Score-Based Density Estimation from Pairwise Comparisons

ICLR 2026

Petrus Mikkola Luigi Acerbi Arto Klami
University of Helsinki

Fechnerian Random Utility Model (RUM): Expert chooses $\mathbf{x}$ if $u(\mathbf{x}) + W > u(\mathbf{x}') + W'$, and $\mathbf{x}'$ otherwise, where u is a utility function and W, W' are i.i.d noise terms

- **Assumption**: $u(\mathbf{x}) = \log p(\mathbf{x})$ where $p(\mathbf{x})$ represents expert's belief density
- **Goal**: Learn $p(\mathbf{x})$, where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$
- **Data**: Pairwise comparisons $\{\mathbf{x}, \mathbf{x}'\}$ and corresponding judgments $\mathbf{x} \succ \mathbf{x}'$

Fechnerian Random Utility Model (RUM): Expert chooses $\mathbf{x}$ if $u(\mathbf{x}) + W > u(\mathbf{x}') + W'$, and $\mathbf{x}'$ otherwise, where u is a utility function and W, W' are i.i.d noise terms

- **Assumption**: $u(\mathbf{x}) = \log p(\mathbf{x})$ where $p(\mathbf{x})$ represents expert's belief density
- **Goal**: Learn $p(\mathbf{x})$, where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$
- **Data**: Pairwise comparisons $\{\mathbf{x}, \mathbf{x}'\}$ and corresponding judgments $\mathbf{x} \succ \mathbf{x}'$

Motivation:

- Prior elicitation: joint prior distribution from expert pairwise comparisons
- LLM elicitation: LLM's probabilistic belief from prompting pairwise comparisons
- Reward modeling: probabilistic reward model from annotator's comparisons

- Standard inference [1,2,3] is based on MLE by maximizing the choice probabilities given by RUM
- Instead, we cast the problem as a (score-based) density estimation:
  - Marginal Winner Density $p_w(\mathbf{x})$ is a density for random vector $\mathbf{x} \sim p_w$ defined by $\log p(\mathbf{x}) + W > \log p(\mathbf{x}') + W'$ for $\mathbf{x}', \mathbf{x}'' \overset{iid}{\sim} \lambda$ from a sampling density λ.
  - Score $\nabla \log p_w(\mathbf{x})$ can be estimated from winner-loser pairs
  - We can recover $\nabla \log p(\mathbf{x}) = \tau(\mathbf{x}) \nabla \log p_w(\mathbf{x})$ (our main theorem)
  - Given estimates $\nabla \log p_w(\mathbf{x})$ and $\tau(\mathbf{x})$,
    - $\mathbf{x} \sim p(\mathbf{x})$ using ALD or diffusion
    - $\mathbf{x} \mapsto p(\mathbf{x})$ using diffusion (probability ODE)

[1] Rank analysis of incomplete block designs [Bradley&Terry, 1952].

[2] Training language models to follow instructions with human feedback [Ouyang *et al.*, NeurIPS22].

[3] Preferential Normalizing Flows [Mikkola *et al.*, NeurIPS24]

We call a function $\tau(\mathbf{x})$ a **tempering field** between densities p and q if

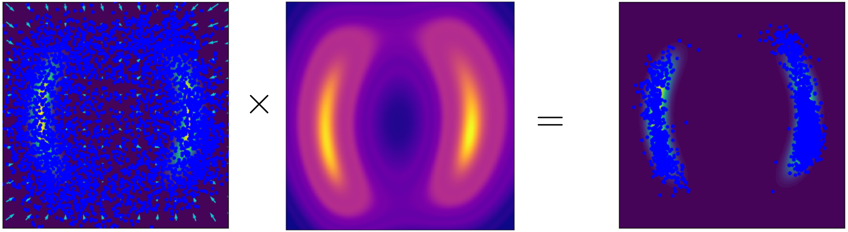
$$\nabla \log p(\mathbf{x}) = \tau(\mathbf{x}) \nabla \log q(\mathbf{x}).$$

Theorem

Assume the expert follows the Bradley-Terry model with noise level $s > 0$ and comparisons are sampled uniformly. A tempering field $\tau(\mathbf{x})$ exists between the belief density p and the MWD p_w, and it is given by the formula,

$$\tau(\mathbf{x}) = s \left(\frac{\int_{\mathcal{X}} \frac{1}{1+r_s(\mathbf{x},\mathbf{x}')} d\mathbf{x}'}{\int_{\mathcal{X}} \frac{r_s(\mathbf{x},\mathbf{x}')}{(1+r_s(\mathbf{x},\mathbf{x}'))^2} d\mathbf{x}'} \right),$$

where $r_s(\mathbf{x},\mathbf{x}') := p^{\frac{1}{s}}(\mathbf{x}') p^{-\frac{1}{s}}(\mathbf{x})$ is the $1/s$-tempered density ratio.



MWD $p_w(\mathbf{x})$ $\times$ Tempering field $\tau(\mathbf{x})$ $=$ 'Tempered' MWD $\approx p(\mathbf{x})$