Property Elicitation on Imprecise Probabilities

James Bailie[1*] & Rabanus Derr[2*]

*Authors listed alphabetically [1]Harvard University, USA [2]University of Tübingen and Tübingen AI Center, Germany

Motivation

- Machine learning ≈ minimizing *expected* loss.
- Robustification against set of possible probability distributions: multi-distribution learning (MDL) ≈ minimizing *worst-case expected* loss.
- Relatedly, distributionally robust optimization (DRO) = *worst-case* stochastic programming.
- **What *can* a model learn when trained via MDL?**
- **What *does* a model learn when trained via MDL?**
- (Asymptotically) an MDL model learns the IP-property *elicited* by its loss function.
- Thus, property elicitation on IP is a useful theory for answering the above two questions.
- **Read this poster if:** you like *penguins*!

Background: Property Elicitation

Definition

Let $\mathcal{R} \subseteq \mathbb{R}^p$ be some set of property values. A **property** is a function $f : \mathbb{P} \to 2^{\mathcal{R}}$ for some $\mathbb{P} \subseteq \Delta(\mathcal{Z})$. For $\theta \in \mathcal{R}$, the **level set** $\mathcal{L}_\theta$ of a property is defined as

$$\mathcal{L}_\theta = \{P \in \mathbb{P} \colon \theta \in f(P)\}.$$

A property f is **elicitable** if there exists a loss function $\ell : \mathcal{R} \times \mathcal{Z} \to \mathbb{R}$ such that for all $P \in \mathbb{P}$,

$$f(P) = \arg\min_{\theta \in \mathcal{R}} \mathbb{E}_{Z\sim P}[\ell(\theta, Z)].$$

Bayes Pair. Let $\ell \colon \mathcal{R} \times \mathcal{Z} \to \mathbb{R}$ be a loss function. The property pair (Θ, L_Θ) is called a **Bayes pair** if Θ is the property elicited by ℓ and, for all $P \in \mathbb{P}$,

$$L_\Theta(P) = \min_{\theta \in \mathcal{R}} \mathbb{E}_{Z\sim P}[\ell(\theta, Z)].$$

Example (Squared error elicits the mean and variance Bayes pair)**.** Suppose $\mathcal{Z} \subseteq \mathbb{R}$ and $\mathcal{R} \subseteq \mathcal{Z}$. Let $\ell(\theta, z) = (\theta - z)^2$ be the loss function.

Suppose that $\mathbb{P}$ is the set of all distributions with finite second moments.

For $P \in \mathbb{P}$, the property $\Theta(P)$ of the corresponding Bayes pair is the mean of P, because

$$\mathbb{E}_{Z\sim P}[(\theta - Z)^2] = \theta^2 - 2\theta\mathbb{E}_{Z\sim P}[Z] + \mathbb{E}_{Z\sim P}[Z^2]\,,$$

is minimized at $\theta = \mathbb{E}_{Z\sim P}[Z]$.

The corresponding $L_\Theta(P)$ is the variance of P since

$$\min_{\theta \in \mathcal{R}} \mathbb{E}_{Z\sim P}[\ell(\theta, Z)^2] = \mathbb{E}_{Z\sim P}[(\mathbb{E}_{Z\sim P}[Z] - Z)^2]\,.$$

IP-Properties

Definition

Let $\mathcal{R} \subseteq \mathbb{R}^p$ be a set of property values. An **IP-property** is a function $f : \mathscr{P} \to 2^{\mathcal{R}}$ for some $\mathscr{P} \subseteq 2^{\Delta(\mathcal{Z})}$ with $\emptyset \notin \mathscr{P}$. For $\theta \in \mathcal{R}$, the **level set** $\mathscr{L}_\theta$ of a property is defined as

$$\mathscr{L}_\theta = \{\mathcal{P} \in \mathscr{P} : \theta \in f(\mathcal{P})\},$$

An IP-property f is **elicitable** if there exists a loss function $\ell : \mathcal{R} \times \mathcal{Z} \to \mathbb{R}$ such that, for all $\mathcal{P} \in \mathscr{P}$,

$$f(\mathcal{P}) = \arg\min_{\theta \in \mathcal{R}} \sup_{P \in \mathcal{P}} \mathbb{E}_{Z\sim P}[\ell(\theta, Z)].$$

Γ-minimax and DRO. Let ℓ be the loss of a decision maker, with $\mathcal{R}$ a finite set of possible decisions. Then f is the corresponding Γ-minimax—or equivalently DRO—decision rule (cf. Troffaes, 2007; Rahimian and Mehrotra, 2022).

Given an imprecise probability $\mathcal{P}$, e.g. a credence about nature's state, $f(\mathcal{P})$ is the decision which minimizes the worst-case risk.

Proper Scoring Rules. When ℓ is a strictly proper scoring rule for forecasts, $f(\mathcal{P})$ can be interpreted as the set of forecasts which are Γ-minimax admissible under $\mathcal{P}$ (as first introduced in Schervish et al., 2025, Definition 2.1).

Mean and Variance. Consider the binary outcome set $\mathcal{Z} = \{0, 1\}$. We denote probabilities $p \in \Delta(\mathcal{Z})$ as $p \in [0, 1]$. Let $\ell(\theta, z) = (\theta - z)^2$ be the loss function which elicits f. Then, for $\mathcal{P} = [0, 0.5]$,

$$f(\mathcal{P}) = \arg\min_{\theta \in \mathbb{R}} \sup_{p \in [0,0.5]} \theta^2 - p(2\theta - 1) = 0.5.$$

Note that, for $\mathcal{Q} = \{0\}$, $f(\mathcal{Q}) = \arg\min_{\theta\in\mathbb{R}} \theta^2 = 0$, while for $\mathcal{Q}' = \{0.5\}$, $f(\mathcal{Q}') = \arg\min_{\theta\in\mathbb{R}} \theta^2 = 0.5$. Hence, there exist $Q, Q' \in \mathcal{P}$ such that $f(\{Q\}) \neq f(\mathcal{P}) = f(\{Q'\})$. In particular, observe that $f(\mathcal{P}) = f(\{Q'\})$ for Q' being the maximal variance distribution in the set $\mathcal{P}$.

Necessary Conditions for Elicitability

Proposition

Let $f \colon \mathscr{P} \to 2^{\mathcal{R}}$ be a full,* elicitable IP-property. Then, the following four statements hold:

(I) For all $\mathcal{P} \in \mathscr{P}$, $f(\mathcal{P}) = f(\overline{\mathrm{co}}\mathcal{P})$.

(II) The level sets of f are convex; i.e. for all $\mathcal{P}, \mathcal{Q} \in \mathscr{L}_\theta$ and $\alpha \in [0, 1]$, the IP

$$\alpha\mathcal{P}+(1-\alpha)\mathcal{Q} = \{\alpha P+(1-\alpha)Q : P \in \mathcal{P}, Q \in \mathcal{Q}\}$$

is also in $\mathscr{L}_\theta$.

(III) The level sets of f are closed under arbitrary unions. That is, if $\{\mathcal{P}_i\}_{i\in I} \subset \mathscr{L}_\theta$, then $\bigcup_{i\in I} \mathcal{P}_i \in \mathscr{L}_\theta$.

(IV) The precise restriction $\hat{f}$ of f is elicitable.

Counterexample to closure under intersection. Consider the outcome set $\mathcal{Z} = \{0, 1, 2\}$. We denote probabilities $p \in \Delta(\mathcal{Z})$ as $(p_0, p_1, p_2) \in [0, 1]^3$. Let $\ell(\theta, z) = (\theta - z)^2$ be the loss function which elicits f.

Then, for $\mathcal{P} = \{(1, 0, 0), (0.5, 0, 0.5)\}$,

$$f(\mathcal{P}) = 1,$$

the mean of $p = (0.5, 0, 0.5)$ which is the maximum variance distribution in $\overline{\mathrm{co}}\mathcal{P}$.

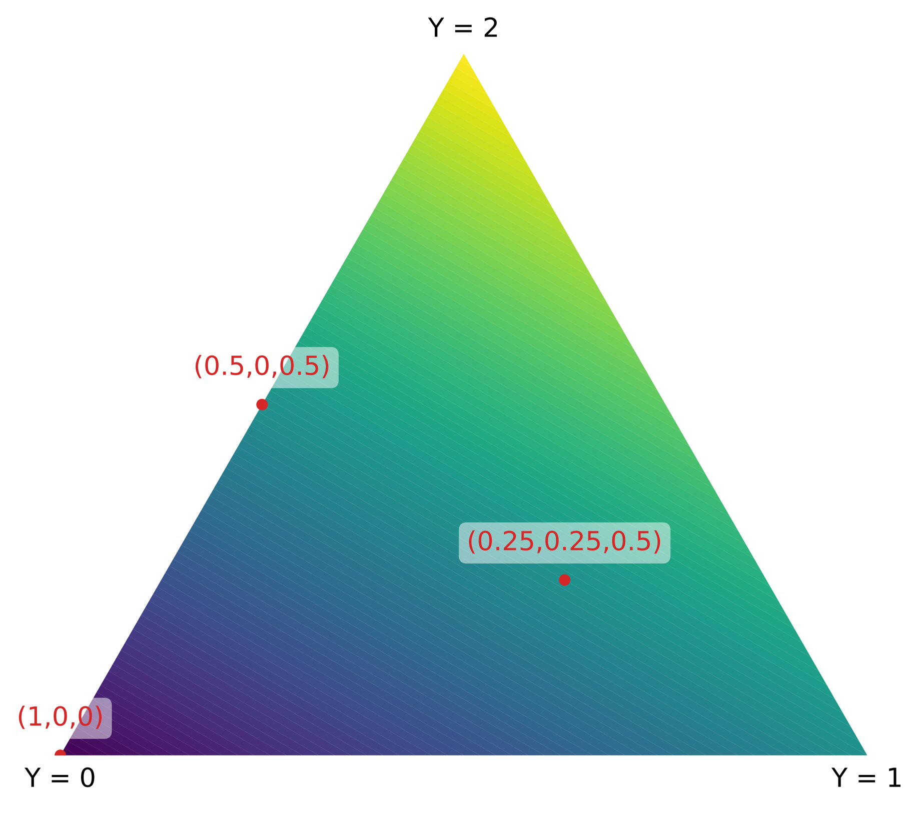
For $\mathcal{Q} = \{(1, 0, 0), (0.25, 0.5, 0.25)\}$, we obtain,

$$f(\mathcal{Q}) = 1,$$

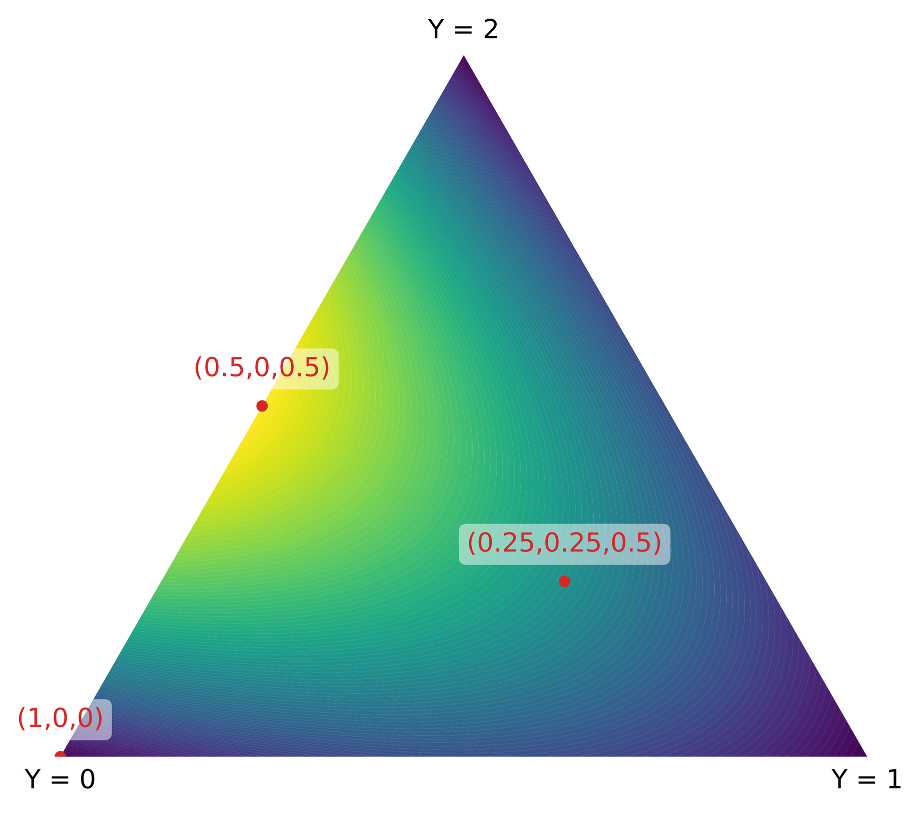
the mean of $p = (0.25, 0.5, 0.25)$ which is the maximum variance distribution in $\overline{\mathrm{co}}\mathcal{Q}$.

However, it is easy to see that,

$$f(\mathcal{P} \cap \mathcal{Q}) = f(\{(1, 0, 0)\}) = 0.$$



Mean on simplex. Distributions marked in red are taken from the counterexample on the left.



Variance on simplex. Distributions marked in red are taken from the counterexample on the above left.

Minimax Solutions for IP-Elicitation

(Informal) Proposition

Let (Θ, L_Θ) be the Bayes pair corresponding to the loss function ℓ. Then, the IP-property $f \colon \mathscr{P} \to 2^{\mathcal{R}}$ elicited by ℓ satisfies $f(\mathcal{P}) \subseteq \Theta(P^*)$ for all $P^* \in \arg\max_{P\in\mathcal{P}} L_\Theta(P)$.

Related results in Grünwald and Dawid, 2004; Fröhlich and Williamson, 2024; Schervish et al., 2025.

Example. If $\ell(z, \theta) = (z - \theta)^2$ and $\mathcal{Z} = [-C, C]$ for some constant $C > 0$, then, for all $\mathcal{P} \in 2^{\Delta(\mathcal{Z})}$, ℓ elicits the mean of maximum variance distribution in $\mathcal{P}$ (Embrechts et al., 2021, Example 1.(iii)).

Technical conditions:

1. The loss function ℓ is lower semi-continuous and convex in θ for every $z \in \mathcal{Z}$, and upper semi-continuous in $z \in \mathcal{Z}$ for every $\theta \in \mathcal{R}$.
2. The set of property values $\mathcal{R}$ is convex.
3. Let (Θ, L_Θ) be the Bayes pair corresponding to the loss function ℓ.
4. Let $\mathcal{P} \in \mathscr{P}$ be closed and convex, and suppose Θ is defined for every element in $\mathcal{P}$.

Then the IP-property $f \colon \mathscr{P} \to 2^{\mathcal{R}}$ elicited by ℓ satisfies $f(\mathcal{P}) \subseteq \Theta(P^*)$ for all $P^* \in \arg\max_{P\in\mathcal{P}} L_\Theta(P)$.

ARXIV.ORG/ABS/2507.05857

Summary

- We provide a list of necessary conditions for IP-properties to be elicitable.
  - Only properties which satisfy these conditions *can* be learned in MDL.
- We further show that an IP-property $f(\mathcal{P})$ elicited through Γ-maximin is equal to the standard property elicited on the maximum Bayes risk distribution in $\mathcal{P}$.
  - MDL *learns* the property of the maximum Bayes risk distribution.

Open Questions.

1. For (precise) probabilities, elicitability can be related to identifiability. What is the analogue of identifiability in the imprecise case, and how does it relate to elicitability (Steinwart et al., 2014)?
2. Comparison of IP-properties for loss functions which elicit the same property in the precise setting (e.g. Bregmann divergences).
3. Sufficient conditions for IP-elicitability, and (eventually) a full characterization of elicitable IP-properties.

Selected References

Grünwald and Dawid (2004). “Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory”. In: *The Annals of Statistics* 32.4, pp. 1367–1433.

Embrechts et al. (2021). “Bayes risk, elicitability, and the expected shortfall”. In: *Mathematical Finance* 31.4, pp. 1190–1217.

Fröhlich and Williamson (2024). *Scoring rules and calibration for imprecise probabilities*. arXiv:2410.23001.

Schervish et al. (2025). “Elicitation for sets of probabilities and distributions”. In: *The 14th Intl. Symposium on Imprecise Probability: Theories and Applications*. Vol. 290. PMLR, pp. 242–251.