

APPROXIMATING f -DIVERGENCES WITH RANK STATISTICS



joint work with



José Manuel de Frutos, University Carlos III of Madrid



Viktor Stein (TUM), ICML 2026

New loss functional on probability measures

↪ compares probability distributions in a robust way for

- hypothesis testing
- generative modeling

Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$. The CDF of μ is R_μ and its quantile is Q_μ .

Draw i.i.d. samples $y \sim \mu$, $x_1, \dots, x_K \sim \nu$.

Consider the **number of samples x_j that lie below y** ,

$$A_{\mu|\nu}^{(K)} := \#\{j \in [K] : x_j \leq y\}.$$

$$\mathbb{P}(A_{\mu|\nu}^{(K)} = \cdot \mid y) \sim \text{Bin}(K, R_\nu(y))$$

By the law of total probability, the **“rank-histogram”** is (the last equality only holds if $\mu \ll \nu$ and ν has no atoms)

$$Q_{\mu|\nu}^{(K)}(n) := \mathbb{P}(A_{\mu|\nu}^{(K)} = n) = \int_{\mathbb{R}} b_{n,K}(R_\nu(y)) d\mu(y) = \int_0^1 b_{n,K}(s) \left(\frac{d\mu}{d\nu} \circ Q_\nu \right) (s) ds, \quad \forall n \in [K].$$

The Bernstein polynomials are $b_{n,K} : [0, 1] \rightarrow [0, \infty)$, $u \mapsto \binom{K}{n} u^n (1-u)^{K-n}$, $n \in [K] := \{0, \dots, K\}$.

THEOREM (RANK STATISTICS FOR NON-PARAMETRIC HYPOTHESIS TESTING (ELVIRA ET AL. 2021))

$$\mu = \nu \iff Q_{\mu|\nu}^{(K)}(n) \equiv \frac{1}{K+1} \quad \forall K \in \mathbb{N}.$$

Core idea: quantify discrepancy between μ and ν by measuring deviation of the “rank histogram” $Q_{\mu|\nu}^{(K)}$ from the uniform distribution $\mathcal{U}[K]$ on $[K]$.

\rightsquigarrow We will measure this discrepancy using f -divergences and extend this to $\mathcal{P}(\mathbb{R}^d)$ by **slicing**.

DEFINITION (BERNSTEIN f -DIVERGENCE (S., DE FRUTOS, '26))

The Bernstein f -divergence of order $K \in \mathbb{N}$ to $\nu \in \mathcal{P}(\mathbb{R})$ is

$$D_{f,\nu}^{(K)} : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty], \quad \mu \mapsto D_f(Q_{\mu|\nu}^{(K)} | \mathcal{U}([K])) = \frac{1}{K+1} \sum_{n=0}^K f\left((K+1)Q_{\mu|\nu}^{(K)}(n)\right).$$

Using the Bernstein theorem, we prove

THEOREM (BERNSTEIN APPROXIMATION (S., DE FRUTOS, '26))

Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $\mu \ll \nu$. If $r := \frac{d\mu}{d\nu} \circ Q_\nu \in \mathcal{C}([0, 1])$ and f is L_f -Lipschitz on $\text{ran}(r)$, then

$$0 \leq D_{f,\nu}(\mu) - D_{f,\nu}^{(K)}(\mu) \begin{cases} \rightarrow 0, & r \in \mathcal{C}([0, 1]), \\ \in O(K^{-\frac{\alpha}{2}}), & r \in C^{0,\alpha}([0, 1]), \\ \in O(K^{-1}), & r \in \mathcal{C}^2([0, 1]), \text{ or } r \text{ Lipschitz, } f \in \mathcal{C}^2([0, \infty)). \end{cases} \quad \text{as } K \rightarrow \infty.$$

THEOREM (PROPERTIES OF THE BERNSTEIN f -DIVERGENCE)

Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $\mu \ll \nu$, ν be atomless, and $K \in \mathbb{N}$.

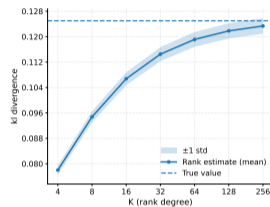
- $D_{f,\nu}^{(K)}$ is convex and weakly lower semicontinuous on $\mathcal{P}(\mathbb{R})$.
- If f is L_f -Lipschitz on $[0, K + 1]$, then

$$\mathbb{E} \left[\left| D_{f,\hat{\nu}_M}^{(K)}(\hat{\mu}_N) - D_{f,\nu}^{(K)}(\mu) \right| \right] \leq L_f(K + 1) \sqrt{2\pi} \left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{M}} \right).$$

and for any $\delta > 0$, with probability at least $1 - \delta$, we have

$$\left| D_{f,\hat{\nu}_M}^{(K)}(\hat{\mu}_N) - \mathbb{E} \left[D_{f,\hat{\nu}_M}^{(K)}(\hat{\mu}_N) \right] \right| \leq CL_f(K + 1) \sqrt{\log(1/\delta) \left(\frac{1}{N} + \frac{1}{M} \right)}.$$

- Asymptotic normality result, $K_N = N^\beta$ with $1/2 < \beta < 1$.

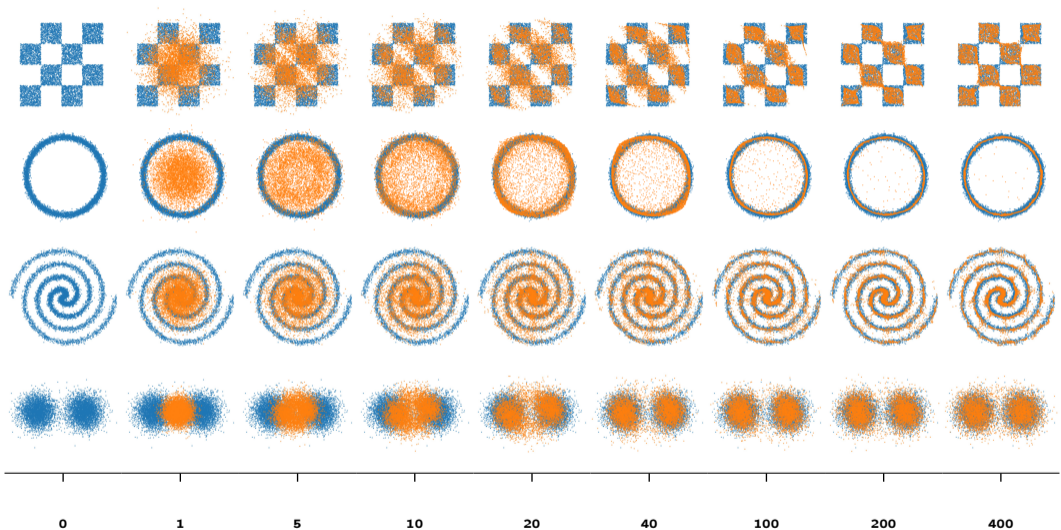


$D_{f_{\text{KL}}}^{(K)}(\mathcal{N}(0, 1) \mid \mathcal{N}(1, 1))$ for different K .

We prove

$$D_{f,\nu}^{(K)}(\mu) \leq D_{f,\nu}^{(K+1)}(\mu).$$

NUMERICAL EXAMPLES - TOY DATA SETS



Rank-Proximal Transport on 2D toy targets. Using SGD to minimize the rank-statistic KL with $L = 10$ random projections

Bernstein f -divergences are

- **efficient and interpretable estimators** of (sliced) f -divergences.
- **easy to evaluate** using (HMC) slicing and rank comparisons



European Research Council

Established by the European Commission

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This project has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement No. 101198055, project acronym NEITALG).

Thank you for your attention!