Statistical Bounds for Entropic Optimal Transport: Sample Complexity and the Central Limit Theorem

Gonzalo Mena*,Jonathan Niles-Weed**

*Statistics Department and Data Science Initiative, Harvard University, **Courant Institute and Center for Data Science, New York University

Introduction

Optimal transport (OT) has become a popular analysis tool for large datasets in high dimension, and **entropic regularization** has shown to provide computationally efficient approximations (Cuturi, 2013).

However, it also appears to have useful **statistical** properties. For instance Genevay et al. (2019) established that even though standard OT suffers from the **curse of dimensionality**, entropic OT always converges at the **parametric** $1/\sqrt{n}$ for compactly supported probability measures.

Definitions

Let $P, Q \in \mathcal{P}(\mathbb{R}^d)$ be probability measures and let P_n, Q_n be their empirical versions. Their squared Wasserstein distance is

$$W_2^2(P,Q) := \inf_{\pi \in \Pi(P,Q)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|x-y\|^2 \, d\pi(x,y) \right]$$

where $\Pi(P,Q)$ is the set of joints with marginals P and Q.We focus on an entropic version ($I(\pi)$ is the mutual information):

$$S(P,Q) := \inf_{\pi \in \Pi(P,Q)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|x-y\|^2 \, d\pi(x,y) + \epsilon I(\pi) \right]$$

Finally, a distribution $P \in \mathcal{P}(\mathbb{R}^d)$ is σ^2-subgaussian if $E_P e^{\frac{\|X\|^2}{2d\sigma^2}} \leq 2$

Main Results

- **Theorem 1:** New **sample complexity** bounds, extending the results of Genevay et al. (2019) to the subgaussian case.
- **Theorem 2. Central Limit Theorem** for the fluctuations of the empirical version of entropic optimal transport around its expected value, extending the results of Del Barrio and Loubes (2019) and Bigot et al. (2018).
- **Theorem 3.** As an application, we show how **entropic OT** can be used to estimate the entropy of random variables corrupted by subgaussian noise.

Sample Complexity

Genevay et al, 2018 for measures defined on a bounded domain with diameter D

$$E|S(P,Q) - S(P_n,Q_n)| \leq \frac{K_{d,D}}{\sqrt{n}} e^{D^2/\epsilon}$$

Our result: for σ^2-subgaussian measures

$$E|S(P,Q) - S(P_n,Q_n)| \leq \frac{K_d}{\sqrt{n}} \left(1 + \frac{\sigma^{O(d)}}{\epsilon^{O(d)}}\right)$$

Remark: $W_2(P,Q)$ is cursed by dimensionality (Dudley, 1969)

$$E_{P,Q}|W_2(P,Q) - W_2(P_n,Q_n)| \leq O(n^{-1/d})$$

References:

1. Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. NeurIPS
2. Csiszar, I. (1975), I-divergence geometry of probability distributions and minimization problems. The Annals of Probability
3. Del Barrio, E. and Loubes, J.M (2019). Central limit theorems for empirical transportation cost in general dimension. The Annals of Probability.
4. Bigot, J., Cazelles, E. and Papadakis, N. (2018). Central limit theorems for Sinkhorn divergence between probability distributions on finite spaces and statistical applications. arXiv.
5. Dudley, R.M. (1969). The Speed of Mean Glivenko-Cantelli Convergence. The Annals of Mathematical Statistics.
6. Genevay, A., Chizat, L., Bach, F., Cuturi, M., and Peyré, G. (2019). Sample Complexity of Sinkhorn divergences. AISTATS.
7. Goldfeld, Z., Greenewald, K., Polyanskiy, Y. and Weed, J., (2019). Convergence of Smoothed Empirical Measures with Applications to Entropy Estimation. arXiv.
8. Berrett, Thomas B., Richard J. Samworth, and Ming Yuan. (2019) Efficient multivariate entropy estimation via k-nearest neighbour distances. *The Annals of Statistics*

Central Limit Theorem

Consider independent samples P_n, Q_n from subgaussian P, Q. Then

$$\sqrt{\frac{mn}{m+n}} \left(S(P_n,Q_m) - E(S(P_n,Q_m)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2_{P,Q}\right)$$

- $\sigma^2_{P,Q}$ can be computed explicitly.
- One sample versions are also available.
- Results and technique extend from Del Barrio and Loubes (2019)

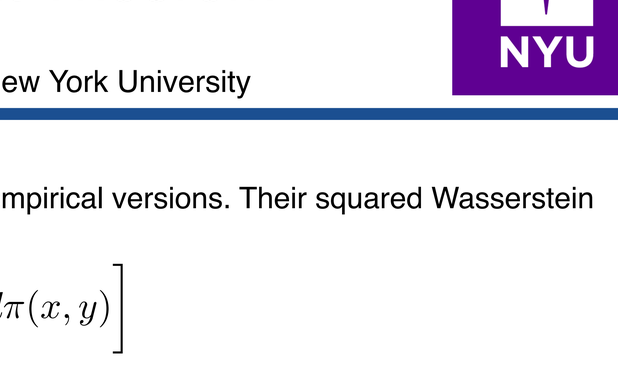

Fig 1. Shades are CLT std predictions. Bars are sample s

Entropy Estimation

Recent work (Goldfeld et al. 2019, Berrett et al., 2019) shows that the differential entropy $h(P * \mathcal{N}(0,I))$ of random variables corrupted by gaussian noise can be estimated at the rate $1/\sqrt{n}$

Our approach

1. Prove that $h\left(P * \mathcal{N}(0,I)\right) = S(P, P * \mathcal{N}(0,I)) + \frac{d}{2}\log(2\pi)$
2. Estimate $\hat{h}_1(P * \mathcal{N}(0,I)) := S(P_n, (P * \mathcal{N}(0,I)_n) + \frac{d}{2}\log(2\pi)$

$\hat{h}_1$ better than $\hat{h}_2$, the best available (from Goldfeld et al., 2019) and has distributional limits

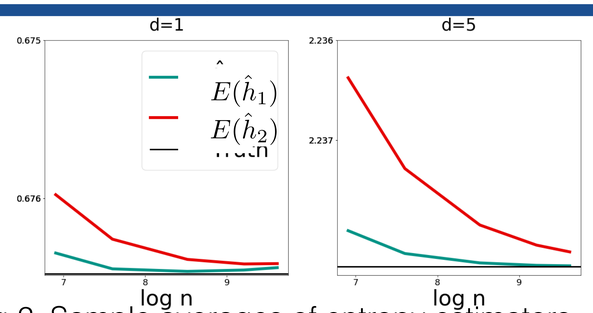

Fig 2. Sample averages of entropy estimators

gomena@fas.harvard.edu, jnw@cims.nyu.edu

This work was supported by the Harvard Data Science Initiative (GM) and the Josephine de Kármán (JNW) fellowships