Victor M. Panaretos
Yoav Zemel

An Invitation to Statistics in Wasserstein Space

Proof of Theorem 2.2.1

March 19, 2020

Springer

Proof (Proof of Theorem 2.2.1). We only prove the equivalence of the first four conditions, and for $\mathscr{X}=\mathbb{R}^d$. Note that 1) and 2) clearly imply 5). The converse implication is shown by Le Gouic and Loubes [2, Lemma 14]. Our proof follows that outlined in Bickel and Freedman [1].

(1 implies 2) Let $X_n \sim \mu_n$ and $X \sim \mu$ be defined on the same generic probability space $(\Omega,\mathscr{F},\mathbb{P})$, and such that (X_n,X) attain the infimum defining $W_p(\mu_n,\mu)$. We shall sometimes use the notation $W_p(X_n,X)=W_p(\mu_n,\mu)$. Suppose that $W_p(X_n,X)\to 0$. Let $f:\mathbb{R}^d\to\mathbb{R}$ be a 1-Lipschitz function. Then $|\mathbb{E}f(X)|\le\mathbb{E}\|X\|\le(\mathbb{E}[\|X\|^p])^{1/p}<\infty$ and

$$|\mathbb{E}f(X_n)-\mathbb{E}f(X)|\le\mathbb{E}\|X_n-X\|\le[\mathbb{E}\|X_n-X\|^p]^{1/p}=W_p(X_n,X)\to 0.$$

This implies that $X_n\to X$ in distribution. The triangle inequality for W_p now gives

$$|\mathbb{E}\|X_n\|^p-\mathbb{E}\|X\|^p|=|W_p(X_n,0)-W_p(X,0)|\le W_p(X_n,X)\to 0.$$

(2 implies 3) Fix $\delta>0$. By the dominated convergence theorem there exists $R_\delta\ge 1$ such that

$$\forall R\ge R_\delta \quad \int_{\|x\|\ge R}\|x\|^p\,\mathrm{d}\mu(x)\le\delta.$$

For $R\ge R_\delta$ define the continuous function

$$f(r)=\begin{cases}0 & r\le R\\ (1+R)^p(r-R) & R\le r\le R+1\\ r^p & r>R+1.\end{cases}$$

Then $r\mapsto r^p-f(r)$ is continuous and bounded on $[0,\infty)$, so that the upper bound

$$\int_{\|x\|\ge R+1}\|x\|^p\,\mathrm{d}\mu_n(x)\le\int f(\|x\|)\,\mathrm{d}\mu_n(x)=\int\|x\|^p\,\mathrm{d}\mu_n(x)+\int[f(\|x\|)-\|x\|^p]\,\mathrm{d}\mu_n(x)$$

converges as $n\to\infty$ to

$$\begin{aligned}\int f(\|x\|)\,\mathrm{d}\mu(x)&\le\int_{\|x\|\ge R+1}\|x\|^p\,\mathrm{d}\mu(x)+(R+1)^p\mu(\{R\le\|X\|\le R+1\})\\&\le\delta+\frac{(R+1)^p}{R^p}\int_{\|x\|\ge R+1}\|x\|^p\,\mathrm{d}\mu(x)\le(1+2^p)\delta\end{aligned}$$

since $R\ge 1$. Hence for all $R\ge R_\delta+1$,

$$\limsup_{n\to\infty}\int_{\|x\|\ge R}\|x\|^p\,\mathrm{d}\mu_n(x)\le(2+2^p)\delta.$$

By increasing R_δ if necessary the limsup can be replaced by a sup.

(3 implies 4) Let g satisfy the growth condition and fix $\varepsilon>0$. Set $M=\sup_x|g(x)|/(1+\|x\|^p)<\infty$ and notice that the truncation

$$g_R(x) = \min(R, \max(-R, g(x))$$

is continuous and bounded and therefore

$$\int g_R(x)\,\mathrm{d}\mu_n(x) \to \int g_R(x)\,\mathrm{d}\mu(x), \qquad n \to \infty.$$

If $|g(x)| > R$ then $\|x\|^p \geq (R/M) - 1$. Thus

$$\sup_n \int |g(x) - g_R(x)|\,\mathrm{d}\mu_n(x) \leq \sup_n M \int_{\|x\| > [R/M-1]^{1/p}} [1 + \|x\|^p]\,\mathrm{d}\mu_n(x) \to 0, \qquad R \to \infty.$$

by (3). Since the same holds with μ_n replaced by μ, this implies (4).

Clearly 4) implies 2). Hence, it suffices to show that 3) implies 1). For simplicity we assume that $\mathscr{X} = \mathbb{R}^d$; the Hilbert case can be obtained with an additional step, using the tightness of the sequence $\{\mu_n\}$ and intersecting the ball $\{x : \|x\| \leq R\}$ with a compact set K satisfying $\mu_n(K) \geq 1 - \varepsilon$ and $\mu(K) \geq 1 - \varepsilon$ for arbitraily small $\varepsilon > 0$. The notation is simpler when using random variables instead of measures. Let $X_n \sim \mu_n$, $X \sim \mu$ and define the truncation $Z_n = X_n 1\{\|X_n\| \leq R\}$. Fix $\delta > 0$. Then for $R = R_\delta > 0$ sufficiently large, for all n

$$W_p^p(Z_n, X_n) \leq E\|X_n\|^p 1\{\|X_n\| > R\} = \int_{\|x\| > R} \|x\|^p\,\mathrm{d}\mu_n(x) \leq \delta^p,$$

and similarly for $W_p(Z, X)$. Hence we can replace X_n by Z_n and X by Z with negligible error ($\delta > 0$ is arbitrary). We may also choose R_δ such that $\mathbb{P}(\|X\| = R_\delta) = 0$.

The next step is to discretise the space and approximate Z and Z_n by point masses. Denote the compact set $\{x : \|x\| \leq R\}$ by C, and fix $\varepsilon > 0$. Cover C with disjoint measurable sets $B_1, \ldots, B_{N_\varepsilon}$ of diameter $\leq \varepsilon$ and such that $\mathbb{P}(Z \in \partial B_i) = 0$ for all i. This is possible since we can vary the radii of the balls over the uncountable set $[\varepsilon/2, \varepsilon]$ and only countably many such radii may have positive probability for the boundary. Let $y_i \in B_i$ and define the discrete random variables

$$Z^\varepsilon = \sum_{i=1}^{N_\varepsilon} y_i 1\{Z \in B_i\}, \qquad Z_n^\varepsilon = \sum_{i=1}^{N_\varepsilon} y_i 1\{Z_n \in B_i\}.$$

The discretisation error is easily bounded:

$$W_p^p(Z^\varepsilon, Z) \leq \mathbb{E}\|Z - Z^\varepsilon\|^p = \sum_{i=1}^{N_\varepsilon} \mathbb{E}\|Z - Z^\varepsilon\|^p 1\{Z \in B_i\} = \sum_{i=1}^{N_\varepsilon} \mathbb{E}\|Z - y_i\|^p 1\{Z \in B_i\} \leq \sum_{i=1}^{N_\varepsilon} \mathbb{E}\varepsilon^p 1\{Z \in B_i\} = \varepsilon^p.$$

The same bound holds for $W_p^p(Z_n^\varepsilon, Z_n)$. To bound $W_p(Z^\varepsilon, Z_n^\varepsilon)$ we construct an explicit coupling π such that Z^ε and Z_n^ε are equal with high probability. Define $p_i = \mathbb{P}(Z \in B_i)$ and $q_i = \mathbb{P}(Z_n \in B_i)$. We construct a joint distribution π for Z^ε and Z_n^ε that makes them exactly equal with high probability. Clearly, the best we can do is to set $\pi(Z^\varepsilon = y_i, Z_n^\varepsilon = y_i) = \min(p_i, q_i)$. The remaining events can be chosen arbitrarily to have the correct marginal distributions $\pi(Z^\varepsilon = y_i) = p_i$ and $\pi(Z_n^\varepsilon = y_i) = q_i$.

For concreteness we build them independently as follows. Let $I = \{i : p_i \geq q_i\}$, $J = \{i : p_i < q_i\}$, and set

$$\pi(Z^\varepsilon = y_i, Z_n^\varepsilon = y_j) = \begin{cases} q_i & i = j \in I \\ p_i & i = j \in J \\ \alpha_i \beta_j & i \in I, j \in J \\ 0 & \text{otherwise,} \end{cases} \qquad \alpha_i = p_i - q_i, \quad \beta_j = \frac{q_j - p_j}{\sum_{j \in J} q_j - p_j}.$$

A simple calculation verifies that this has the correct marginal distributions. (If J is empty, then $p_i = q_i$ for all i and the coupling is such that $Z^\varepsilon = Z_n^\varepsilon$ with π-probability one.) Then

$$\pi(Z^\varepsilon = Z_n^\varepsilon) = \sum_{i=1}^{N_\varepsilon} \pi(Z^\varepsilon = Z_n^\varepsilon = y_i) = \sum_{i \in J} p_i + \sum_{i \in I} p_i + q_i - p_i = 1 - \sum_{i \in I} p_i - q_i = 1 - \frac{1}{2} \sum_{i=1}^{N_\varepsilon} |p_i - q_i|.$$

(This is one minus the total variation between Z^ε and Z_n^ε.) Since $\sup_{x,y \in C} \|x - y\| \leq 2R_\delta$ we have

$$W_p^p(Z^\varepsilon, Z_n^\varepsilon) \leq 2^p R_\delta^p \frac{1}{2} \sum_{i=1}^{N_\varepsilon} |\mathbb{P}(Z \in B_i) - \mathbb{P}(Z_n \in B_i)| \to 0, \qquad n \to \infty.$$

Now take $\varepsilon \to 0$:

$$\limsup_{n\to\infty} W_p(Z_n, Z) \leq \limsup_{\varepsilon \to 0} \limsup_{n \to \infty} W_p(Z_n, Z_n^\varepsilon) + W_p(Z, Z^\varepsilon) + W_p(Z^\varepsilon, Z_n^\varepsilon) \leq \limsup_{\varepsilon \to 0} 2\varepsilon = 0.$$

Note that Z_n and Z depend on δ through R_δ. Letting $\delta \to 0$ gives

$$\limsup_{n\to\infty} W_p(X_n, X) \leq \limsup_{\delta \to 0} \limsup_{n \to \infty} W_p(X_n, Z_n) + W_p(X, Z) + W_p(Z_n, Z) \leq \limsup_{\delta \to 0} 2\delta = 0.$$

We have thus established that $W_p(\mu_n, \mu) \to 0$.

References

1. P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981.
2. T. Le Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *Prob. Theory and Related Fields*, 168(3-4):901–917, 2017.