Asymptotic Guarantees for Learning Generative Models with the Sliced-Wasserstein Distance

Kimia Nadjahi[1] Alain Durmus[2] Umut Şimşekli[1,3] Roland Badeau[1]

[1] Télécom Paris [2] ENS Paris-Saclay [3] University of Oxford

Minimum Distance Estimation

$$\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \mathbf{D}(\hat{\mu}_n, \mu_\theta)$$

$\mathbf{D}$: distance between distributions

$\hat{\mu}_n$: empirical distribution of **data points** $Y_1, \dots, Y_n$ i.i.d from $\mu_\star$

μ_θ: distribution parametrized by $\theta \in \Theta$

Minimum Distance Estimation

$$\hat{\theta}_n = \mathrm{argmin}_{\theta \in \Theta} \mathbf{D}(\hat{\mu}_n, \mu_\theta)$$

$\mathbf{D}$: distance between distributions

$\hat{\mu}_n$: empirical distribution of **data points** $Y_1, \dots, Y_n$ i.i.d from $\mu_\star$

μ_θ: distribution parametrized by $\theta \in \Theta$

Example: **Generative Modeling**

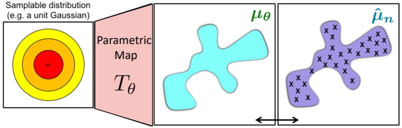


Figure courtesy of Soheil Kolouri

Minimum Expected Distance Estimation

Directly optimizing $\boldsymbol{\mu_\theta}$ is often **not possible** (e.g. GANs)

$$\hat{\theta}_{n,m} = \text{argmin}_{\theta \in \Theta} \, \mathbb{E}\left[\mathbf{D}(\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\mu}}_{\boldsymbol{\theta},\boldsymbol{m}}) \mid Y_{1:n}\right]$$

$\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta},\boldsymbol{m}}$: empirical distribution of a sample $Z_1, \dots, Z_m$ i.i.d. from $\boldsymbol{\mu_\theta}$

Minimum Wasserstein Estimation

Choose $\mathbf{D} = \mathbf{W}_p$ (Wasserstein distance of order $p \geq 1$)

- ✓ Robust and increasingly popular estimators: Wasserstein GAN [1], Wasserstein auto-encoders [2]
- ✓ Asymptotic guarantees [3]

[1] Arjovsky et al., 2017 [2] Tolstikhin et al., 2018 [3] Bernton et al., 2019

Minimum Wasserstein Estimation

Choose $\mathbf{D} = \mathbf{W}_p$ (Wasserstein distance of order $p \geq 1$)

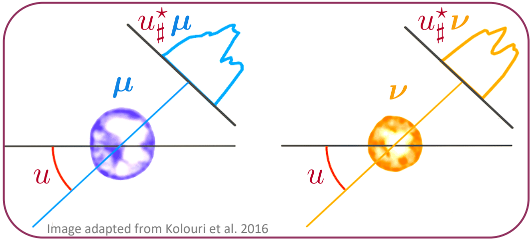
- ✓ Robust and increasingly popular estimators: Wasserstein GAN [1], Wasserstein auto-encoders [2]
- ✓ Asymptotic guarantees [3]

[1] Arjovsky et al., 2017 [2] Tolstikhin et al., 2018 [3] Bernton et al., 2019

- × $\mathbf{W}_p$: expensive + curse of dimensionality
- × Central limit theorem in [3] valid in 1D

In 1D, $\mathbf{W}_p$ has an analytical form $\Rightarrow$ Motivates a practical alternative:

$$\mathbf{SW}_p^p(\mu, \nu) = \int_{\mathbb{S}^{d-1}} \mathbf{W}_p^p(u_\sharp^\star \mu, u_\sharp^\star \nu) \mathrm{d}\boldsymbol{\sigma}(u)$$



Minimum Sliced-Wasserstein Estimation

$$\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \, \mathbf{SW}_p(\hat{\mu}_n, \mu_\theta)$$

$$\hat{\theta}_{n,m} = \text{argmin}_{\theta \in \Theta} \, \mathbb{E}\left[\mathbf{SW}_p(\hat{\mu}_n, \hat{\mu}_{\theta,m}) \mid Y_{1:n}\right]$$

Successful in generative modeling applications (e.g., SW-GAN, Deshpande et al., 2018)

Minimum Sliced-Wasserstein Estimation

$$\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \mathbf{SW}_p(\hat{\mu}_n, \mu_\theta)$$

$$\hat{\theta}_{n,m} = \text{argmin}_{\theta \in \Theta} \mathbb{E}\left[\mathbf{SW}_p(\hat{\mu}_n, \hat{\mu}_{\theta,m}) \mid Y_{1:n}\right]$$

Successful in generative modeling applications (e.g., SW-GAN, Deshpande et al., 2018)

Our contributions:

- Convergence in $\mathbf{SW}_p \Rightarrow$ weak convergence of probability measures
- Existence and consistency of $\hat{\theta}_n$, $\hat{\theta}_{n,m}$
- Central limit theorem for $\hat{\theta}_n$: $\sqrt{n}$ convergence rate for any dimension

Thank you!

Our Poster: East Exhibition Hall B + C #226