Lecture 16, Monday, 5/2/2022

Today: Wasserstein metric.

Set up X, Y: Polish. $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$

$c : X \times Y \to [0, \infty]$: measurable.

$$\mathcal{A}(\mu, \nu) \triangleq \{ \gamma \in \mathcal{P}(X \times Y) : \pi^X_\# \gamma = \mu, \ \pi^Y_\# \gamma = \nu \}$$

$$E_K[\gamma] = \int_{X \times Y} c(x, y) \, d\gamma(x, y), \quad \gamma \in \mathcal{A}(\mu, \nu).$$

<u>Existence Theorem</u> If c is lower semicontinuous, then $\exists \ \hat{\gamma} = \arg\min_{\gamma \in \mathcal{A}(\mu,\nu)} E_K[\gamma]$.

Now, consider

This is the d for $X \times X$.

- (X, d) : a Polish space.
- $\mathcal{P}_2(X) = \{ \mu \in \mathcal{P}(X) : \exists \ x_0 \in X \text{ s.t. } \int_X d^2(x, x_0) \, d\mu(x) < \infty \}$.

Remark. Let $\mu \in \mathcal{P}(X)$. Then $\exists \ x_0 \in X$ s.t. $\int_X d^2(x, x_0) \, d\mu(x) < \infty$
$\iff \forall \ x_0' \in X$ s.t. $\int_X d^2(x, x_0') \, d\mu(x) < \infty$. Since, $\forall x_0' \in X$, by the triangle inequality, $d^2(x, x_0') \le 2 d^2(x, x_0) + 2 d^2(x_0, x_0')$.

- $\forall \mu, \nu \in \mathcal{P}_2(X)$, define (This d is the d for $X \times Y$.)

$$W_2(\mu, \nu) = \min_{\gamma \in \mathcal{A}(\mu,\nu)} \left[\int_{X \times X} d^2(x, y) \, d\gamma(x, y) \right]^{1/2}.$$

Note that the "min" is attained by the existence thm.

<u>Theorem</u> Let (X, d) be a Polish space. Then $(\mathcal{P}_2(X), W_2)$ is a metric space.

<u>Remarks</u> - Call W_2 the 2-Wasserstein or just the

Wasserstein metric (W-metric).

⊙ The result can be extended to the p-Wasserstein metric for $1 \le p < \infty$. But, we shall just focus on the case $p=2$.

⊙ The following lemma is proved in Dudley, Real Anal. and Probability. Cambridge Univ. Press, 2002:

Lemma Let X_i be Polish spaces and $\mu_i \in \mathcal{P}_2(X_i)$ $(i=1,2,3)$. Let $\gamma^{1,2} \in \mathcal{A}(\mu_1, \mu_2)$ and $\gamma^{2,3} \in \mathcal{A}(\mu_2, \mu_3)$. Then there exists $\gamma \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that

$$\pi^{1,2}_{\#}\gamma = \gamma^{1,2} \quad \text{and} \quad \pi^{2,3}_{\#}(\gamma) = \gamma^{2,3}.$$

Here, $\pi^{1,2}(x_1, x_2, x_3) = (x_1, x_2)$ and $\pi^{2,3}(x_1, x_2, x_3) = (x_2, x_3)$.

This lemma can be used to prove the triangle inequality of the W-metric; see the proof below. But the proof of this lemma requires more tools (e.g., disintegration). For completeness. I will present an elementary proof of the triangle inequality, related to the discrete case, by Clement & Desch (Proc. AMS, 136:1, 333-339, 2008).

Proof of Theorem

⊙ $W_2(\mu, \nu)$ is finite, Fix $x_0 \in X$. For any $x, y \in X$:
$d^2(x,y) \le 2d^2(x, x_0) + 2d^2(x_0, y)$. So, $\forall \gamma \in \mathcal{A}(\mu, \nu)$,

$$\left(E_K[\gamma]\right)^2 \le 2\int_{X\times Y} d^2(x, x_0)\,d\gamma + 2\int_{X\times Y} d^2(x_0, y)\,d\gamma$$

$$= 2\int_X d^2(x, x_0)\,d\mu(x) + 2\int_Y d^2(x_0, y)\,d\nu(y) < \infty.$$

Hence, $W_2(\mu,\nu) < \infty$.

(i) Clearly, $W_2(\mu,\nu) \geq 0$ $\forall \mu,\nu \in \mathcal{P}_2(X)$. We show that $W_2(\mu,\nu) = 0 \Leftrightarrow \mu = \nu$. Suppose $\mu = \nu$. We construct $\gamma \in \mathcal{A}(\mu,\mu)$ such that $\gamma(S) = 0$, where $S = \{(x,y) \in X \times X : x \neq y\}$. This implies that

$$E_c[\gamma] = \int_{X\times X} d^2(x,y)\,d\gamma(x,y) = \int_{D_{X\times X}} d^2(x,y)\,d\gamma(x,y) = 0.$$

Hence $W_2(\mu,\mu) = 0$.

Define $Id \times Id : X \to X \times X$ by $(Id \times Id)(x) = (x,x)$. and $\gamma = (Id \times Id)_{\#}\mu$. Since $Id \times Id$ is continuous, γ is a Borel measure on $X \times Y$. Moreover, $\gamma(X \times X) = \mu((Id \times Id)^{-1}(X \times X) = \mu(X) = 1$. So, $\gamma \in \mathcal{P}(X \times X)$.

If $A \in \mathcal{B}(X)$ then $\gamma(A \times X) = \mu((Id \times Id)^{-1}(A \times X)) = \mu(A)$.

$[a \in (Id \times Id)^{-1}(A \times X) \Leftrightarrow (Id \times Id)(a) = (a,a) \in A \times X \Leftrightarrow a \in A]$

Similarly, $\gamma(X \times A) = \mu(A)$. Hence, $\gamma \in \mathcal{A}(\mu,\mu)$. Finally, for $S = \{(x,y) \in X \times Y : x \neq y\}$, we have $(Id \times Id)^{-1}(S) = \phi \subseteq X$. Hence, $\gamma(S) = \mu(\phi) = 0$.

Now, assume $\mu,\nu \in \mathcal{P}_2(X)$ and $W_2(\mu,\nu) = 0$. Let $\hat{\gamma} \in \mathcal{A}(\mu,\nu)$ be such that $W_2(\mu,\nu) = E_c[\hat{\gamma}] = 0$ i.e.,

$$\int_{X\times X} d^2(x,y)\,d\hat{\gamma}(x,y) = 0.$$

Since $d(x,y) > 0$ $\forall x,y \in X$, $x \neq y$. $\hat{\gamma}(S) = 0$, where $S = \{(x,y) \in X \times Y : x \neq y\}$, i.e., $\hat{\gamma}$ is concentrated on the diagonal $D = \{(x,x) : x \in X\}$. Now, for any bounded Borel function $f: X \to \mathbb{R}$, since $\pi^X_{\#}\gamma = \mu$, $\pi^Y_{\#}\gamma = \nu$, we have

$$\int_X f(x)\,d\mu(x) = \int_{X\times X} f(x)\,d\hat{\gamma}(x,y) = \int_D f(x)\,d\hat{\gamma}(x,y)$$
$$= \int_{X\times X} f(y)\,d\hat{\gamma}(x,y) = \int_X f(y)\,d\nu(y).$$

Hence $\mu = \nu$.

⊙ $\forall \mu, \nu \in \mathcal{P}_2(X)$. $W_2(\mu,\nu) = W_2(\nu,\mu)$.

Let $\gamma \in \mathcal{A}(\mu,\nu)$. Define $\gamma^T : \mathcal{B}(X\times Y) \to \mathbb{R}$ by $\gamma^T(A\times B) = \gamma(B\times A)$. Then, $\gamma^T \in \mathcal{P}(X\times Y)$. Moreover,

$(\pi^X_{\#}\gamma^T)(A) = \gamma^T((\pi^X)^{-1}(A)) = \gamma^T(A\times Y) = \gamma(Y\times A) = \nu(A)$,

$(\pi^Y_{\#}\gamma^T)(B) = \mu(B)$. Hence $\gamma^T \in \mathcal{A}(\nu,\mu)$. Also, $d(x,y) = d(x,y)$ $\forall x,y \in X$. Using the change of variables $T(x,y) = (y,x)$, we have $T_{\#}\gamma = \hat{\gamma}$, and

$$\int_{X\times X} d^2(x,y)\,d\gamma(x,y) = \int_{X\times X} d^2(y,x)\,d\gamma(x,y) = \int_{X\times X} (d^2\circ T)(x,y)\,d\gamma(x,y)$$
$$= \int_{X\times X} d^2(x,y)\,d(T_{\#}\gamma)(x,y) = \int_{X\times X} d^2(x,y)\,d\hat{\gamma}(x,y).$$

Hence, $W_2(\mu,\nu) = W_2(\nu,\mu)$

⊙ The triangle inequality. $\forall \mu_i \in \mathcal{P}_2(X)$. Show that
$$W_2(\mu_1,\mu_3) \le W_2(\mu_1,\mu_2) + W_2(\mu_2,\mu_3).$$

Let $\gamma^{1,2} \in \mathcal{A}(\mu_1,\mu_2)$ and $\gamma^{2,3} \in \mathcal{A}(\mu_2,\mu_3)$ be such

that $W_2^2(\mu_1,\mu_2)=E_K[\gamma^{1,2}]$ and $W_2^2(\mu_2,\mu_3)=E_K[\gamma^{2,3}]$.
By the lemma, $\exists\,\gamma\in\mathcal{P}(X\times X\times X)$ such that $\pi^{1,2}_{\#}\gamma=\gamma^{1,2}$ and $\pi^{2,3}_{\#}(\gamma)=\gamma^{2,3}$. Thus

$$
\begin{aligned}
W_2(\mu_1,\mu_3) &\le \|d(x_1,x_3)\|_{L^2(\gamma)} \quad \text{(1st marginal of } \gamma \text{ is } \mu_1\text{, 3rd marginal of } \gamma \text{ is } \mu_3\text{)}\\
&\le \|d(x_1,x_2)\|_{L^2(\gamma)} + \|d(x_2,x_3)\|_{L^2(\gamma)}\\
&= \|d(x_1,x_2)\|_{L^2(\gamma^{1,2})} + \|d(x_2,x_3)\|_{L^2(\gamma^{2,3})}\\
&= W_2(\mu_1,\mu_2)+W_2(\mu_2,\mu_3).
\end{aligned}
$$

QED

Proposition Let (X,d) be a Polish space.
(1) Let $x_0\in X$ and $\nu\in\mathcal{P}_2(X)$. If $\gamma\in\mathcal{A}(\delta_{x_0},\nu)$ then

$$\int_{X\times X} d^2(x,y)\,d\gamma(x,y)=\int_X d^2(x_0,x)\,d\nu(x),$$

and hence $W_2^2(\delta_{x_0},\nu) = \int_X d^2(x_0,x)\,d\nu(x)$.
If $y_0\in X$ then $W_2(\delta_{x_0},\delta_{y_0})=d(x_0,y_0)$. $\quad(*)$
(2) The mapping $x\mapsto\delta_x$ is an isometric from X to $(\mathcal{P}_2(X),W_2)$.

Proof (1) If $\gamma\in\mathcal{A}(\delta_{x_0},\nu)$ then $\gamma(\{x_0\}\times X)=\delta_{x_0}(\{x_0\})=1$. So,

$$
\begin{aligned}
\int_{X\times X} d^2(x,y)\,d\gamma(x,y) &= \int_{X\times X} d^2(x_0,y)\,d\gamma(x,y)\\
&= \int_X d^2(x_0,y)\,d\nu(y) = \int_X d^2(x_0,x)\,d\nu(x).
\end{aligned}
$$

Hence, $W_2(\delta_{x_0},\nu)=\left[\int_X d^2(x_0,x)\,d\nu(x)\right]^{1/2}$.
If $y_0\in X$ then setting $\nu=\delta_{y_0}$, we get $(*)$.
(2) Clearly, $x\mapsto\delta_x$ is an injection from X to $\mathcal{P}_2(X)$.
Moreover, $\forall x_0,y_0\in X$, $\forall\gamma\in\mathcal{A}(\delta_{x_0},\delta_{y_0})$:

$$\gamma((X\setminus\{x_0\})\times X)=\delta_{x_0}(X\setminus\{x_0\})=0,$$

hence, $\gamma((X\setminus\{x_0\})\times\{y_0\})=0$. Thus,

$$\gamma(\{x_0\}\times\{y_0\})=\gamma(X\times\{y_0\})-\gamma((X\setminus\{x_0\})\times\{y_0\})$$
$$=\delta_{y_0}(\{y_0\})-0=1.$$

Hence, $\int_{X\times X} d^2(x,y)\,d\gamma(x,y)=\int_{X\times X} d^2(x_0,y_0)\,d\gamma(x,y)=d^2(x_0,y_0)$,

and $W_2(\delta_{x_0},\delta_{y_0})=d(x_0,y_0)$. QED

Example Let X be a (separable) Hilbert space, $a\in X$, and $\mu\in\mathcal{P}_2(X)$. Then, by the above Proposition,

$$W_2(\mu,\delta_a)=\left[\int_X \|x-a\|^2\,d\mu(x)\right]^{1/2}.$$

Let $m=\int_X x\,d\mu(x)\in X$, the mean of μ. We have

$$\|m\|\le\left(\int_X \|x\|^2\,d\mu(x)\right)^{1/2}<\infty.$$

Claim $m\in X$ is the unique minimizer for

$$\inf_{a\in X} W_2(\delta_a,\mu).$$

Moreover, $W_2(\delta_m,\mu)$ is the variance of μ.

Proof $W_2^2(\delta_a,\mu)-W_2^2(\delta_m,\mu)=\int_X\left(\|x-a\|^2-\|x-m\|^2\right)d\mu(x)$

$$=\int_X\left(\|a\|^2+\|m\|^2-2\langle x,a-m\rangle\right)d\mu(x)$$
$$=\|a\|^2+\|m\|^2-2\left\langle\int_X x\,d\mu(x),a-m\right\rangle$$
$$=\|a\|^2+\|m\|^2-2\langle m,a-m\rangle$$
$$=\|a-m\|^2.$$

Hence, $W_2(\delta_a,\mu)$ is uniquely minimized at $a=m$.

Now,
$$W_2(\delta_m, \mu) = \left(\int_X \|x - m\|^2 \, d\mu(x) \right)^{1/2}$$
is the variance of μ.