Lecture 1 | Rates of estimation I

Let $(\mathcal{X}, d)$ be metric space, μ a prob. measure on $\mathcal{X}$.

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \quad , \ X_i \overset{iid}{\sim} \mu \qquad \text{empirical measure}$$

View μ_n as random element of $\mathcal{P}_p(\mathcal{X})$. How does it behave?

N.B. $\sup_{\nu \in \mathcal{P}_p(\mathcal{X})} |W_p(\mu_n, \nu) - W_p(\mu, \nu)| = \underbrace{W_p(\mu_n, \mu)}_{\text{object of study}}$

Can show $W_p(\mu_n, \mu) \xrightarrow{a.s.} 0$ (HW). Rates?

Two strategies:

1. exhibit coupling w/ small cost ("primal")
2. use dual problem ← surprisingly powerful!

Warm up: finite metric spaces

Prop: Let $(\mathcal{X}, d)$ be finite: $|\mathcal{X}| = K$, $\max d(x_i, x_j) \le D$.
Let $\mu \in \mathcal{P}(\mathcal{X})$ be arbitrary. Then

$$\mathbb{E}\, W_p(\mu_n, \mu) \le \frac{D}{2^{1/p}} \left(\frac{K}{n}\right)^{\frac{1}{2p}} \asymp n^{-1/2p}$$

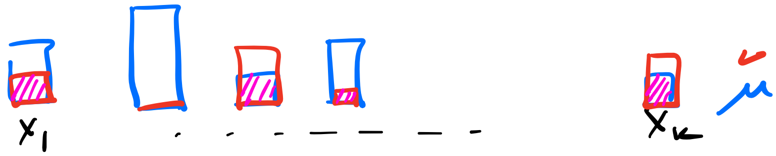
We will show that for any $\mu, \nu \in \mathcal{P}(\mathcal{X})$,

$$W_p^p(\mu, \nu) \le \frac{D^p}{2} \sum_{i=1}^{K} |\underbrace{\mu_i}_{\mu(\{x_i\})} - \underbrace{\nu_i}_{\nu(\{x_i\})}| \qquad (*)$$

To conclude, use $\left(\mathbb{E}\,|(\hat{\mu}_n)_i - \mu_i|\right)^2 \le \mathbb{E}\,((\hat{\mu}_n)_i - \mu_i)^2 = \frac{\mu_i(1-\mu_i)}{n}$.

So $\mathbb{E}\, W_p^p(\mu,\hat{\mu}) \le \frac{D^p}{2} \sum_{i=1}^{k} \sqrt{\frac{\mu_i}{n}} \le \frac{D^p}{2}\sqrt{\frac{k}{n}}$.

Pf. of $(*)$: "Primal version"



leave shared mass fixed, distribute rest independently

Let $\rho = \mu \wedge \nu$, mass $\lambda \in [0,1]$. If $\lambda = 1$, $\mu = \nu$ and bound holds. Otherwise set

$$\gamma(x_i, x_j) = \rho_i \mathbb{1}_{i=j} + (1-\lambda)^{-1}(\mu_i - \rho_i)\cdot(\nu_j - \rho_j)$$

Claim: $\gamma \in \Gamma(\mu,\nu)$. Clearly $\gamma \ge 0$

Compute:

$$\begin{aligned}
\gamma(x_i \times \mathcal{X}) &= \sum_j \gamma(x_i, x_j) \\
&= \rho_i + (1-\lambda)^{-1}(\mu_i - \rho_i)\sum_{j=1}^{k}(\nu_j - \rho_j) \\
&= \mu_i.
\end{aligned}$$

So left marginal is μ, same computation shows right marginal is ν.

So

$$\begin{aligned}
W_p^p(\mu,\nu) &\le \sum_{i,j} d^p(x_i,x_j)\,\gamma(x_i,x_j) \\
&= D^p \quad (1-\lambda)^{-1} \sum_{i,j}(\mu_i - \rho_i)(\nu_j - \rho_j) \\
&= D^p(1-\lambda).
\end{aligned}$$

But $1-\lambda = \frac{1}{2}\left(\sum_i (\mu_i - \rho_i) + (\nu_i - \rho_i)\right)$

$$= \frac{1}{2}\left(\sum_i \mu_i + \nu_i - 2\,\mu_i \wedge \nu_i\right)$$

$$= \frac{1}{2}\sum_i |\mu_i - \nu_i|.$$

$\square$

<u>Pf</u> of $(*)$ "dual version":

Duality says

$$W_p^p(\mu,\nu) = \sup_{\substack{f,g \in L_1 \\ f(x)+g(y) \le d^p(x,y)}} \int f\,d\mu + \int g\,d\nu$$

finite space $\rightsquigarrow$ $$= \sup_{\mathcal{D} =: \left\{\substack{f,g \in \mathbb{R}^{|X|} \\ f_i + g_j \le d^p(x_i,x_j)}\right.} \langle f,\mu\rangle + \langle g,\nu\rangle$$

Since $d(x_i,x_i)=0$, we also have $g_i \le -f_i \;\; \forall i$, so

$\langle g,\nu\rangle \le -\langle f,\nu\rangle$, and

$$W_p^p(\mu,\nu) \le \sup_{(f,g)\in\mathcal{D}} \langle f, \mu-\nu\rangle$$

($= +\infty$ unless $\mu=\nu$, we've lost too much)

Idea: Choose $\bar{\mathcal{D}} \subseteq \mathcal{D}$ s.t. we can guarantee

$$\sup_{\mathcal{D}} \langle f,\mu\rangle + \langle g,\nu\rangle = \sup_{\bar{\mathcal{D}}} \langle f,\mu\rangle + \langle g,\nu\rangle$$

Then same proof shows

$$W_p^p(\mu,\nu) \leq \sup_{(f,g)\in\bar{\mathcal{D}}} \langle f, \mu-\nu\rangle$$

How to pick $\bar{\mathcal{D}}$?

Recall that we can always assume $f_i = \min_j d^p(x_i,x_j) - g_j$.

$g_j = \min_i d^p(x_i,x_j) - f_i$

(Otherwise, improve!)

Also, by shifting, we can assume

$$\max_i f_i = D^p/2$$

So

$$g_j = \min_i d^p(x_i,x_j) - f_i \leq D^p - D^p/2 = D^p/2.$$

$$f_i = \min_j d^p(x_i,x_j) - g_j \geq -D^p/2. \implies \|f\|_\infty \leq D^p/2.$$

So we can choose

$$\bar{\mathcal{D}} = \mathcal{D} \cap \{ f : \|f\|_\infty \leq D^p/2 \}.$$

Hence

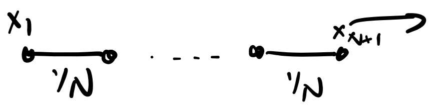
$$W_p^p(\mu,\nu) \leq \sup_{(f,g)\in\bar{\mathcal{D}}} \langle f, \mu-\nu\rangle$$

$$\leq \sup_{f:\|f\|_\infty \leq D^p/2} \langle f, \mu-\nu\rangle = \frac{D^p}{2}\|\mu-\nu\|_1. \qquad \square$$

Message: $\mathbb{E}\, W_p(\mu, \mu_n) \lesssim n^{-1/2p}$ if $|\mathcal{X}| < \infty$

Sharp? Yes (HW) $\to$ no general improvement possible. But not always tight!

Important example: $\mathcal{X}$ = graph metric on P_N



path w/ $N+1$ vertices
edges weighted by $1/N$, so $D = 1$.

Suppose μ is uniform measure ($1/N+1$ on each atom).

Our naive bound gives

$$\mathbb{E}\, W_1(\mu, \mu_n) \lesssim \tfrac{1}{2}\, \mathbb{E}\, \|\mu - \mu_n\|_1$$

If $N \gg n$, then $\mathbb{E}\, \|\mu - \mu_n\|_1 \approx 2$. So we predict μ, μ_n far in W_1.

Prop: For μ uniform measure on P_N,

$$\mathbb{E}\, W_1(\mu, \mu_n) \lesssim n^{-1/2}.$$

independent of N.

Once again, based on deterministic bound:

$$W_1(\mu, \nu) \leq \frac{1}{N} \sum_{i=1}^{N} \Big| \sum_{j=1}^{i} \mu_j - \nu_j \Big|$$

(actually, an equality!)

If $\nu = \mu_n$, then

$$\mathbb{E}\left|\sum_{j=1}^{i} \mu_j - (\mu_n)_j\right| \lesssim \sqrt{\frac{i/(N+1)}{n}}$$

So

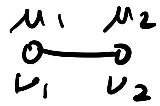
$$\mathbb{E}\, W_1(\mu,\hat{\mu}) \lesssim N^{-3/2} \underbrace{\sum_{i=1}^{N} \sqrt{i}}_{\lesssim \int_0^{N+1} \sqrt{x}\, dx} \cdot n^{-1/2} \lesssim n^{-1/2}$$

<u>Primal proof</u>:

For convenience, prove "unweighted" version, w/o $1/N$ factor.

We show the existence of a suitable coupling by induction.

$N=1$ case:

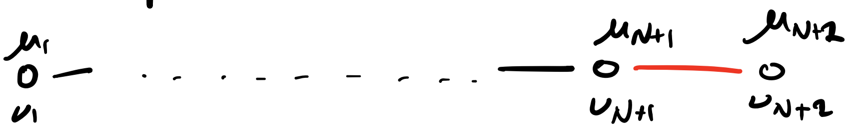


if , send excess mass to ν_2.

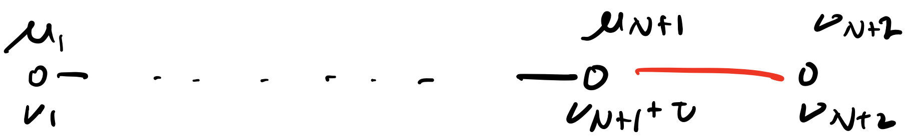
Otherwise, $\mu_2 > \nu_2$, send excess mass to ν_1.

Total cost: $|\mu_1 - \nu_1|$.

Induction step:



Suppose $\mu_{N+2} - \nu_{N+2} =: \tau \geq 0$. Then send extra mass along red edge to get



By induction, can find coupling of $N+1$ prefix w/ cost

$$\sum_{i=1}^{N} \left| \sum_{j=1}^{i} \mu_j - \nu_j \right|$$

So total cost is $\displaystyle\sum_{i=1}^{N} \left| \sum_{j=1}^{i} \mu_j - \nu_j \right| + \tau = \sum_{i=1}^{N+1} \left| \sum_{j=1}^{i} \mu_j - \nu_j \right|.$

If $\mu_{N+2} - \nu_{N+2} < 0$, send mass from ν_{N+2} to μ_{N+1} instead.

Dual proof:

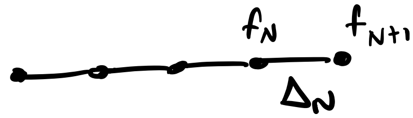
Recall that

$$W_1(\mu,\nu) = \sup_{f \text{ lip.}} \int f \, d\mu - d\nu$$

$$= \sup_{\substack{f \in \mathbb{R}^{N+1} \\ f_i - f_j \le \frac{|i-j|}{N}}} \langle f, \mu - \nu \rangle .$$

Given $f \in \mathbb{R}^{N+1}$, let $\Delta \in \mathbb{R}^N$ be given by

$$\Delta_i = f_{i+1} - f_i$$



Then

$$\langle f, \mu - \nu \rangle = \sum_{j=1}^{N+1} f_j (\mu_j - \nu_j)$$

$$= \sum_{j=1}^{N+1} \left(f_{N+1} - \sum_{i=j}^{N} \Delta_i \right) (\mu_j - \nu_j)$$

$$= f_{N+1} \cdot \sum_{j=1}^{N+1} (\mu_j - \nu_j) - \sum_{i=1}^{N} \Delta_i \left(\sum_{j=1}^{i} \mu_j - \nu_j \right)$$

"summation by parts"

$$= \sum_{i=1}^{N} (-\Delta_i) \left(\sum_{j=1}^{i} \mu_j - \nu_j \right)$$

if $f \in \mathrm{Lip}$, then $|\Delta_i| \leq 1/N$. So

$$\sup_{f \in \mathrm{Lip}} \langle f, \mu - \nu \rangle \leq \sup_{\substack{\Delta \in \mathbb{R}^N \\ \|\Delta\|_\infty \leq 1/N}} \sum_{i=1}^{n} (-\Delta_i) \left(\sum_{j=1}^{i} \mu_j - \nu_j \right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left| \sum_{j=1}^{i} \mu_j - \nu_j \right| .$$

(And actually this is achievable by choosing $\Delta_i = -\mathrm{sign}(\sum_j \mu_j - \nu_j)$).

$\square$

Note that in $N \to \infty$ limit, suggests that if μ, ν are measures on $[0,1]$, then

$$W_1(\mu, \nu) = \int_0^1 \left| \int_0^t d\mu - d\nu \right| dt$$

and

$$\mathbb{E}\, W_1(\mu, \mu_n) \lesssim n^{-1/2}. \qquad \text{(HW)}$$

Moral: structure of metric space matters a lot!