Tsallis Regularized Optimal Transport and Ecological Inference

Boris Muzellec, Richard Nock, Giorgio Patrini and Frank Nielsen

École polytechnique and Data61

Overview

A new framework for optimal transport which

- Unifies Monge-Kantorovich and Sinkhorn-Cuturi
- Interpolates a broad class of classic distortions
- Has metric properties
- Can be efficiently solved for algorithmically
- Can be successfully applied to ecological inference

Optimal Transport

A powerful framework for computing distances between probability distributions. Two approaches:

- **Monge-Kantorovich** (unregularized):

$$d_M(r,c) = \min_{P \in U(r,c)} \langle P, M \rangle \qquad \text{(OT)}$$

  - Problem: $O(n^3 \ln(n))$ complexity...
- **Sinkhorn-Cuturi** (entropic regularization):

$$d_M^{\lambda}(r,c) = \min_{P \in U(r,c)} \langle P, M \rangle - \frac{1}{\lambda} h(P) \qquad \text{(SC)}$$

  - Fast convergence with Sinkhorn's algorithm!

where $r, c \in \Sigma_n$, and $U(r,c) = \{P \in \mathbb{R}^{n \times n} : P\mathbb{1} = r, P^\top \mathbb{1} = c\}$ is the *transportation polytope*

Tsallis Entropy

- Generalization of Shannon's entropy:

$$H_q(P) := \sum_i p_i \ln_q\left(\frac{1}{p_i}\right)$$

$$K_q(P,S) := -\sum_i p_i \ln_q\left(\frac{s_i}{p_i}\right)$$

  - Generalized logarithm $\ln_q(u) := \frac{1}{1-q}(u^{1-q} - 1) \underset{q \to 1}{\to} \ln(q)$

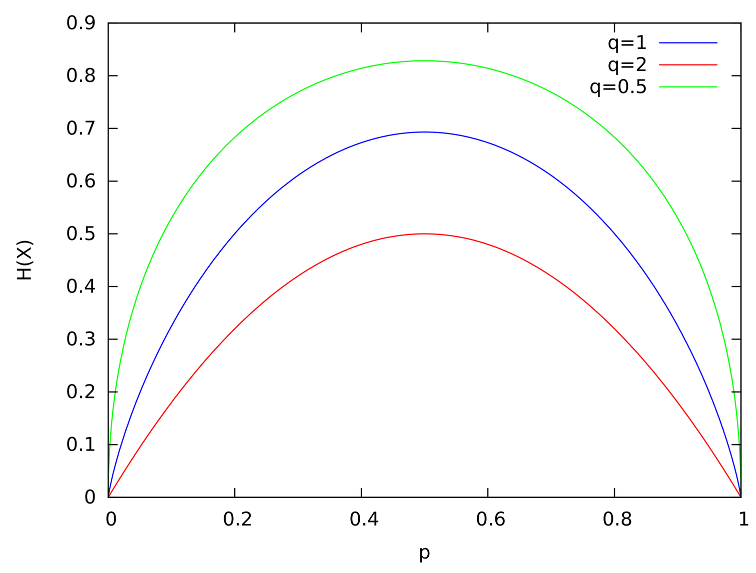

Figure: Tsallis entropy of a Binomial random variable for various values of q

- Tsallis Relative Entropy spans classic divergences:
  - **Kullback-Leibler divergence** ($q \to 1$): $\mathrm{KL}(P\|S) = \sum_i p_i \ln(\frac{p_i}{s_i})$
  - **Pearson's χ^2** ($q = 2$): $\sum_i \frac{(p_i - s_i)^2}{s_i}$
  - **Neyman's χ^2** ($q = 2$): $\sum_i \frac{(s_i - q_i)^2}{s_i}$
  - **Hellinger distance** ($q = 1/2$): $\sum_i (\sqrt{p_i} - \sqrt{s_i})^2$

TROT

Add a Tsallis regularizer to OT:

$$d_M^{\lambda}(r,c) = \min_{P \in U(r,c)} \langle P, M \rangle - \frac{1}{\lambda} H_q(P) \qquad \text{(TROT)}$$

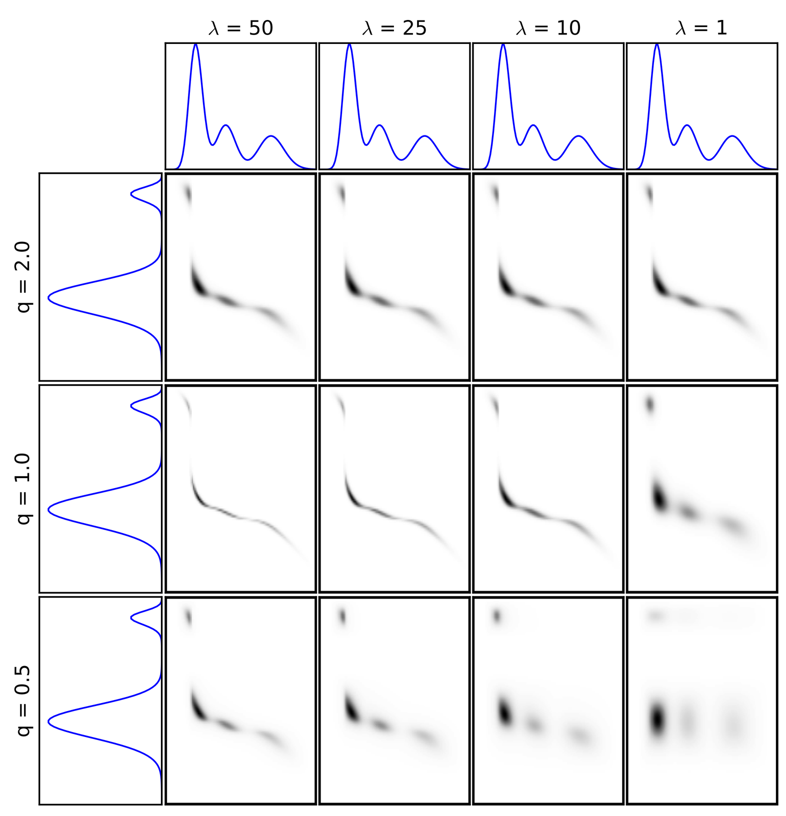

Figure: Examples of optimal transportation plans for varying q, λ

TROT Properties

- **Has a unique solution**
- **Interpolates** Monge-Kantorovich ($q \to 0$) and Sinkhorn-Cuturi ($q \to 1$)
- **Unrecoverability**: $(q, \lambda) \neq (q', \lambda') \Rightarrow \mathrm{TROT}(q, \lambda) \neq \mathrm{TROT}(q', \lambda')$
- **Efficient solving** by approximate matrix balancing
- **Metric properties** ($q \geq 1$): e.g. triangle inequality

Ecological Inference

- **Def**: Reconstruction of *joint distributions* from *marginal distributions*
  - e.g. find the vote share within social groups
- **Ill-posed problem**: zero or infinite number of solutions
- **Usual approach**: use an information theoretic criterion
- **TROT**: information theory + encode *a priori* information in the cost matrix M

Ecological Inference with TROT

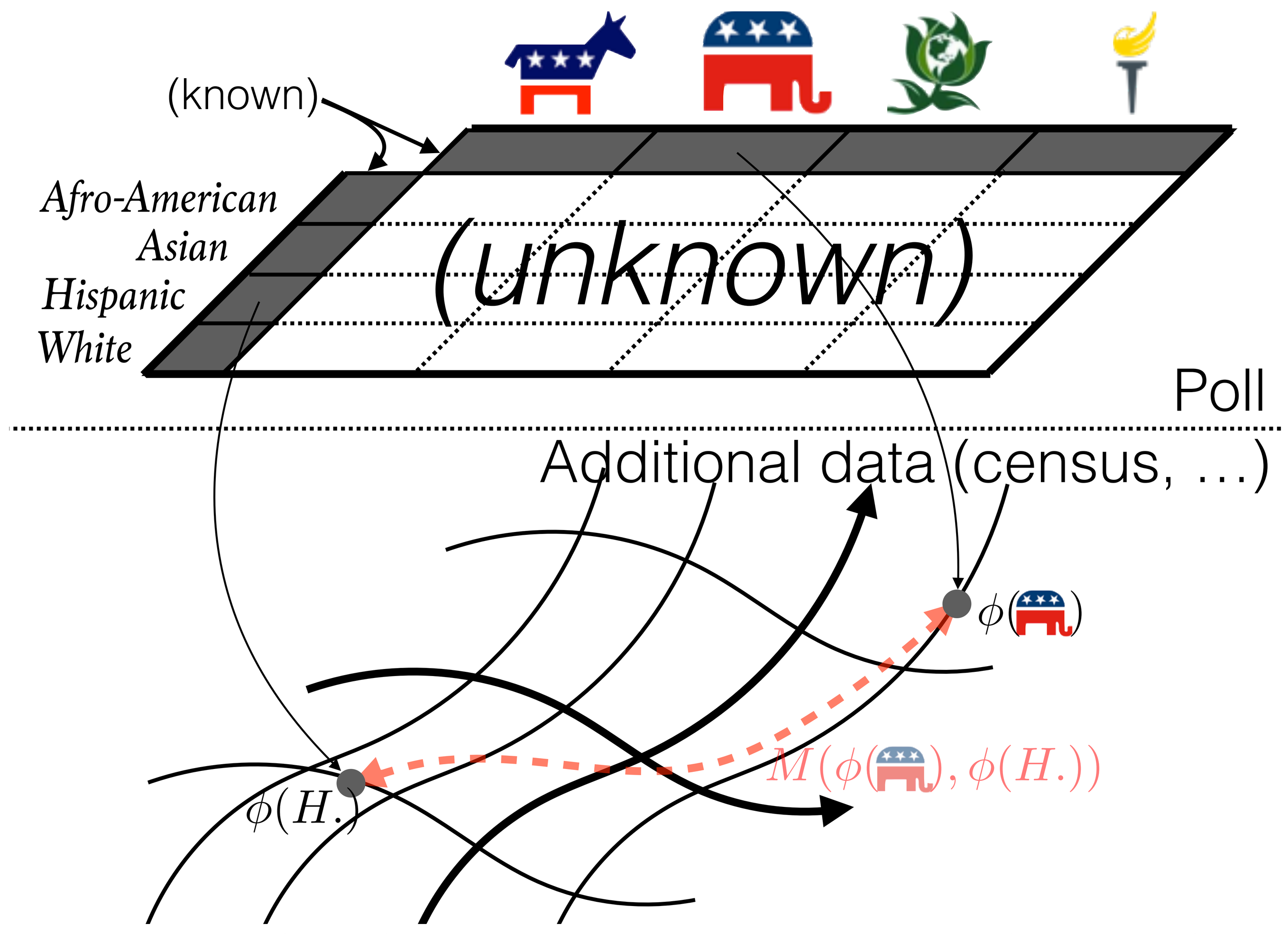

Figure: Top: suppose we know (in grey) marginals for the US presidential election (topmost row) and ethnic breakdowns in the US population (leftmost column). Can we recover an estimated joint distribution (white cells)? If side information is available such as individual level census data (bottom, as depicted on a Hilbert manifold with ϕ-coordinates), then distances can be computed within the supports (dashed red), and optimal transport can provide an estimation of the joint distribution.

Experiments

Experiments on 2012 Florida US vote data. Two approaches:

- Demographic information only: $m_{ij}^{\mathrm{RBF}} = \sqrt{2 - 2\exp(-\gamma \cdot \|\mu_i^{\mathrm{p}} - \mu_j^{\mathrm{e}})\|_2)}$
- Survey statistics: $m_{ij}^{\mathrm{sur}} = 1 - p_{ij}^{\mathrm{sur}}$

Algorithm	Cost matrix	q	λ	KL-divergence ± SD	Abs. error ± SD
Florida-Average	-	-	-	0.251 ± 0.187	0.025 ± 0.011
Simplex	M^{RBF}	-	-	0.280 ± 0.108	0.023 ± 0.008
Simplex	survey	-	-	0.136 ± 0.098	0.013 ± 0.009
Sinkhorn	M^{RBF}	$1.0^\dagger$	10^0	0.054 ± 0.036	0.009 ± 0.005
Sinkhorn	survey	$1.0^\dagger$	10^1	0.035 ± 0.027	0.007 ± 0.004
TROT	M^{RBF}	1.0	10^0	0.054 ± 0.036	0.009 ± 0.005
TROT	survey	2.5	10^1	0.007 ± 0.009	0.003 ± 0.002
TROT	$\mathbb{1}\mathbb{1}^\top$	0.8	10^0	0.076 ± 0.048	0.011 ± 0.005

Table: Average KL-divergence and absolute error with standard deviation (SD) of algorithms inferring joint distributions of all Florida counties. Parameters noted with † are not cross-validated but defined by the algorithm.

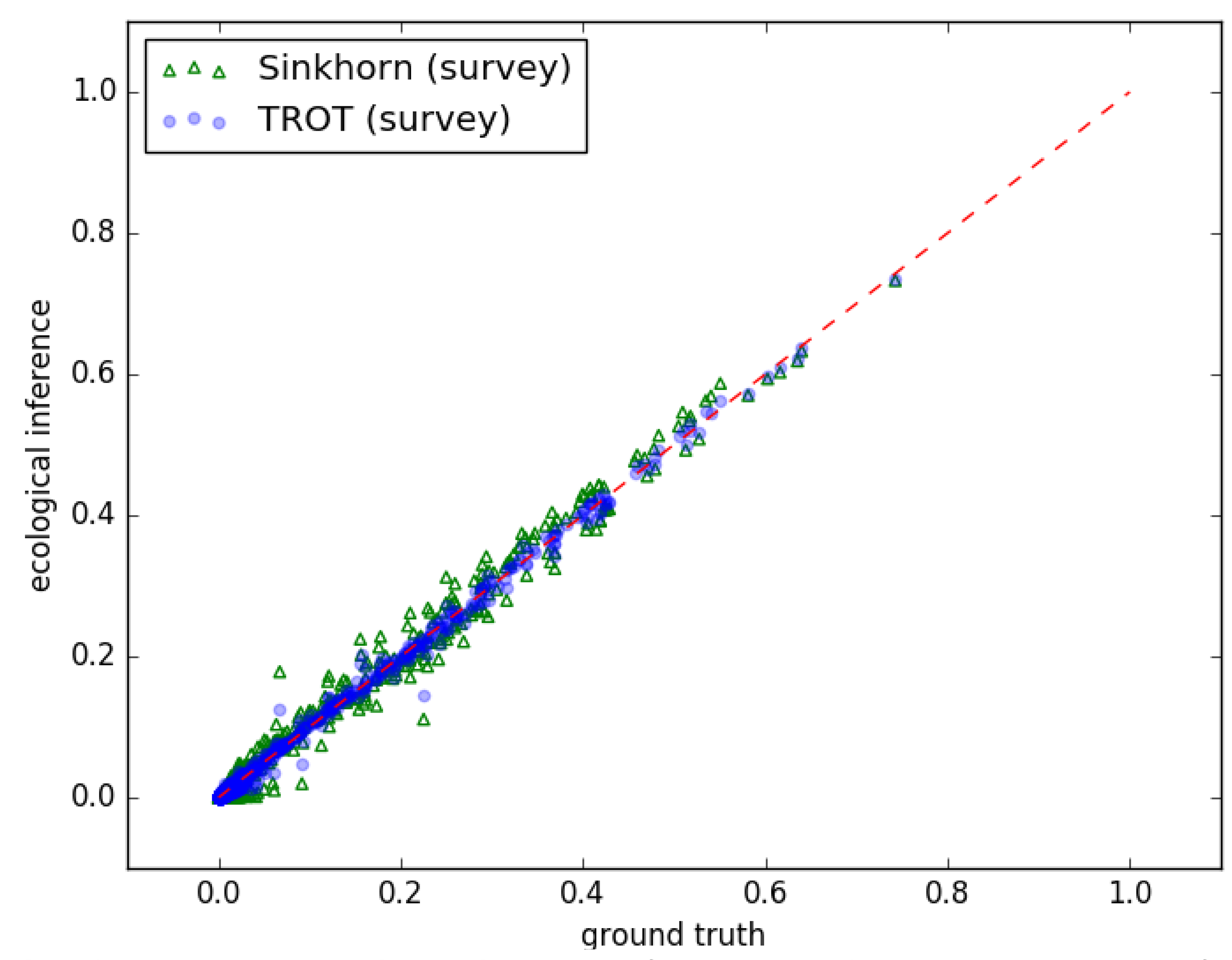

Figure: Correlation between TROT vs Sinkhorn inferred probabilities and ground truth for all Florida counties (the closer to y = x, the better).

Acknowledgements

We wish to thank Seth Flaxman and Wendy K. Tam Cho for numerous stimulating discussions. Work done while Boris Muzellec was visiting Nicta / Data61. Nicta was funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Center of Excellence Program.

References

[1] Boris Muzellec, Richard Nock, Giorgio Patrini, and Frank Nielsen.
Tsallis regularized optimal transport and ecological inference.
arXiv preprint arXiv:1609.04495, 2016.

Code repository: `https://github.com/BorisMuzellec/TROT`