Existence and consistency of Wasserstein barycenters

Thibaut Le Gouic, Jean-Michel Loubes
Aix Marseille Université, Université de Toulouse

2021. 2. 26

Outlines

Barycenter: Definition

Geodesic space：

- (E, d) is a complete metric space
- Every two points $x, y \in E$ have a mid-point $z \in E$. (Mid-point: Given two points x, y in a metric space (E, d), their mid-point is any point $z \in E$ such that $d(x, z) = d(z, y) = 0.5 * d(x, y)$.)

Barycenter:

- Set $p \geq 1$ and let (E, d) be a geodesic space and μ a probability measure on (E, d) such that

$$\int d^p(x, x_0)\, d\mu(x) < \infty \tag{1}$$

  for some(and thus any) $x_0 \in E$.

- A point $x_0 \in E$ is called a p-barycenter of μ if

$$\int d^p(x, x_0)\, d\mu(x) = \inf\left\{\int d^p(x, y) d\mu(x); y \in E\right\} \tag{2}$$

- The set of all probability measures satisfying (1) is denoted $\mathcal{W}_p(E)$.

Barycenter: Properties

- Barycenters do not always exist.
- Hopf–Rinow–Cohn–Vossen theorem states that, on locally compact geodesic spaces, every closed ball is compact. Consequently, the infimum in (2) can be taken on a compact ball, and thus existence of a barycenter is ensured.

定理 1.1

Set $p \geq 1$ and let (E, d) be a locally compact geodesic space and $\mu \in \mathcal{W}_p(E)$ Then, there exists a barycenter of μ.

Wasserstein space: Definition

- Set $p \geq 1$ and let (E, d) be a metric space.Given two measures μ, v in $\mathcal{W}_p(E)$, we denote by $\Gamma(\mu, v)$ the set of all probability measures π over the product set EE with first, resp. second, marginal , resp. .
- The transportation cost with cost function d^p between two measures μ, v in $\mathcal{W}_p(E)$, is defined as $\mathcal{T}_p(\mu, v) = \inf_{\pi \in \Gamma(\mu, v)} \int d^p(x, y) d\pi$.
- The transportation cost allows to endow the set $\mathcal{W}_p(E)$ with a metric W_p defined by $W_p(\mu, v) = \mathcal{T}_p(\mu, v)^{1/p}$.

This metric is known as the p-Wasserstein distance and the metric space $(\mathcal{W}_p(E), W_p)$ is called the Wassertein space of (E, d).

Wasserstein space: Properties

- NPC spaces: A complete metric space (E,d) is called a global NPC space if for each pair of points $x_0, x_1 \in E$, there exists $y \in E$ such that for all $z \in E$, $d^2(z, y) \leq \frac{1}{2} d^2(z, x_0) + \frac{1}{2} d^2(z, x_1) - \frac{1}{4} d^2(x_0, x_1)$.
- NPC spaces are geodesic spaces and every probability measure on such spaces that satisfies $\int d^2(x, x_0)\, d\mu(x) < \infty$ for some $x_0 \in E$ has a unique 2-barycenter.
- Wasserstein spaces are not NPC spaces in general. Two probability measures μ_0, μ_1 can have more than one mid-point in $(\mathcal{W}_p(E), W_p)$: each mid-point is a barycenter of $\frac{1}{2}\left(\delta_{\mu_0} + \delta_{\mu_1}\right) \in \mathcal{W}_p(\mathcal{W}_p(E))$.

Wasserstein space: Properties

- Consider a random probability measure $\tilde{\mu}$ in $\mathcal{W}_p(E)$, following a distribution $\mathbb{P}$.
- $\mathbb{P}$ is chosen in $\mathcal{W}_p(\mathcal{W}_p(E))$ endowed with the metric W_p.
- For all $v \in \mathcal{W}_p(E)$,

$$W_p^p(\delta_v, \mathbb{P}) = \mathbb{E}\left(W_p^p(\nu, \tilde{\mu})\right) = \int W_p^p(\nu, \mu) d\mathbb{P}(\mu) \tag{3}$$

- For a probability $\mathbb{P} \in \mathcal{W}_p(\mathcal{W}_p(E))$, consider a minimizer over $v \in \mathcal{W}_p(E)$ of $v \mapsto \mathbb{E}\left[W_p^p(v, \tilde{\mu})\right] = W_p^p(\delta_v, \mathbb{P})$, where $\tilde{\mu}$ is a random probability of $\mathcal{W}_p(E)$ with distribution $\mathbb{P}$.
- If exists, this probability measure is a barycenter of $\mathbb{P}$.

Existence of a Wasserstein Barycenter

定理 2.1

Set $p \geq 1$ and let (E, d) be a separable locally compact geodesic space. Hence, for $\mathbb{P} \in \mathcal{W}_p(\mathcal{W}_p(E))$, there exists a barycenter $\bar{\mu}_{\mathbb{P}}$ defined as

$$\bar{\mu}_{\mathbb{P}} \in \arg \min_{v \in \mathcal{W}_p(E)} \mathbb{E}\left[W_p^p(v, \tilde{\mu}) \right], \tag{4}$$

for $\tilde{\mu}$ a random measure with distribution $\mathbb{P}$.

Using the expression (3), we can see that Theorem 2.1 can be reformulated as stating the existence of the metric projection of $\mathbb{P}$ onto the subset of $\mathcal{W}_p(\mathcal{W}_p(E))$ of Dirac measures.

Consistency of the barycenter of a sequence of measures

定理 2.2

Set $p \geq 1$ and let (E, d) be a separable locally compact geodesic space. Let $\left(\mathbb{P}_j\right)_{j\geq 1} \subset \mathcal{W}_p\left(\mathcal{W}_p(E)\right)$ be a sequence of probability measures on $\mathcal{W}_p(E)$ and set μ_j a barycenter of $\mathbb{P}_j$, for all $j \in \mathbb{N}$. Suppose that for some $\mathbb{P} \in \mathcal{W}_p\left(\mathcal{W}_p(E)\right)$, we have that $W_p\left(\mathbb{P}, \mathbb{P}_j\right) \xrightarrow{j\to+\infty} 0$. Then, the sequence $(\mu_j)_{j\geq 1}$ is precompact in $\mathcal{W}_p(E)$ and any limit is a barycenter of $\mathbb{P}$. Corollary:

- The set of all barycenters of a given measure $\mathbb{P} \in \mathcal{W}_p\left(\mathcal{W}_p(E)\right)$ is compact.
- Suppose $\mathbb{P} \in \mathcal{W}_p\left(\mathcal{W}_p(E)\right)$ has a unique barycenter. Then for any sequence $\left(\mathbb{P}_j\right)_{j\geq 1} \subset \mathcal{W}_p\left(\mathcal{W}_p(E)\right)$ converging to $\mathbb{P}$, any sequence $(\mu_j)_{j\geq 1}$ of their barycenters converges to the barycenter of $\mathbb{P}$.

Example, $E = \mathbb{R}^d$ and $p = 2$

- Let $\mathbb{P} \in \mathcal{W}_2\left(\mathcal{W}_2(\mathbb{R})\right)$ such that there exists a set $A \subset \mathcal{P}_2\left(\mathbb{R}^d\right)$ of measures such that for all $\mu \in A$,

$$B \in \mathcal{B}\left(\mathbb{R}^d\right), \dim(B) \leq d-1 \Longrightarrow \mu(B) = 0, \tag{5}$$

  and $\mathbb{P}(A) > 0$, then, $\mathbb{P}$ admits a unique barycenter.

- For any sequence $\left(\mathbb{P}_j\right)_{j \geq 1}$ converging to $\mathbb{P}$ in $\mathcal{W}_2\left(\mathcal{W}_2(\mathbb{R})\right)$, the barycenters of Pjconverge to the barycenter of P.

- Proof: if v satisfies (5), then $\mu \mapsto W_2(\mu, v)$ is strictly convex, so is $\mu \mapsto \mathbb{E}\, W_2^2(\mu, \tilde{\mu})$.

Two statistical frameworks

- When confronted to the statistical analysis of a collection of probability measures in $\mathcal{W}_p(E), \mu_1, ..., \mu_J$, it is natural to define a notion of variability as $\mathrm{V}_J(\mu_1, \ldots, \mu_J) = \inf_{v \in \mathcal{W}_p(E)} \frac{1}{J} \sum_{j=1}^{J} W_p^p\left(\nu, \mu_j\right)$.
- In this work, we can extend this definition. $\mathrm{V}(\mu) = \inf_{v \in \mathcal{W}_p(E)} \mathbf{E}\left(W_p^p(v, \tilde{\mu})\right)$, where $\tilde{\mu}$ is a random probability measure in $\mathcal{W}_p(E)$.
- Two different frameworks: whether the number of probabilities goes to infinity or whether the probabilities are not observed directly but through empirical samples. Theorem (2.2) handles both of these settings.

Two statistical frameworks

- First: The distribution $\mathbb{P} \in \mathcal{W}_p\left(\mathcal{W}_p(E)\right)$ is approximated by a growing discrete distribution $\mathbb{P}_J$ supported on J elements, with J growing to infinity. Assume that $\mathbb{P}_J$ converges to some measure $\mathbb{P}$ with respect to Wasserstein distance. Hence Theorem (2.2) states that the barycenter (or any barycenter if not unique) of $\mathbb{P}_J$ converges to the barycenter of $\mathbb{P}$ (provided $\mathbb{P}$ has a unique barycenter).
- Second: The measures μ_j are unknown but approximated by a sequence of measures μ_j^n converging with respect to the Wasserstein distance to measures μ_j when n grows to infinity. Extracting a subsequence $\mathbb{P}_n = \sum_{j=1}^{J} \lambda_j \delta_{\mu_j^n}$, where $\mu_j^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_{i,j}}$ is empirical measure. We can still get the similar conclusion.