Lecture 4 | Inference

So far — close analysis of <u>expected</u> behavior of (regularized) OT distances.

This lecture — finer behavior

Warm up: entropic OT

We know

$$\mathbb{E}\,|S(\mu_n,\nu) - S(\mu,\nu)| \lesssim n^{-1/2}$$

Does

$$\sqrt{n}\,\big(S(\mu_n,\nu) - S(\mu,\nu)\big) \xrightarrow{d} \text{non-degenerate R.V.}$$

Motivation: confidence sets

Suppose

$$\sqrt{n}\big(S(\mu_n,\nu) - S(\mu,\nu)\big) \xrightarrow{d} W$$

If A_α is open set such that $\mathbb{P}\{W \in A_\alpha\} \geq 1-\alpha$, then

$$\liminf_{n\to\infty} \mathbb{P}\{S(\mu,\nu) \in \underbrace{(-A_\alpha + S(\mu_n,\nu))/\sqrt{n}}_{\text{confidence set for } S(\mu,\nu)}\} \geq 1-\alpha$$

Theorem: If μ,ν are compactly supported,

$$\sqrt{n}\,\big(S(\mu_n,\nu) - S(\mu,\nu)\big) \xrightarrow{d} N(0,\sigma^2)$$

$$\sigma^2 = \mathrm{Var}_\mu(\underbrace{f^*}_{\text{optimal dual solution}})$$

Pf:

$$S(\mu_n,\nu)-S(\mu,\nu) = \sup_{f',g'} \int f'\, d\mu_n + \int g'\, d\nu - \varepsilon \int e^{\frac{f'\oplus g'-c}{\varepsilon}}\, d\mu_n d\nu$$

$$- \sup_{f,g} \int f\, d\mu + \int g\, d\nu - \varepsilon \int e^{\frac{f\oplus g-c}{\varepsilon}}\, d\mu d\nu$$

$$\geq \int f^*\, d(\mu_n-\mu)$$

$$\leq \int f_n\, d(\mu_n-\mu)$$

$$= \int f^*\, d(\mu_n-\mu) + \int (f^*-f_n)\, d(\mu_n-\mu).$$

So
$$\left| \big(S(\mu_n,\nu)-S(\mu,\nu)\big) - \int f^*\, d(\mu_n-\mu) \right| \leq \underbrace{\left| \int (f^*-f_n)\, d(\mu_n-\mu) \right|}_{\Delta}$$

By standard CLT,
$$\sqrt{n} \int f^*\, d(\mu_n-\mu) \xrightarrow{d} N(0,\sigma^2).$$

So it is enough to show that $\Delta = o_p(n^{-1/2})$

(means: $\sqrt{n}\,\Delta \xrightarrow{P} 0$)

Recalling notation from last lecture, we have

$$\int (f^*-f_n)\, d(\mu_n-\mu) = \int (\varphi_n - \varphi^*)\, d(\mu_n-\mu)$$

$\varphi = \frac{1}{2}\|\cdot\|^2 - f$

$$= \int (\overline{\Phi}_n - \overline{\Phi}^*)(\xi) \left(\int e^{2\pi i \langle x,\xi\rangle}\, d(\mu_n-\mu)(x) \right) d\xi$$

$$\lesssim \|\overline{\varphi}_n - \varphi^*\|_{H^s} \underbrace{\left(\int \left| \int e^{2\pi i\langle x,\xi\rangle}\, d(\mu_n-\mu)(x) \right|^2 d\xi \right)^{1/2}}_{O_p(n^{-1/2})}$$

means "stoch. bounded"
$\forall \varepsilon>0\ \exists R: \sup_n \mathbb{P}\{|\cdot| \geq R n^{-1/2}\} \leq \varepsilon$

We have reduced to the following question: if φ_n, φ^* are optimal dual solutions for (μ_n, ν) and (μ, ν), does $\overline{\varphi}_n \to \overline{\varphi}^*$ in H^s as $n \to \infty$?

Need to be careful: φ_n, φ^* only defined up to additive constant!

Lemma: Let φ_n, φ^* be normalized such that $\int \vec{\varphi}\, d\mu = \int \varphi^* d\mu = 0$.
Then $\overline{\varphi}_n \to \overline{\varphi}^{**}$ in H^s almost surely.

Strategy: suffices to show $\| D^\alpha \varphi_n - D^\alpha \varphi^* \|_{L_\infty(B_2)} \xrightarrow{a.s.} 0 \quad \forall \alpha$

Step 1: extracting a converging subsequence, by induction.

(HW)

Base: Compact support implies φ_n are pointwise bounded, as are $\nabla \varphi_n \rightarrow$ so φ_n^* are equicontinuous.

Arzela-Ascoli implies that (passing to subsequence), $\exists\, \varphi_\infty$ st $\| \varphi_n - \varphi_\infty \|_{L_\infty(B_2)} \to 0$.

Induction step: $D^\alpha \varphi_n$ are uniformly bounded and (by same fact for higher derivative) equicontinuous. Passing to a further subsequence,
$\| D^\alpha \varphi_n - D^\alpha \varphi_\infty \|_{L_\infty}(B_2) \to 0$.

Step 2: Optimality of φ_∞

Define $\psi_n(y) = \varepsilon \log \int e^{\frac{\langle x,y \rangle - \varphi_n(x)}{\varepsilon}} d\mu_n(x)$

Optimality for (μ_n, ν) implies

$$\varphi_n(x) = \varepsilon \log \int e^{\frac{\langle x,y \rangle - \psi_n(y)}{\varepsilon}} d\nu(y).$$

Uniform convergence $\varphi_n \to \varphi_\infty$ + law of large numbers implies

$$\psi_n(y) \xrightarrow{a.s.} \psi_\infty(y) = \varepsilon \log \int e^{\frac{\langle x,y\rangle - \varphi_\infty(x)}{\varepsilon}} d\mu(x)$$

And equicontinuity implies convergence is uniform on B_2.

Hence $$\varphi_\infty(x) = \varepsilon \log \int e^{\frac{\langle x,y\rangle - \psi_\infty(y)}{\varepsilon}} d\nu(y)$$

So $(\varphi_\infty, \psi_\infty)$ are optimal duals for (μ,ν).

Step 3: Conclude by uniqueness.

Subject to the constraint $\int \varphi \, d\mu = 0$, the dual problem has a unique solution (HW). So $\varphi_\infty = \varphi^*$.

Examining the above argument, we see that for any subsequence n_j, we can find a further subsequence n_{j_k} along which

$$\| \varphi_{n_{j_k}} - \varphi^* \|_{H^s} \xrightarrow{a.s.} 0$$

Hence the whole sequence φ_n converges to φ^* in H^s, a.s. $\square$ (of lemma)

We obtain that

$$\Delta = \underbrace{\| \overline{\varphi}_n - \overline{\varphi}^* \|_{H^s}}_{\to 0} \Big(\int \Big| \underbrace{\int e^{2\pi i \langle x,\xi\rangle} d(\mu_n-\mu)(x)}_{O_p(n^{-1/2})} \Big|^2 d\xi \Big)^{1/2}$$

which implies $\Delta = o_p(n^{-1/2})$. (HW)

$\square$

Examining argument, two big pieces

1. uniqueness of optimal duals
2. sufficiently fast convergence of $\int (f^* - f_n) \, d(\mu_n - \mu)$

<u>Either</u> can fail in general (non-regularized) context.

<u>Finite case</u> (problem 1)

Let's return to $(\mathcal{X}, d)$, $|\mathcal{X}| < \infty$.

We know
$$\mathbb{E}\,|W_p^p(\mu_n,\nu) - W_p^p(\mu,\nu)| \lesssim n^{-1/2}$$

If above proof worked, then
$$\sqrt{n}\left(W_p^p(\mu_n,\nu) - W_p^p(\mu,\nu)\right) \xrightarrow{d} \text{Gaussian}$$

But this is nonsense (in particular, $\sqrt{n}\,W_1(\mu_n,\mu) \geq 0$ a.s.)

Problem: no unique dual in general

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

In prior proof, we showed that this is close to $\langle f^*, \mu_n - \mu\rangle$ where f^* is optimal for (μ,ν). No strict convexity $\to$ no uniqueness!

<u>Claim:</u>
$$W_p^p(\mu_n,\nu) - W_p^p(\mu,\nu) = \sup_{f \in \mathcal{D}^*} \langle f, \mu_n - \mu\rangle + R_n$$

where $\mathcal{D}^* = \{(f,g) \in \mathcal{D} : W_p^p(\mu,\nu) = \langle f,\mu\rangle + \langle g,\nu\rangle\}$

and $R_n = o_p(n^{-1/2})$.

Conclusion, by continuous mapping theorem (HW)

$$\sqrt{n}\left(W_p^p(\mu_n,\nu)-W_p^p(\mu,\nu)\right)\xrightarrow{d}\sup_{f^*\in\mathcal{D}^*}\langle f^*, G\rangle$$

$G \sim N(0,\Sigma)$

$\Sigma_{ii} = \mu_i(1-\mu_i)$

Not Gaussian unless $\mathcal{D}^*$ is trivial.

In particular, limit of $\sqrt{n}\, W_p^p(\mu_n,\mu)$ is a.s. positive. (HW)

Proof of claim: Adopt normalization that $\langle f,\mu\rangle = 0 \;\forall f\in\mathcal{D} \Rightarrow$ $\mathcal{D}^*$ compact.

Continuing as in entropic case, we have

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

$$\leq \sup_{f_n\in\mathcal{D}_n}\langle f_n, \mu_n-\mu\rangle$$

$$\leq \sup_{f^*\in\mathcal{D}^*}\langle f^*,\mu_n-\mu\rangle + \sup_{f_n\in\mathcal{D}_n}\inf_{f^*\in\mathcal{D}^*}\langle f_n-f^*,\mu_n-\mu\rangle$$

Let $\quad S_n := \sup_{f_n\in\mathcal{D}_n}\inf_{f^*\in\mathcal{D}^*}\|f_n-f^*\|_\infty$.

Then

$$W_p^p(\mu_n,\nu)-W_p^p(\mu,\nu) = \sup_{f^*\in\mathcal{D}^*}\langle f^*,\mu_n-\mu\rangle + R_n,$$

$$|R_n| \leq S_n\cdot\underbrace{\|\mu_n-\mu\|_1}_{O_p(n^{-1/2})}$$

Suffices to show $S_n \xrightarrow{a.s.} 0$. Let $(f_n)_{n\geq 1}$ be a sequence, $f_n\in\mathcal{D}_n$.

Since $\mathcal{D}_n \subseteq [-D^p/2, D^p/2]^k$, can extract a subsequence converging to f_∞. HW: $f_\infty \in \mathcal{D}^\alpha$ a.s.

We obtain that $S_n \xrightarrow{a.s.} 0$, as desired. $\square$

<u>$\mathbb{R}^d$ case</u> (problem 2):

Working over $\mathbb{R}^d$, suppose μ, ν are absolutely continuous, and that optimal dual solutions $W_2^2(\mu_n, \nu)$, $W_2^2(\mu, \nu)$ are unique up to constant shifts. (HW)

Then by our existing argument,

$$W_2^2(\mu_n, \nu) - W_2^2(\mu, \nu) = 2\int \varphi^\alpha (d\mu - d\mu_n)$$
$$+ 2\int (\varphi_n - \varphi^*)(d\mu - d\mu_n)$$

Now problem is <u>second</u> term. Indeed, we know

$$\int \varphi^* (d\mu - d\mu_n) = O_p(n^{-1/2}), \text{ but}$$

$$\mathbb{E}\, |W_2^2(\mu_n, \nu) - W_2^2(\mu, \nu)| \gtrsim n^{-2/d}$$

Conclusion: Second term must explode.

Is there a way to get a non-trivial limit theorem?

<u>Fact</u> (HW): even though $\mathbb{E}\, W_2^2(\mu_n, \mu) \gtrsim n^{-2/d}$,

$\mathrm{Var}(W_2^2(\mu_n, \mu)) \lesssim n^{-1}$.

Fluctuations much smaller than "bias"

<u>Theorem</u>: Let μ, ν be compactly supported and absolutely continuous, and assume ν is positive on int supp ν, which is convex. Then

$$\sqrt{n}\left(W_2^2(\mu_n, \nu) - \mathbb{E}\, W_2^2(\mu_n, \nu)\right) \xrightarrow{d} N(0, \sigma^2),$$

$$\sigma^2 = \mathrm{Var}_\mu(2\varphi^*)$$

Pf idea: even though $\int (\varphi_n - \varphi^*)(d\mu - d\mu_n)$ is <u>not</u> $o_p(n^{-1/2})$, its <u>fluctuations</u> are small:

$$\mathrm{Var}\left(\int (\varphi_n - \varphi^*)\, d(\mu - \mu_n)\right) = o(n^{-1}).$$

(more details on HW)