Recap: $\underline{r_n}(\hat{T} - T_0) \xrightarrow{d} \sim\varphi$

Statistical OT setting: given $\left.\begin{array}{l} X_1, \dots X_n \sim \mu \\ Y_1, \dots Y_n \sim \nu \end{array}\right]$

Estimate: 1. OT map T_0 ✓

2. $W_p(\mu, \nu)$

$\star$ 3. Construct $\hat{\mu}$, so that $W_p(\hat{\mu}, \mu)$ is small.

$\left[\begin{array}{l} 4. \quad \bar{\mu} = \underset{\mu}{\operatorname{argmin}} \sum_{j=1}^{k} W_2(\mu, \mu_j) \\ \quad \vdots \end{array}\right]$

→ Talked about estimating T_0.

Dual estimators

$$\underline{\underline{\hat{\varphi}}} = \left[\underset{\varphi \in \Phi}{\arg\min} \int \varphi \, d\mu_n + \int \varphi^* d\nu_n\right]$$

$$\underbrace{\hat{T} = \nabla \hat{\varphi}}$$

Primal / plugin estimators

→ Construct $\hat{\mu}, \hat{\nu}$

$\underline{\hat{T} = \text{OTMap}(\hat{\mu}, \hat{\nu})}$

→ Use $\underline{\mu_n}, \underline{\nu_n}$

→ $T_n = \text{OT Map}(\mu_n, \nu_n)$

Extend T_n to obtain

$\underbrace{\hat{T} \text{ by 1-NN principle.}}$

Risk: $\underbrace{R(\hat{T}, T_0)} = \int \|\hat{T}(x) - T_0(x)\|_2^2 \, d\mu(x)$

Analysis: Dual estimators:

$$\rightarrow S(\varphi) = \int \varphi \, d\mu + \int \varphi^* \, d\nu . \leftarrow$$

We showed $S(\varphi)$ grows quadratically around minimizer φ_0 if φ is smooth & strongly convex (i.e.)

$$\|\nabla\varphi - \nabla\varphi_0\|^2_{L^2(\mu)} \lesssim S(\varphi) - S(\varphi_0) \lesssim \|\nabla\varphi - \nabla\varphi_0\|^2_{L^2(\mu)}$$

Immediately implies that

$$\left[\hat{\varphi} = \underset{\varphi \in \Phi}{\arg\min} \int \varphi \, d\mu_n + \int \varphi^* \, d\nu_n . \right]$$

is a reasonable idea. With prob at least $1-\delta$.

$$\|\nabla\hat{\varphi} - \nabla\varphi_0\|^2_{L^2(\mu)} \lesssim \underbrace{\sqrt{\frac{\log(M/\delta)}{n}}}$$

Plugin estimators

<u>Lemma:</u> $\underline{(\nabla\varphi_0)_{\#}\mu = \nu}$ $\underbrace{\varphi_0}$ is smooth & strongly convex. Then for any $\underline{\underline{\hat{\mu}, \hat{\nu}}}$ we can construct $\underline{\underline{\hat{T}}}$

$$\underbrace{\|\hat{T} - T_0\|^2_{L^2(\hat{\mu})}} \lesssim \underbrace{W_2^2(\hat{\mu}, \mu)} + \underbrace{W_2^2(\hat{\nu}, \nu)}.$$

$$X_1, \dots X_n \sim \mu.$$

→ construct $\hat{\mu}$ so that $W_p(\hat{\mu}, \mu)$ is small.

→ $\underbrace{\mu_n}$ is already consistent ....

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

<u>Theorem:</u> $X_1, \dots X_n \sim \mu$ supported on $[0,1]^d$

$$\mathbb{E} W_p(\mu_n, \mu) \lesssim \begin{cases} n^{-1/2p} & d < 2p \\ \left(\frac{\log n}{n}\right)^{1/2p} & d = 2p \\ n^{-1/d} & d > 2p \end{cases}$$

$$p=1,\ \mathbb{E}W_1(\mu_n,\mu) \lesssim \begin{cases} 1/\sqrt{n} & d=1 \\ \sqrt{\frac{\log n}{n}} & d=2 \\ n^{-1/d} & d>2 \end{cases}$$

$\rightarrow$ "curse of dimensionality".

"Intrinsic dimension variants"

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

$$\limsup_{\epsilon \to 0} \frac{\log N_\epsilon(\operatorname{supp}(\mu))}{-\log \epsilon} = d^* .$$

Smooth case: μ s-Hölder.

$\|\mu^s\|_\infty \lesssim L$

$$\mathbb{E}W_1(\hat{\mu},\mu) \lesssim \begin{cases} n^{-1/2} & d=1 \\ \sqrt{\frac{\log n}{n}} & d=2 \\ n^{-\frac{s+1}{2s+d}} & d>2 \end{cases}$$

$\rightarrow$ "usual non-parametric rate" $\sim n^{-\frac{s}{2s+d}}$.

→ empirical measure cannot take advantage of smoothness

→ kernel / wavelet estimator

(higher-order kernel)

$$\hat{\mu} = \mu_n * K_h.$$

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

Implications for transport map est.

1. $\|\hat{T}_{\mathrm{rw}} - T_0\|^2_{L^2(\mu)} \lesssim n^{-\frac{2(s+1)}{2s+d}}$.

$\left\{ \hat{T} = \mathrm{OTMap}(\hat{\mu}, \hat{\nu}) \right\}$. if μ, ν are smooth.

2. $\|\hat{T}_{NN} - T_0\|^2_{L^2(\mu)} \lesssim \underbrace{n^{-2/d}}$.

if μ, ν have bdd moments.

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

$\tilde{X}_1, \ldots, \tilde{X}_N \sim \hat{\mu}$.

Focus on W_1:

$$W_1(\mu_n, \mu) = \sup_{f \in 1\text{-Lip}} \int f \, d\mu - \int f \, d\mu_n .$$

$$= \sup_{f \in 1\text{-Lip}} \left[\underbrace{\mathbb{E} f - \frac{1}{n} \sum_{i=1}^{n} f(X_i)} \right].$$

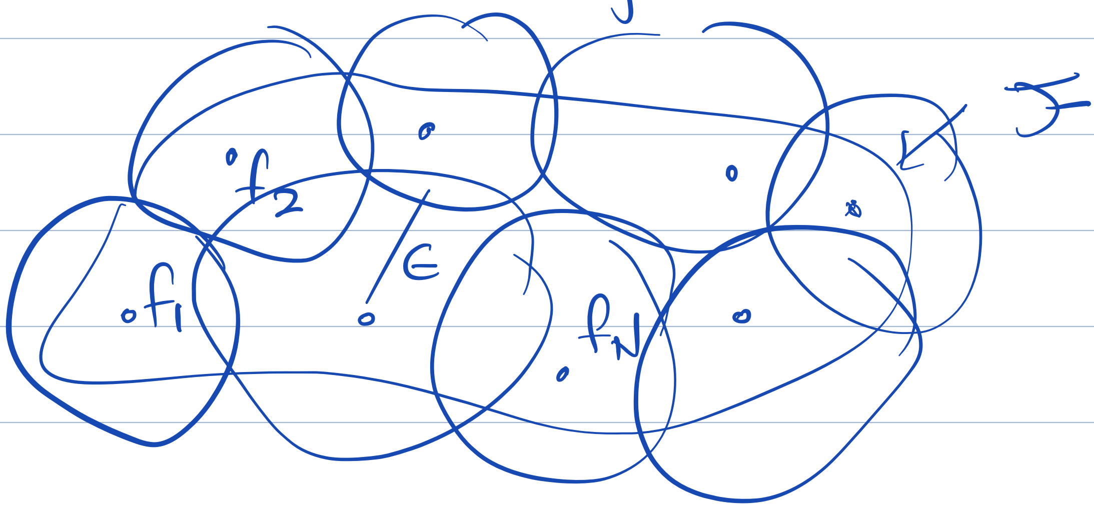
Dudley's chaining bound: $\mathcal{F}$, $\|f\|_\infty \leq R$

$$\underbrace{\mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E} f\right]} \lesssim \inf_{\tau > 0} \left[\tau + \frac{1}{\sqrt{n}} \int_{\tau}^{R} \sqrt{\log N(\epsilon, \mathcal{F})} \, d\epsilon \right]$$

$$\underline{N(\epsilon, \mathcal{F})} = \min \left\{ |F| : F \text{ forms an } \epsilon\text{-cover} \right\}.$$

$$F = \{ f_1, \dots, f_N \}.$$

for any $f \in \mathcal{F}$, $\min_j \|f - f_j\|_\infty \leq \epsilon$



$\to \mu$ supported on unit cube $[0,1]^d$

f - 0 at $(0,0,\dots 0)$.

$\|f\|_\infty \lesssim \underline{\underline{R}}$.

<u>Lemma</u>: $\log \underline{N(\epsilon, \mathcal{F})} \lesssim \left(\frac{1}{\epsilon}\right)^d \leftarrow$

<u>Do computation</u>:

$d=1$, $\tau = 0$

$$\mathbb{E} W_1(\mu_n, \mu) \lesssim \frac{1}{\sqrt{n}} \underbrace{\int_0^R \frac{d\epsilon}{\sqrt{\epsilon}}} \lesssim \frac{1}{\sqrt{n}}.$$

$d=2$, $\tau = n^{-1/d}$.

$$\mathbb{E} W_1(\mu_n, \mu) \lesssim \underline{\underline{n^{-1/d}}} + \frac{1}{\sqrt{n}} \underbrace{\int_{n^{-1/d}}^R \frac{d\epsilon}{\epsilon}}$$

AKT

$$\lesssim \frac{\log n}{\sqrt{n}}. \quad \Bigg\} \quad \frac{\sqrt{\log n}}{\sqrt{n}}.$$

$d \geq 2$ $\tau = n^{-1/d}$

$$\mathbb{E} W_1(\mu_n, \mu) \lesssim n^{-1/d} + \frac{1}{\sqrt{n}} \int_{n^{-1/d}}^R \frac{d\epsilon}{\epsilon^{d/2}} \Big]$$

$$\lesssim n^{-1/d} + \frac{1}{\sqrt{n}} \cdot \frac{1}{n^{-1/2+1/d}}$$

$$\lesssim \underline{\underline{n^{-1/d}}}.$$

Lower bound for empirical measure:

$\mu \sim$ uniform dist. on $[0,1]^d$.

$\underline{W_1(\mu_n, \mu)} \gtrsim n^{-1/d}$. $d > 2$.

$$\geq \mathbb{E} f - \underbrace{\frac{1}{n}\sum_{i=1}^{n} f(x_i)}.$$

$\rightarrow f(x) = \|x - nn(x)\|$

$= \min_j \underbrace{\|x - X_j\|}_{\rightarrow \text{1-Lipschitz.}}$

$W_1(\mu_n, \mu) \geq \mathbb{E}f \gtrsim n^{-1/d}$.

$\mathbb{E}$ "nearest neighbor dist"

$$W_1(\mu_n, \mu) = \sup_{f \in 1\text{-Lip}} \mathbb{E} f - \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

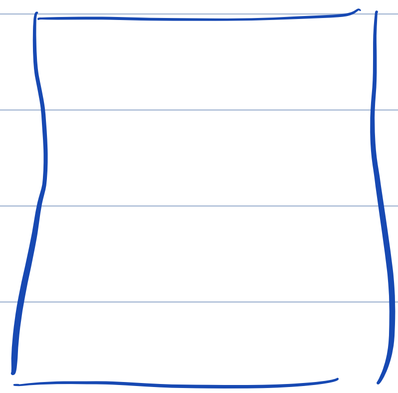
$$\geq \mathbb{E}\tilde{f} - \frac{1}{n}\sum_{i=1}^{n} \tilde{f}(X_i) \qquad \tilde{f} \text{ is } 1\text{-Lip}.$$

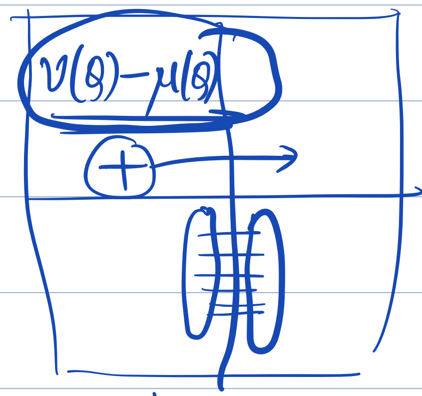
Dyadic partitioning upper bounds:

Philosophy: $W_p(\mu_n, \mu) \lesssim \epsilon$.

Sufficient to construct a coupling between μ_n & μ to certify an upper bound.

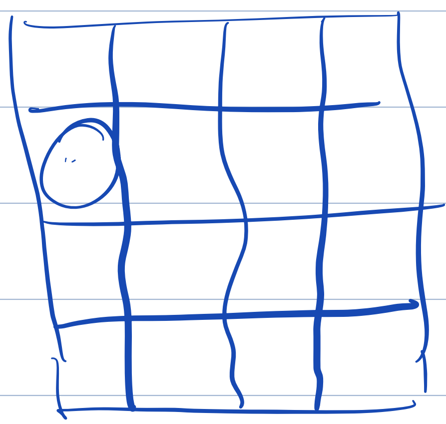
Lemma:





$k=1$

$\underline{\mathcal{Q}^k}$



K

$k=2$

$$\to W_p^p(\mu,\nu) \lesssim 2^{-Kp} + \sum_{k=1}^{K} 2^{-kp} \sum_{Q \in \mathcal{Q}^k} |\mu(Q) - \nu(Q)| .$$

$\mu_n \qquad \mu.$ By Bernstein.

$d>2, \ \mu$ unif, $\quad |\mu_n(Q)-\mu(Q)| \lesssim \sqrt{\frac{\mu(Q)}{n}}.$

$$W_1(\mu_n,\mu) \lesssim \underbrace{2^{-K}} + \sum_{k=1}^{K} 2^{-k} \times \underbrace{\sqrt{\frac{2^{-kd}}{n}}} \times \frac{1}{2^{-kd}}$$

$$K = \frac{\log n}{d} \qquad \frac{\log n}{\sqrt{n}}.$$

$$\lesssim n^{-1/d}$$

$$\frac{2^{K\left(\frac{d}{2}-1\right)}}{\sqrt{n}}$$

$$\mathbb{P}\left(W_2^2(\mu,\nu) \in C_\alpha \right) \geq 1-\alpha .$$

$$r_n\left(\hat{W}_2^2 - W_2^2(\mu,\nu)\right) \xrightarrow{d} N(0,1)$$

$$\lim_{n\to\infty} \mathbb{P}\left(W_2^2(\mu,\nu) \in \left[\hat{W} \pm \frac{z_{\alpha/2}}{r_n} \right] \right) \geq 1-\alpha .$$

IPM :

$$\rho(\mu,\nu) = \sup_{f\in\mathcal{F}} \left| \mathbb{E}_\mu f - \mathbb{E}_\nu f \right| .$$