Lecture 6. Friday, 4/8/2022

⊙ OT: Measure-theoretical descriptions

Consider two probability spaces $(X, P(X), \mu)$ and $(Y, P(Y), \nu)$:

$X = \{x_1, \cdots, x_m\}$ and $Y = \{y_1, \cdots, y_n\}$.

$P(X) = \{$all subsets of $X\}$, $P(Y) = \{$all subsets of $Y\}$.

$\mu : P(X) \to [0,1]$ and $\nu : P(Y) \to [0,1]$ are probability measures (additive, and the total measure $= 1$).

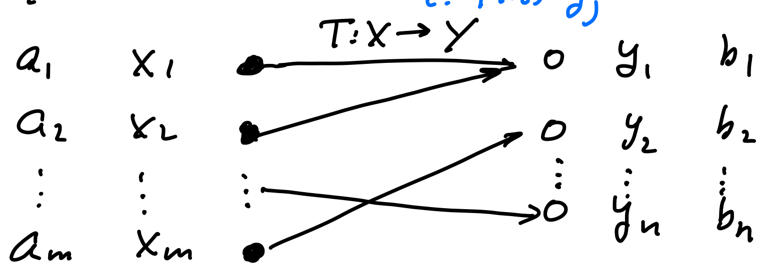
Note. μ and ν are completely determined by $\mu(\{x_i\})$ $(i = 1, \cdots, m)$ and $\nu(\{y_j\})$ $(j = 1, \cdots, n)$, respectively.

Denote $a = (\mu(\{x_i\})) \in \mathbb{R}^m$ and $b = (\nu(\{y_j\})) \in \mathbb{R}^n$.

Since $\mu(\{x_i\}) \geq 0$ $\forall i$ and $\sum_{i=1}^m \mu(\{x_i\}) = \mu(X) = 1$, we have $a \in \mathcal{P}_m$. Similarly, $b \in \mathcal{P}_n$.

Recall: In Monge's formulation of the discrete OT prob., we define

$$\mathcal{J}(a,b) = \{ T : X \to Y : \ b_j = \sum_{i:\, T(x_i) = y_j} a_i, \ \forall j \in \{1, \cdots, n\} \}.$$



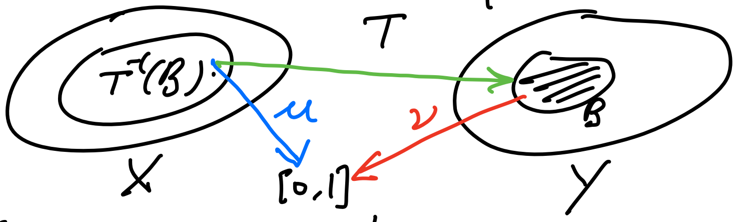
Since $a = \mu(\{x_i\})$ and $b = \nu(\{y_j\})$, the constraints $b_j = \sum_{i:\, T(x_i) = y_j} a_i$ $\forall j$, for $T : X \to Y$, really means

$$\nu(\{y_j\}) = \sum_{i:\, T(x_i) = y_j} \mu(\{x_i\}) = \mu\Big(\bigcup_{i:\, T(x_i) = y_j} \{x_i\} \Big) \ \forall j.$$

Fix j: $\bigcup_{i:\, T(x_i) = y_j} \{x_i\} = T^{-1}(\{y_j\})$.

Recall: $T: X \to Y \Longrightarrow T^{-1}: P(Y) \to P(X)$.

$T^{-1}(B) = \{\alpha \in X : T(\alpha) = \beta\}$, pre-image of B.



T^{-1} is additive:
$$T^{-1}(\{y_{i_1}, \dots, y_{i_k}\}) = \bigcup_{\ell=1}^{k} T^{-1}(\{y_{i_\ell}\}).$$

$$\nu(\{y_j\}) = \mu(T^{-1}(\{y_j\})) \quad \forall j.$$

Since both μ and T^{-1} are additive (and X and Y are finite), we have
$$\forall B \subseteq Y: \quad \nu(B) = \mu(T^{-1}(B)) = (\mu \circ T^{-1})(B)$$

Definition Given two measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, μ a measure on $(X, \mathcal{A})$, and $T: X \to Y$. Define $T_\# \mu : \mathcal{B} \to \mathcal{A}$ by $T_\#\mu(B) = (\mu \circ T^{-1})(B) = \mu(T^{-1}(B))$. $\forall B \in \mathcal{B}$, i.e. $T_\#\mu = \mu \circ T^{-1}$.

One can verify that $T_\#\mu$ is a measure on $(Y, \mathcal{B})$. Call $T_\#\mu = \mu \circ T^{-1}$ the push-forward measure.

Denote $\mathcal{M}(X) = \mathcal{M}(X, \mathcal{A})$ and $\mathcal{M}(Y) = \mathcal{M}(Y, \mathcal{B})$ the sets of all measures on $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ resp.

Call the operator $T_\# : \mathcal{M}(X) \to \mathcal{M}(Y)$, $T_\#\mu = \mu \circ T^{-1}$ $\forall \mu \in \mathcal{M}(X)$, the push-forward operator.

Now, back to $X = \{x_1, \dots, x_m\}$, $Y = \{y_1, \dots, y_n\}$, and μ, ν are prob. measures on $(X, P(X))$, $(Y, P(Y))$, resp.

Denote $\mathcal{J}(\mu, \nu) = \{T : X \to Y : T_\#\mu = \nu\}$.

Mass conservation!

Given $c : X \times Y \to [0,\infty)$. It's completely determined by $c(x_i, y_j)$ $(i=1,\cdots,m;\ j=1,\cdots,n)$. Define ~~the~~ cost

$$E_M[T] = \int_X c(x, T(x))\, d\mu(x) \quad \forall T \in \mathcal{T}(\mu, \nu)$$

<u>Monge's OT:</u> Find $\hat{T} \in \mathcal{T}(\mu,\nu)$ such that

$$E_M[\hat{T}] = \min_{T \in \mathcal{T}(\mu,\nu)} E_M[T].$$

$$E_M[T] = \int_X c(x, T(x))\, d\mu(x) = \int_{\bigcup_{i=1}^m \{x_i\}} c(x, T(x))\, d\mu(x)$$

$$= \sum_{i=1}^m \int_{\{x_i\}} c(x_i, T(x_i))\, d\mu = \sum_{i=1}^m c(x_i, T(x_i))\, \mu(\{x_i\})$$

$$= \sum_{i=1}^m a_i c(x_i, T(x_i)), \quad \text{where } a_i = \mu(\{x_i\}) \ \forall i.$$ $\boxed{\text{Same as before!}}$

Now, consider measure-theoretic Kantorovich's OT formulation.

Given $X = \{x_1, \cdots, x_m\}$, μ: a prob. measure on X,
$Y = \{y_1, \cdots, y_n\}$, ν: a prob. measure on Y.
$c : X \times Y \to [0, \infty)$, a cost function.

Denote $a = (a_i) = (\mu(\{x_i\})) \in \mathcal{P}_m$, $b = (b_j) = (\nu\{y_j\}) \in \mathcal{P}_n$.

Recall $\mathcal{A}(a,b) = \{ P = [P_{ij}] \in \mathbb{R}^{m\times n} : P \geq 0,\ \sum_j P_{ij} = a_i\ \forall i,\ \sum_i P_{ij} = b_j\ \forall j \}$.

$P \in \mathcal{A}(a,b) \Longrightarrow P_{ij} \geq 0$, $\sum_{i,j} P_{ij} = 1$. So, each $P \in \mathcal{A}(a,b)$ defines a prob. measure γ on $X \times Y$, or $(X\times Y, \mathcal{P}(X\times Y))$:

$\gamma((x_i, y_j)) = P_{ij}\ \forall i,j$. Or: for any

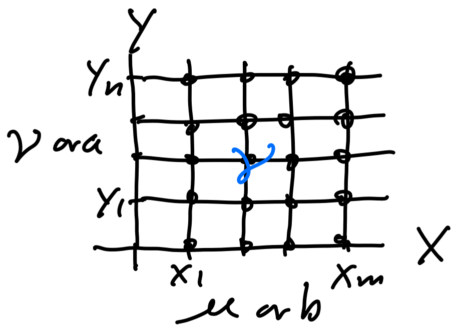
$S = \{(x_i, y_j) \in X\times Y : i \in I,\ j \in J\} \subseteq X \times Y$, $I \subseteq \{1,\cdots,m\}$, $J \subseteq \{1,\cdots,n\}$

$$\gamma(S) = \sum_{i\in I} \sum_{j \in J} P_{ij}.$$

Row sum of $P = a \iff \forall i:\ a_i = \sum_{j=1}^{n} P_{ij}$. i.e.,

$$\mu(\{x_i\}) = \sum_{j=1}^{n} \gamma((x_i, y_j)) = \gamma\Big(\bigcup_{j=1}^{n} (x_i, y_j)\Big) = \gamma(\{x_i\}\times Y)$$

Hence, for any $A \subseteq X$, $\mu(A) = \gamma(A \times Y)$. Similarly, $\forall B \subseteq Y$. $\nu(B) = \gamma(X \times B)$. Moreover $\gamma(A\times B) \neq \mu(A)\nu(B)$.



Row sum of $P = a \Rightarrow \gamma|_X = \mu$

Col. sum of $P = b \Rightarrow \gamma|_Y = \nu$.

Define $\pi_X : X\times Y \to X$ and $\pi_Y : X \times Y \to Y$

$\pi_X((x,y)) = x$, $\pi_Y((x,y)) = y$ $\forall (x,y) \in X\times Y$

$A \subseteq X \Longrightarrow \mu(A) = \gamma(A\times Y)$. But $A = \pi_X(A\times Y)$. So,

$A\times Y = \pi_X^{-1}(A)$, and $\mu(A) = \gamma(\pi_X^{-1}(A)) = (\pi_X)_{\#}\gamma(A)$.

$\mu = (\pi_X)_{\#}\gamma$ and $\nu = (\pi_Y)_{\#}\gamma$: marginal probabilities.

Define $\mathcal{A}(\mu,\nu) = \{$probability measures γ on $X\times Y$ with marginal measures on X, Y being μ, ν, resp.$\}$

Given $X = \{x_1, \dots, x_m\}$, μ: a prob. measure on X,
$Y = \{y_1, \dots, y_n\}$, ν: a prob. measure on Y.
$c : X\times Y \to [0,\infty)$. a cost function.

Define $\mathcal{A}(\mu,\nu)$ as above.

Define ~~the~~ cost $E_K[\gamma] = \int_{X\times Y} c(x,y)\,d\gamma(x,y)$.

Kantorovich's (discrete) OT prob.

$$\min_{\gamma\in\mathcal{A}(\mu,\nu)} E_K[\gamma].$$

Given $\gamma \in \mathcal{A}(\mu,\nu)$. What is the cost $E_K[\gamma]$ in terms of a and b?

$$E_K[\gamma] = \int_{X\times Y} c(x,y)\,d\gamma(x,y) = \sum_{i,j} \int_{\{(x_i,y_j)\}} c(x,y)\,d\gamma(x,y)$$

$$= \sum_{i,j} c(x_i, y_j)\,\gamma(\{(x_i,y_j)\}) = \sum_{i,j} C_{ij} P_{ij} = \langle C, P \rangle, \quad P \in \mathcal{A}(a,b).$$

Thus, the discrete OT can be formulated using measures.

Plan:
- Regularization of discrete OT problems
- Sinkhorn's algorithm and related topics.