GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Distributional limits for optimal transport on finite spaces

Mass Transportation Theory: Opening Perspectives in Statistics, Probability and Computer Science

Marcel Klatt Axel Munk Max Sommerfeld Carla Tameling

June 5, 2018

Institute for Mathematical Stochastics, University Göttingen

(Empirical) optimal transport

Statistical framework:

(i) Let $(\mathcal{X}, d)$ be a Polish metric space, $p \in [1, +\infty)$ and μ, ν two probability measures on $\mathcal{X}$. The optimal transport distance (OT-distance a.k.a. pth-Wasserstein distance) between μ and ν is defined as

$$W_p(\mu, \nu) := \left\{ \min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y) \, d\pi(x, y) \right\}^{1/p} .$$

(ii) The empirical OT-distance is defined as

$$W_p(\hat{\mu}_n, \nu)$$

(resp. $W_p(\hat{\mu}_n, \hat{\nu}_m)$), where the empirical measure $\hat{\mu}_n$ (resp. $\hat{\nu}_m$) is generated by a sample $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu$ (resp. $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} \nu$).

(Empirical) optimal transport

Statistical framework:

(i) Let $(\mathcal{X}, d)$ be a Polish metric space, $p \in [1, +\infty)$ and μ, ν two probability measures on $\mathcal{X}$. The optimal transport distance (OT-distance a.k.a. pth-Wasserstein distance) between μ and ν is defined as

$$W_p(\mu, \nu) := \left\{ \min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y)\, d\pi(x, y) \right\}^{1/p} .$$

(ii) The empirical OT-distance is defined as

$$W_p(\hat{\mu}_n, \nu)$$

(resp. $W_p(\hat{\mu}_n, \hat{\nu}_m)$), where the empirical measure $\hat{\mu}_n$ (resp. $\hat{\nu}_m$) is generated by a sample $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu$ (resp. $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} \nu$).

Central question for (empirical) OT-distances

How does the random quantity $W_p(\hat{\mu}_n, \nu)$ relate to $W_p(\mu, \nu)$?

- Rates of convergence and concentration results for the empirical OT-distance
  - ▷ del Barrio & Matrán (2013)
  - ▷ Boissard & Le Gouic (2014)
  - ▷ Fournier & Guilin (2014)
  - ▷ Weed & Bach (2017)

Central question for (empirical) OT-distances

How does the random quantity $W_p(\hat{\mu}_n, \nu)$ relate to $W_p(\mu, \nu)$?

- Rates of convergence and concentration results for the empirical OT-distance
  - ▷ del Barrio & Matrán (2013)
  - ▷ Boissard & Le Gouic (2014)
  - ▷ Fournier & Guilin (2014)
  - ▷ Weed & Bach (2017)
- **Distributional limits for the empirical OT-distance**
  - ▷ Dimension $D = 1$ ($\mathbb{R}$, the real line):
    - ○ Munk & Czado (1998)
    - ○ del Barrio et al. (1999, 2005, 2015)
    - ○ Freitag & Munk (2005)

Central question for (empirical) OT-distances

How does the random quantity $W_p(\hat{\mu}_n, \nu)$ relate to $W_p(\mu, \nu)$?

- Rates of convergence and concentration results for the empirical OT-distance
  - ▷ del Barrio & Matrán (2013)
  - ▷ Boissard & Le Gouic (2014)
  - ▷ Fournier & Guilin (2014)
  - ▷ Weed & Bach (2017)
- **Distributional limits for the empirical OT-distance**
  - ▷ Dimension $D = 1$ ($\mathbb{R}$, the real line):
    - Munk & Czado (1998)
    - del Barrio et al. (1999, 2005, 2015)
    - Freitag & Munk (2005)
  - ▷ Dimension $D \geq 2$ ($\mathbb{R}^D$):
    - elliptical distributions: Rippl et al. (2016)
    - general results for $\mu \neq \nu$: del Barrio & Loubes (2017)

Central question for (empirical) OT-distances

How does the random quantity $W_p(\hat{\mu}_n, \nu)$ relate to $W_p(\mu, \nu)$?

- Rates of convergence and concentration results for the empirical OT-distance
  - ▷ del Barrio & Matrán (2013)
  - ▷ Boissard & Le Gouic (2014)
  - ▷ Fournier & Guilin (2014)
  - ▷ Weed & Bach (2017)
- **Distributional limits for the empirical OT-distance**
  - ▷ Dimension $D = 1$ ($\mathbb{R}$, the real line):
    - Munk & Czado (1998)
    - del Barrio et al. (1999, 2005, 2015)
    - Freitag & Munk (2005)
  - ▷ Dimension $D \geq 2$ ($\mathbb{R}^D$):
    - elliptical distributions: Rippl et al. (2016)
    - general results for $\mu \neq \nu$: del Barrio & Loubes (2017)

Optimal transport on finite metric spaces

Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be a finite space with metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ and

$$\Delta_N := \{r \in \mathbb{R}_+^N \mid \sum_{i=1}^{N} r_i = 1\}$$

be the N-dimensional simplex of probability measures on $\mathcal{X}$. The OT-distance between $r, s \in \Delta_N$ is given by the optimal value of a finite dimensional linear program, i.e.,

$$W_p(r,s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p},$$

where the feasible set in a finite setting is given by

$$\Pi(r,s) = \{\pi \in \mathbb{R}_+^{N \times N} \mid \sum_{j=1}^{N} \pi_{ij} = r_i, \sum_{i=1}^{N} \pi_{ij} = s_j\}.$$

Optimal transport on finite metric spaces

Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be a finite space with metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ and

$$\Delta_N := \{r \in \mathbb{R}_+^N \mid \sum_{i=1}^{N} r_i = 1\}$$

be the N-dimensional simplex of probability measures on $\mathcal{X}$.The OT-distance between $r, s \in \Delta_N$ is given by the optimal value of a finite dimensional linear program, i.e.,

$$W_p(r, s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p},$$

where the feasible set in a finite setting is given by

$$\Pi(r, s) = \{\pi \in \mathbb{R}_+^{N \times N} \mid \sum_{j=1}^{N} \pi_{ij} = r_i, \sum_{i=1}^{N} \pi_{ij} = s_j\}.$$

Limit laws for finite metric spaces

Theorem (Sommerfeld & Munk (2017))

With the sample size n approaching infinity, it holds that

- *One sample* ($r = s$):

$$n^{1/2p} W_p(\hat{r}_n, r) \xrightarrow{\mathfrak{D}} \left\{ \max_{f \in \Phi^*(r,r)} \langle \mathbf{G}, f \rangle \right\}^{1/p}$$

- *One sample* ($r \neq s$):

$$\sqrt{n}\{W_p(\hat{r}_n, s) - W_p(r, s)\} \xrightarrow{\mathfrak{D}} \frac{1}{p} W_p^{1-p}(r, s) \left\{ \max_{f \in \Phi^*(r,s)} \langle \mathbf{G}, f \rangle \right\}$$

- $\hat{r}_n$ empirical measure generated by $X_1, \ldots, X_n \overset{i.i.d.}{\sim} r$
- $\Phi^*(r, s)$ set of dual solutions
- $\mathbf{G}$ the Gaussian limit of $\sqrt{n}(\hat{r}_n - r)$

Limit laws for finite metric spaces

Theorem (Sommerfeld & Munk (2017))

With the sample size n approaching infinity, it holds that

- *One sample* ($r = s$):

$$n^{1/2p} W_p(\hat{r}_n, r) \xrightarrow{\mathfrak{D}} \left\{ \max_{f \in \Phi^*(r,r)} \langle \mathbf{G}, f \rangle \right\}^{1/p}$$

- *One sample* ($r \neq s$):

$$\sqrt{n}\{W_p(\hat{r}_n, s) - W_p(r, s)\} \xrightarrow{\mathfrak{D}} \frac{1}{p} W_p^{1-p}(r, s) \left\{ \max_{f \in \Phi^*(r,s)} \langle \mathbf{G}, f \rangle \right\}$$

- $\hat{r}_n$ empirical measure generated by $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} r$
- $\Phi^*(r, s)$ set of dual solutions
- $\mathbf{G}$ the Gaussian limit of $\sqrt{n}(\hat{r}_n - r)$

Proof strategy:

- Consider the OT-distance as the optimal value of a **finite dimensional linear program**

$$(r, s) \mapsto \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij}\,.$$

- Prove that this functional is **directionally Hadamard differentiable** (Gal & Greenberg (1997))

Limit laws for empirical OT-distances

Proof strategy:

- Consider the OT-distance as the optimal value of a **finite dimensional linear program**

$$(r, s) \mapsto \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} \,.$$

- Prove that this functional is **directionally Hadamard differentiable** (Gal & Greenberg (1997))
- Apply (generalized) **delta method** (Römisch (2004))

Proof strategy:

- Consider the OT-distance as the optimal value of a **finite dimensional linear program**

$$(r, s) \mapsto \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} \,.$$

- Prove that this functional is **directionally Hadamard differentiable** (Gal & Greenberg (1997))
- Apply (generalized) **delta method** (Römisch (2004))

Limit laws for empirical OT-distances – Extensions

- Can be extended to $W_p(\hat{r}_n, \hat{s}_m)$
- Explicit limit distributions, e.g. for tree metrics, non-degeneracy
- m out of n bootstrap (need $m = o(n)$; n out of n bootstrap fails)
- Limit laws for countable metric spaces $\mathcal{X} = \{x_1, x_2, \ldots\}$ (Tameling, Sommerfeld & Munk (2017))
  - ▷ requires a careful calibration of the norm
  - ▷ only for measures r with $\sum_{i=1}^{\infty} d^p(x_i, x_0)\sqrt{r_i} < \infty$, where $x_0 \in \mathcal{X}$ arbitrary

Limit laws for empirical OT-distances – Extensions

- Can be extended to $W_p(\hat{r}_n, \hat{s}_m)$
- Explicit limit distributions, e.g. for tree metrics, non-degeneracy
- m out of n bootstrap (need $m = o(n)$; n out of n bootstrap fails)
- Limit laws for countable metric spaces $\mathcal{X} = \{x_1, x_2, \dots\}$ (Tameling, Sommerfeld & Munk (2017))
  - ▷ requires a careful calibration of the norm
  - ▷ only for measures r with $\sum_{i=1}^{\infty} d^p(x_i, x_0)\sqrt{r_i} < \infty$, where $x_0 \in \mathcal{X}$ arbitrary

Computational burden of OT-distances

In general, the computational cost to calculate the OT-distance

$$W_p(r, s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p}$$

is of order $\mathcal{O}(N^3 \log(N))$.

Workarounds:

- Exploiting the underlying metric structure (Ling & Okada (2007))
- Graph sparsification (Pele & Werman (2009))
- Specialized algorithms, e.g. shortlist (Gottschlich & Schuhmacher (2014)), shielding neighborhood (Schmitzer (2016))
- Subsampling methods (Sommerfeld, Schrieber & Munk (2018))

Computational burden of OT-distances

In general, the computational cost to calculate the OT-distance

$$W_p(r, s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p}$$

is of order $\mathcal{O}(N^3 \log(N))$.

Workarounds:

- Exploiting the underlying metric structure (Ling & Okada (2007))
- Graph sparsification (Pele & Werman (2009))
- Specialized algorithms, e.g. shortlist (Gottschlich & Schuhmacher (2014)), shielding neighborhood (Schmitzer (2016))
- Subsampling methods (Sommerfeld, Schrieber & Munk (2018))
- **Regularization methods** (Cuturi (2013), Dessein et al. (2016))

Computational burden of OT-distances

In general, the computational cost to calculate the OT-distance

$$W_p(r,s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p}$$

is of order $\mathcal{O}(N^3 \log(N))$.

Workarounds:

- Exploiting the underlying metric structure (Ling & Okada (2007))
- Graph sparsification (Pele & Werman (2009))
- Specialized algorithms, e.g. shortlist (Gottschlich & Schuhmacher (2014)), shielding neighborhood (Schmitzer (2016))
- Subsampling methods (Sommerfeld, Schrieber & Munk (2018))
- **Regularization methods** (Cuturi (2013), Dessein et al. (2016))

Regularized optimal transport

Let $f\colon \mathbb{R}^{N\times N} \to \mathbb{R}$ be the negative entropy

$$f(\pi) := \begin{cases} \sum_{i,j=1}^{N} \pi_{ij} \log(\pi_{ij}) & \text{for } \pi \in \mathbb{R}_{+}^{N\times N}, \\ +\infty & \text{otherwise.} \end{cases}$$

For two measures r, s on the finite metric space $\mathcal{X} = \{x_1, \ldots, x_N\}$ and $\lambda > 0$ find the regularized transport plan

$$\pi_\lambda(r,s) = \underset{\pi\in\Pi(r,s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i,x_j)\pi_{ij} + \lambda f(\pi)\,.$$

Regularized optimal transport

Let $f\colon \mathbb{R}^{N\times N} \to \mathbb{R}$ be the negative entropy

$$f(\pi) := \begin{cases} \sum_{i,j=1}^{N} \pi_{ij}\log(\pi_{ij}) & \text{for } \pi \in \mathbb{R}_{+}^{N\times N}, \\ +\infty & \text{otherwise.} \end{cases}$$

For two measures r, s on the finite metric space $\mathcal{X} = \{x_1, \dots, x_N\}$ and $\lambda > 0$ find the regularized transport plan

$$\pi_\lambda(r, s) = \underset{\pi\in\Pi(r,s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} + \lambda f(\pi).$$

Define the regularized OT-distance (a.k.a. Sinkhorn divergence, rot mover's distance) as

$$W_{\lambda,p}(r, s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_\lambda(r, s)_{ij} \right\}^{1/p}.$$

Regularized optimal transport

Let $f\colon \mathbb{R}^{N\times N} \to \mathbb{R}$ be the negative entropy

$$f(\pi) := \begin{cases} \sum_{i,j=1}^{N} \pi_{ij}\log(\pi_{ij}) & \text{for } \pi \in \mathbb{R}_+^{N\times N}, \\ +\infty & \text{otherwise.} \end{cases}$$

For two measures r, s on the finite metric space $\mathcal{X} = \{x_1, \dots, x_N\}$ and $\lambda > 0$ find the regularized transport plan

$$\pi_\lambda(r,s) = \underset{\pi\in\Pi(r,s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i,x_j)\pi_{ij} + \lambda f(\pi)\,.$$

Define the regularized OT-distance (a.k.a. Sinkhorn divergence, rot mover's distance) as

$$W_{\lambda,p}(r,s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i,x_j)\pi_\lambda(r,s)_{ij} \right\}^{1/p}.$$

Limit laws for empirical regularized transport plans

The empirical regularized transport plan is defined as

$$\pi_\lambda(\hat{r}_n, s) = \underset{\pi \in \Pi(\hat{r}_n, s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} + \lambda f(\pi) .$$

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_{N^2}(0, \Sigma_\lambda(r|s)) .$$

Limit laws for empirical regularized transport plans

The empirical regularized transport plan is defined as

$$\pi_\lambda(\hat{r}_n, s) = \underset{\pi \in \Pi(\hat{r}_n, s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} + \lambda f(\pi) \, .$$

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_{N^2}(0, \Sigma_\lambda(r|s)) \, .$$

Remark

Limit distributions for the (nonregularized) transport plan ($\lambda = 0$) are not known.

Limit laws for empirical regularized transport plans

The empirical regularized transport plan is defined as

$$\pi_\lambda(\hat{r}_n, s) = \underset{\pi \in \Pi(\hat{r}_n, s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} + \lambda f(\pi).$$

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_{N^2}(0, \Sigma_\lambda(r|s)).$$

Remark

Limit distributions for the (nonregularized) transport plan ($\lambda = 0$) are not known.

Proof strategy:

- We think of $\pi_\lambda(r,s)$ as a vector and consider the **functional**

$$\phi_\lambda : (r,s) \mapsto \underset{\pi \in \mathbb{R}^{N^2}}{\arg\min} \ \langle d^p, \pi \rangle + \lambda f(\pi)$$
$$\text{s.t.} \quad A_\star \pi = \left[r \, , \, s_\star \right]^T .$$

Advantage to (nonregularized) OT: Uniqueness of $\pi_\lambda(r,s)$

- **Sensitivity analysis** of the optimal solution (Fiacco (1983)):

Proof strategy:

- We think of $\pi_\lambda(r, s)$ as a vector and consider the **functional**

$$\phi_\lambda : (r, s) \mapsto \underset{\pi \in \mathbb{R}^{N^2}}{\arg\min} \ \langle d^p, \pi \rangle + \lambda f(\pi)$$
$$\text{s.t.} \quad A_\star \pi = \left[r \, , \, s_\star \right]^T .$$

  Advantage to (nonregularized) OT: Uniqueness of $\pi_\lambda(r, s)$

- **Sensitivity analysis** of the optimal solution (Fiacco (1983)):
  - ▷ State optimality conditions for $\pi_\lambda(r, s)$ (a.k.a. KKT-conditions)
  - ▷ Apply the implicit function theorem
  - ⇒ The function ϕ_λ is **differentiable**

  Advantage to (nonregularized) OT: Non-Sparsity of $\pi_\lambda(r, s)$

Proof strategy:

- We think of $\pi_\lambda(r,s)$ as a vector and consider the **functional**

$$\phi_\lambda : (r,s) \mapsto \underset{\pi \in \mathbb{R}^{N^2}}{\arg\min} \; \langle d^p, \pi \rangle + \lambda f(\pi)$$
$$\text{s.t.} \quad A_\star \pi = \left[r \,,\, s_\star \right]^T .$$

Advantage to (nonregularized) OT: Uniqueness of $\pi_\lambda(r,s)$

- **Sensitivity analysis** of the optimal solution (Fiacco (1983)):
  - ▷ State optimality conditions for $\pi_\lambda(r,s)$ (a.k.a. KKT-conditions)
  - ▷ Apply the implicit function theorem
  - ⇒ The function ϕ_λ is **differentiable**

Advantage to (nonregularized) OT: Non-Sparsity of $\pi_\lambda(r,s)$

- Apply (multivariate) **delta method** (van der Vaart (2000))

Proof strategy:

- We think of $\pi_\lambda(r, s)$ as a vector and consider the **functional**

$$\phi_\lambda : (r, s) \mapsto \underset{\pi \in \mathbb{R}^{N^2}}{\arg\min} \ \langle d^p, \pi \rangle + \lambda f(\pi)$$

$$\text{s.t.} \quad A_\star \pi = \left[r \, , \, s_\star \right]^T .$$

  Advantage to (nonregularized) OT: Uniqueness of $\pi_\lambda(r, s)$

- **Sensitivity analysis** of the optimal solution (Fiacco (1983)):
  - ▷ State optimality conditions for $\pi_\lambda(r, s)$ (a.k.a. KKT-conditions)
  - ▷ Apply the implicit function theorem
  - ⇒ The function ϕ_λ is **differentiable**

  Advantage to (nonregularized) OT: Non-Sparsity of $\pi_\lambda(r, s)$

- Apply (multivariate) **delta method** (van der Vaart (2000))

The covariance matrix $\Sigma_\lambda(r|s)$

According to the **implicit function theorem** we obtain that

$$\nabla\phi_\lambda(r,s) = D A_\star^T [A_\star \, D \, A_\star^T]^{-1} \,.$$

- $A_\star$ is the coefficient matrix encoding the marginal constraints
- D is a diagonal matrix with diagonal $\pi_\lambda(r,s)$

Hence, the (multivariate) delta method tells us that

$$\Sigma_\lambda(r|s) = \nabla_r\phi_\lambda(r,s)\,\Sigma(r)\,\nabla_r\phi_\lambda(r,s)^T \,.$$

The covariance matrix $\Sigma_\lambda(r|s)$

According to the **implicit function theorem** we obtain that

$$\nabla\phi_\lambda(r,s) = DA_\star^T[A_\star\, D\, A_\star^T]^{-1}\,.$$

- $A_\star$ is the coefficient matrix encoding the marginal constraints
- D is a diagonal matrix with diagonal $\pi_\lambda(r,s)$

Hence, the (multivariate) delta method tells us that

$$\Sigma_\lambda(r|s) = \nabla_r\phi_\lambda(r,s)\,\Sigma(r)\,\nabla_r\phi_\lambda(r,s)^T\,.$$

Limit laws for empirical regularized transport distances

The empirical regularized OT-distance is defined as

$$W_{\lambda,p}(\hat{r}_n, s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_\lambda(\hat{r}_n, s)_{ij} \right\}^{1/p}.$$

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\left\{W_{\lambda,p}(\hat{r}_n, s) - W_{\lambda,p}(r, s)\right\} \xrightarrow{\mathfrak{D}} \mathcal{N}_1(0, \sigma^2_\lambda(r|s)).$$

The empirical regularized OT-distance is defined as

$$W_{\lambda,p}(\hat{r}_n, s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_\lambda(\hat{r}_n, s)_{ij} \right\}^{1/p} .$$

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{W_{\lambda,p}(\hat{r}_n, s) - W_{\lambda,p}(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_1(0, \sigma^2_\lambda(r|s)) .$$

Limit laws for empirical regularized transport distances

One can also consider directly the **optimal value** of the regularized transport problem, i.e.,

$$p_\lambda(r,s) = \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i,x_j)\pi_{ij} + \lambda f(\pi).$$

Theorem (Bigot, Cazelles & Papadakis (2017))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\{p_\lambda(\hat{r}_n,s) - p_\lambda(r,s)\} \xrightarrow{\mathfrak{D}} \langle \mathbf{G}, \lambda \log(u) \rangle$$

- The vector u is the left scaling for the regularized transport plan

$$\pi_\lambda(r,s) = \mathrm{diag}(u) \exp\left(-\frac{d^p}{\lambda}\right) \mathrm{diag}(v)$$

- $\mathbf{G}$ the Gaussian limit of $\sqrt{n}(\hat{r}_n - r)$

One can also consider directly the **optimal value** of the regularized transport problem, i.e.,

$$p_\lambda(r,s) = \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i,x_j)\pi_{ij} + \lambda f(\pi)\,.$$

Theorem (Bigot, Cazelles & Papadakis (2017))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\{p_\lambda(\hat{r}_n,s) - p_\lambda(r,s)\} \xrightarrow{\mathfrak{D}} \langle \mathbf{G},\, \lambda \log(u) \rangle$$

- The vector u is the left scaling for the regularized transport plan

$$\pi_\lambda(r,s) = \text{diag}(u)\,\exp\left(-\frac{d^p}{\lambda}\right)\,\text{diag}(v)$$

- $\mathbf{G}$ the Gaussian limit of $\sqrt{n}(\hat{r}_n - r)$

Speed of convergence

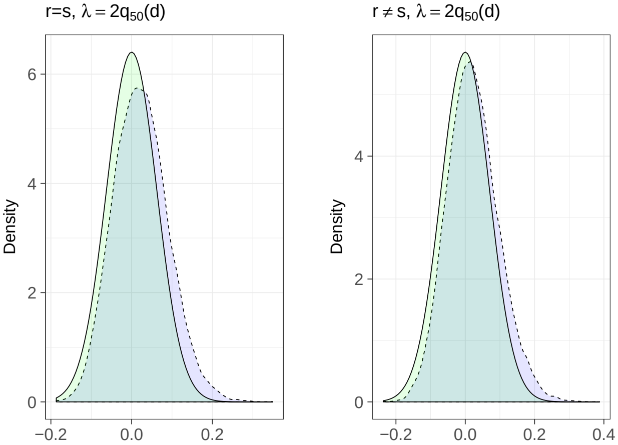

Figure 1: Density for the sample distribution (dashed line, $n = 10$ samples) for $r = s$ (left) and $r \neq s$ (right) and the density of the corresponding normal limit (solid line).

Speed of convergence

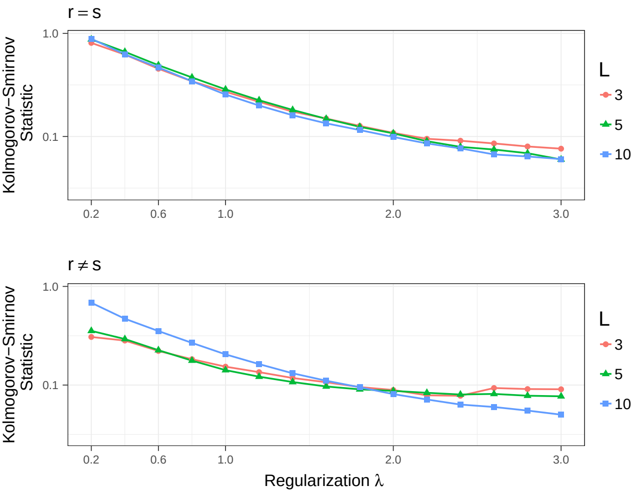

Figure 2: Kolmogorov-Smirnov distance on a logarithmic scale between the finite sample distribution ($n = 25$) and the theoretical normal distribution averaged over five measures.

As a byproduct of the **delta method**, we obtain **consistency** of the n out of n bootstrap:

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sup_{h \in BL_1(\mathbb{R})} \Big| \mathbb{E}[h(\sqrt{n}\{W_{\lambda,p}(\hat{r}_n^*, s) - W_{\lambda,p}(\hat{r}_n, s)\})|X_1, \ldots, X_n] - \mathbb{E}[h(\sqrt{n}\{W_{\lambda,p}(\hat{r}_n, s) - W_{\lambda,p}(r, s)\})] \Big| \xrightarrow{\mathbb{P}} 0\,.$$

- $\mathrm{BL}_1(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid \|f\|_\infty \leq 1, |f(z_1) - f(z_2)| \leq |z_1 - z_2|\}$
- $\hat{r}_n^*$ bootstrapped version of $\hat{r}_n$

As a byproduct of the **delta method**, we obtain **consistency** of the n out of n bootstrap:

Theorem (K., Tameling & Munk (2018))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sup_{h \in BL_1(\mathbb{R})} \left| \mathbb{E}[h(\sqrt{n}\{W_{\lambda,p}(\hat{r}_n^*, s) - W_{\lambda,p}(\hat{r}_n, s)\}) | X_1, \dots, X_n] - \mathbb{E}[h(\sqrt{n}\{W_{\lambda,p}(\hat{r}_n, s) - W_{\lambda,p}(r, s)\})] \right| \xrightarrow{\mathbb{P}} 0\,.$$

- $\mathrm{BL}_1(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid \|f\|_\infty \leq 1,\ |f(z_1) - f(z_2)| \leq |z_1 - z_2|\}$
- $\hat{r}_n^*$ bootstrapped version of $\hat{r}_n$

Limit laws for empirical regularized transport – Extensions

- Limit laws hold for **more general regularizers**:

  Let f be twice continuously differentiable on the interior of its domain with positive definite Hessian $\nabla^2 f$. Moreover, assume that

(A1) f of Legendre type,

(A2) $\mathbb{R}_{-}^{N^2} \subset \operatorname{dom} f^*$,

(A3) $(0,1)^{N^2} \subseteq \operatorname{dom} f$,

(A4) $\operatorname{dom} f \subseteq \mathbb{R}_{+}^{N^2}$.

Limit laws for empirical regularized transport – Extensions

- Limit laws hold for **more general regularizers**:
  Let f be twice continuously differentiable on the interior of its domain with positive definite Hessian $\nabla^2 f$. Moreover, assume that

  (A1) f of Legendre type,

  (A2) $\mathbb{R}_{-}^{N^2} \subset \operatorname{dom} f^*$,

  (A3) $(0,1)^{N^2} \subseteq \operatorname{dom} f$,

  (A4) $\operatorname{dom} f \subseteq \mathbb{R}_{+}^{N^2}$.

  Examples are given by Dessein et al. (2016), e.g.

  ▷ Burg entropy: $f(\pi) = \sum_{i,j=1}^{N^2} \log(\pi_{ij}) - \pi_{ij} - 1$

  ▷ l_p quasi norm: $f(\pi) = \sum_{i,j=1}^{N^2} {\pi_{ij}}^p$, $0 < p < 1$

Limit laws for empirical regularized transport – Extensions

- Limit laws hold for **more general regularizers**:
  Let f be twice continuously differentiable on the interior of its domain with positive definite Hessian $\nabla^2 f$. Moreover, assume that

(A1) f of Legendre type,

(A2) $\mathbb{R}_-^{N^2} \subset \operatorname{dom} f^*$,

(A3) $(0,1)^{N^2} \subseteq \operatorname{dom} f$,

(A4) $\operatorname{dom} f \subseteq \mathbb{R}_+^{N^2}$.

Examples are given by Dessein et al. (2016), e.g.

- Burg entropy: $f(\pi) = \sum_{i,j=1}^{N^2} \log(\pi_{ij}) - \pi_{ij} - 1$
- l_p quasi norm: $f(\pi) = \sum_{i,j=1}^{N^2} {\pi_{ij}}^p$, $0 < p < 1$

Summary: (nonregularized) OT vs. regularized OT

- **Different limit laws** under equality of measures (non-normal vs. normal)

(Nonregularized) OT-distance	regularized OT-distance
$n^{1/2p} W_p(\hat{r}_n, r)$	$\sqrt{n}\,\{W_{\lambda,p}(\hat{r}_n, r) - W_{\lambda,p}(r, r)\}$
$\big\downarrow \mathfrak{D}$	$\big\downarrow \mathfrak{D}$
$\left\{ \max_{f \in \Phi^*(r,r)} \langle \mathbf{G}, f \rangle \right\}^{1/p}$	$\mathcal{N}_1(0, \sigma^2_\lambda(r\|r))$

- **Different scaling behavior**, i.e., for regularized OT-distance the scaling behavior is independent of p

Summary: (nonregularized) OT vs. regularized OT

- **Different limit laws** under equality of measures (non-normal vs. normal)

(Nonregularized) OT-distance	regularized OT-distance
$n^{1/2p} W_p(\hat{r}_n, r)$	$\sqrt{n}\,\{W_{\lambda,p}(\hat{r}_n, r) - W_{\lambda,p}(r, r)\}$
$\big\downarrow \mathfrak{D}$	$\big\downarrow \mathfrak{D}$
$\left\{ \max_{f \in \Phi^*(r,r)} \langle \mathbf{G}, f \rangle \right\}^{1/p}$	$\mathcal{N}_1(0, \sigma^2_\lambda(r\|r))$

- **Different scaling behavior**, i.e., for regularized OT-distance the scaling behavior is independent of p
- We find that

$$\lim_{\lambda \searrow 0} \sigma^2_\lambda(r|r) = 0\,.$$

Summary: (nonregularized) OT vs. regularized OT

- **Different limit laws** under equality of measures (non-normal vs. normal)

(Nonregularized) OT-distance	regularized OT-distance
$n^{1/2p} W_p(\hat{r}_n, r)$	$\sqrt{n}\,\{W_{\lambda,p}(\hat{r}_n, r) - W_{\lambda,p}(r, r)\}$
$\Big\downarrow \mathfrak{D}$	$\Big\downarrow \mathfrak{D}$
$\left\{ \max_{f \in \Phi^*(r,r)} \langle \mathbf{G}, f \rangle \right\}^{1/p}$	$\mathcal{N}_1(0, \sigma^2_\lambda(r\vert r))$

- **Different scaling behavior**, i.e., for regularized OT-distance the scaling behavior is independent of p
- We find that

$$\lim_{\lambda \searrow 0} \sigma^2_\lambda(r|r) = 0\,.$$