

# Influence de l'Echantillonnage sur la Distance de Wasserstein Empirique

June 21, 2023

## Semaine 7 du 21 Juin au 28 Juin

On rappelle que la sample complexity de l'estimateur plug-in de la distance de Wasserstein dépend de la dimension [CRL<sup>+</sup>20]:

$$|\mathbb{E}W_2^2(\mu_n, \nu_n) - W_2^2(\mu, \nu)| \asymp n^{-2/d} \quad d > 4 \quad (1)$$

avec  $\mu_n = \frac{1}{n} \sum \delta_{x_i}$  (resp.  $\nu_n = \frac{1}{n} \sum \delta_{y_i}$ ),  $x_i \sim \mu$  i.i.d. (resp.  $y_i \sim \nu$  i.i.d.).

Une question est de savoir si l'on peut augmenter artificiellement la taille des distributions empiriques (i.e. augmenter le nombre d'atomes), pour aider la convergence des mesures empiriques.

Pour cela on peut rajouter un petit bruit artificiel autour de chaque atome. Pour  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ , on peut définir son augmentation:

$$\bar{\mu}_{nm} = \sum_{i=1}^n \sum_{j=1}^m \delta_{x_{ij}} \quad (2)$$

avec les  $x_{i1}, \dots, x_{im} \sim \mathcal{N}(x_i, \epsilon)$ , i.i.d.

Une fois les distributions augmentées, on s'intéresse à leur proximité avec la distance de Wasserstein:

$$|\mathbb{E}W_2^2(\bar{\mu}_{nm}, \bar{\nu}_{nm}) - W_2^2(\mu, \nu)| \quad (3)$$

En particulier est-elle meilleur la version non augmenté?

### Faire quelques expérimentations avec les Notebook

- Compléter le notebook pour avoir la distance entre la vraie distance de Wasserstein et la version augmenté
- (Optionel) Tester la différence entre la vraie distance de Wasserstein et celle pour des distributions augmentées, selon la façon de sampler de la semaine dernière

**Petit résumé de maximum 1 page.** Les points clés sont:

- Observation de l'effet du sampling sur l'estimation de  $W_2^2(\mu, \nu)$

## References

- [CRL<sup>+</sup>20] Lenaïc Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré. Faster wasserstein distance estimation with the sinkhorn divergence. *Advances in Neural Information Processing Systems*, 33:2257–2269, 2020.