Lecture 16, Monday, 5/2/2022
Today: Wasserstein metric.

Set up $X, Y$: Polish. $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$
$C : X \times X \rightarrow [0, \infty]$: measurable,
$\mathcal{A}(\mu, \nu) \equiv \{ \gamma \in \mathcal{P}(X \times Y) : \pi_X^\# \gamma = \mu, \pi_Y^\# \gamma = \nu \}$
$E_k[\gamma] = \int_{X \times Y} c(x, y) d\gamma(x, y), \forall \gamma \in \mathcal{A}(\mu, \nu)$.

Existence Theorem: If $C$ is lower semicontinuous, then $\exists \hat{\gamma} = \arg \min \{ E_k[\gamma] : \gamma \in \mathcal{A}(\mu, \nu) \}$.

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

Remark: Let $\mu \in \mathcal{P}(X)$. Then $\exists x_0 \in X \text{ s.t. } \int d^2(x, x_0) d\mu(x) < \infty \iff \forall x' \in X \text{ s.t. } \int d^2(x, x') d\mu(x) < \infty$.
Since, $\forall x' \in X$.

by the triangle inequality, $d^2(x, x') \leq 2d^2(x, x_0) + 2d^2(x_0, x')$.

- $\forall \mu, \nu \in \mathcal{P}_2(X)$, define

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

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

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

Remarks
- Call $W_2$ the 2-Wasserstein or just the
Wasserstein metric (W-metric).

- The result can be extended to the $p$-Wasserstein metric for $1 \leq p < \infty$. 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 $\gamma^{12}, \gamma^{13} \in \mathcal{A}(\mu_1, \mu_2)$ and $\gamma^{23} \in \mathcal{A}(\mu_2, \mu_3)$. Then there exists $\gamma \in \mathcal{P}(X_1 \times X_2 \times X_3)$ such that

$$\pi^{1,2}_\# \gamma = \gamma^{12} \quad \text{and} \quad \pi^{2,3}_\# (\gamma) = \gamma^{23}.$$  

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 $x, y \in X$,

$$d^2(x, y) \leq 2d^2(x, x_0) + 2d^2(x_0, y).$$

So, $\forall \gamma \in \mathcal{A}(\mu, \nu)$,

$$\langle \mathbb{E} \chi [\gamma] \rangle \leq 2 \int_X d^2(x, x_0) d\gamma + 2 \int_{X \times X} d^2(x, y) d\gamma \quad \forall x, y \in X,

= 2 \int_X d^2(x, x_0) d\mu(x) + 2 \int_Y d^2(x_0, y) d\nu(y) < \infty.$$
Hence, \( W_2(\mu, \nu) < \infty \).

Clearly, \( W_2(\mu, \nu) \geq 0 \) for all \( \mu, \nu \in P_2(X) \). We show that \( W_2(\mu, \nu) = 0 \iff \mu = \nu \). Suppose \( \mu = \nu \). We construct \( \gamma \in A(\mu, \mu) \) such that \( \gamma(S) = 0 \) where 

\[ S = \{(x, y) \in X \times X : x \neq y \} \]

This implies that 

\[ E_\mu[\gamma] = \int_X \int_Y d^2(x, y) \, d\mu(x, y) = \int_X \int_Y d^2(x, y) \, d\nu(x, y) = 0 \]

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

Define \( Id \times Id : X \to X \times X \) by \( (Id \times Id)(x) = (x, x) \) and \( \gamma = (Id \times Id)^\tau \mu \). Since \( Id \times Id \) is continuous, \( \gamma \) is a Borel measure on \( X \times X \). Moreover, 

\[ \gamma(X \times X) = \mu \left( (Id \times Id)^\tau \gamma \right) (X \times X) = \mu(X) = 1 \]

So, \( \gamma \in P(X \times X) \).

If \( A \in P(X) \) then 

\[ \gamma(A \times X) = \mu \left( (Id \times Id)^\tau \gamma \right) (A \times X) = \mu(A) \]

\[ a \in (Id \times Id)^\tau \gamma(A \times X) \iff (Id \times Id)(a) = (a, a) \in A \times X \iff a \in \gamma(A) \]

Similarly, \( \gamma(X \times A) = \mu(A) \). Hence, \( \gamma \in A(\mu, \mu) \). Finally, for \( S = \{(x, y) \in X \times Y : x \neq y \} \), we have \( (Id \times Id)^\tau \gamma(S) = 0 \) \( \iff S \subset X \). Hence, \( \gamma(S) = \mu(S) = 0 \).

Now, assume \( \mu, \nu \in P_2(X) \) and \( W_2(\mu, \nu) = 0 \). Let \( \gamma \in A(\mu, \nu) \) be such that \( W_2(\mu, \nu) = E_\mu[\gamma] = 0 \), i.e., 

\[ \int_X \int_Y d^2(x, y) \, d\gamma(x, y) = 0 \]

Since \( d(x, y) \geq 0 \) for all \( x, y \in X \times Y \), \( \gamma(S) = 0 \), where 

\[ S = \{(x, y) \in X \times Y : x \neq y \} \]

i.e., \( \gamma \) is concentrated
on the diagonal $D = \{(x, x) : x \in X\}$. Now, for any bounded Borel function $f : X \to \mathbb{R}$, since $\nu_\# \gamma = \mu$, $\mu_\# \gamma = \nu$, we have
\[
\int_X f(x) \, d\mu(x) = \int_X f(x) \, d\nu(x) = \int_D f(x) \, d\gamma(x, y) = \int_X f(y) \, d\nu(y).
\]
Hence $\mu = \nu$.

\(\Box\)

Let $\gamma \in \mathcal{P}(X \times Y)$. Define $\gamma^T : \mathcal{P}(X \times Y) \to \mathbb{R}$ by $\gamma^T(A \times B) = \gamma(B \times A)$. Then, $\gamma^T \in \mathcal{P}(X \times Y)$. Moreover,
\[
(\tau^X_\# \gamma^T)(A) = \gamma^T((\tau^X_\#)^{-1}(A)) = \gamma^T(A \times Y) = \gamma(Y \times A) = \nu(A),
\]
and
\[
(\tau^Y_\# \gamma^T)(B) = \mu(B).
\]
Hence $\gamma^T \in \mathcal{A}(\nu, \mu)$. Also, $d(x, y) = d(y, x)$ for all $x, y \in X$. Using the change of variables $T(x, y) = (y, x)$, we have $T_\# \gamma = \tilde{\gamma}$, and
\[
\int_{X \times X} d^2(x, y) \, d\gamma(x, y) = \int_{X \times X} d^2(y, x) \, d\gamma(x, y) = \int_{X \times X} d^2(y, x) \, d\tilde{\gamma}(x, y) = \int_{X \times X} d^2(x, y) \, d\tilde{\gamma}(x, y).
\]
(Hence, $W_2(\mu, \nu) = W_2(\nu, \mu)$)

\(\Box\)

The triangle inequality. Let $\gamma, \gamma_1, \gamma_2 \in \mathcal{P}_2(X)$. Show that
\[
W_2(\mu_1, \mu_2) \leq W_2(\mu_1, \mu_3) + W_2(\mu_2, \mu_3).
\]
Let $\gamma_{12} \in \mathcal{A}(\mu_1, \mu_2)$ and $\gamma_{23} \in \mathcal{A}(\mu_2, \mu_3)$ be such
that \( W_2^2(\mu_1, \mu_2) = E_K[\gamma^{1,2}] \) and \( W_2^3(\mu_1, \mu_3) = E_K[\gamma^{2,3}] \).

By the lemma, \( \exists \gamma \in \mathcal{P}_2(X \times X \times X) \) such that \( \gamma^{1,2} = \gamma^{1,2} \) and \( \gamma^{2,3} = \gamma^{2,3} \). Thus

\[
W_2(\mu_1, \mu_3) \leq \| d(x_1, x_3) \|_{L^2(\gamma)} \qquad (1st \, marginal \, of \, \gamma \, is \, \mu_1)
\]

\[
\leq \| d(x_1, x_2) \|_{L^2(\gamma)} + \| d(x_2, x_3) \|_{L^2(\gamma)}
\]

\[
= \| d(x_1, x) \|_{L^2(\gamma^{1,2})} + \| d(x_2, x) \|_{L^2(\gamma^{2,3})}
\]

\[
= W_2(\mu_1, \mu_1) + W_2(\mu_2, \mu_3). \quad \square
\]

**Proposition** Let \((X, d)\) be a Polish space.

1. Let \( x_0 \in X \) and \( \nu \in \mathcal{P}_2(X) \). If \( \gamma \in \mathcal{A}(d_{x_0}, \nu) \) then

\[
\int_{X \times X \times X} d^2(x, y) \, d\gamma(x, y) = \int_{X \times X} d^2(x, x) \, d\nu(x),
\]

and hence \( W_2^1(\delta_{x_0}, \nu) = \int_X d^2(x, x) \, d\nu(x) \).

2. The mapping \( x \mapsto \delta_x \) is an isometric from \( X \) to \( \mathcal{P}_2(X, W_2) \).

**Proof** (1) If \( \gamma \in \mathcal{A}(d_{x_0}, \nu) \) then \( \gamma(\{x_0\} \times X) = d_{x_0}(x_0) \)

\(= 1 \). So,

\[
\int_{X \times X \times X} d^2(x, y) \, d\gamma(x, y) = \int_{X \times X} d^2(x, x) \, d\nu(x, y)
\]

\[
= \int_X d^2(x, y) \, d\nu(y) = \int_X d^2(x, x) \, d\nu(x).
\]

Hence, \( W_2(\delta_{x_0}, \nu) = \left( \int_X d^2(x, x) \, d\nu(x) \right)^{1/2} \).

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, \( \forall x_0, y_0 \in X, \forall \gamma \in \mathcal{A}(d_{x_0}, d_{y_0}) \)
\[ \gamma((X \setminus \{x_0\}) \times X) = \delta_{x_0}(X \setminus \{x_0\}) = 0, \]
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,
\[ \int \gamma^2(x, y) d\gamma(x, y) = \int_{X \times X} \gamma^2(x_0, y_0) d\delta(x, y) = \|x_0\|^2, \]
and \[ W_2(\delta_{x_0}, \delta_{y_0}) = \|x_0 - y_0\|. \quad Q.E.D. \]

**Example** Let \( X \) be a (separable) Hilbert space, \( a \in X \), and \( \mu \in P_2(X) \). 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 \( \mu \). We have
\[ \|m\| \leq \left( \int X x^2 \, d\mu(x) \right)^{1/2} < \infty. \]

**Claim** \( m \in X \) is the unique minimizer for \( \min_{a \in X} W_2(\delta_a, \mu). \)

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

**Proof**
\[ W_2^2(\delta_a, \mu) - W_2^2(\delta_m, \mu) = \int X (\|x-a\|^2 - \|x-m\|^2) \, d\mu(x) \]
\[ = \int X (\|a\|^2 + \|m\|^2 - 2 \langle x, a-m \rangle) \, d\mu(x) \]
\[ = \|a\|^2 + \|m\|^2 - 2 \| m \| \langle m, a-m \rangle \]
\[ = \|a-m\|^2. \]

(Hence, \( W_2(\delta_a, \mu) \) is uniquely minimized at \( a = m \).
Now, \[ W_2(\mu, \nu) = \left( \int_{\mathbb{X}} \| x - m \|^2 \, d\nu(x) \right)^{\frac{1}{2}} \]
is the variance of \( \mu \).