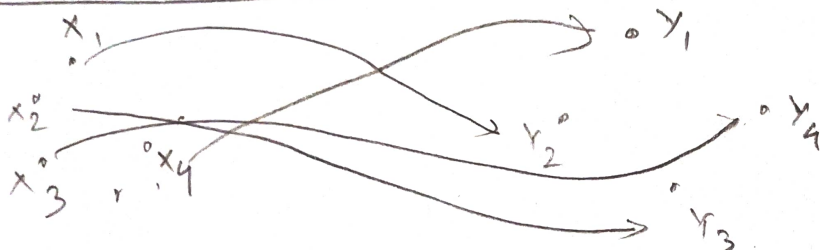


Day-3 Wh.

Primal Estimators



"Compute an insample matching?" — Hungarian alg. / Auction Alg.

② We want some extension of insample map

— Use (NN) nearest neighbor principle.

$$\hat{T}(x_{\text{new}}) = \hat{T}(x_{\text{nn}}(x_{\text{new}}))$$

— What stat properties does it have?

Plug-in Estimators

$x_1, \dots, x_n \sim \mu$ $y_1, \dots, y_n \sim \nu$
Construct density estimates of $\hat{\mu}, \hat{\nu}$

$$\hat{T} = OT \text{ Map}(\hat{\mu}, \hat{\nu})$$

What properties does it have?

Risk

$$R(\hat{T}, T_0) = \mathbb{E} \left[\int \| \hat{T}(x) - T_0(x) \|_2^2 d\mu \right] \sim L^2 \text{ measure}$$

Analysis of Stability

Brenier regularity

(i) $\nu = (\nabla \phi_0)_\# \mu$

(ii) ϕ_0 — strongly convex & smooth

$$\Rightarrow \underline{\lambda} Id \leq \nabla^2 \phi_0 \leq \lambda Id$$

The above $\wedge$ assumption holds in the following examples
(1) μ, ν are supported on Ω , $\mu, \nu > 0$.

Ω — connected, convex set with smooth boundaries

Then "Caffarelli regularity" theory will certify Brenier regularity.

(*) Monge-Ampere $\mu = \nu(\nabla \phi_0(x)) |\det \nabla^2 \phi_0(x)|$

(2) μ, ν are smoothly log-concave.

$$|\det \nabla^2 \phi_0| = f$$

Dual stability

$$S_{\mu, \nu}(\phi) = \int \phi d\mu + \int \phi^* d\nu$$

Lemma: Functional has quadratic growth for ϕ is smooth & strongly convex

$$\| \nabla \phi - \nabla \phi_0 \|_{L^2(\mu)}^2 \leq S_{\mu, \nu}(\phi) - S_{\mu, \nu}(\phi_0) \lesssim \| \nabla \phi - \nabla \phi_0 \|_{L^2(\mu)}^2$$

Statistical Analysis

$$\hat{\phi} = \underset{\phi \in \underline{\Phi}}{\operatorname{argmin}} S_{\mu_n, \nu_n}(\phi)$$

Transport map/potential selection

$$\rightarrow \underline{\Phi} \in \{\phi_1, \dots, \phi_M\} \text{ \& } \phi_0 \in \underline{\Phi}$$

$$\rightarrow \text{bounded domain } \|\phi_j\|_{\infty} \leq B, \|\phi_j^*\|_{\infty} \leq B$$

$\rightarrow$ All potential are smooth & strongly convex.

Basic inequality

$$S_{\mu_n, \nu_n}(\hat{\phi}) \leq S_{\mu_n, \nu_n}(\phi_0)$$

$$S_{\mu, \nu}(\hat{\phi}) - S_{\mu, \nu}(\phi_0)$$

$$= \frac{S_{\mu, \nu}(\hat{\phi}) - S_{\mu_n, \nu_n}(\hat{\phi}) + S_{\mu_n, \nu_n}(\hat{\phi}) - S_{\mu_n, \nu_n}(\phi_0) + S_{\mu_n, \nu_n}(\phi_0) - S_{\mu, \nu}(\phi_0)}{\leq 0}$$

$$\leq S_{\mu, \nu}(\hat{\phi}) - S_{\mu_n, \nu_n}(\hat{\phi}) + S_{\mu_n, \nu_n}(\phi_0) - S_{\mu, \nu}(\phi_0)$$

$$\leq 2 \max_{j \in \{1, \dots, M\}} |S_{\mu, \nu}(\phi_j) - S_{\mu_n, \nu_n}(\phi_j)|$$

We will show

$$\|\nabla \hat{\phi} - \nabla \phi_0\|_{L^2(\mathcal{H})}^2 \leq 2 \max_{j \in \{1, \dots, M\}} |S_{\mu, \nu}(\phi_j) - S_{\mu_n, \nu_n}(\phi_j)|$$

$$\int \phi d\mu + \int \phi^* d\nu - \int \phi d\mu_n - \int \phi^* d\nu_n$$

Using Hoeffding's inequality with prob at least $(1-\delta)$

$$\max_{j \in \{1, \dots, M\}} |S_{\mu, \nu}(\phi_j) - S_{\mu_n, \nu_n}(\phi_j)|$$

$$\leq 2B \sqrt{\frac{\log(M/\delta)}{2n}}$$

Proof of quadratic growth $\phi = \text{smooth \& strongly convex}$

$$S_{\mu, \nu}(\phi) - S_{\mu, \nu}(\phi_0)$$

$$= \int \phi d\mu + \int \phi^* d\nu - \int \phi_0 d\mu - \int \phi_0^* d\nu$$

Facts about Fenchel conjugate

$$\phi^*(y) = \sup_x [x^T y - \phi(x)]$$

$$\Rightarrow \phi(x) + \phi^*(y) \geq x^T y \text{ for any } (x, y)$$

If $y = \nabla \phi(x)$ then equality holds.

$$\phi(x) + \phi^*(\nabla \phi(x)) = x^T \nabla \phi(x)$$

$$(\nabla \phi_0)_\# \mu = \nu$$

$$\int \phi_0^* d\nu = \int \phi_0^*(\nabla \phi_0(x)) d\mu(x)$$

$$S_{\mu, \nu}(\phi_0) = \int \phi_0 d\mu + \int \phi_0^* d\nu$$

$$= \int [\phi_0(x) + \phi_0^*(\nabla \phi_0(x))] d\mu$$

$$= \int \langle x, \nabla \phi_0(x) \rangle d\mu$$

Thus

$$S_{\mu, \nu}(\phi) - S_{\mu, \nu}(\phi_0)$$

$$= \int \phi d\mu + \int \phi^* d\nu - \int \langle x, \nabla \phi_0(x) \rangle d\mu \quad \leftarrow$$

$$= \int [\phi^*(\nabla \phi(x)) + \phi(x) - \langle x, \nabla \phi_0(x) \rangle] d\mu \quad (\text{Fenchel equality})$$

$$= \int [\phi^*(\nabla \phi_0(x)) - \phi^*(\nabla \phi(x))] d\mu$$

$$= \int \langle x, \nabla \phi_0(x) - \nabla \phi(x) \rangle d\mu$$

$$x = \nabla(\phi^*(\nabla \phi(x))) \text{ putting ave.}$$

$$= \int d\phi^*(\nabla \phi_0, \nabla \phi) d\mu$$

Using strong convexity & smoothness

$$\|\nabla\phi_0 - \nabla\phi\|_2^2 \leq d_{\phi^*}(\nabla\phi_0, \nabla\phi) \lesssim \|\nabla\phi_0 - \nabla\phi\|_2^2$$

Primal Stability (plugin Estimator)

$$\hat{T} = OT(\hat{\mu}, \hat{\nu}) \quad T_0 = OT(\mu, \nu)$$

① When is $\hat{T} \approx T_0$? ② maybe $\hat{\mu} \approx \mu$, $\hat{\nu} \approx \nu$ in what distance?

Lemma: μ, ν $\nu = (\nabla\phi_0)_\# \mu$, ϕ_0 is smooth & strongly convex

then for any $\hat{\mu}, \hat{\nu}$, we can construct $\hat{T}$ such that

$$\int \|\hat{T}(x) - T_0(x)\|^2 d\hat{\mu} \lesssim W_2^2(\hat{\mu}, \mu) + W_2^2(\hat{\nu}, \nu)$$

OT Coupling $\hat{\delta}$, Barycentric projection of $\hat{\delta}$

$$\hat{T}(x) = \mathbb{E}_{(X,Y) \sim \hat{\delta}} [Y | X]$$

Wasserstein density estimation

$$W_2^2(\hat{\mu}, \mu) / W_1(\mu, \mu) \leftarrow \text{small}$$

Day-4

$\rightarrow X_1, \dots, X_n \sim \mu$ construct $\hat{\mu}$ so that $W_p(\hat{\mu}, \mu)$ is small

$\rightarrow \mu_n$ is already consistent.

$\rightarrow$ Is $\hat{\mu} = \mu_n$ best possible?

Theorem $X_1, \dots, X_n \sim \mu$ supported on $[0,1]^d$.

$$\mathbb{E}[W_p(\mu_n, \mu)] \leq \begin{cases} n^{-1/2p} & d < 2p \\ (\log n)^{1/2p} & d = 2p \\ n^{-1/d} & d \geq 2p \end{cases}$$

$$p=1 \quad \mathbb{E}[W_1(\mu_n, \mu)] \leq \begin{cases} 1/\sqrt{n} & d=1 \\ \sqrt{\frac{\log n}{n}} & d=2 \\ n^{-1/d} & d \geq 2 \end{cases} \rightarrow \text{Curse of dimensionality.}$$

(Intrinsic dimension variants)

→ can replace the dimension d by d^* "intrinsic dimension" of μ .

$$\limsup_{\varepsilon \rightarrow 0} \frac{\log N_{\varepsilon}(\text{supp}(\mu))}{-\log \varepsilon} = d^*$$

→ Covering no.

Smooth case μ s -Hölder, $\Rightarrow \|f^s\|_{\infty} \leq L$

$$\mathbb{E} N_{\varepsilon}(\hat{\mu}, \mu) \lesssim \begin{cases} n^{-1/2} & d=1 \\ \sqrt{\frac{\log n}{n}} & d=2 \\ n^{-\frac{s+1}{2sd}} & d>2 \end{cases}$$

"Usual non-parametric rate" $\sim n^{-\frac{s}{2sd}}$.

→ empirical measure can't take advantage of smoothness.

→ Kernel/Wavelet estimator

$$\hat{\mu}_n = \mu_n * k_n$$

$$\Rightarrow \hat{\mu}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\|x - x_i\|}{h}\right)$$

Implications for transport map est.

$$1. \|\hat{T} - T_0\|_{L^2(\mu)}^2 \lesssim n^{-2(s+1)/2sd} \rightarrow \hat{T} = \text{OTMap}(\hat{\mu}, \hat{\nu})$$

$$2. \|\hat{T}_{NN} - T_0\|_{L^2(\mu)}^2 \lesssim n^{-2/d} \text{ if } \mu, \nu \text{ have bounded moments.}$$

$\hat{\mu}$ by kernel density estimator

Focus on W_1

$$W_1(\mu_n, \mu) = \sup_{f \in 1\text{-Lip}} \int f d\mu - \int f d\mu_n \quad (\text{dual formulation})$$
$$= \sup_{f \in 1\text{-Lip}} \left[\mathbb{E}[f] - \frac{1}{n} \sum_{i=1}^n f(x_i) \right]$$