On concentration of the empirical measure for general transport costs

Jonghwa Park, Carnegie Mellon University

Joint work with Martin Larsson(Carnegie Mellon University) and Johannes Wiesel(Columbia University)

June 6th 2023
SIAM Conference on Financial Mathematics and Engineering

Wasserstein distance

- Let μ be a probability measure on $\mathbb{R}^d$. Approximate μ by its empirical measure $\mu_N = \frac{1}{N}\sum_{j=1}^{N}\delta_{X_j}$ where $X_1, \cdots, X_N$ are i.i.d samples of μ.
- It is well-known that μ_N converges to μ as $N \to \infty$, e.g. LLN.
- Wasserstein distance is commonly used in quantitative analysis.

Definition (The p-Wasserstein distance)

Let $p \geq 1$ and μ, ν be probability measures on $\mathbb{R}^d$.

$$\mathcal{W}_p(\mu,\nu) = \left(\inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \, \pi(dx,dy) \right)^{1/p}$$

where $\mathrm{Cpl}(\mu,\nu)$ is a collection of couplings between μ and ν.

- Optimal transport problem: minimizing the cost of the transport.
- $\mathcal{W}_p$ metrizes d-convergence.
- Application in data science, machine learning as well as finance.

Concentration estimates

- Interested in how fast $\mathcal{W}_p(\mu, \mu_N)$ deviates from 0. In other words, we study deviation estimates of the form: for all $x > 0$ and $N \in \mathbb{N}$,

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \leq \alpha(N, x)$$

  such that $\alpha(N, x) \to 0$ as $N \to \infty$.
- Existing results on $\mathbb{R}^d$: Bolley–Guillin–Villani('07), Golzan–Léonard('07), Boissard('11), Fournier–Guillin('15).
- Existing results on general spaces: Dedecker–Fan('15), Weed–Bach('19), Lei('20).

Fournier–Guillin rates

Theorem (Fournier–Guillin, '15).

Suppose μ is compactly supported. Then for all $N \geq 1$ and $x > 0$,

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \leq Ce^{-cN\varphi(x)} \mathbb{1}_{\{x \leq A\}}$$

where $\varphi(x) = x^2$ if $p > d/2$, $\varphi(x) = (x/\log(2 + 1/x))^2$ if $p = d/2$ and $\varphi(x) = x^{d/p}$ if $p \in [1, d/2)$.

- Constants $c, C, A > 0$ depend on d, p and $\text{supp}(\mu)$.
- For small dimension d, $\sqrt{N}\mathcal{W}_p^p(\mu, \mu_N)$ has subgaussian tails.
- As the dimension d gets larger, the estimate becomes weaker.
- It implies

$$\mathbb{E}[\mathcal{W}_p^p(\mu, \mu_N)] \leq C \begin{cases} N^{-1/2} & \text{if } p > d/2 \\ \log(N+1)N^{-1/2} & \text{if } p = d/2 \\ N^{-p/d} & \text{if } p \in [1, d/2) \end{cases}$$

which are known to be optimal (up to logarithmic terms when $p = d/2$).

Extension

How to extend the Fournier–Guillin rate to unbounded probability measure μ?

$$\mathbb{P}(\mathcal{W}_p^p(\mu,\mu_N) > x) \le Ce^{-cN\varphi(x)}\mathbb{1}_{\{x\le A\}} + \text{“error term”}.$$

- Existing results suggest that

$$\mathbb{P}(\mathcal{W}_p^p(\mu,\mu_N) > x) \le Ce^{-cN\varphi(x)}\mathbb{1}_{\{x\le A\}} + \mathbb{P}(M_p(\mu_N) - M_p(\mu) > x)$$

  for $M_p(\mu) = \int_{\mathbb{R}^d} |y|^p \, \mu(dy)$ and $M_p(\mu_N) = \int_{\mathbb{R}^d} |y|^p \, \mu_N(dy)$.
- Consistent with the compact case because

$$\mathbb{P}(M_p(\mu_N) - M_p(\mu) > x) \le Ce^{-cNx^2}\mathbb{1}_{\{x\le A\}} \le Ce^{-cN\varphi(x)}\mathbb{1}_{\{x\le A\}}.$$

- $M_p^{1/p}(\mu_N) - M_p^{1/p}(\mu)$ is a lower bound of $\mathcal{W}_p(\mu,\mu_N)$, e.g. when $p=1$,

$$\mathbb{P}(M_p(\mu_N) - M_p(\mu) > x) \le \mathbb{P}(\mathcal{W}_p^p(\mu,\mu_N) > x).$$

- The mean-deviation probability $\mathbb{P}(M_p(\mu_N) - M_p(\mu) > x)$ is well-studied.

Main result

Consider the optimal transport cost $\mathscr{D}$ which is

$$\mathscr{D}(\mu,\nu) = \inf_{\pi \in \mathrm{Cpl}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} f(|x-y|)\pi(dx,dy).$$

e.g when $f(r) = r^p$, $\mathscr{D} = \mathcal{W}_p^p$.

Growth condition

$f \geq 0$ is lower semicontinuous and satisfies $\sup_{0<r\leq R} \frac{f(r)}{r^p} < \infty$ for all $R > 0$.

Theorem

Under certain moment conditions on μ, for all $N \geq 1$ and $x > 0$,

$$\mathbb{P}(\mathscr{D}(\mu,\mu_N) > x)$$
$$\leq Ce^{-cN\varphi(x)}\mathbb{1}_{\{x\leq A\}} + \mathbb{P}\left(\int_{\mathbb{R}^d} G(|y|)\mu_N(dy) - \int_{\mathbb{R}^d} G(|y|)\mu(dy) > x\right).$$

where G is a nondecreasing function such that

$$f(|x-y|) \leq G(|x|) + G(|y|) \text{ for all } x, y \in \mathbb{R}^d.$$

- The moment condition mentioned above can be made explicit.

Examples: $\mathscr{D} = \mathcal{W}_p^p$

It provides estimates for $\mathcal{W}_p$ when $M_q(\mu) < \infty$ for some $q > 2p$.

Corollary

Suppose $M_q(\mu) < \infty$ for some $q > 2p$.

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \leq Ce^{-cN\varphi(x)}\mathbb{1}_{\{x \leq A\}} + \mathbb{P}(M_p(\mu_N) - M_p(\mu) > x).$$

The mean-deviation probability $\mathbb{P}(M_p(\mu_N) - M_p(\mu) > x)$ is well-studied.

- Let $X_1, \cdots, X_N$ be i.i.d samples of μ.
- Then

$$M_p(\mu_N) = \frac{1}{N}\sum_{j=1}^{N} |X_j|^p, \quad M_p(\mu) = \mathbb{E}[|X_1|^p].$$

- Deviation of the empirical mean from the true mean.
- Good estimates are well-known, e.g. under the same assumption,

$$\mathbb{P}(M_p(\mu_N) - M_p(\mu) > x) \leq C\left(e^{-cNx^2} + \frac{N}{(Nx)^{q/p}}\right).$$

Comparison with existing results

Examples

Suppose $M_q(\mu) < \infty$ for some $q > 2p$.

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \le Ce^{-cN\varphi(x)} \mathbb{1}_{\{x \le 1\}} + C \frac{N}{(Nx)^{q/p}}.$$

Theorem (Fournier–Guillin, '15)

Under the same assumption, for $\varepsilon \in (0, q/p)$.

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \le Ce^{-cN\varphi(x)} \mathbb{1}_{\{x \le 1\}} + C \frac{N}{(Nx)^{q/p-\varepsilon}}.$$

- Constants are not necessarily the same.
- It improves the estimate by Fournier–Guillin for fixed $x_0 > 0$,

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x_0) \le \frac{C}{N^{(q-p)/p}}, \quad \mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x_0) \le \frac{C}{N^{(q-p)/p-\varepsilon}}.$$

- Improves the existing results under different assumptions.

Examples: $\mathscr{D} = \mathcal{W}_p^p$

There is a different version of Theorem that works under more relaxed moment conditions.

- As a special case, it gives estimates for $\mathcal{W}_p$ when $M_q(\mu) < \infty$ for some $q > p$.
- No known results for $p < q \leq 2p$.

Corollary $(p > d/2)$

Suppose $M_q(\mu) < \infty$ for some $p < q \leq 2p$. Fix $\varepsilon \in (0, 1 - p/q)$.

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \leq Ce^{-cN^{2(1-p/q-\varepsilon)}x^2} \mathbb{1}_{\{x \leq 1\}} + C\frac{N}{(Nx)^{q/p}}.$$

- When $q > 2p$, recall that

$$\mathbb{P}(\mathcal{W}_p^p(\mu, \mu_N) > x) \leq Ce^{-cN\varphi(x)} \mathbb{1}_{\{x \leq 1\}} + C\frac{N}{(Nx)^{q/p}}.$$

- Under relaxed assumptions, the estimate becomes weaker:

$$Ce^{-cN\varphi(x)} \mathbb{1}_{\{x \leq 1\}} \leq Ce^{-cN^{2(1-p/q-\varepsilon)}x^2} \mathbb{1}_{\{x \leq 1\}}.$$

Examples: $\mathscr{D} \neq \mathcal{W}_p^p$

Let $p \geq 1$. Consider

$$\mathcal{E}_p(\mu, \mu_N) = \inf_{\pi \in \mathrm{Cpl}(\mu, \mu_N)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(e^{|x-y|^p} - 1 \right) \pi(dx, dy).$$

- $f(r) = e^{r^p} - 1$ satisfies the growth assumption.
- No known results.

Corollary (deviation estimates)

Suppose $\int_{\mathbb{R}^d} e^{b|y|^p} \mu(dy) < \infty$ for some

$$b > \begin{cases} 2^{2p+1} & \text{if } p \geq d/2 \\ \dfrac{4^p d}{d-p} & \text{if } p \in [1, d/2). \end{cases}$$

Then

$$\mathbb{P}(\mathcal{E}_p(\mu, \mu_N) > x) \leq Ce^{-cN\varphi(x)} \mathbb{1}_{\{x \leq 1\}} + C \frac{N}{(Nx)^{b/2^p}}.$$

Concentration inequalities imply following moment bounds:

Corollary (moment bounds)

Under the same assumptions,

$$\mathbb{E}[\mathcal{E}_p(\mu, \mu_N)] \leq C \begin{cases} N^{-1/2} & \text{if } p > d/2 \\ \log(N+1) N^{-1/2} & \text{if } p = d/2 \\ N^{-p/d} & \text{if } p \in [1, d/2). \end{cases}$$

- These bounds can be achieved (up to logarithmic term when $p = d/2$).
- As far as general laws are concerned, the moment bound mentioned above are sharp.

Idea of proof

Combine the Fournier–Guillin rate for compactly supported measures with empirical process theory.

1. Partition $\mathbb{R}^d$ into some compact sets $\{A^k\}_{k\geq 1}$ and apply Fournier–Guillin rates to each A^k.
2. Use bounds on the uniform deviation of self-normalized empirical process to control error terms.

Lemma (extension of Vapnik–Chervonenkis, '74)

For all $\delta > 0$, $N \geq 1$ and $x > 0$,

$$\mathbb{P}\left(\sup_{k\geq 1} 2^{-k\delta} \frac{(\mu(A^k) - \mu_N(A^k))_+}{\sqrt{\mu(A^k)}} > x\right) \leq Ce^{-cNx^2}.$$

Thank you !

arXiv: 2305.18636