GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

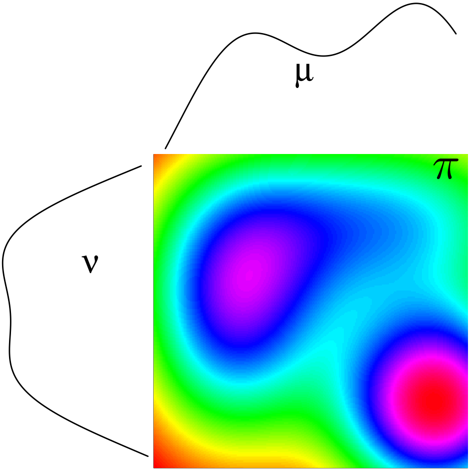
COMPUTATIONAL STRATEGIES FOR STATISTICAL INFERENCE BASED ON EMPIRICAL OPTIMAL TRANSPORT

Axel Munk   Carla Tameling

June 4, 2018

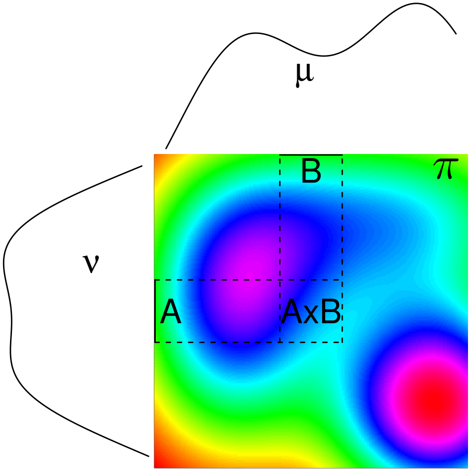
Institute for Mathematical Stochastics, University Göttingen

Wasserstein Distance as Optimal Transport Distance



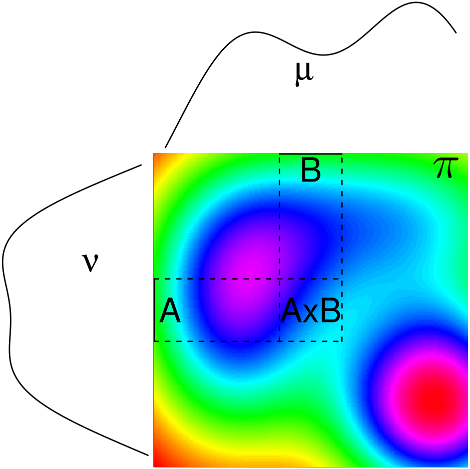
- Let μ and ν be two probability measures on a metric space $(\mathcal{X}, d)$
- $\Pi(\mu, \nu) = \{$probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and $\nu\}$
- $\pi \in \Pi(\mu, \nu)$ is a **transport plan** from ν to μ.

Wasserstein Distance as Optimal Transport Distance



- Let μ and ν be two probability measures on a metric space $(\mathcal{X}, d)$
- $\Pi(\mu, \nu) = \{$probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and $\nu\}$
- $\pi \in \Pi(\mu, \nu)$ is a **transport plan** from ν to μ.
- $\pi(A \times B) =$ mass transported from A to B

Wasserstein Distance as Optimal Transport Distance



- incorporates **ground distance** from the space in question
- intuitive interpretation as **amount of 'work'** to transform one probability measure into another
- performs exceptionally well at capturing **human perception** of similarity

The Wasserstein Distance

Definition

a) The p-th Wasserstein distance (WD) between μ and ν is defined as

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

where $\Pi(\mu, \nu)$ are all probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν (couplings).

b) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu$. The empirical Wasserstein distance (EWD) is defined as

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

$\hat{\mu}_n$ empirical measure.
Same scheme for two sample case yields $W_p(\hat{\mu}_n, \hat{\nu}_n)$.

How to do inference?

Inference

We want to quantify the **fluctuation behavior** of $W_p(\hat{\mu}_n, \nu)$ around $W_p(\mu, \nu)$. Typical questions are:

- testing whether two samples $\hat{\mu}_n$ and $\hat{\nu}_n$ stem from the same distribution, i.e., $W_p(\hat{\mu}_n, \hat{\nu}_n) = 0$
- testing a given reference measure $\nu = \nu_0$
- giving confidence statements for $W_p(\hat{\mu}_n, \nu)$

$$\Downarrow$$

Limit distributions

Limit Theorems

The One Dimensional Case

In the case $\mathcal{X} = \mathbb{R}$ and $d(x, y) = |x - y|$ there is an **explicit** solution for the Wasserstein distance:

$$W_p(\mu, \nu) = \left\{ \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt \right\}^{1/p},$$

where F^{-1} and G^{-1} are the quantile functions of μ and ν, respectively.

$$\Downarrow$$

- for $\mu = \nu$ asymptotic behavior of $W_p(\hat{\mu}_n, \nu)$ boils down to analysis of **L_p-norm of quantile process**, limit distribution can be calculated **explicitly**, del Barrio et al. (1999, 2005), Samworth & Johnson (2004)
- for $\mu \neq \nu$ limit is a **centered normal** with variance depending on μ and ν which can be **consistently estimated** or **bootstrapped**, Munk & Czado (1998), Freitag et al. (2007), Berthe et al. (2017)

The General Case

- **explicit** solution known for elliptical distributions (e.g. multivariate normal)
  - ▷ $\mu \neq \nu$: **normal limit**, Rippl et al. (2016)
  - ▷ $\mu = \nu$: explicit description of limit distribution **difficult**, limit distribution can be **bootstrapped (m out of n)**, Rippl et al. (2016)
- for **general distributions** (positive densities on convex support, $4 + \delta$ moment) for $\mu \neq \nu$ and $p = 2$ the limit is **normal**, del Barrio & Loubes (2017)

In the case where $\mathcal{X}$ is a finite or countable space the Wasserstein distance is a finite/infinite dimensional **linear program**.

Theorem (Sommerfeld, Munk & T.)

For $\mu = \nu$ it holds for $n \to \infty$

$$n^{\frac{1}{2p}} W_p(\hat{\mu}_n, \mu) \xrightarrow{\mathscr{D}} \left\{ \max_{\lambda \in \mathcal{S}^*(\mu,\mu)} \langle G, \lambda \rangle \right\}^{\frac{1}{p}}$$

under the assumption that $\sum_{x\in\mathcal{X}} d^p(x, x_0)\sqrt{\mu_x} < \infty$ in the case that $\mathcal{X}$ is countable.

- $\mathcal{S}^*(\mu, \mu)$ set of dual solutions
- G Gaussian limit of $\sqrt{n}(\hat{\mu}_n - \mu)$

Discrete Spaces (cont.)

The limit distribution

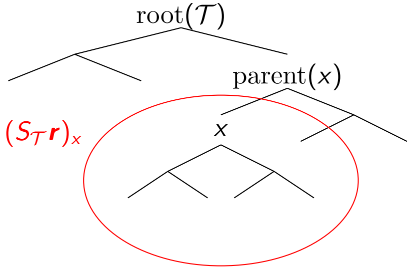
$$\left\{ \max_{\lambda \in \mathcal{S}^*(\mu,\mu)} \langle G, \lambda \rangle \right\}^{\frac{1}{p}}$$

- can **not** be calculated explicitly in general
- **computational very demanding** to sample from this limit distribution as for each realization a linear program has to be solved

Approximating the Limit Distribution by Trees

Explicit Limiting Distribution for Tree Metrics

$\mathcal{X}$ = vertices of a tree $\mathcal{T}$
$d_{\mathcal{T}}$ = path length in $\mathcal{T}$
x_0 = root of $\mathcal{T}$



Theorem (Sommerfeld, Munk & T.)

Again under the condition $\sum_{x\in\mathcal{X}} d^p(x,x_0)\sqrt{\mu_x} < \infty$ *for countable* $\mathcal{X}$

$$n^{\frac{1}{2p}} W_p(\hat{\mu}_n, \mu) \xrightarrow{\mathscr{D}} \left\{ \sum_{x\in\mathcal{X}} |(S_{\mathcal{T}} G)_x| d_{\mathcal{T}}(x, \mathrm{parent}(x))^p \right\}^{\frac{1}{p}} .$$

Distributional Bound for the Limiting Distribution

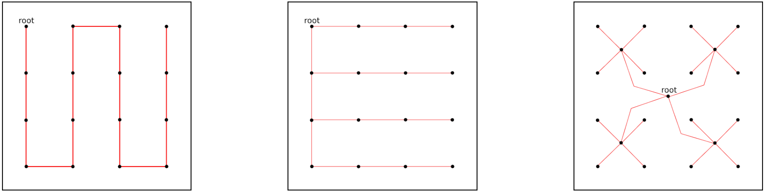
Upper bound:

Theorem (Sommerfeld, Munk & T.)

$$\limsup_{n\to\infty} P\left[n^{1/2p} W_p(\hat{\mu}_n, \mu) \geq z\right] \leq P\left[Z_{\mathcal{T},p}(G) \geq z\right]$$

- $Z_{\mathcal{T},p}(G) = \left\{\sum_{x\in\mathcal{X}} |(S_{\mathcal{T}} G)_x| d_{\mathcal{T}}(x, \mathrm{parent}(x))^p\right\}^{\frac{1}{p}}$
  $\rightarrow$ can be computed **explicitly**
- $\mathcal{T}$ **spanning tree** of $\mathcal{X}$, i.e., a rooted tree with elements of $\mathcal{X}$ as vertices and tree metric $d_{\mathcal{T}}$ given by the length of the unique path joining two elements
- G Gaussian limit as before

Spanning Trees



(a) 'Chain' (b)'Fork' (c) 'Dyadic Partition'

Figure 1: Three different spanning trees on a 4×4 grid. The black rectangle depicts the unit square $[0,1]^2$, the dots indicate the locations which represent the pixels, i.e., $\mathcal{X}$ (the vertices of the tree) and the red lines indicate the edges.

Probability Measures

We numerically investigate the upper bound

$$Z_{\mathcal{T},p}(G) = \left\{ \sum_{x \in \mathcal{X}} |(S_{\mathcal{T}} G)_x| d_{\mathcal{T}}(x, \mathrm{parent}(x))^p \right\}^{\frac{1}{p}}$$

for the following probability distributions:

- **uniform** distribution
- **random** measure, i.e., a generalization of the Dirichlet distribution
- tow **discretized bivariate Gaussians** with mean $(0.5, 0.5)$ and covariances

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$$

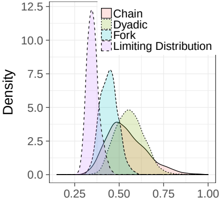
The Discretized Gaussians

(a) Gaussian 1

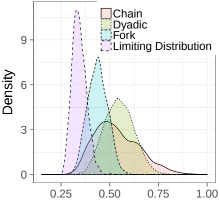
(c) Gaussian 2

Figure 2: Discretized Gaussians. The probability weights of two discretized Gaussians with mean $(0.5, 0.5)$ and different covariances.

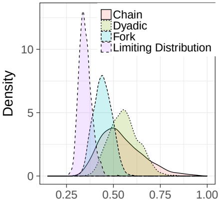
Results I: Dependency on Measure on a 8×8 Grid



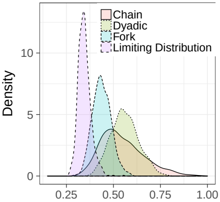
(a) Uniform



(b) Random

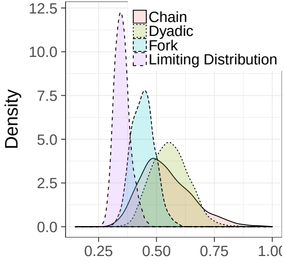


(c) Gaussian 1

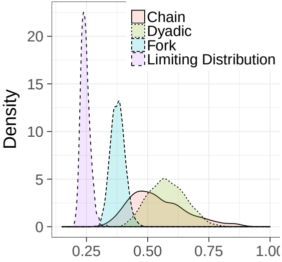


(d) Gaussian 2

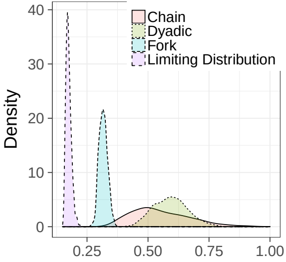
Results II: Dependency on Grid Size of Uniform Distribution



(a) 8×8 grid



(b) 16×16 grid



(c) 32×32 grid

Conclusion

- Review of limit results for the Wasserstein distance from a computational point of view
- Investigated upper bound for limit distribution on finite and countable spaces
  - The spanning tree **'Fork'** gives the **best approximation** no matter which measure
  - Approximation **better** for **small** grid sizes
- Future research: Optimize the approximation for a given number of nodes