Monotone Density Estimation with Wasserstein Projection

Lessard, Charles-Étienne; Liu, Zirui; Shen, Ruihang; Shirani, Parsa
Supervised by Wong, Leonard; Gunasingam, Madhu
Department of Statistical Sciences, University of Toronto

Statistical Sciences
UNIVERSITY OF TORONTO

Statistical Sciences
UNIVERSITY OF TORONTO

Abstract

This project investigated a novel approach to nonparametric density estimation under a monotonicity constraint, using optimal transport theory. While the Grenander estimator offers a classical solution by maximizing likelihood over non-increasing densities on $\mathbb{R}_+$, this project proposed an alternative method based on projecting the empirical distribution onto a convex set of probability measures whose quantile functions are convex and anchored at zero. This allows reformulation as a 2-Wasserstein projection in a Hilbert space, enabling efficient computation via convex optimization. This project implemented Wasserstein estimator, evaluated its performance via simulations. The project also analyzed its rate of convergence in comparison with the Grenander estimator, and proved the Wasserstein estimator is consistent.

Setting

In this project we consider i.i.d. samples from a density f on $\mathbb{R}_+$ such that:

- f is non-increasing on $\mathbb{R}_+$: for all $x \le y$, we have $f(x) \ge f(y)$.
- f has a finite second moment: $\int_0^\infty x^2 f(x)\,dx < \infty$.

We let $\mathcal{F}_0$ be the set of all such densities. Given observations $x_1, x_2, ..., x_n > 0$, we say $\hat{f}$ is the **Grenander estimator** with respect to $\mathcal{F}_0$ if

$$\hat{f} \in \arg\max_{f \in \mathcal{F}_0} \sum_{i=1}^{n} \log f(x_i).$$

If μ is a probability distribution, we let $Q_\mu(u) = \inf\{x : F_\mu(x) \ge u\}$ be its quantile function. For any probability distributions μ, ν, we define the 2-**Wasserstein distance** $\mathcal{W}_2(\mu,\nu)$ by the L^2 distance between their quantile functions:

$$\mathcal{W}_2(\mu,\nu) = \|Q_\mu - Q_\nu\|_{L^2([0,1])}.$$

Let $\mathcal{F}$ be a set of distributions μ on $\mathbb{R}_+$ such that their quantile functions Q_μ satisfy:

(i) Q_μ is non-decreasing.
(ii) $Q_\mu(0) = \lim_{u \to 0^+} Q_\mu(u) = 0$.
(iii) Q_μ is convex.

It can be shown that $\mathcal{F}$ is the $\mathcal{W}_2$-closure of $\mathcal{F}_0$. Define $\mathcal{Q} = \{Q_\mu : \mu \in \mathcal{F}\}$ as the set of quantile functions of distributions in $\mathcal{F}$.

(**Wasserstein Projection**): For $\mu \in \mathcal{P}_2(\mathbb{R}_+)$, there exists unique $\mu^* \in \mathcal{F}$ such that

$$\mu^* = \arg\min_{\nu \in \mathcal{F}} \mathcal{W}_2(\nu,\mu).$$

We write $\mu^* = \mathrm{proj}_{\mathcal{F}}\,\mu$. This is equivalent to finding

$$Q^* = \min_{Q \in \mathcal{Q}} \|Q - Q_\mu\|^2_{L^2([0,1])}.$$

(**Wasserstein Density Estimator**): Given data points $x_1, \ldots, x_n > 0$, consider the empirical distribution $\mu_n = \sum_{i=1}^n \frac{1}{n}\delta_{x_i}$. The Wasserstein density estimator $\hat{\mu}_n$ with respect to $\mathcal{F}$ is defined by the Wasserstein projection onto $\mathcal{F}$:

$$\hat{\mu}_n = \mathrm{proj}_{\mathcal{F}}\,\mu_n \text{ with quantile function } Q_{\hat{\mu}_n} = \min_{Q \in \mathcal{Q}} \|Q - Q_{\hat{\mu}_n}\|^2_{L^2([0,1])}.$$

Implementation

- **Wasserstein density estimator:** We implemented the Wasserstein density estimator in three steps:
  1. **Empirical quantile.** On the uniform grid $0 = u_0 < u_1 < \cdots < u_n = 1$, compute the empirical quantile function Q_{emp} based on the data.
  2. **Projection via quadratic programming.** Using the `cvxpy` package in Python, solve
  $$\min_{Q \in \mathbb{R}^n} \frac{1}{n} \sum_{j=0}^{n} \big(Q_j - Q_{\text{emp}}(u_j)\big)^2$$
  subject to $Q_0 = 0$, $Q_j \le Q_{j+1}$ (monotonicity), and $Q_{j+2} - 2Q_{j+1} + Q_j \ge 0$ (convexity) to obtain Q^*.
  3. **Density recovery.** Invert the projected quantile Q^* to get an approximate cumulative distribution function and hence the density.
- **Grenander estimator:** We used the R package `fdrtool`.

Findings

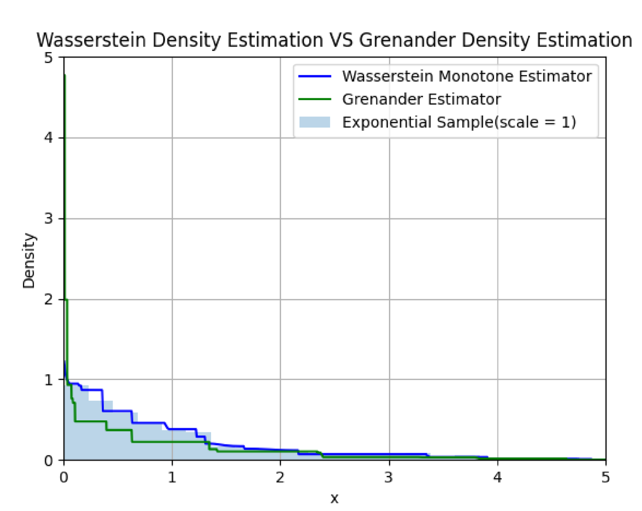

Figure 1: Wasserstein Estimator and Grenander Estimator estimating the exponential distribution

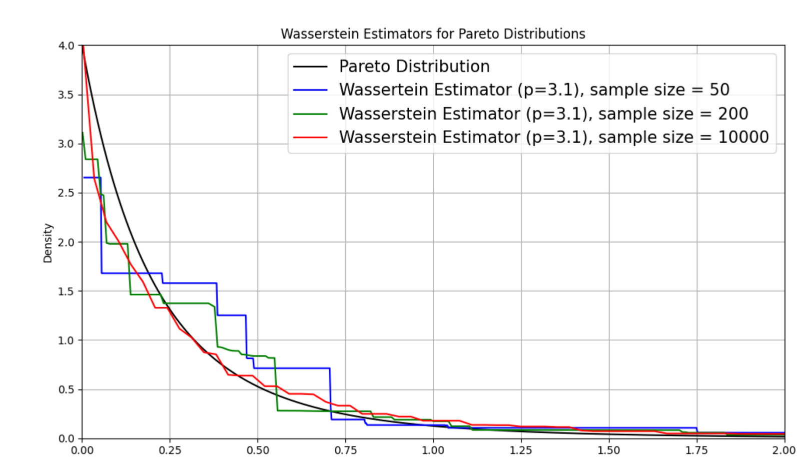

Figure 2: Wasserstein Estimator using different sample sizes estimating the Pareto distribution

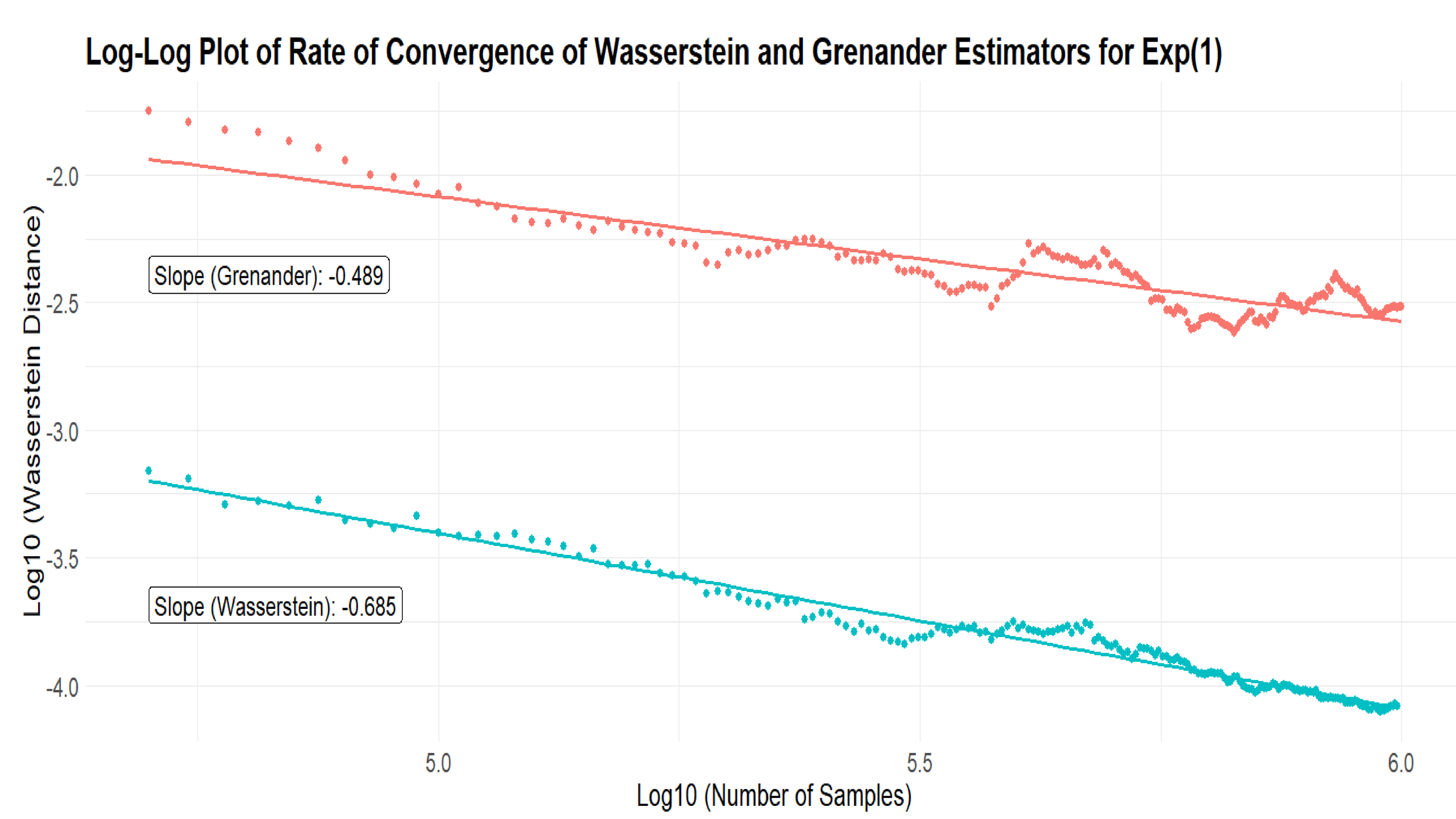

Figure 3: Log-log plot comparison between Wasserstein and Grenander Density Estimators estimating the exponential distribution (rate = 1)

Theoretical properties of the Wasserstein density estimator

Theorem: The Wasserstein density estimator is consistent.

$$\mathcal{W}_2(\hat{\mu}_n, \mu^*) \xrightarrow{a.s.} 0$$

Proof: Observe that

$$\mathcal{W}_2(\hat{\mu}_n, \mu^*) = \mathcal{W}_2\left(\mathrm{proj}_{\mathcal{F}}(\mu_n), \mathrm{proj}_{\mathcal{F}}(\mu)\right) \le \mathcal{W}_2(\mu_n, \mu) \qquad (1)$$

by 1-Lipschitz continuity of the projection operator. Since the empirical measure is $\mathcal{W}_2$-consistent we have $\mathcal{W}_2(\mu_n,\mu) \xrightarrow{a.s.} 0$ and thus $\mathcal{W}_2(\hat{\mu}_n, \mu^*) \xrightarrow{a.s.} 0$ as desired.

Theorem: Suppose μ^* is log-concave, then there exists $C > 0$ such that

$$\mathbb{E}\left[\mathcal{W}_2^2(\hat{\mu}_n, \mu^*)\right] \le \frac{C \log n}{n}.$$

Proof: From (1), we have

$$\mathbb{E}\left[\mathcal{W}_2^2(\hat{\mu}_n, \mu^*)\right] \le \mathbb{E}\left[\mathcal{W}_2^2(\mu_n, \mu)\right].$$

And by [1], $\mathbb{E}\left[\mathcal{W}_2^2(\mu_n, \mu)\right] \le \frac{C \log n}{n}$, where $C = C'\sigma^2$ with σ^2 the variance of μ and $C' > 0$.

Conclusions

On the exponential distribution with rate 1, we observe that:

- The Wasserstein distance of the Wasserstein estimator is significantly smaller than the Wasserstein distance of the Grenander estimator.
- The Wasserstein estimator converges faster than the Grenander estimator.

The Wasserstein estimator also provides accurate estimates for Pareto distributions. Finally, it possesses strong theoretical properties such as consistency and rapid convergence in expectation.

References

[1] Sergey Bobkov and Michel Ledoux.
One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances, volume 261 of *Memoirs of the American Mathematical Society*.
American Mathematical Society, Providence, RI, 2019.

[2] René Carmona and François Delarue.
Probabilistic Theory of Mean Field Games with Applications I: Mean Field FBSDEs, Control, and Games, volume 83 of *Probability Theory and Stochastic Modelling*, pages 1–695.
Springer Nature, United States, 2018.

[3] Ulf Grenander.
On the theory of mortality measurement: part ii.
Scandinavian Actuarial Journal, 1956(2):125–153, 1956.

Contact

Charles-Étienne Lessard: charles-etienne.lessard@mail.mcgill.ca
Zirui Liu: rayxxx.liu@mail.utoronto.ca
Ruihang Shen: ruihang.shen@mail.mcgill.ca
Parsa Shirani: parsa.shirani@mail.mcgill.ca