Distribution Free M-estimation

Felipe Areces

Based on joint work with John Duchi

Stanford University

AISTATS Spotlight, May 4, 2026

Problem formulation

The Fundamental Question:
In the convex risk minimization setting, what is the dividing line between problems that are solvable **without distributional assumptions** and problems that are not?

Problem formulation

The Fundamental Question:
In the convex risk minimization setting, what is the dividing line between problems that are solvable **without distributional assumptions** and problems that are not?

The Minimax Framework

Given a triplet $(\ell, \mathcal{Z}, \Theta)$ consisting of:

- A convex loss function $\ell_z(\theta)$, parameter space Θ, and data space $\mathcal{Z}$

$$\mathfrak{M}_n(\ell, \mathcal{Z}, \Theta) := \inf_{\widehat{\theta}_n} \sup_{P \in \mathcal{P}(\mathcal{Z})} \mathbb{E}_{P^n} \left[L_P(\widehat{\theta}_n) - L_P^\star(\Theta) \right].$$

Problem formulation

The Fundamental Question:
In the convex risk minimization setting, what is the dividing line between problems that are solvable **without distributional assumptions** and problems that are not?

The Minimax Framework

Given a triplet $(\ell, \mathcal{Z}, \Theta)$ consisting of:

- A convex loss function $\ell_z(\theta)$, parameter space Θ, and data space $\mathcal{Z}$

$$\mathfrak{M}_n(\ell, \mathcal{Z}, \Theta) := \inf_{\widehat{\theta}_n} \sup_{P \in \mathcal{P}(\mathcal{Z})} \mathbb{E}_{P^n} \left[L_P(\widehat{\theta}_n) - L_P^\star(\Theta) \right].$$

Objective: Identify properties of $\ell, \mathcal{Z}, \Theta$ to separate

$$\lim_n \mathfrak{M}_n(\ell, \mathcal{Z}, \Theta) > 0 \quad \text{from} \quad \lim_n \mathfrak{M}_n(\ell, \mathcal{Z}, \Theta) = 0$$

The bright dividing line

Assume $\Theta \subset \mathbb{R}^d$ is compact

> **Condition C.1:** For each compact subset $\Theta_0 \subset \operatorname{int}\Theta$, the functions $\ell_z(\cdot)$ restricted to Θ_0 are uniformly Lipschitz: there exists $M = M(\Theta_0) < \infty$ such that for each $z \in \mathcal{Z}$, the function $\ell_z(\cdot)$ is M-Lipschitz continuous on Θ_0.

Theorem

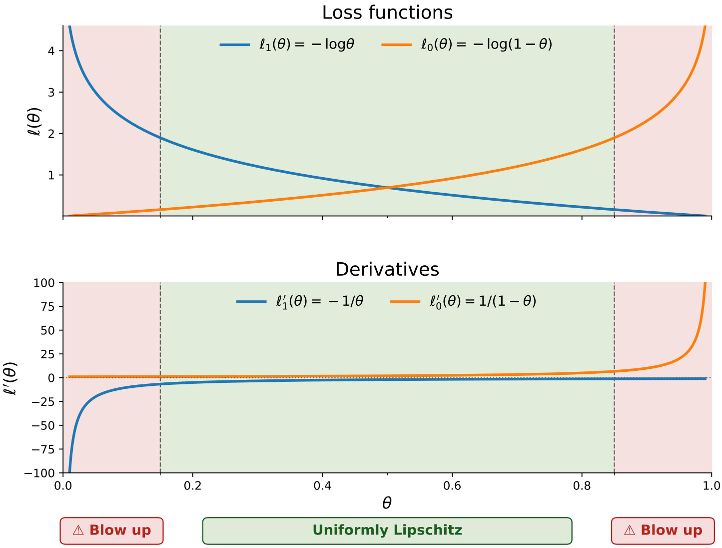
If Condition C.1 fails, then

$$\lim_n \mathfrak{M}_n(\ell, \mathcal{Z}, \Theta) > 0.$$

If Condition C.1 holds, then

$$\lim_n \mathfrak{M}_n(\ell, \mathcal{Z}, \Theta) = 0.$$

An example: Log loss with $\mathcal{Z} \in \{0, 1\}$ and $\Theta = [0, 1]$



Further results

A key takeaway:

Uniform Lipschitz continuity is **not necessary** for distribution-free learnability in convex M-estimation.

Further results

A key takeaway:

Uniform Lipschitz continuity is **not necessary** for distribution-free learnability in convex M-estimation.

Beyond the compact case

- When the domain is **unbounded**, uniform Lipschitz continuity is **not sufficient**.
- What can we say about **achievability** and algorithms that work for solvable instances?
- How do these results relate to the task of finding **stationary points** with similar distribution-free guarantees?

Further results

A key takeaway:

Uniform Lipschitz continuity is **not necessary** for distribution-free learnability in convex M-estimation.

Beyond the compact case

- When the domain is **unbounded**, uniform Lipschitz continuity is **not sufficient**.
- What can we say about **achievability** and algorithms that work for solvable instances?
- How do these results relate to the task of finding **stationary points** with similar distribution-free guarantees?

Stop by Poster **194** to chat!