Last time I defined the Kantorovich metric
P, Q are probability distributions

$$K(P,Q) = \sup_{f \in Lip} \left| \int f\,dP - \int f\,dQ \right|$$

where $f : (X,d) \to \mathbb{R}$ satisfies

$$\forall\, x, x' \in X \quad |f(x) - f(x')| \leq d(x,x').$$

We say f is 1-Lipschitz or nonexpansive.
It is very easy to verify that this is a metric.

Here is an apparently different metric also due to Kantorovich

$$W_1(P,Q) = \inf_{\pi \in \mathcal{C}(P,Q)} \int_{X \times X} d(x,y)\,d\pi$$

Let us deconstruct this to see what it means. Recall that $\mathcal{C}(P,Q)$ is the set of couplings of P, Q: measures on $X \times X$ s.t. the marginals are P and Q. We think of each distribution P, Q as representing a "pile of sand" on X. We want to move sand around so that the pile P is transformed into the pile Q. There are many possible "transportation plans". A coupling is exactly such a transport plan. Thinking discretely for a moment $\pi(x,y)$ tells you how much to move from x to y. The cost of moving something depends on how far they are so if we use the plan π the total cost is $\int d(x,y)\,d\pi$. Hence W_1 is the minimum cost of any plan

The Kantorovich - Rubinstein duality theorem :

$$K = W_1$$

I will not prove this but discuss some examples based on finite spaces.

If X is a finite set equipped with a metric d then we can write K as a linear program: We assume $X = \{x_1, \ldots x_i, \ldots x_n\}$. We introduce variables a_i $i = 1 \ldots n$ and we seek to maximize

$$\max \sum_{i=1}^{n} a_i \left(P(x_i) - Q(x_i) \right)$$

subject to the constraints $\forall i, j$

$$0 \leq a_i \leq 1 \qquad |a_i - a_j| \leq d(x_i, x_j)$$

Dual form new variables $l_{ij}, \alpha_i, \beta_j$ $i, j = 1 \ldots n$

$$\min \sum_{i,j=1}^{n} l_{ij} d(x_i, x_j) + \sum_{i=1}^{n} \alpha_i + \sum_{j=1}^{n} \beta_j$$

subject to

$$\forall i \quad \sum_{j=1}^{n} l_{ij} + \alpha_i = P(x_i)$$

$$\forall j \quad \sum_{i=1}^{n} l_{ij} + \beta_j = Q(x_j)$$

$$\forall i, j. \; l_{ij}, x_i, y_j \geq 0$$

(Example I) $W_1(P, P) = 0$

set $l_{ij} = \delta_{ij} P(x_i)$ $x_i = y_j = 0$

then the sum we are minimizing is 0 & clearly we cannot go below 0 so this must be the minimal value so $W_1(P, P) = 0$.

<u>EXAMPLE 2.</u> Let x, y be two points in X, let δ_x be the probability measure concentrated at x & similarly for δ_y and suppose $d(x,y) = r$.

From the dual form we get an upper bound

choose $l_{xy} = 1$ & all other $l_{ij} = 0$ and $x_i = 0$ & $y_j = 0$

$$\sum l_{ij}\, d(x_i, x_j) = d(x,y) = r.$$

From the primal we get a lower bound

Choose $a_x = 0$, $a_y = r$ and all others to match the constraints then we have

$$\sum (\delta_y(x_i) - \delta_x(x_i))\, a_i$$
$$= r$$

Thus $W_1(\delta_x, \delta_y) = r = d(x,y)$

Thus (X, d) is isometrically embedded in the space of probability distributions.

The x_i & y_j are used in case we are dealing with subprobability distributions.

—— x ——

The version of duality can be greatly generalized. The cost does not have to be a metric and the sup can involve 2 different functions. The theorems can be stated for complete separable metric spaces. See "Optimal Transport, Old & New" by Cedric Villani.

—— x ——

Related distances

$$W_p(P,Q) = \left(\inf_{\pi \in \mathcal{C}(P,Q)} \int d(x,y)^p \, d\pi \right)^{1/p}$$

W_1, W_2 are both commonly used.