Consistency Conditions for Differentiable Surrogate Losses

NeurIPS 2025

Drona Khurana, Anish Thilagar, Dhamma Kimpara, Rafael Frongillo

University of Colorado Boulder

November 6, 2025

What is a surrogate loss?

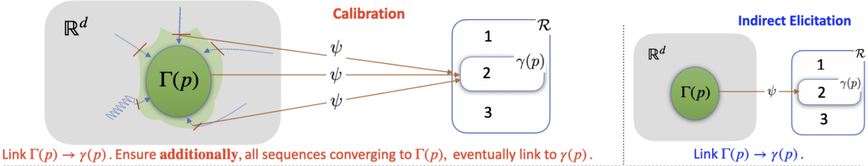
- Discrete prediction tasks specify a natural notion of error called the **target loss** ℓ. Eg: 0-1 loss in classification.
- Let $\mathcal{Y}$ and $\mathcal{R}$ denote (discrete) spaces of labels and predictions respectively.
- $\ell : \mathcal{Y} \times \mathcal{R}$ is discrete, non-convex $\implies$ optimizing ℓ is generally NP-hard.
- Work-around: Replace ℓ with a convex **surrogate loss** $L : \mathcal{Y} \times \mathbb{R}^d \to \mathbb{R}$ and a **link function** $\psi : \mathbb{R}^d \to \mathcal{R}$. Eg: Hinge loss, cross-entropy loss etc.

So, does any convex surrogate do the trick?

- Convexity of L enables ease of optimization, but is L a good proxy for ℓ?
- Not necessarily - (L, ψ) must be **statistically consistent** with respect to ℓ.
- Statistical consistency $\approx$ minimizing the surrogate L enables recovery of the target ℓ minimizer via link ψ provided sufficient data.
- Consistency is hard to verify in practice.
- Bartlett et al. (2006): Under finite $\mathcal{Y} = [n], \mathcal{R} = [k]$, consistency reduces to a simpler condition called **calibration**
- Calibration has served as the key technical condition for designing statistically sound surrogates.

Calibration vs. Indirect Elicitation

- Calibration requires ensuring that for each $p \in \Delta_n$ *all* sequences converging to $\Gamma(p) := \text{argmin}_{u \in \mathbb{R}^d} \mathbb{E}_{y \sim p}[L(u, y)]$ eventually link to $\gamma(p) := \text{argmin}_{r \in \mathcal{R}} \mathbb{E}_{y \sim p}[\ell(r, y)]$.



Notation: $\mathbb{R}^d$- surrogate reports; $\Gamma(p)$- surrogate minimizers; $\mathcal{R}$- target reports; $\gamma(p) = \{2\}$- target minimizers; ψ- link

- Directly verifying calibration can be cumbersome in 2 or more dimensions owing to the need to analyze an uncountable set of sequences converging to $\Gamma(p)$.
- Motivated by this challenge, Finocchiaro et al. (2019) establish equivalence between **Indirect Elicitation (IE)** and Calibration for the class of polyhedral surrogate losses.
- IE is weaker than calibration in general. IE only requires ensuring that surrogate minimizers link to target minimizers making it much simpler to verify.

Research Questions

- While IE and calibration are equivalent for the class of polyhedral losses, no verifiable conditions that yield consistency have been identified for the broad and practically important class of differentiable surrogates.

Question

Does calibration reduce to indirect elicitation for differentiable surrogate losses?

Question

If not, are there other easy to verify conditions that imply calibration for the class of differentiable surrogates?

IE $\iff$ Calibration for 1-d differentiable surrogate losses

Theorem

Let $L : \mathbb{R} \to \mathbb{R}^n$ be a convex, differentiable surrogate that indirectly elicits ℓ. Under a mild technical assumption, L is calibrated with respect to ℓ.

- This IE-calibration equivalence enables us to construct consistent 1-d differentiable surrogates for *any* target loss satisfying a condition called orderability. Targets not satisfying orderability have no consistent surrogates with domain dimension 1 Finocchiaro et al. (2020).

Theorem

Given an orderable target $\ell : \mathbb{R} \to \mathbb{R}^n$, there exists a convex, differentiable surrogate $L : \mathbb{R} \to \mathbb{R}^n$ which is calibrated with respect to ℓ.

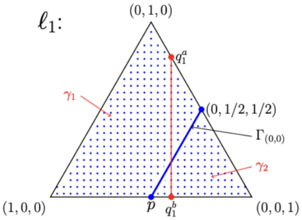
- **Application**: Novel 1-d differentiable surrogate consistent for the ordinal regression loss (see Example 5 in paper)

IE and Calibration are not equivalent in higher dimensions

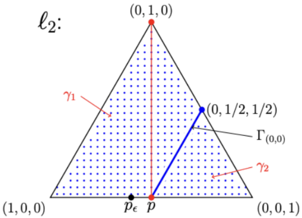
Example

Let $\mathcal{Y} = \{1, 2, 3\}$, $\mathcal{R} = \{1, 2\}$ and consider $L_{\mathrm{CE}} : \mathbb{R}^2 \to \mathbb{R}^{\mathcal{Y}}$, where

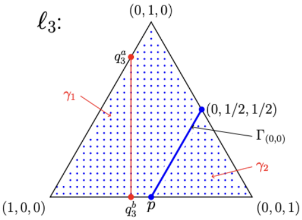
$$L_{\mathrm{CE}}(u) = \begin{pmatrix} u_1^2 + u_1 + u_2^2 + 2u_2 \\ 2u_1^2 + u_1 + 2u_2^2 + 2u_2 \\ 3u_1^2 - u_1 + u_2^2 - 2u_2 \end{pmatrix}.$$



Violates IE and thus violates calibration



Satisfies IE, but still violates calibration



Satisfies calibration

Strong Indirect Elicitation

- Our example suggests that bounding the surrogate level-sets *away* from the target-boundaries recovers calibration. We formalize this via a novel condition, a tightening of IE we term *strong indirect elicitation*

Definition (Strong Indirect Elicitation)

A surrogate L *strongly indirectly elicits* ℓ if for every $u \in \mathbb{R}^d$ and $p, q \in \Gamma_u$, $\gamma(p) = \gamma(q)$.

- Despite being stronger than IE, strong IE remains easy-to-verify as it it characterized entirely by surrogate and target minimizers.

Strong IE $\implies$ Calibration for differentiable surrogates

Theorem

Let L be a convex, differentiable surrogate that strongly indirectly elicits ℓ. Under a mild technical assumption, L is calibrated with respect to ℓ.

- For the practically significant class of strongly convex, differentiable surrogates, strong IE is fundamental as it is both necessary and sufficient for calibration.

Theorem

Let $L : \mathbb{R}^d \to \mathbb{R}^n$ be a surrogate, such that for each $y \in [n]$, the component $L(\cdot)_y : \mathbb{R}^d \to \mathbb{R}$ is strongly convex and differentiable. Then L is calibrated with respect to ℓ if and only if it strongly indirectly elicits ℓ.

IE and Strong IE: Ease of Verification

- IE and strong IE are both completely characterized by the relationship between surrogate and target minimizers.
- Importantly, unlike calibration, neither condition requires analyzing sequences converging to surrogate minimizers. This makes IE and strong IE substantially simpler to verify than calibration.
- To illustrate this ease of verification more concretely, we provide a couple of examples demonstrating how proofs of consistency via calibration can be made substantially simpler and shorter by proceeding via strong IE instead.

Example 3 (Universally calibrated surrogate). *Lemma 11 of Ramaswamy and Agarwal [2016] proposes a $n-1$-dimensional, strongly convex, differentiable surrogate that is calibrated for all discrete targets. After the first claim in their proof (see pages 29-30 Ramaswamy and Agarwal [2016]):*

Proof via strong IE

Fix $p \in \Delta_n$. Minimizing $\langle p, L(u)\rangle = \sum_{j=1}^{n-1} \left(p_j(u_j-1)^2 + (1-p_j)u_j^2\right)$ yields the unique minimizer $u^* = (p_1, \ldots, p_{n-1})^\top$. Hence $|\Gamma(p)| = 1$ and $\Gamma_u = \{p\}$. Immediately, L satisfies strong IE, and thus L is calibrated by Theorem 2.

Our approach shortens the proof from an entire page to a few lines. We also obviate the need for subtle arguments regarding the convergence of sequences that were required in the original proof.

Example 4 (Subset-ranking surrogates). *Theorem 3 of Ramaswamy et al. [2013] proposes a low-dimensional calibrated surrogate for subset-ranking targets common in information retrieval. Our results significantly shorten their calibration proof (see pages 3-4, Ramaswamy et al. [2013]):*

Proof via strong IE

The surrogate is strongly convex and differentiable, so strong IE suffices for calibration. Pick any $\mathbf{u} \in \mathbb{R}^d$ and any $\mathbf{p}, \mathbf{q} \in \Gamma_{\mathbf{u}}$. To prove strong IE, it suffices to show that $\gamma(\mathbf{p}) = \gamma(\mathbf{q})$. $\mathbf{u}$ is the unique minimizer for $\langle \mathbf{p}, L(\cdot)\rangle$ and $\langle \mathbf{q}, L(\cdot)\rangle \implies (*)\ \underline{\mathbf{u}^{\mathbf{p}} = \mathbf{u}^{\mathbf{q}} = \mathbf{u}}$. By line 1 of page 4, $\mathbf{p}^T \boldsymbol{\ell}_t = (\mathbf{u}^{\mathbf{p}})^\top \boldsymbol{\beta}_t + c$. Similarly, $\mathbf{q}^\top \boldsymbol{\ell}_t = (\mathbf{u}^{\mathbf{q}})^\top \boldsymbol{\beta}_t + c$. By $(*)$, $\mathbf{p}^\top \boldsymbol{\ell}_t = \mathbf{q}^\top \boldsymbol{\ell}_t$ for any $t \in \mathcal{T}$ (target reports). Thus, $\text{argmin}_{t\in\mathcal{T}} \mathbf{p}^\top \boldsymbol{\ell}_t = \text{argmin}_{t\in\mathcal{T}} \mathbf{q}^\top \boldsymbol{\ell}_t$. So, $\gamma(\mathbf{p}) = \gamma(\mathbf{q})$.

This bypasses all subsequent proof steps (25 lines) following the first line of page 4 wherein intricate reasoning to show all sequences converging to minimizer sets are appropriately linked.

Future Directions

- We believe *strong* IE's ease of verification can ultimately aid in the design of novel surrogates for open target problems.
- Since strong IE is necessary for calibration to hold for strongly convex, differentiable losses, we believe it can be used to obtain lower bounds on the prediction dimension of consistent surrogates from this class.

Thank you!

For questions/ more details visit our poster!
(4th December, 16:30 - 19:30, Exhibit Hall C,D,E #3003 San Diego Convention Center)

Scan for the arXiv page

arxiv.org/abs/2505.13760

References

Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2006). Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156.

Finocchiaro, J., Frongillo, R., and Waggoner, B. (2019). An embedding framework for consistent polyhedral surrogates. *Advances in neural information processing systems*, 32.

Finocchiaro, J., Frongillo, R., and Waggoner, B. (2020). Embedding dimension of polyhedral losses. In *Conference on Learning Theory*, pages 1558–1585. PMLR.