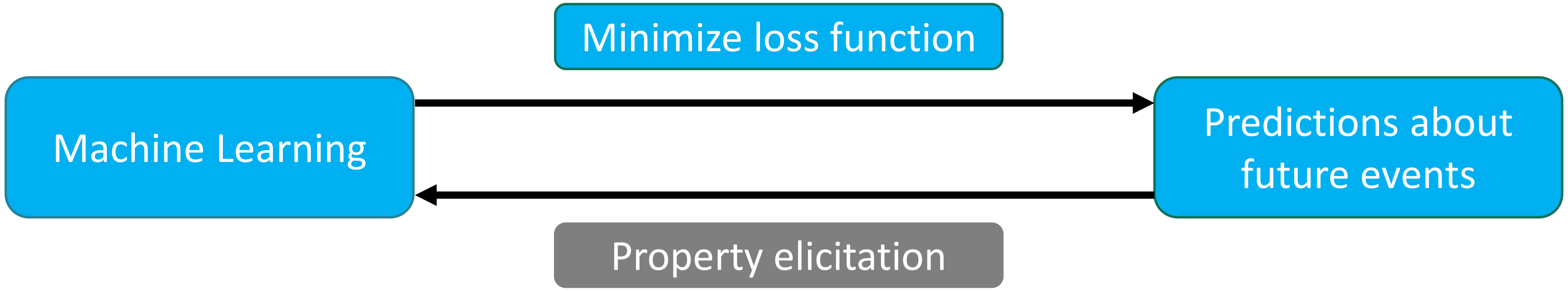
Property Elicitation as a tool for Understanding Consistent Polyhedral Losses

Jessie Finocchiaro

Based on joint work with Rafael Frongillo and Bo Waggoner

December 15, 2020

Loss functions and property elicitation



Outline

- Elicitation Introduction
- Embeddings [**F**FW19]
- Embedding Dimension [**F**FW20]

Goal: use elicitation to study when convex surrogates are calibrated.

Elicitation tools are arguably a bit easier to work with.

Properties

A **property** $\Gamma: \Delta_{\mathcal{Y}} \to \mathcal{R}$ maps probability distributions to optimal reports.

ex: Expected value, mode, median, top-k classifier, ranking.

A property $\Gamma: \Delta_{\mathcal{Y}} \to \mathcal{R}$ is **elicitable** if there is a loss function $L: \mathcal{R} \times \mathcal{Y} \to \mathbb{R}_+$ such that, for all $p \in \Delta_{\mathcal{Y}}$,

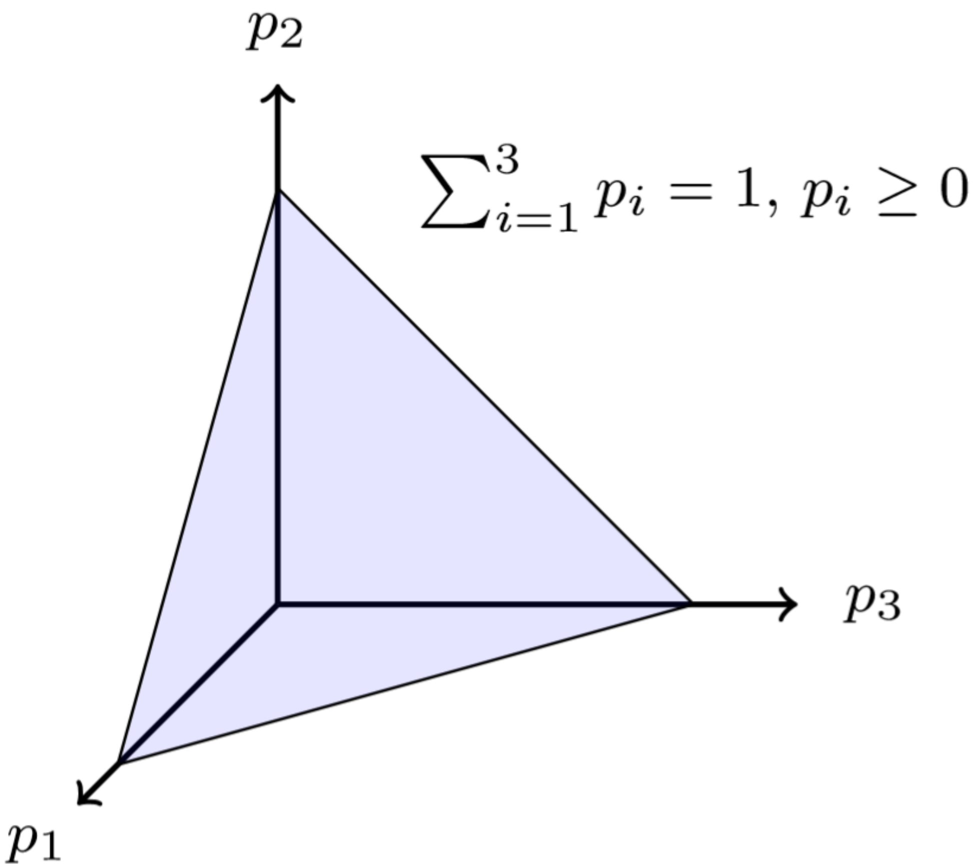
$$\Gamma(p) = \operatorname{argmin}_r \mathbb{E}_{Y \sim p} L(r, Y).$$

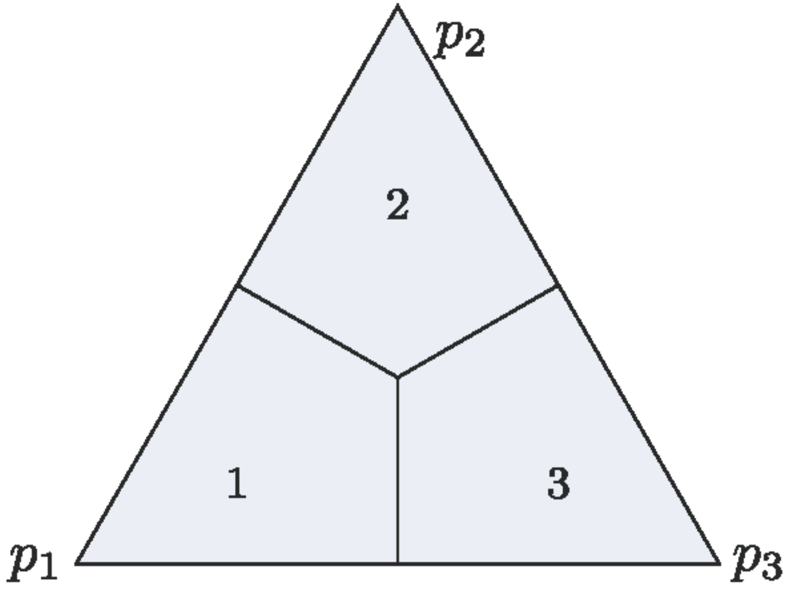
Here, we say the loss L elicits Γ.

“Drawing” a property

n-simplex in $(n-1)$-dimensional space.

Example: $n = 3$





Elicitable properties have convex level sets. [LS09]

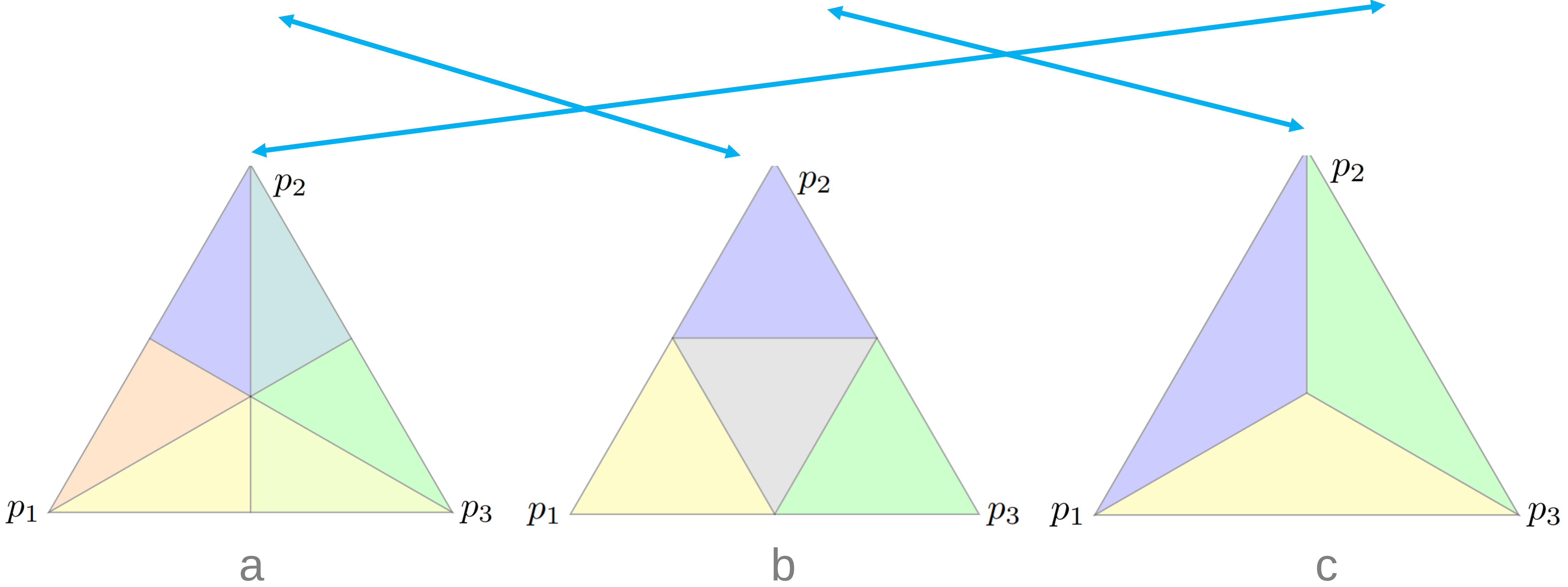
$\Gamma_r = \{p \in \Delta_{\mathcal{Y}} : r \in \Gamma(p)\}$

Pop quiz

1) Abstain

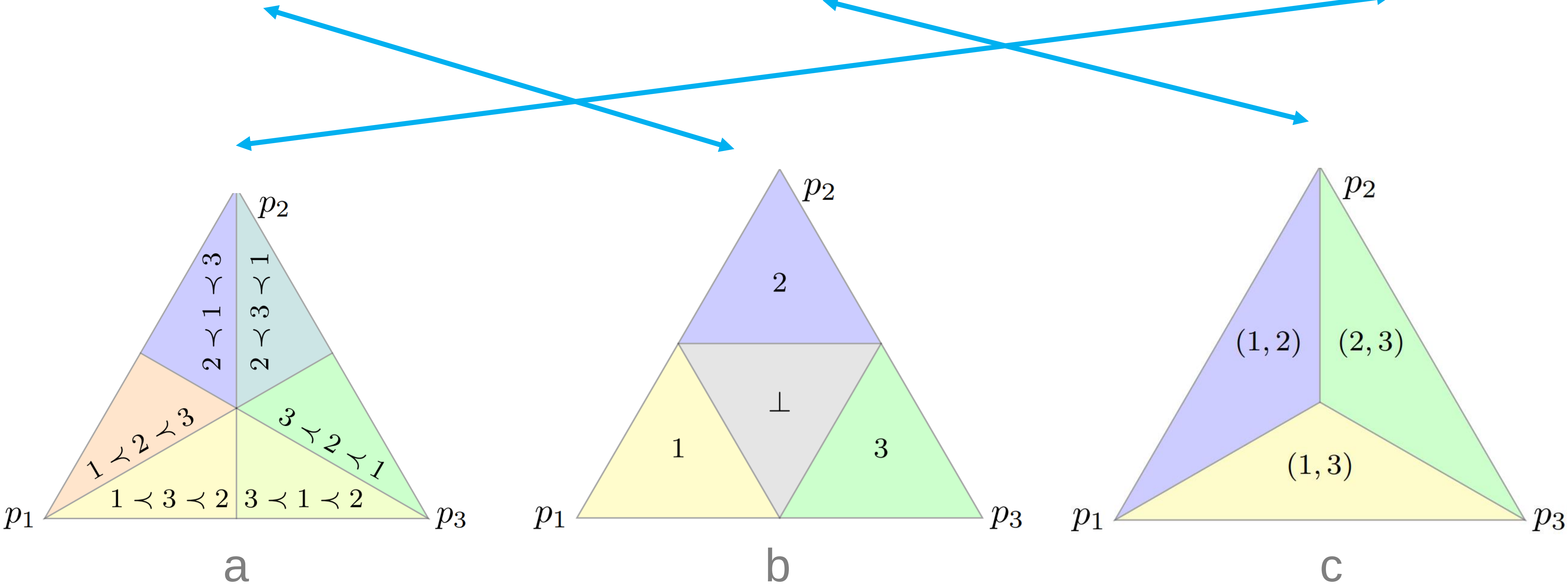
2) Top-2

3)Ranking

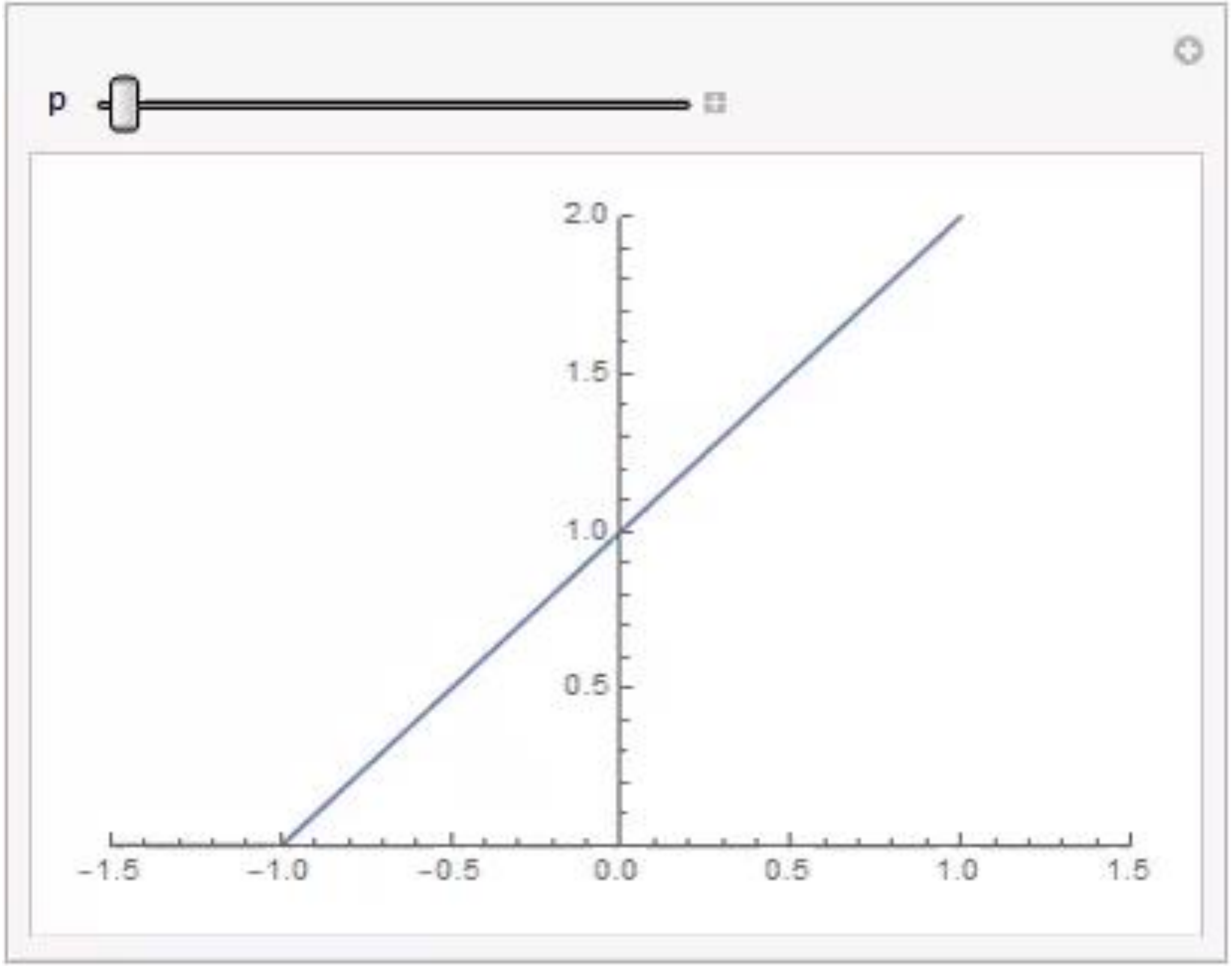


Pop quiz

1) Abstain　　2) Top-2　　3)Ranking



Eliciting a loss: Hinge and the mode



What is this? You promised me a talk about consistency

A surrogate loss and link pair (L, ψ) is calibrated with respect to a target loss ℓ if

$$\forall p \in \Delta_{\mathcal{Y}}, \quad \inf_{u \in \mathbb{R}^d : \psi(u) \in \gamma(p)} \mathbb{E}_p L(u, Y) > \inf_{u \in \mathbb{R}^d} \mathbb{E}_p L(u, Y)$$

Inf over reports not linked to the property values elicited by the target loss.

A surrogate loss and link pair (L, ψ) is consistent with respect to a target loss ℓ if

$$\forall\, D \in \Delta_{\mathcal{X} \times \mathcal{Y}}, msbl\, \{f_m : \mathcal{X} \to \mathcal{R}\}$$

$$\mathbb{E}_D L(f_m(X), Y) \to \inf_f \mathbb{E}_D L(f(X), Y) \Rightarrow \mathbb{E}_D \ell(\psi \circ f_m(X), Y) \to \inf_f \mathbb{E}_D \ell(\psi \circ f(X), Y)$$

Relating elicitation, consistency, and calibration

[BJM06, TB07, AA15, RTA16]

If $\mathcal{Y}$ is a finite set, calibration $\Leftrightarrow$ consistency

[**F**FW, forthcoming]

If $\mathcal{Y}$ is a finite set, consistency $\Rightarrow$ indirect elicitation

The research questions

Given a property γ (elicited by target loss ℓ),

1) When can we construct a convex loss L that also elicits* γ?

Today: focus on embedding for finite prediction problems.

2) When/how can we construct a surrogate that has low prediction dimension?

*We're okay with indirect elicitation, meaning we elicit some other property and can compute γ from that.

Outline

- Elicitation Introduction
- Embeddings [**F**FW19]
- Embedding Dimension [**F**FW20]

Research Questions

1. When can we construct a convex loss L that also indirectly elicits γ?
2. When/how can we construct a surrogate that has low prediction dimension?

Setting for embeddings

Finite prediction problems

- Binary classification
- Multiclass classification (with reject option)
- Quantiles (chunked)
- Ordered partitions

Notation

- Finite outcomes $|\mathcal{Y}| \coloneqq n$
- Report set $\mathcal{R}$
- Probability Distribution over $\mathcal{Y}$: $p \in \Delta_{\mathcal{Y}}$
  - $p_y = \Pr[Y = y]$
- Bayes Risk of loss $\underline{L}(p) = \inf_u \mathbb{E}_p L(u, Y)$

Embedding conditions

A polyhedral (piecewise linear and convex) loss $L: \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}$ embeds a target loss ℓ if there exists an embedding $\varphi: \mathcal{R} \to \mathbb{R}^d$

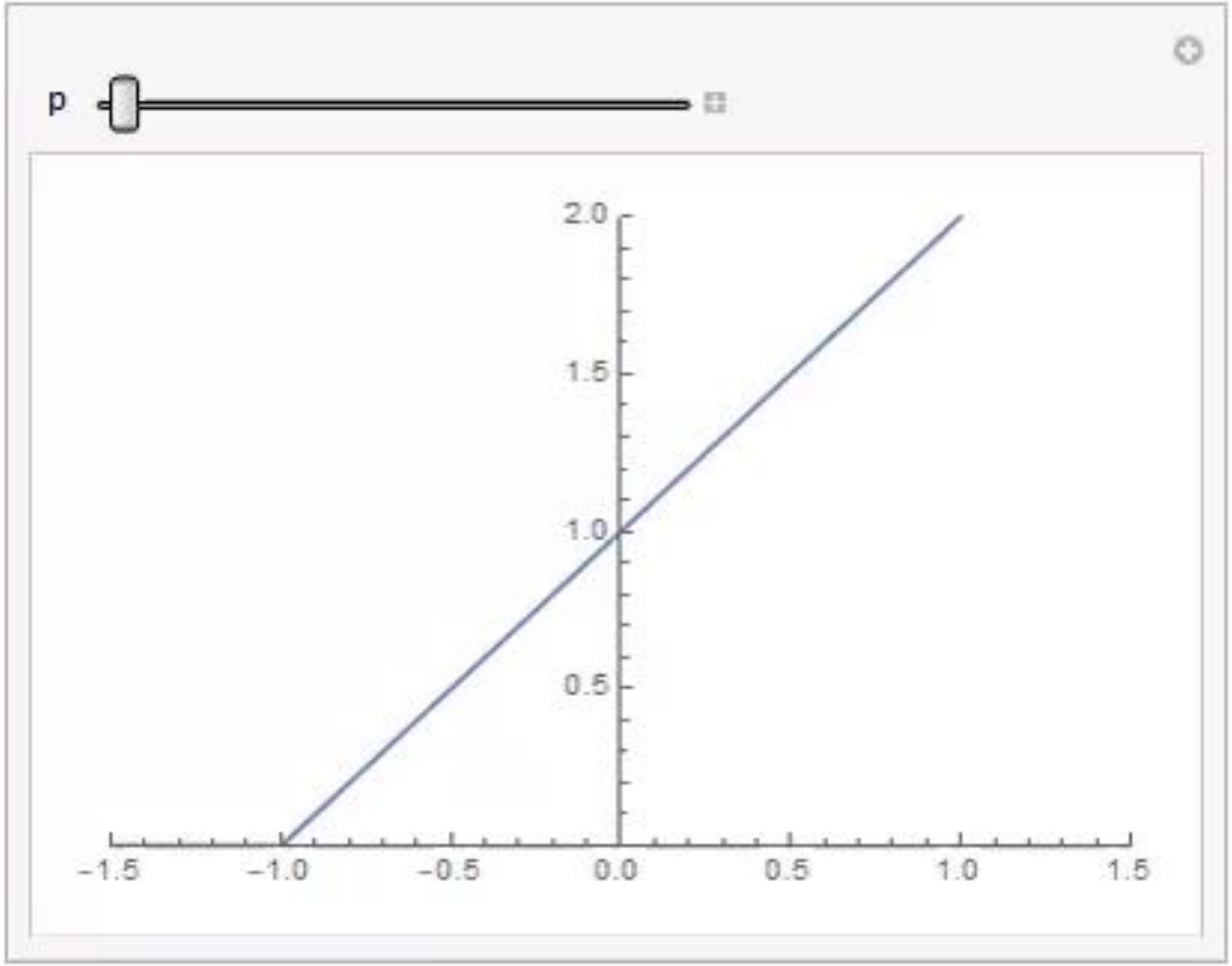
1) For all $r \in \mathcal{R}, y \in \mathcal{Y}$, we have $L(\varphi(r), y) = \ell(r, y)$
2) For all $p \in \Delta_{\mathcal{Y}}, r \in \mathcal{R}$, we have

$$r \in prop[\ell](p) \Leftrightarrow \varphi(r) \in \text{prop}[L](p)$$

[**F**FW19 Proposition 1]

L embeds ℓ if and only if $\underline{L} = \underline{\ell}$

Matching losses



Embedding 0-1 loss: property conditions

The hinge loss L embeds (twice) 0-1 loss ℓ via the embedding $\varphi(r) = r$ if

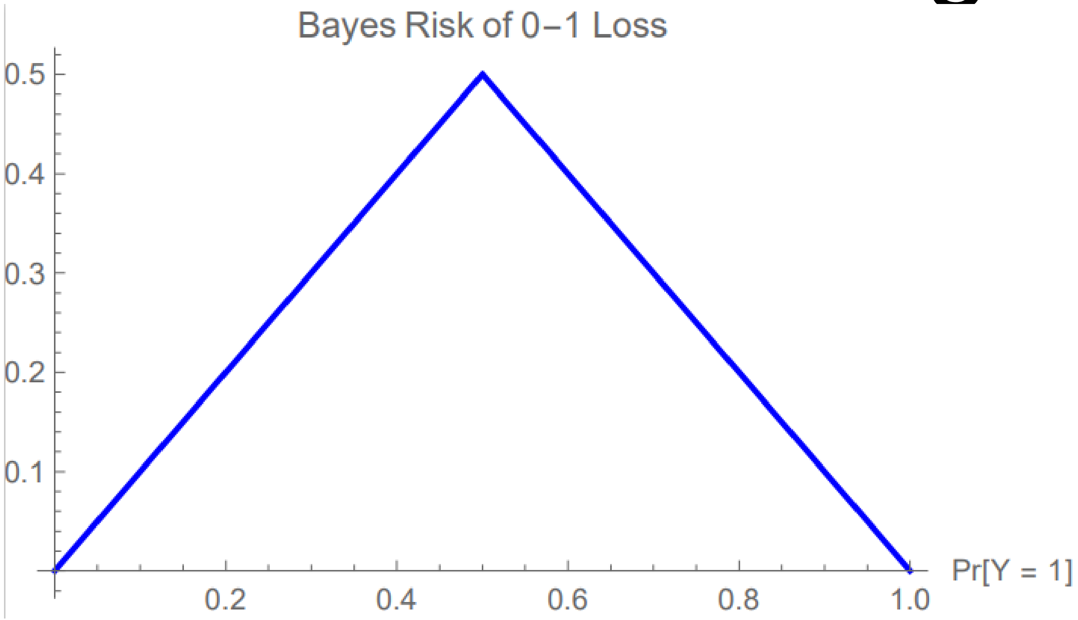
2) For all $p \in \Delta_{\mathcal{Y}}, r \in \mathcal{R}$, we have

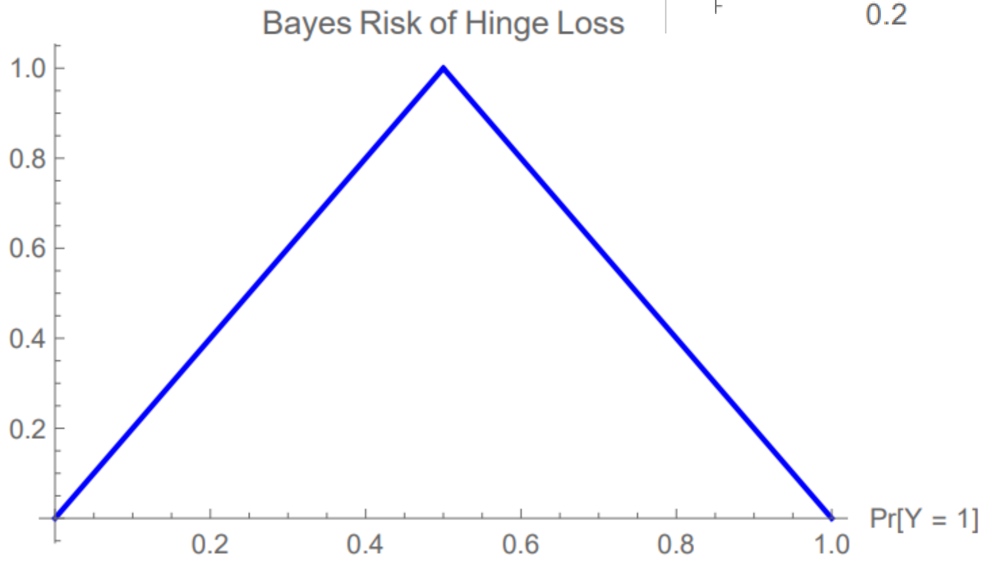
$$r \in prop[\ell](p) \Leftrightarrow \varphi(r) \in \mathrm{prop}[L](p)$$

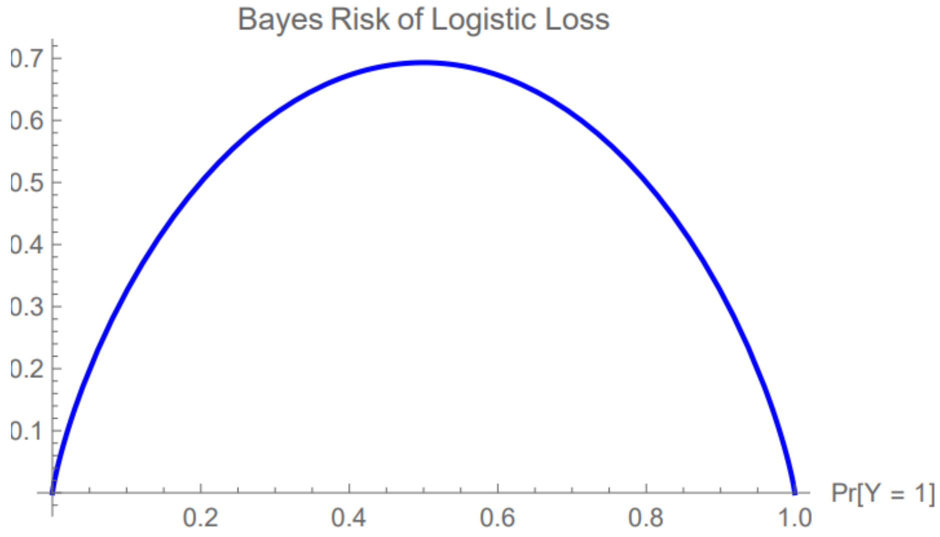
$$prop[0-1](p) = \begin{cases} 1 & p_1 > \frac{1}{2} \\ \{-1,1\} & p_1 = \frac{1}{2} \\ -1 & p_1 < \frac{1}{2} \end{cases}$$

$$prop[hinge](p) = \begin{cases} [1,\infty) & p_1 = 1 \\ 1 & p_1 \in (\frac{1}{2}, 1) \\ [-1,1] & p_1 = \frac{1}{2} \\ -1 & p_1 \in (0, \frac{1}{2}) \\ (-\infty, 1] & p_1 = 0 \end{cases}$$

Embedding 0-1 loss: Matching Risks

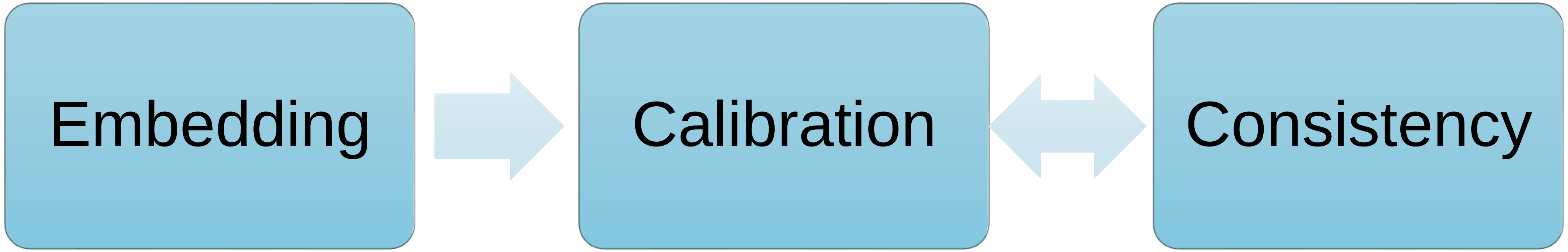






Embedding and calibration

If a surrogate loss L embeds a property γ elicited by target loss ℓ, then there is a link ψ so that (L, ψ) are calibrated with respect to ℓ. [**F**FW Theorem 3]



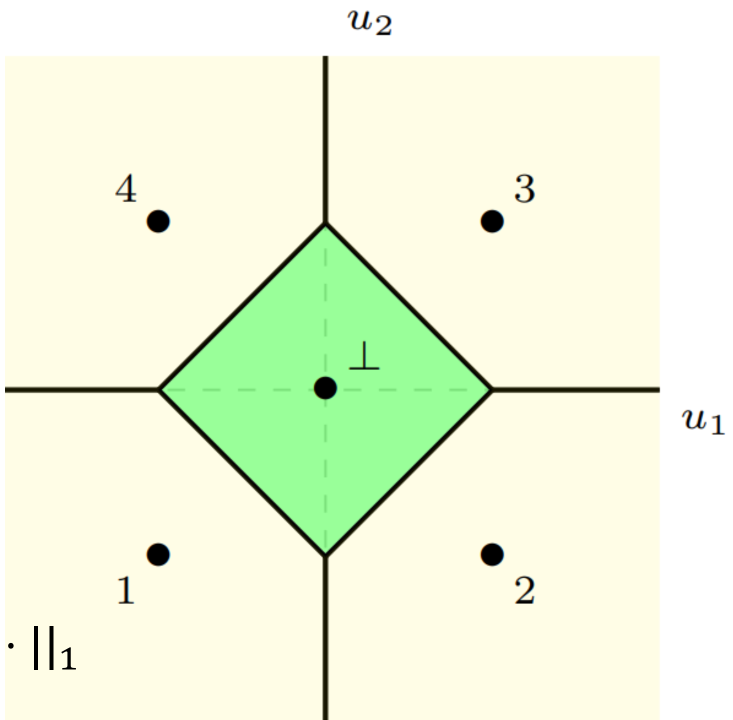
Calibrated link construction

Given embedding, norm, and ϵ

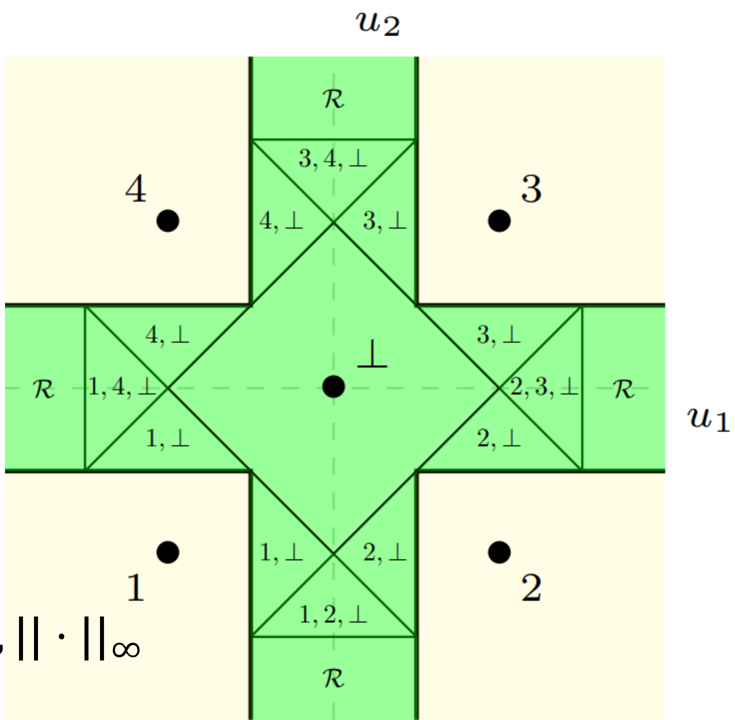
For all $u \in \mathbb{R}^d$, assign $\psi^*(u) = \mathcal{R}$

Go through possible optimal sets $\mathrm{R_U}$, take the (normed) ball of radius ϵ around this set, reassign $\psi^*(u) = \psi^*(u) \cap R_U$

Take any choice of ψ^* for ψ



$\epsilon = 1, || \cdot ||_1$



$\epsilon = \frac{1}{2}, || \cdot ||_\infty$

Outline

- Elicitation Introduction
- Embeddings [**F**FW19]
- Embedding Dimension [**F**FW20]

Research Questions

1. When can we construct a convex loss L that also indirectly elicits γ?
    1. For all finite prediction problems, via embeddings
2. When/how can we construct a surrogate that has low prediction dimension?

Moving to “good” embeddings

Can any finite property (target loss) be embedded?

Yes! Using classic constructions in the market design literature. [**F**FW19 Theorem 2]

Is there a notion of a “best” embedding?

Look at dimension- $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}$

Low d improves the complexity of the optimization problem, while embedding says we are still learning the task we want to learn.

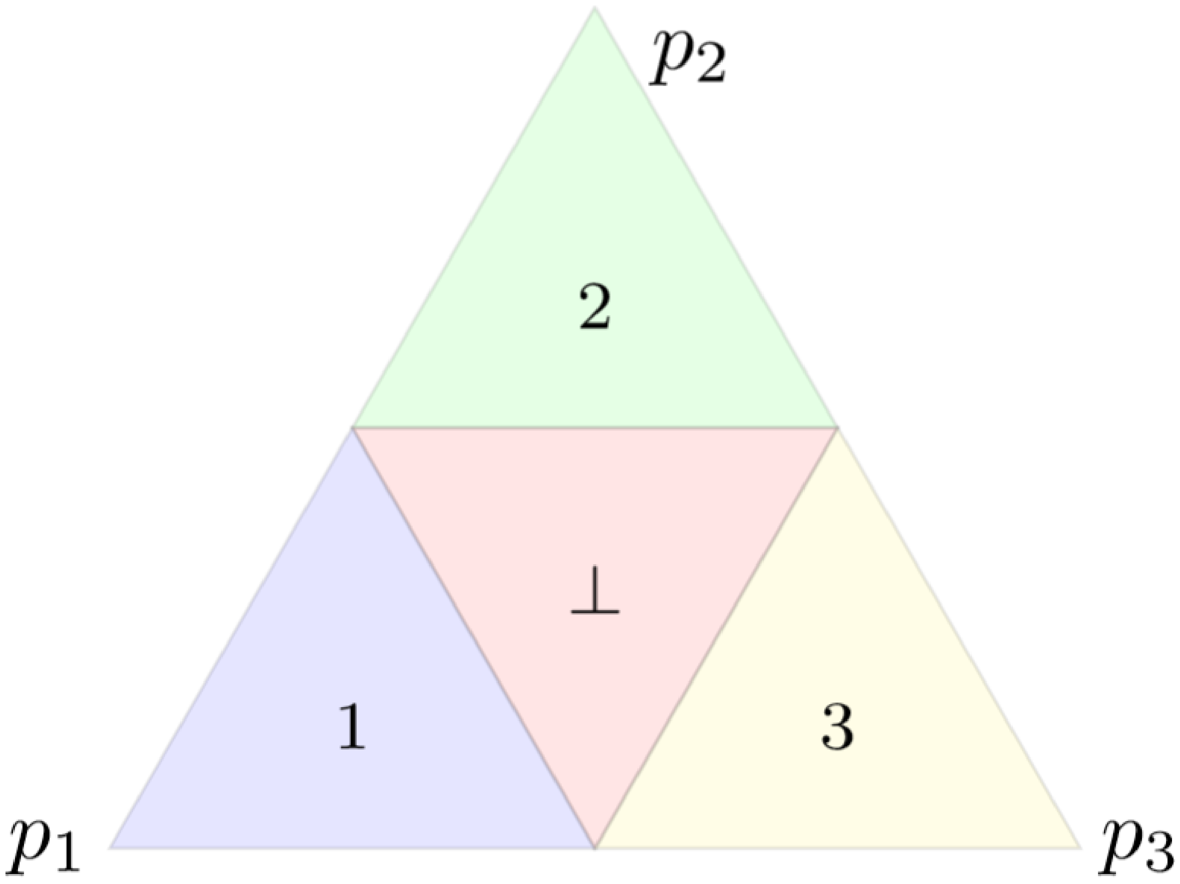
Case Study: Abstain

Situations where the cost of misclassification is high.

- College admissions.
- Medical diagnoses.

Discrete loss for this problem:

$$\ell^{\frac{1}{2}}(r, y) = \begin{cases} 0, & r = y \\ \frac{1}{2}, & r = \bot \\ 1, & r \notin \{y, \bot\} \end{cases}$$



Real-valued embeddings: $d = 1$

[FFW20 Theorem 12]

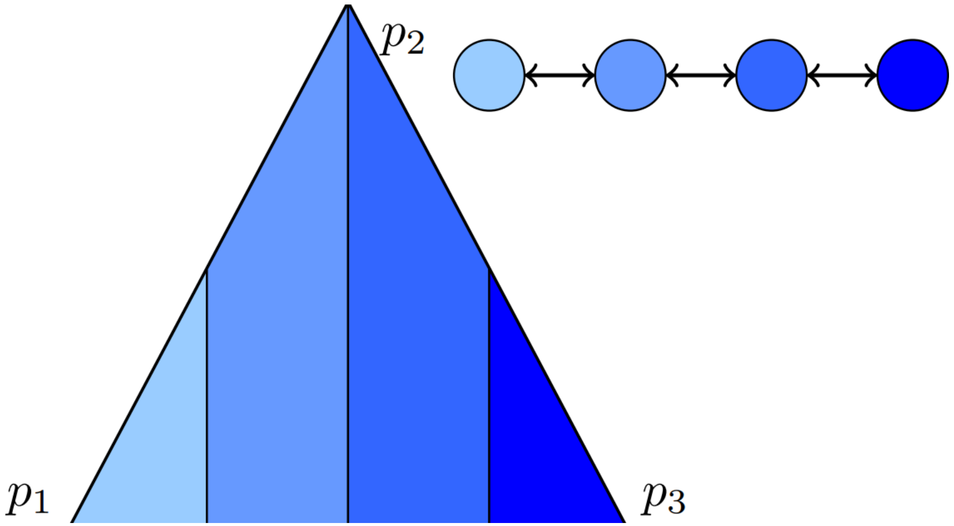
Given a finite property γ elicited by the target loss ℓ, the following conditions are equivalent:

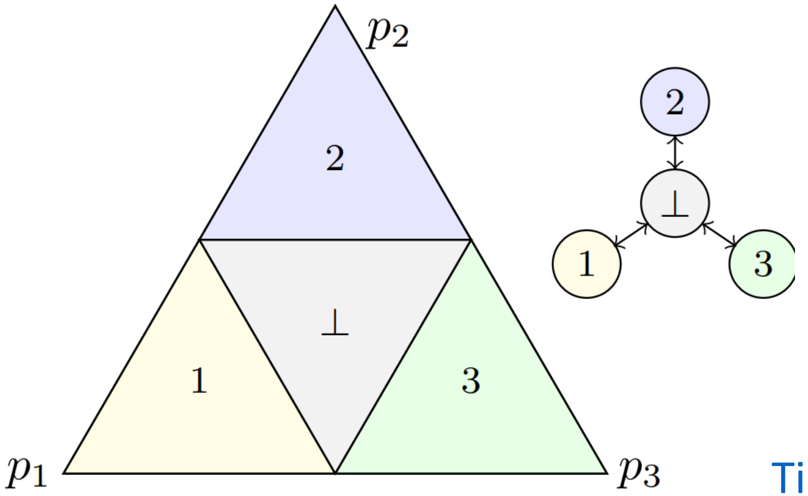
a) γ is 1-embeddable

b) The intersection graph of γ is a path

c) There is a polyhedral calibrated surrogate loss calibration for ℓ

d) There is a convex calibrated surrogate loss calibration for ℓ



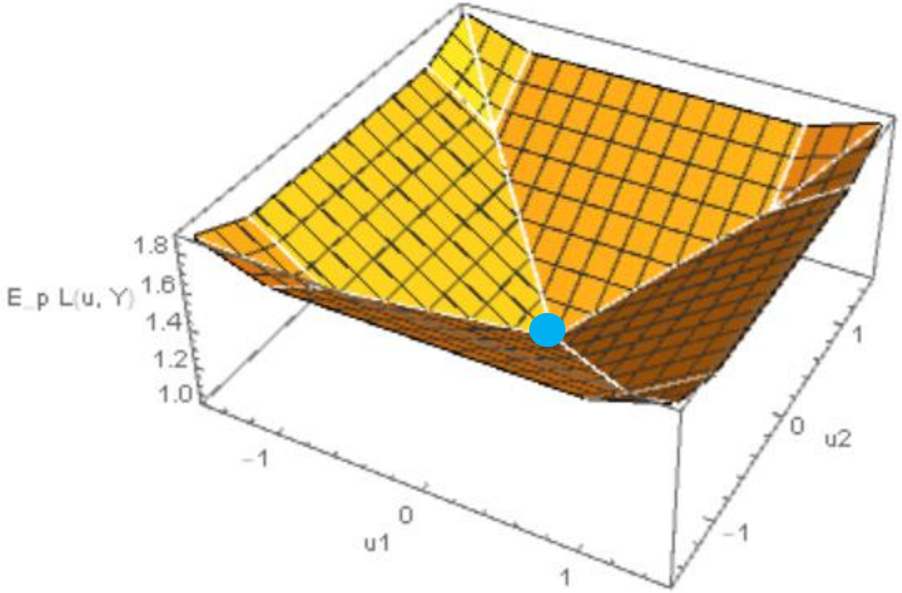


Time crunch ☺

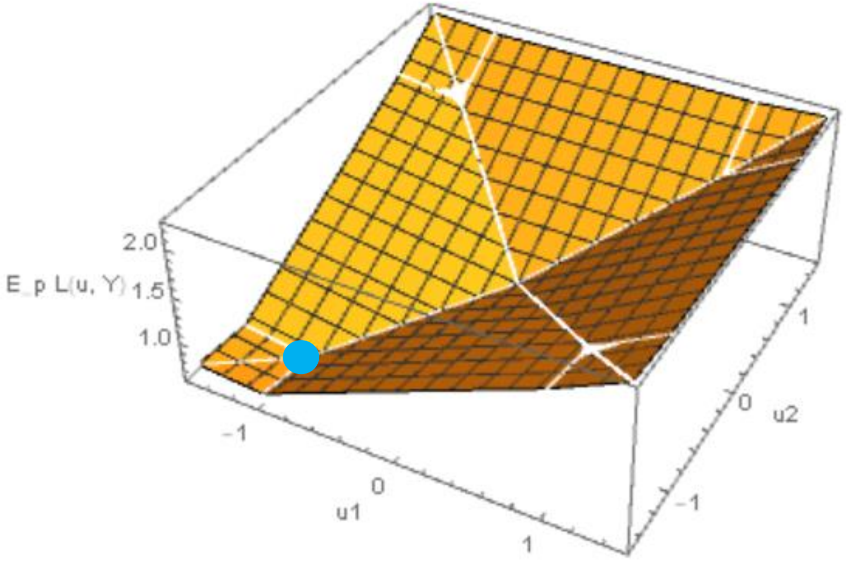
Abstain $d > 1; n \geq 3$

$$L^{1/2}(u, y) = \max_{j \in [d]} (B_j(y) u_j + 1)_+$$

$$\psi^{1/2} = \begin{cases} \perp & \min_{1 \leq i \leq d} |u_i| \leq \tau \\ B^{-1}(\text{sgn}(-u)) & \text{otherwise} \end{cases}$$



$p = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$



$p = (\frac{7}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10})$

Ramaswamy et al. (2018) Section 4

Necessary and sufficient conditions for embeddings

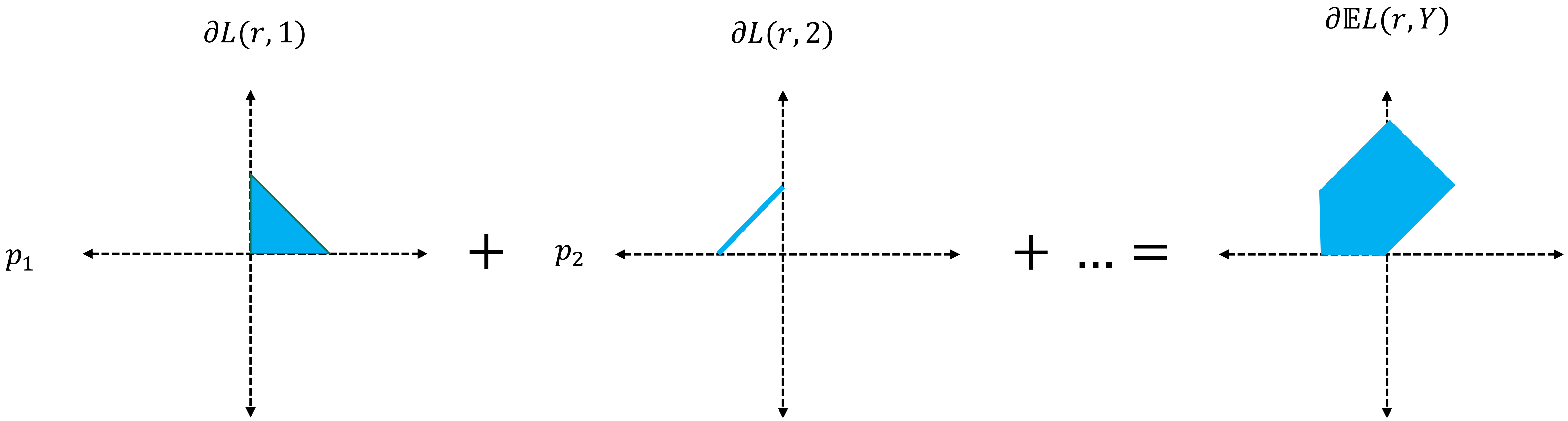
Looking at subgradient sets at embedding points, we find two conditions.

Monotonicity: There must be some adjacency of subgradient sets in the intersection of level sets.
Optimality: the embedding points must minimize expected loss.

Give a Quadratic Feasibility Program equivalent to satisfying optimality. [**F**FW20 Definition 17]

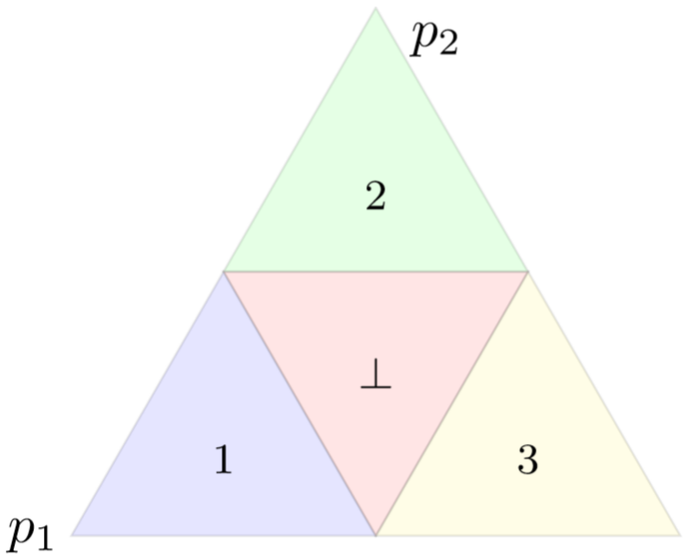
QFP for Optimality

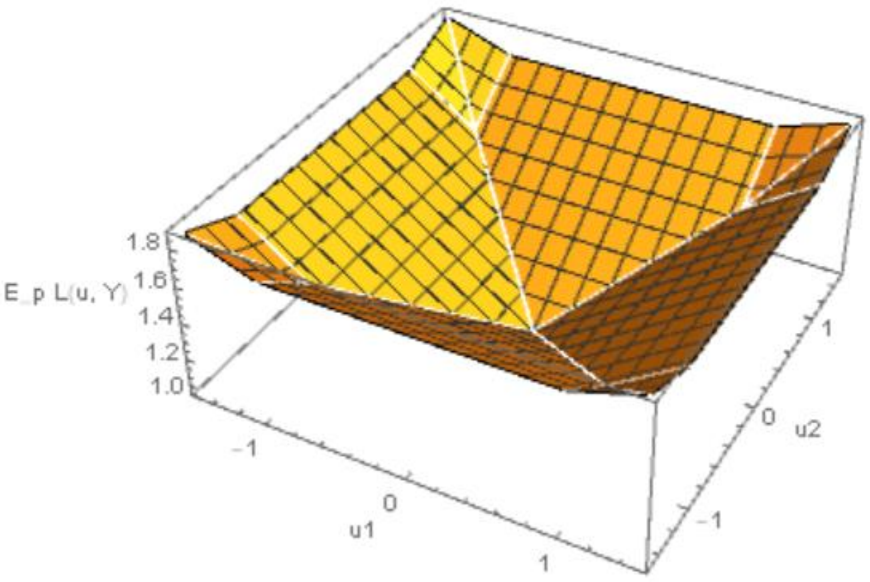
The intuition: look at the subdifferentials at embedded points and take weighted Minkowski sums.



Summary

$$\ell^{\frac{1}{2}}(r,y) = \begin{cases} 0, & r = y \\ \frac{1}{2}, & r = \bot \\ 1, & r \neq y \text{ or } \bot \end{cases}$$





Thank you, questions?