned}\underline{P}(\{\mathrm{B},\mathrm{C}\}) &= \min_{r \in \mathcal{K}_{p,\frac{1}{3}}} r(\{\mathrm{B},\mathrm{C}\}) = \min_{r \in \mathcal{K}_{p,\frac{1}{3}}} r(\mathrm{B}) + r(\mathrm{C}) \\ &= \min_{q \in \Delta^3} \tfrac{2}{3}(p_\mathrm{B} + p_\mathrm{C}) + \tfrac{1}{3}(q_\mathrm{B} + q_\mathrm{C}) \\ &= \min_{q \in \Delta^3} \tfrac{2}{3}\left(\frac{1}{4} + \frac{2}{4}\right) + \tfrac{1}{3}(q_\mathrm{B} + q_\mathrm{C}) = \frac{1}{2}\end{aligned}$$

The last step holds because the minimum is attained at $q = 0$.

Example

- $\mathcal{K}_{\boldsymbol{p},\epsilon} \coloneqq \{(1-\epsilon)p + \epsilon q : q \in \Delta^3\}$ 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 $f = (1,0,-1)$



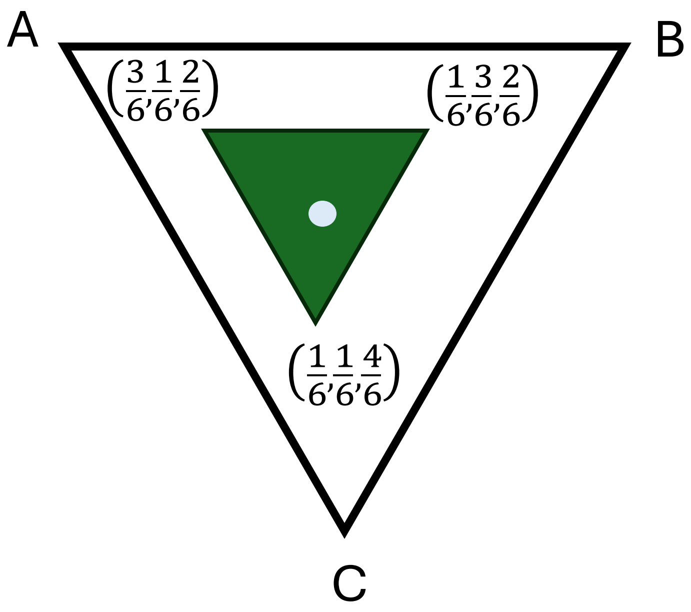
First, we can write the credal set as $\mathcal{K}_{p,\frac{1}{3}} \coloneqq \left\{\frac{2}{3}p + \frac{1}{3}q : q \in \Delta^3\right\}$:

$$\begin{aligned}
\underline{\mathbb{E}}(f) &= \min_{r \in \mathcal{K}_{p,\frac{1}{3}}} \underline{\mathbb{E}}_r(f) = \min_{r \in \mathcal{K}_{p,\frac{1}{3}}} r_\mathrm{A} - r_\mathrm{C} \\
&= \min_{q \in \Delta^3} \tfrac{2}{3}(p_\mathrm{A} - p_\mathrm{C}) + \tfrac{1}{3}(q_\mathrm{A} - q_\mathrm{C}) \\
&= \min_{q \in \Delta^3} \tfrac{2}{3}\left(\frac{1}{4} - \frac{2}{4}\right) + \tfrac{1}{3}(q_\mathrm{A} - q_\mathrm{C}) \\
&= \min_{q \in \Delta^3} \; -\tfrac{2}{3} \times \tfrac{1}{4} + \tfrac{1}{3}(q_\mathrm{A} - q_\mathrm{C}) = -\frac{1}{2}
\end{aligned}$$

The minimum is attained when $q_\mathrm{A} = 0$ and $q_\mathrm{C} = 1$.

Example

- $\mathcal{K}_{p,\epsilon} \coloneqq \{(1-\epsilon)p + \epsilon q : q \in \Delta^3\}$ 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 $f = (1, 0, -1)$



Now, let's compute $\underline{P}(\{\mathrm{B}, \mathrm{C}\})$ using the extreme points $\mathrm{ext}[\mathcal{K}]$:

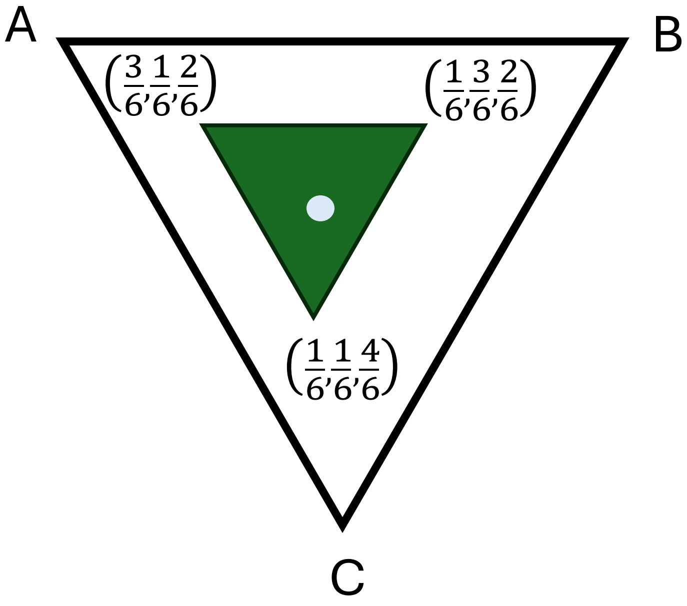
$$\underline{P}(\{\mathrm{B}, \mathrm{C}\}) = \min_{\boldsymbol{r} \in \mathrm{ext}[\mathcal{K}]} r(\{\mathrm{B}, \mathrm{C}\}) = \min_{\boldsymbol{r} \in \mathrm{ext}[\mathcal{K}]} r(\mathrm{B}) + r(\mathrm{C})$$

$$= \min\left(\frac{1}{6} + \frac{2}{6}, \frac{3}{6} + \frac{2}{6}, \frac{1}{6} + \frac{4}{6}\right) = \min\left(\frac{1}{2}, \frac{5}{6}, \frac{5}{6}\right) = \frac{1}{2}$$

which coincides with what we have computed.

Example

- $\mathcal{K}_{p,\epsilon} \coloneqq \{(1-\epsilon)p + \epsilon q : q \in \Delta^3\}$ 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 $f = (1, 0, -1)$



Now, the extreme points $\mathrm{ext}[\mathcal{K}]$ are given:

$$\underline{\mathbb{E}}(f) = \min_{r \in \mathrm{ext}[\mathcal{K}]} \underline{\mathbb{E}}_r(f) = \min_{r \in \mathrm{ext}[\mathcal{K}]} r_\mathrm{A} - r_\mathrm{C}$$

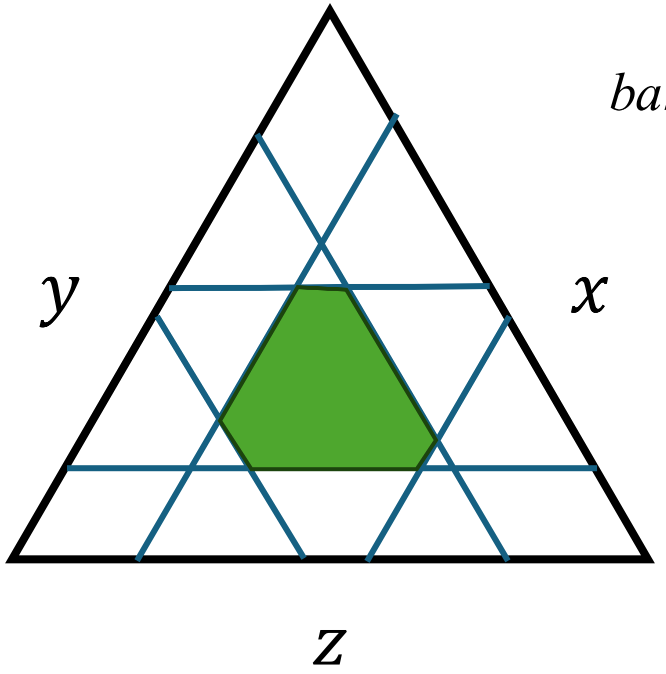
$$= \min\left(\frac{3}{6} - \frac{2}{6}, \frac{1}{6} - \frac{2}{6}, \frac{1}{6} - \frac{4}{6}\right) = \min\left(\frac{1}{6}, -\frac{1}{6}, -\frac{3}{6}\right) = -\frac{1}{2}$$

which coincides with what we have computed.

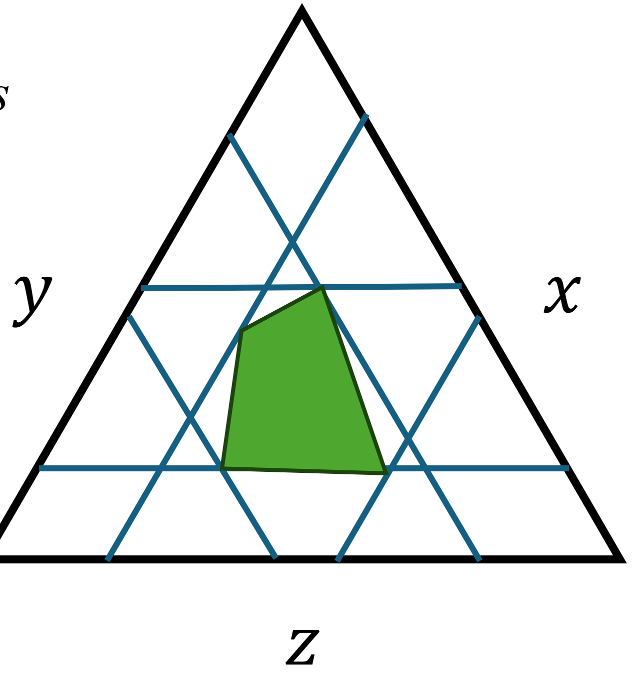
Connections to IP Models

Interval of Probabilities

Several sets of distributions may generate the same lower envelope.



baricentric coordinates



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

Lower Expectations

- One can create a credal set from a lower expectation model $\underline{\mathbb{E}}[\cdot]$:

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

- $\mathcal{K}$ consists of distributions p whose $\mathbb{E}_p[f]$ **dominates** $\underline{\mathbb{E}}[f]$ for all f.
- If $\underline{\mathbb{E}}[\cdot]$ is **superadditive** and **affinely homogeneous**:

$$\underline{\mathbb{E}}[f + g] \geq \underline{\mathbb{E}}[f] + \underline{\mathbb{E}}[g], \qquad \underline{\mathbb{E}}[\alpha f + \beta] = \alpha\underline{\mathbb{E}}[f] + \beta, \alpha > 0, \beta \in R$$

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 ω_1, ω_2
  4. $\overline{p}(\omega_1 \cup \omega_2) \leq \overline{p}(\omega_1) + \overline{p}(\omega_2)$ for any disjoint ω_1, ω_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 ω_1, ω_2
- There is a correspondence between the 2-monotone lower probability and the set of all probability distributions that **dominate** it.

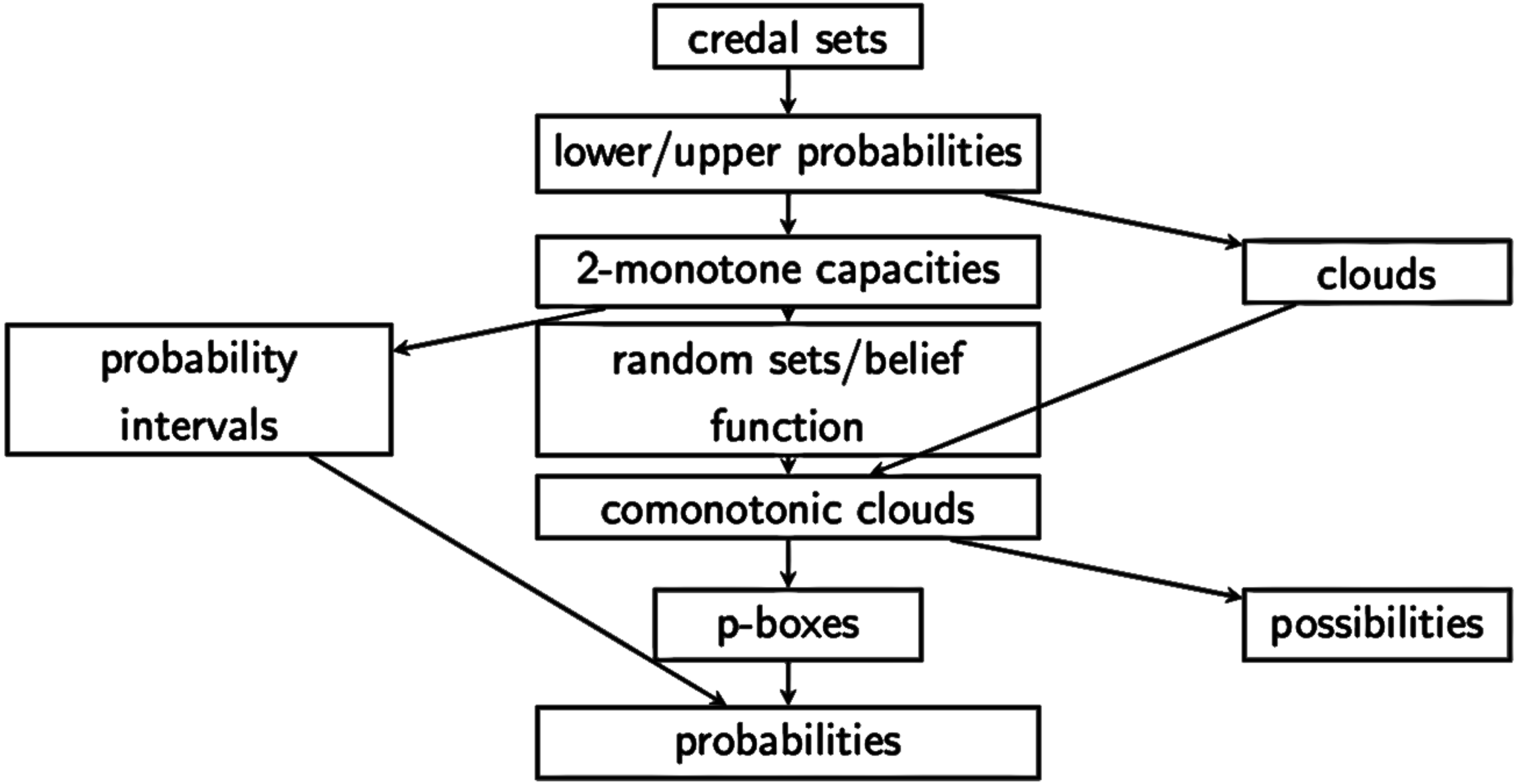
Choquet Capacities

- 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 $A_1, \dots, A_k$, the following inequality holds:

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

- 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 of IP Relationships



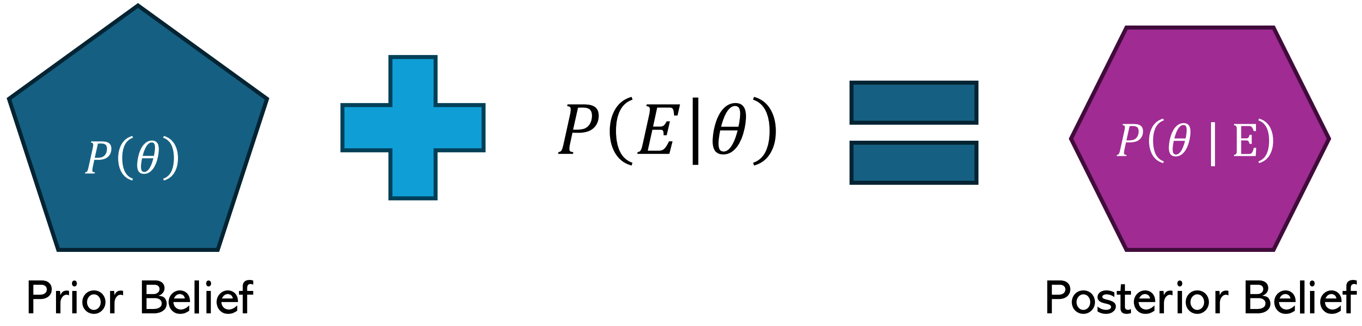
Learning and Reasoning

Marginal and Conditional Credal Sets

- A Bayesian agent can update their belief using standard **Bayes rule**:

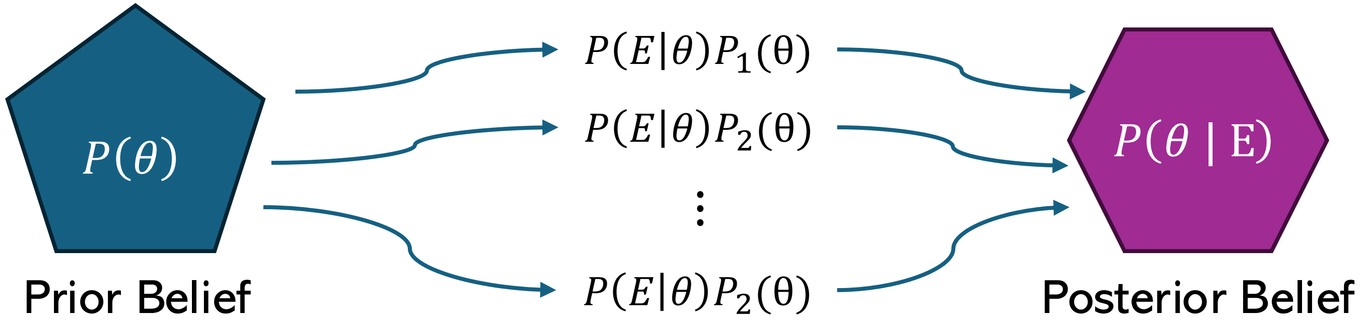
$$P(\theta|E) \propto P(E|\theta)P(\theta)$$

- Here, $P(E|\theta)$ and $P(\theta)$ are the **likelihood function** of the evidence E and the agent's **prior belief** over some parameter θ.
- Is there an equivalent way for a quasi-Bayesian agent to update their belief?



Generalised Bayes Rule

- A quasi-Bayesian agent maintains a **convex of posterior distributions**, each of which is obtained using standard Bayes rule:



Caveat: No general expression like Bayes rule for conditional lower expectations, lower envelopes, lower probabilities, or Choquet capacities.

Independence

- Two events X and Y are **stochastically independent** if $P(XY) = P(X)P(Y)$
- Is there a concept of independence for the theory of credal sets? **Yes, but there are many of them:**
  1. **Complete independence:** X and Y are *completely independent* when for all $P \in \mathcal{K}(X, Y)$, $P(X = x, Y = y) = P(X = x) \times P(Y = y)$.
  2. **Strong independence**: X and Y are *strongly independent* when $\mathcal{K}(X, Y)$ is the convex hull of a set of distributions satisfying complete independence.
  3. and many more.

Caveat: No consensus on the right definition of credal independence!

Epistemic Independence

- Epistemic independence is defined with respect to Walley's symmetrized **irrelevance** between X and Y.
  - **Levi's confirmational irrelevance**: Y is *confirmationally irrelevant* to X when $\mathcal{K}(X \mid Y = \mathrm{y}) = \mathcal{K}(X)$.
  - **Walley's epistemic irrelevance:** Y is *epistemically irrelevant* to X when for any function $f(X)$, $\underline{\mathbb{E}}[f(X) \mid Y = y] = \underline{\mathbb{E}}[f(X)]$.

- We say that X and Y are ***epistemically independent*** when
  1. Y is epistemically irrelevant to X and
  2. X is epistemically irrelevant to Y .

Recommended Reading

- Introduction to the Theory of Sets of Probabilities by Fabio Cozman
- Introduction to the Theory of Imprecise Probability by Erik Quaeghebeur
- [Video] SIPTA School 2024: Introduction to Imprecise Probabilities by Erik Quaeghebeur