Sample Complexity of Sinkhorn Divergences

Computational Optimal Transport

Simon Queric

Télécom Paris

January 18, 2024

Table of content

Introduction

- **Notations :** samples $X_{1:n}, Y_{1:n}$ drawn from proba measures α, β, empirical measures $\alpha_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}, \beta_n = n^{-1} \sum_{j=1}^{n} \delta_{Y_j}$
- OT suffers form curse of dimensionality : $\mathbb{E}|W(\alpha, \beta) - W(\alpha_n, \beta_n)| = O(n^{-1/d})$ (Dudley, '84)
- MMD doesn't : $\mathbb{E}|\mathrm{MMD}(\alpha, \beta) - \mathrm{MMD}(\alpha_n, \beta_n)| = O(n^{-1/2})$
- Practitioners use entropic regularized distance W_ε
- **How sample complexity of Sinkhorn Divergences behave ?**

Table of content

Theoretical bounds on sample complexity

Sample complexity of sinkhorn divergences (A. Genevay)

$$\mathbb{E}\left[|W_\varepsilon(\alpha,\beta) - W_\varepsilon(\alpha_n,\beta_n)|\right] = O\left(\frac{e^{\kappa/\varepsilon}}{\sqrt{n}}\left(1 + \frac{1}{\varepsilon^{d/2}}\right)\right)$$

Limit cases

1. $\varepsilon \to 0$ then $\mathbb{E}\left[|W_\varepsilon(\alpha,\beta) - W_\varepsilon(\alpha_n,\beta_n)|\right] = O\left(\frac{e^{\kappa/\varepsilon}}{\varepsilon^{d/2}\sqrt{n}}\right)$
2. $\varepsilon \to +\infty$ then $\mathbb{E}\left[|W_\varepsilon(\alpha,\beta) - W_\varepsilon(\alpha_n,\beta_n)|\right] = O(1/\sqrt{n})$

Table of content

Implementation

- Work with W_ε and $\overline{W}_\varepsilon$
- Gaussian, uniform, beta measures
- Sinkhorn in log-domain
- $\sup_{u\mathcal{C}(\mathcal{X}), v\in\mathcal{C}(\mathcal{Y})} \int u(x)\alpha(dx) + \int v(y)\beta(dy)$

- $\varepsilon \int_{\mathcal{X}\times\mathcal{Y}} e^{u(x)+v(y)-c(x,y)}\alpha(dx)\beta(dy) + \varepsilon$

- $\widehat{W}_\varepsilon(\alpha_n, \beta_n) = \frac{1}{n}\sum_{i=1}^{n} u(X_i) + \frac{1}{n}\sum_{j=1}^{n} v(Y_j) + 0$

Results on gaussian measures

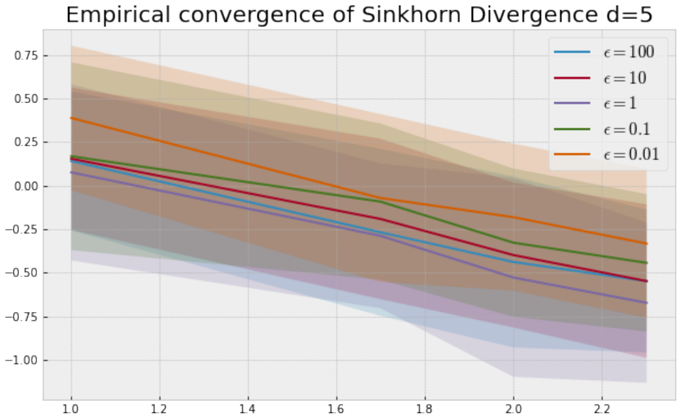

Figure: $|W_\varepsilon(\alpha_n, \beta_n) - W(\alpha, \beta)|$ as a function of n in the log-log space. Experiment for gaussian measures in dimension 5 with $c(x, y) = \|x - y\|_2^2$.

Results on gaussian measures

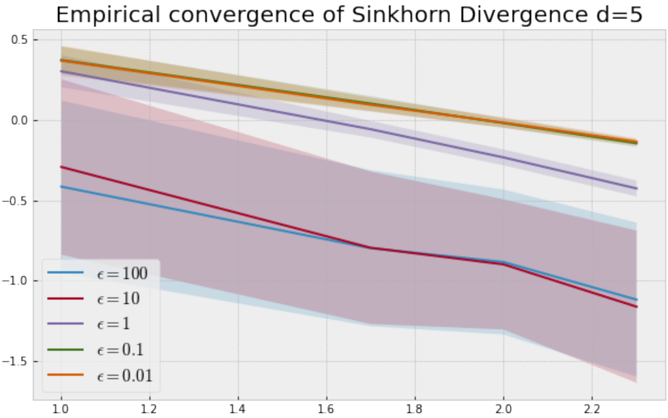

Figure: $\overline{W}_{\varepsilon}(\alpha_n, \alpha'_n)$ as a function of n in the log-log space. Experiment for standard normal distribution in dimension 5 with $c(x, y) = \|x - y\|_2^2$.

Results on uniform distributions

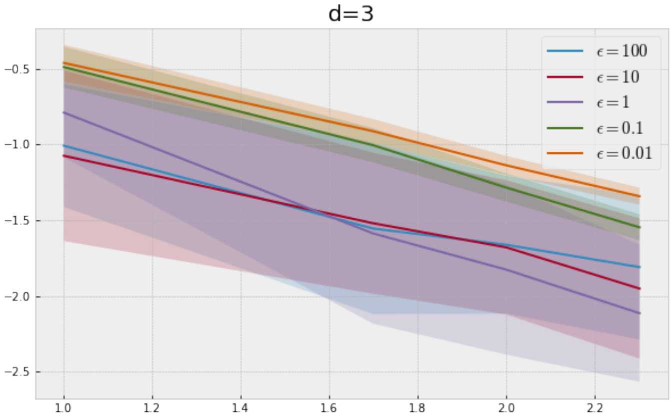

Figure: $\overline{W}_{\varepsilon}(\alpha_n, \alpha'_n)$ as a function of n in the log-log space. Experiment for uniform measures in $[-1, 1]^3$ with $c(x, y) = \|x - y\|_2^2$.

Results on uniform distributions

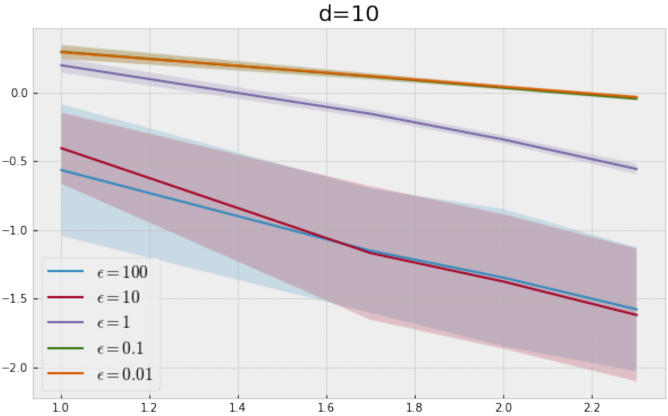

Figure: $\overline{W}_\varepsilon(\alpha_n, \beta_n)$ as a function of n in the log-log space. Experiment for uniform measures in $[-1, 1]^{10}$ with $c(x, y) = \|x - y\|_2^2$.

Results on beta distribution

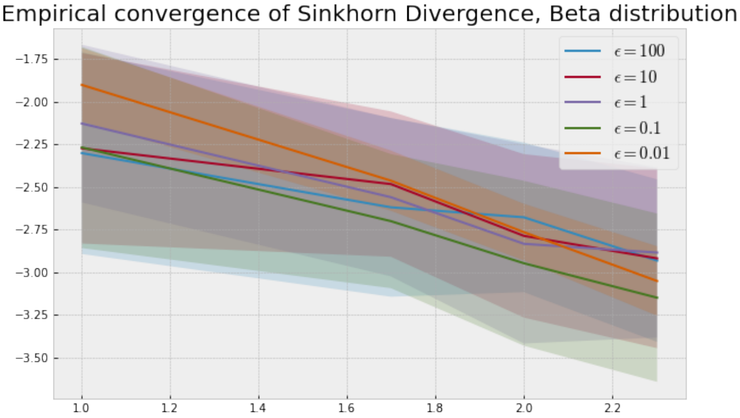

Figure: $\overline{W}_\varepsilon(\alpha_n, \alpha'_n)$ as a function of n in the log-log space. Experiment for Beta$(2, 5)$ in dimension $d = 2$, with $c(x, y) = \|x - y\|_2^2$.

Table of content

Critics

- Strengths
  - Interpretation of entropic OT as an interpolation between OT and MMD
  - Numerical results corroborate theory
- Weaknesses
  - Lack of closed forms to improve theoretical results, get sharper bounds, numerical guarentees

Conclusion & Perspective

- Influence of ε is not clear for two different gaussian measures.
- Influence of dimension, curse of dimensionality.
- Understand approximation error between regularized OT and OT.