Lecture 20, Wednesday, 5/11/2022

Characterization of the convergence w.r.t. the W_2-metric

<u>Theorem</u> (Convergence in $(\mathcal{P}_2(X), W_2)$) Let X be Polish and $\mu_n, \mu \in \mathcal{P}_2(X)$ $(n=1,2,\cdots)$.

(1) $\mu_n \xrightarrow{W_2} \mu \Rightarrow \mu_n \to \mu$ narrowly and $\forall x_0 \in X$

$$\int_X d^2(x, x_0)\, d\mu_n(x) \longrightarrow \int_X d^2(x, x_0)\, d\mu(x). \qquad (*)$$

(2) $\mu_n \longrightarrow \mu$ narrowly and $(*)$ holds true for some $x_0 \Longrightarrow \mu_n \xrightarrow{W_2} \mu$.

<u>Corollary</u> If $\mu_n \to \mu$ narrowly, then $(*)$ is true for some $x_0 \in X \Longleftrightarrow$ it is true for all $x_0 \in X$.

<u>Proof</u> If $(*)$ is true for one x_0, then $\mu_n \xrightarrow{W_2} \mu$ by part (1) of the theorem. Now, $\mu_n \xrightarrow{W_2} \mu \Rightarrow (*)$ is true for all $x_0 \in X$ by part (2). <u>QED</u>

<u>Remarks</u> $\odot$ As shown in previous lectures, this theorem implies other results, e.g., $(\mathcal{P}_2(X), W_2)$ is Polish, and $(\mathcal{P}_2(X), W_2)$ is compact if X is.

$\odot$ There are other characterization of W_2-convergence.

<u>Theorem</u> Let X be a Polish space and $\mu_n, \mu \in \mathcal{P}_2(X)$ $(n=1,2,\cdots)$. The following are equivalent:

(1) $\mu_n \xrightarrow{W_2} \mu$;

(2) $\mu_n \longrightarrow \mu$ narrowly and $(*)$ is true for some $x_0 \in X$;

(3) $\mu_n \longrightarrow \mu$ narrowly and for some $x_0 \in X$

$$\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d(x_0,x)\geq R} d^2(x_0, x)\, d\mu_n(x) = 0;$$

(4) If $\varphi: X \to \mathbb{R}$ is continuous and satisfies the growth condition
$$\sup_{x \in X} \frac{|\varphi(x)|}{1 + d^2(x_0, x)} < \infty$$
for some $x_0 \in X$, then
$$\int_X \varphi \, d\mu_n \to \int_X \varphi \, d\mu.$$

We will not prove this theorem.

To prove the main convergence theorem, we need the following lemma, proved in Lecture 13:

<u>Lemma</u> Let X be a metric space, $G \subseteq X$ an open subset, and $f: X \to [0, \infty]$ a lower semi-continuous function. Suppose $\mu_n \to \mu$ narrowly in $\mathcal{P}(X)$, then
$$\liminf_{n \to \infty} \int_G f \, d\mu_n \geq \int_G f \, d\mu.$$
In particular $\liminf_{n \to \infty} \int_X f \, d\mu_n \geq \int_X f \, d\mu$.

<u>Proof of the convergence theorem</u>

(1) $\mu_n \xrightarrow{W_2} \mu \Rightarrow \mu_n \to \mu$ narrowly and $\forall x_0 \in X$
$$\int_X d^2(x, x_0) \, d\mu_n(x) \to \int_X d^2(x, x_0) \, d\mu(x). \qquad (*)$$
Assume $\mu_n \xrightarrow{W_2} \mu$. Note that $\forall \nu \in \mathcal{P}_2(X)$ $\forall x_0 \in X$:
$\gamma \in \mathcal{A}(\delta_{x_0}, \nu) \Rightarrow \gamma(\{x_0\} \times X) = \delta_{x_0}(\{x_0\}) = 1$. Thus,
$$\int_{X \times X} d^2(x, y) \, d\gamma(x, y) = \int_{X \times X} d^2(x_0, x) \, d\gamma(x, y) = \int_X d^2(x_0, x) \, d\nu(x).$$
Hence $$W_2^2(\delta_{x_0}, \nu) = \int_X d^2(x_0, x) \, d\nu(x).$$
therefore,

$$\left| \int_X d^2(x, x_0)\, d\mu_n(x) - \int_X d^2(x, x_0)\, d\mu(x) \right|$$
$$= \left| W_2^2(\delta_{x_0}, \mu_n) - W_2^2(\delta_{x_0}, \mu) \right|$$
$$\le \left| W_2(\delta_{x_0}, \mu_n) - W_2(\delta_{x_0}, \mu) \right| \left| W_2(\delta_{x_0}, \mu_n) + W_2(\delta_{x_0}, \mu) \right|$$
$$\le W_2(\mu_n, \mu) \cdot \left[W_2(\mu_n, \mu) + 2 W_2(\mu, \delta_{x_0}) \right]$$
$$\longrightarrow 0,$$
proving $(*)$.

We now prove $\mu_n \to \mu$ narrowly. It suffices to show that $\forall f \in BL(X)$: $\int_X f\, d\mu_n \to \int_X f\, d\mu$. Let $\hat{\gamma}_n \in \mathcal{A}(\mu_n, \mu)$ be such that $W_2^2(\mu_n, \mu) = \int_{X\times X} d^2(x,y)\, d\hat{\gamma}_n(x,y)$ $(n = 1, 2, \cdots)$. Then, for any $f \in BL(X)$, we have

$$\left| \int_X f\, d\mu_n - \int_X f\, d\mu \right| = \left| \int_X f(x)\, d\mu_n - \int_X f(y)\, d\mu \right|$$
$$= \left| \int_{X\times X} f(x)\, d\hat{\gamma}_n(x,y) - \int_{X\times X} f(y)\, d\hat{\gamma}_n(x,y) \right|$$
$$\le \mathrm{Lip}(f) \int_{X\times X} d(x,y)\, d\hat{\gamma}_n(x,y)$$
$$\le \mathrm{Lip}(f) \left[\int_{X\times X} d^2(x,y)\, d\hat{\gamma}_n(x,y) \right]^{1/2}$$
$$= \mathrm{Lip}(f)\, W_2(\mu_n, \mu)$$
$$\longrightarrow 0 \quad \text{as } n \to \infty.$$

(2) $\mu_n \longrightarrow \mu$ narrowly and $(*)$ holds true for some $x_0 \Longrightarrow \mu_n \xrightarrow{W_2} \mu$.

<u>Step 1</u> Assume X is compact. Fix $z \in X$. Define
$$\mathcal{F} = \{ f \in \mathrm{Lip}(X) : \mathrm{Lip}(f) \le 1,\ f(z) = 0 \}.$$
If $f \in \mathcal{F}$ then $\max_X |f| \le \mathrm{diam}(X)$. Thus $\mathcal{F}$ is uniformly bounded and equicontinuous. Hence, by the

Ascoli-Arzelà Theorem, $\mathcal{F}$ is compact in $C(X) = C_b(X)$.
We claim that

$$\alpha_n := \sup_{f \in \mathcal{F}} \int_X f \, d(\mu_n - \mu) \longrightarrow 0.$$

In fact, $\exists f_n \in \mathcal{F}$ such that

$$\int_X f_n d(\mu_n - \mu) \le \alpha_n \le \int_X f_n d(\mu_n - \mu) + \frac{1}{n}, \quad n = 1, 2, \dots$$

We show that

$$\beta_n := \int_X f_n \, d(\mu_n - \mu) \longrightarrow 0.$$

If not, $\exists \varepsilon_0 > 0$ and subsequences of f_n and μ_n, not relabeled, such that $|\beta_n| \ge \varepsilon_0$ $(n = 1, 2, \dots)$. Since $f_n \in \mathcal{F}$ and $\mathcal{F}$ is compact in $C(X)$, there exists a subseq. $\{f_{n_j}\}$, of $\{f_n\}$ that converges to some $f_\infty \in C(X)$. Clearly $f_\infty \in \mathcal{F}$. Now, $\|f_{n_k} - f\|_\infty \to 0$ and $\mu_{n_k} \to \mu$ narrowly. So

$$|\beta_{n_k}| \le \left| \int_X f_\infty \, d(\mu_{n_k} - \mu) \right| + \left| \int_X (f_{n_k} - f_\infty) \, d(\mu_{n_k} - \mu) \right| \to 0,$$

contradicting $|\beta_n| \ge \varepsilon_0$. Hence, $\beta_n \to 0$, and $\alpha_n \to 0$.

Now, for any $\varphi \in C_b$ with $\mathrm{Lip}(\varphi) \le 1$, $\varphi - \varphi(z) \in \mathcal{F}$. and $\int_X \varphi \, d(\mu_n - \mu) = \int_X (\varphi(x) - \varphi(z)) \, d(\mu_n - \mu)(x)$.
Hence,

$$\sup_{\varphi \in C(X), \, \mathrm{Lip}(\varphi) \le 1} \int_X \varphi \, d(\mu_n - \mu) = \sup_{f \in \mathcal{F}} \int_X f \, d(\mu_n - \mu) \to 0.$$

By the Kantorovich-Rubinstein theorem, $\exists \gamma_n \in \mathcal{A}(\mu_n, \mu)$ optimal w.r.t. to d (not d^2), and

$$\lim_{n \to \infty} \int_{X \times X} d(x, y) \, d\gamma_n(x, y) = \lim_{n \to \infty} \sup_{\varphi \in C(X), \mathrm{Lip}(\varphi) \le 1} \int_X \varphi \, d(\mu_n - \mu) = 0.$$

Therefore,

$$\limsup_{n \to \infty} W_2^2(\mu_n, \mu) \le \limsup_{n \to \infty} \int_{X \times X} d^2(x, y) \, d\gamma_n(x, y)$$

$$\leq \operatorname{diam}(X) \limsup_{n\to\infty} \int_{X\times X} d(x,y)\, d\gamma_n(x,y) = 0,$$
hence $W_2(\mu_n, \mu) \to 0$.

Step 2 The general case (i.e., X is only Polish). Define
$$\sigma_n = \frac{1}{Z_n}\left(1 + d^2(x_0, \cdot)\right)\mu_n \in \mathcal{P}(X),$$
$$\sigma = \frac{1}{Z}\left(1 + d^2(x_0, \cdot)\right)\mu \in \mathcal{P}(X),$$
where Z_n, Z are normalizing constants. By $(*)$, $Z_n \to Z$. Also, since $\mu_n \to \mu$ narrowly, by the lemma,
$$\liminf_{n\to\infty} \int_A d^2(x, x_0)\, d\mu_n(x) \geq \int_A d^2(x,x_0)\, d\mu(x) \quad \forall \text{ open } A \subseteq X.$$
We therefore have
$$\liminf_{n\to\infty} \sigma_n(A) \geq \sigma(A) \quad \forall A \subseteq X: \text{open}.$$
Thus, $\sigma_n \to \sigma$ narrowly.

By Prokhorov's Theorem, we can find a sequence of compact subsets $K_1 \subseteq K_2 \subseteq \cdots$ of X, with $x_0 \in K_1$, s.t.
$$\lim_{k\to\infty} \sup_{n\geq 1} \sigma_n(X \setminus K_k) = 0.$$
Since
$$Z_n \sigma_n(X\setminus K_k) \geq \int_{X\setminus K_k} d^2(x, x_0)\, d\mu_n(x), \quad n = 1, 2, \dots,$$
we get
$$\lim_{k\to\infty} \sup_{n\geq 1} \int_{X\setminus K_k} d^2(x, x_0)\, d\mu_n(x) = 0. \quad (**)$$
Now, define for each $k \in \mathbb{N}$
$$\mu_{n,k} = \mu_n \llcorner K_k + (1 - \mu_n(K_k))\, \delta_{x_0} \in \mathcal{P}(K_k).$$
Since each K_k is compact, $\mathcal{P}_2(K_k) = \mathcal{P}(K_k)$ is compact w.r.t. the narrow topology. Hence, $\{\mu_{n,k}\}_{n=1}^{\infty}$ has a subseq. that converges narrowly in $\mathcal{P}(K_k)$. Since K_k is compact, the distance function is bounded and continuous on $K_k \times K_k$.

So, (*) is true, for any $x_0 \in K_k$, with X replaced by K_k. Thus, by Step 1, and by the diagonal argument, we can find a sub seq. $\{n_j\}_{j=1}^{\infty}$, such that for each $k \geq 1$

$\{\mu_{n_j,k}\}_{j=1}^{\infty}$ converges in $(\mathcal{P}_2(K_k), W_2)$, and hence in $\mathcal{P}(X)$, if we identify $\mu_{n,k}$ as measures on X. (We cannot use the fact that $(\mathcal{P}_2(K_k), W_2)$ is compact as it is proved using this theorem.)

Now, let

$$\gamma_{n,k} = (Id \times Id)_{\#} \mu_n \llcorner K_k + (Id \times x_0)_{\#} \mu_n \llcorner (X \setminus K_k).$$

We can verify that $\gamma_{n,k} \in \mathcal{A}(\mu_n, \mu_{n,k})$. Let $A, B \in \mathcal{B}(X)$.

$$\begin{aligned}
\gamma_{n,k}(A \times X) &= \mu_n \llcorner K_k \left((Id \times Id)^{-1}(A \times X) \right) \\
&\quad + \mu_n \llcorner (X \setminus K_k) \left((Id \times x_0)^{-1}(A \times X) \right) \\
&= \mu_n \llcorner K_k (A) + \mu_n \llcorner (X \setminus K_k)(A) \\
&= \mu_n (K_k \cap A) + \mu_n ((X \setminus K_k) \cap A) = \mu_n(A).
\end{aligned}$$

If $x_0 \notin B$ then

$$\begin{aligned}
\gamma_{n,k}(X \times B) &= \mu_n \llcorner K_k \left((Id \times Id)^{-1}(X \times B) \right) \\
&\quad + \mu_n \llcorner (X \setminus K_k) \left((Id \times x_0)^{-1}(X \times B) \right) \\
&= \mu_n \llcorner K_k (B) + \mu_n \llcorner (X \setminus K_k)(\phi) = \mu_n \llcorner K_k (B).
\end{aligned}$$

$$\begin{aligned}
\mu_{n,k}(B) &= \mu_n \llcorner K_k (B) + (1 - \mu_n(K_k)) \delta_{x_0}(B) \\
&= \mu_n \llcorner K_k (B) = \gamma_{n,k}(X \times B).
\end{aligned}$$

If $x_0 \in B$ then

$$\begin{aligned}
\gamma_{n,k}(X \times B) &= \mu_n \llcorner K_k (B) + \mu_n \llcorner (X \setminus K_k)(X) \\
&= \mu_n (B \cap K_k) + 1 - \mu_n(K_k)
\end{aligned}$$

$$\begin{aligned}
\mu_{n,k}(B) &= \mu_n \llcorner K_k (B) + (1 - \mu_n(K_k)) \delta_{x_0}(B) \\
&= \mu_n (B \cap K_k) + 1 - \mu_n(K_k) = \gamma_{n,k}(X \times B).
\end{aligned}$$

Moreover, we have the uniform estimate

$$W_2^2(\mu_n, \mu_{n,k}) \leq \int_{X \times X} d^2(x,y)\, d\gamma_{n,k}(x,y)$$

$$= \int_{X\times X} d^2(x,y)\, d(Id\times Id)_\# \mu_n \llcorner K_k$$
$$+ \int_{X\times X} d^2(x,y)\, d(Id\times x_0)_\# \mu_n \llcorner (X\setminus K_k)$$
$$= \int_X d^2(Id\times Id(x,y))\, d\mu_n \llcorner K_k$$
$$+ \int_X d^2(Id\times x_0(x,y))\, d\mu_n \llcorner (X\setminus K_k)$$
$$= \int_{X\setminus K_k} d^2(x,x_0)\, d\mu_n(x), \quad \forall n, k.$$

This and $(**)$ lead to
$$\lim_{k\to\infty} \sup_{n\geq 1} W_2(\mu_n, \mu_{n,k}) = 0,$$
particularly,
$$\lim_{k\to\infty} \sup_{j\geq 1} W_2(\mu_{n_j}, \mu_{n_j,k}) = 0. \qquad (***)$$

Since for each k, $\{\mu_{n_j,k}\}$ converges in $\mathcal{P}_2(X)$ w.r.t. W_2, for any $\varepsilon>0$, by $(***)$, there is k_0 large enough such that for any j,l,
$$W_2(\mu_{n_j}, \mu_{n_l}) \leq W_2(\mu_{n_j}, \mu_{n_j,k_0}) + W_2(\mu_{n_l}, \mu_{n_l,k_0})$$
$$+ W_2(\mu_{n_j,k_0}, \mu_{n_l,k_0})$$
$$\leq \varepsilon + \varepsilon + W_2(\mu_{n_j,k_0}, \mu_{n_l,k_0}).$$

Thus, $\limsup_{j,l\to\infty} W_2(\mu_{n_j}, \mu_{n_l}) \leq 2\varepsilon$, and $\{\mu_{n_j}\}$ is a Cauchy sequence w.r.t. W_2. But $(\mathcal{P}_2(X), W_2)$ is complete. So $\{\mu_{n_j}\}$ converges w.r.t. W_2.

The argument can be applied to any subseq. of $\{\mu_n\}$ to get a further subseq. converging to $\tilde{\mu}$ in W_2-metric. By Part (1), this further sequence converges to $\tilde{\mu}$ narrowly. Hence $\tilde{\mu} = \mu$. Therefore, $\mu_n \xrightarrow{W_2} \mu$. QED