Wasserstein density estimation

$W_2^2(\hat{\mu}, \mu) \;/\; W_1(\mu, \mu)$ ← small

Day-4

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

→ μ_n is already consistent.

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

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

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

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

("Intrinsic dimension variants")

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

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

(→ Covering no.)

Smooth case μ s-Hölder. $\Rightarrow \|\mu^s\|_\infty \le L$

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

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

→ emperical measure can't take advantage of smoothness.

→ Kernel/Wavelet estimator

↳ $\hat{\mu} = \mu_n * k_n$

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

Implications for transport map est.

1. $\|\hat{T} - T_0\|^2_{L^2(\mu)} \lesssim n^{-2(s+1)/2s+d} \to \hat{T} = \mathrm{OTMap}(\hat{\mu}, \hat{\nu})$
2. $\|\hat{T}_{NN} - T_0\|^2_{L^2(\mu)} \lesssim n^{-2/d}$ if μ, ν have bounded moments.

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

Focus on W_1

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

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

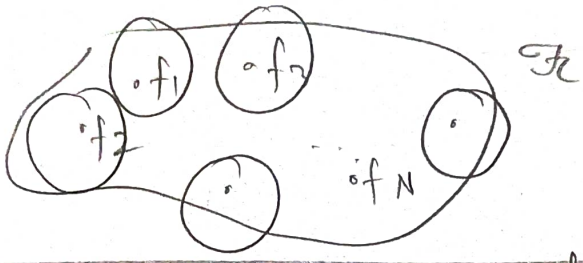
Dudley's Chaining bound

Collection of function $\mathcal{F}$, $\|f\|_\infty \le R$

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

$N(\varepsilon,\mathcal{F}) = \min\{|F| : F \text{ form an } \varepsilon \text{ cover}\}$

$F = \{f_1 \cdots f_N\}$ for any $f\in\mathcal{F}$. $\min_j \|f - f_j\|_\infty \le \varepsilon$



H is supported on a unit cube. $[0,1]^d$

$f = 0$ at $(0,0\cdots 0)$

$\|f\|_\infty \le R$

Lemma $\log N(\varepsilon,\mathcal{F}) \lesssim \left(\frac{1}{\varepsilon}\right)^d$

Do computation $d=1$ $\delta=0$

$$\mathbb{E}[W_1(H_n,H)] \lesssim \int_0^R \frac{d\varepsilon}{\sqrt{\varepsilon}} \lesssim \frac{1}{\sqrt{n}}$$

$d=2$

$$\mathbb{E}\, W_1(H_n,H) \le n^{-1/d} + \frac{1}{\sqrt{n}}\int_{n^{-1/d}}^{R}\frac{d\varepsilon}{\varepsilon}$$

$$\lesssim \frac{\log n}{\sqrt{n}}$$

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

$$\mathbb{E}\, W_1(H_n,H) \lesssim n^{-1/d} + \frac{1}{\sqrt{n}}\int_{n^{-1/d}}^{R}\frac{d\varepsilon}{\varepsilon^{d/2}}$$

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

Lower Bound for empirical measure

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

$W_1(\mu_n, \mu) \geq n^{-1/d} \quad d > 2$

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

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

$= \min_j \|x - x_j\|$ ~~~> 1-Lipschitz

$=$ E "nearest neighbor distance"

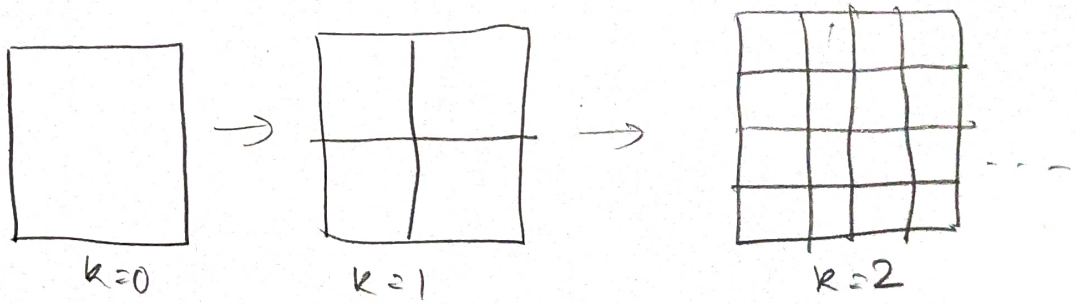
$W_1(\mu_n, \mu) \geq \mathbb{E}[f] \gtrsim n^{-1/d}$

Dyadic Partition upper bound

Philosophy: $W_p(\mu_n, \mu) \leq \varepsilon$

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

→ Lemma:



Q_k = collection of cubes at level k

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