Learning Parametric Distributions from Samples and Preferences

Marc Jourdan, Gizem Yüce, Nicolas Flammarion

EPFL

July 14th, 2025

Problem Statement

Fact: Using preference data outperforms methods based on positive examples only, e.g., supervised fine-tuning vs. alignment phase.

How can preferences explain these empirical performance gains?

Setting: *Estimation for parametric distributions and preferences.*

Samples: $(X_i, Y_i)_{i\in[n]} \sim p_{\theta^\star}^{\otimes[2n]}$ with $\theta^\star \in \Theta \subseteq \mathbb{R}^k$ and $\mathcal{X} \subseteq \mathbb{R}^d$.

Preference $\ell_{\theta^\star}$, e.g., $\ell_\theta(x, y) = r_\theta(x) - r_\theta(y)$ and $r_\theta = \log p_\theta$.

- ☞ **Stochastic**: $Z_i = \begin{cases} 1 & \text{with probability } 1/(1+e^{-\ell_{\theta^\star}(X_i,Y_i)}) \\ -1 & \text{otherwise} \end{cases}.$
- ☞ **Deterministic**: $Z_i = \text{sign}(\ell_{\theta^\star}(X_i, Y_i))$.

Baseline: The **sample-only maximum likelihood estimator** is

$$\widehat{\theta}_n^{\mathrm{SO}} \in \arg\min_{\theta\in\Theta} L_n^{\mathrm{SO}}(\theta) \text{ with } L_n^{\mathrm{SO}}(\theta) := -\sum_{i\in[n]} \log p_\theta^{\otimes 2}(X_i, Y_i) \,. \quad (\textbf{SO MLE})$$

Preference-based M-estimator

The **stochastic preferences MLE** is

$$\widehat{\theta}_n^{\mathrm{SP}} \in \arg\min_{\theta\in\Theta}\{L_n^{\mathrm{SO}}(\theta) + \sum_{i\in[n]} \log\left(1 + e^{-Z_i \ell_\theta(X_i, Y_i)}\right)\} . \qquad \text{(SP MLE)}$$

$\widehat{\theta}_n^{\mathrm{SP}_{\mathrm{det}}}$ defined similarly when preferences are deterministic.

Theorem (Smaller asymptotic variance)

*Under regularity and **geometric assumptions** on p_θ and ℓ_θ:*

- ☞ $\widehat{\theta}_n^{SO}$, $\widehat{\theta}_n^{SP}$ *and* $\widehat{\theta}_n^{SP_{det}}$ *are asymptotically normal estimators,*
- ☞ *with asymptotic variance* $V_{\theta^\star}^{SP_{det}} \preceq V_{\theta^\star}^{SP} \preceq V_{\theta^\star}^{SO}$.

Beyond M-estimators for deterministic preferences

Minimizers of the **empirical** 0-1 **loss** are

$$\mathcal{C}_n := \arg\min_{\theta \in \Theta} \sum_{i \in [n]} \mathbb{1}\left(Z_i \ell_\theta(X_i, Y_i) < 0\right) = \{\theta \mid \forall i \in [n], Z_i \ell_\theta(X_i, Y_i) \geq 0\} \ .$$

Any estimator $\widehat{\theta}_n^{\mathrm{AE}} \in \mathcal{C}_n$. The **deterministic preferences MLE** is

$$\widehat{\theta}_n^{\mathrm{DP}} \in \arg\min \left\{ L_n^{\mathrm{SO}}(\theta) \mid \theta \in \mathcal{C}_n \right\} \ . \qquad \text{(DP MLE)}$$

Upper bound on the estimation error

Theorem (Fast estimation rate within $\mathcal{C}_n$)

For Gaussian distributions with known Σ and $r_\theta = \log p_\theta$, for all $n \geq \widetilde{\mathcal{O}}(\log(1/\delta))$, with probability $1-\delta$,

$$\forall \widehat{\theta}_n \in \mathcal{C}_n, \quad \left\| \widehat{\theta}_n - \theta^\star \right\|_\Sigma \leq \mathcal{O}\left(\frac{A_d}{n} \log(1/\delta) \log n \right) \text{ with } A_d =_{+\infty} \mathcal{O}(\sqrt{d}) \,.$$

Theorem also holds under **geometric assumptions** on p_θ and ℓ_θ:

- ☞ **Identifiability** under preferences feedback,
- ☞ **Linearization validity** of the preferences constraints,
- ☞ **Positive and regular** p.d.f. of $\frac{\ell_{\theta^\star}(X_i, Y_i)}{-\langle u, \nabla_{\theta^\star} \ell_{\theta^\star}(X_i, Y_i)\rangle}$ near 0 for all u.

Lower bound on the estimation error

Theorem (Fast estimation rate is minimax optimal)

For Gaussian distributions with known Σ and $r_\theta = \log p_\theta$, for all n,

$$\inf_{\widehat{\theta}_n} \sup_{\theta^\star \in \Theta} \mathbb{E}_{q_{\theta^\star, h_{det}}^{\otimes [n]}} \left[\left\| \widehat{\theta}_n - \theta^\star \right\|_\Sigma \right] \geq \Omega \left(\min \left\{ \frac{A_d \sqrt{d}}{n}, \sqrt{\frac{d}{n}} \right\} \right) .$$

Theorem also holds under **geometric assumptions** on p_θ and ℓ_θ:

- ☞ **Squared Hellinger distance** is bounded by a **quadratic**,
- ☞ The **Bhattacharyya coefficient** restricted to the set of paired observations with disagreeing preference is **Lipschitz**.

Empirical validation for Gaussians and $r_\theta = \log p_\theta$

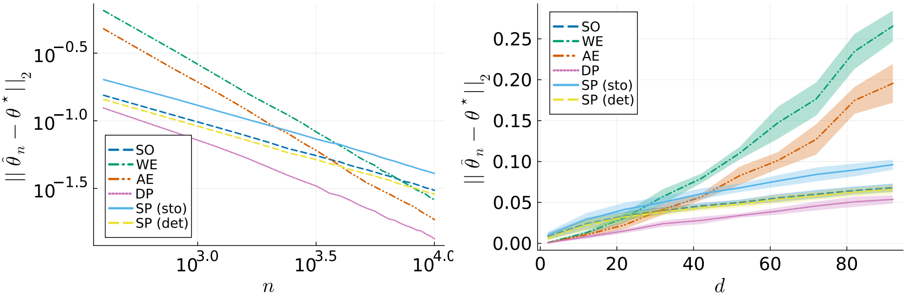

Figure: **Estimation error** with $\mathcal{N}(\theta^\star, I_d)$ where $\theta^\star \sim \mathcal{U}([1,2]^d)$, as a function of (a) the **sample size** n for $d = 20$ and (b) the **dimension** d for $n = 10^4$.

Conclusion

Benefits of additional preference feedback:

1. Preference-based M-estimators have smaller asymptotic variance than sample-only M-estimators.

2. The deterministic preference-based MLE achieves an accelerated estimation error rate of $\mathcal{O}(1/n)$, significantly improving upon the rate $\Theta(1/\sqrt{n})$ of M-estimators.

3. This matches the minimax lower bound of $\Omega(1/n)$, up to problem-specific constants.

Paper & Poster