

Entropic Optimal Transport on Countable Spaces: Statistical Theory and Asymptotics [†]

Shayan Hundrieser ¹, Marcel Klatt ¹ and Axel Munk ^{1,2}

¹ Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstrasse 7, Göttingen, Germany

² Max Planck Fellow at Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, Göttingen, Germany

[†] Presented at the Entropy 2021: The Scientific Tool of the 21st Century, 5–7 May 2021; Available online: <https://sciforum.net/conference/Entropy2021/>.

Published: 5 May 2021

In recent years, the theory of OT has become a fundamental tool in statistical research, e.g., the Wasserstein distance being a prominent method for inferential purposes. The large computational complexity, however, has hindered OT in becoming a routine methodology for the analysis of large scale data sets. This has encouraged the formulation of regularized OT which often turns out to be computationally more accessible. The most prominent proposal is given by entropic regularization (c.f. [1]) that serves to define an entropic OT distance (EOTD) and Sinkhorn divergence.

In the present study, we derive limit distributions for empirical EOTD (i.e., when data are sampled randomly) between probability measures supported on countable discrete spaces. In particular, we consider a general class of cost functions and state conditions on the probability measures to ensure general weak convergence for empirical EOTD. Furthermore, for bounded cost functions we show that the empirical entropic transport plan itself converges weakly in ℓ^1 -sense to a Gaussian process. The theory generalizes results derived by [2] for finite discrete spaces. Moreover, they complement recent findings by [3] for the empirical EOTD between more general probability measures on \mathbb{R}^m with quadratic cost. Our approach is based on a sensitivity analysis of necessary and sufficient optimality conditions for the entropic transport plan. We demonstrate possible application for colocalization analysis of protein networks in biology.

This is joint work with Marcel Klatt and Axel Munk.

References

1. Cuturi, M. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 2292–2300.
2. Klatt, M.; Tameling, C.; Munk, A. Empirical Regularized Optimal Transport: Statistical Theory and Applications. *arXiv* **2018**, arXiv:1810.09880.
3. Mena, G.; Weed, J. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *arXiv* **2019**, arXiv:1905.11882.



© 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).