Math 612: Single Cell Analysis **2019W Term 1**

Lecture 14: October 22

Lecturer: Geoffrey Schiebinger *Scribe: Sibyl Drissler*

Definition 14.1 *A metric is a function that defines a distance. A metric satisfies*

1. *symmetry* $d(x, y) = d(y, x)$
2. $d(x, y) = 0$
3. *triangle inequality* $d(A, C) \leq d(A, B) + d(B, C)$

Distances for $\mathbb{R}^n$

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

$$d(x, y) = ||x - y||_p = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}$$

$$d(x, y) = ||x - y||_\infty = \max_i |x_i - y_i|$$

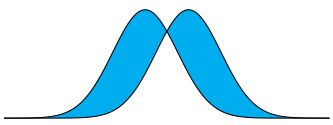
The first two distances above come with an inner product. $||x-y||_\infty$ does not have an inner product. Neither do the distances on probability distributions.

14.1 Distances for Probability Distributions

For probability distributions $\mathbb{P}, \mathbb{Q}$ with densities p, q on $\chi = \{x_1, ..., x_n\}$

Total variation distance

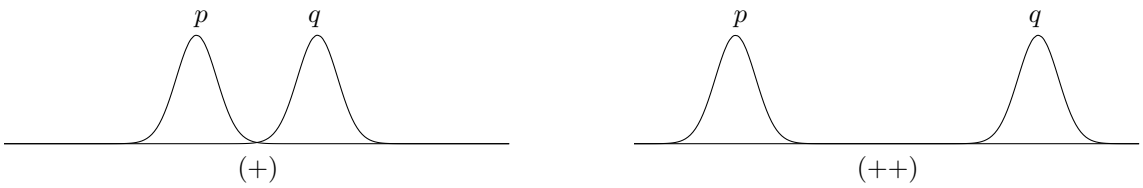
$$d_{T_v}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \sum_{i=1}^{n} |p(x_i) - q(x_i)|$$



Hellinger distance

$$d_H(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \sqrt{\sum_{i=1}^{n} (p(x_i) - q(x_i))^2}$$

The problem with d_{T_v} and d_H is that they cannot tell which of the following two cases are more different.



Optimal Transport

We will measure the distance via optimal transport because it can tell the difference between (+) and (++).

For simplicity, assume the state space is finite $\chi = \{x_1, ..., x_n\}$. Consider a distance d on χ. We will use this as a cost function in optimal transport.

Let D be the matrix of pairwise distances $D_{ij} = d(x_i, x_j)$.

Theorem 14.2 *The optimal transport distance between distributions a and b is*

$$W_p(a,b) = \left(\begin{array}{c} \min_{\pi} \sum_{j=1}^{n} \sum_{i=1}^{n} \pi_{ij} d^p(x_i, x_j) \\ \pi \mathbb{1} = a \\ \pi^T \mathbb{1} = b \\ \pi_{ij} \geq 0 \end{array} \right)^{1/p}$$

This is called the **p-Wasserstein distance**. It works for $p \geq 1$. We will mostly consider $p = 2$.

Metric properties of optimal transport.

It is easy to see that $W_p(a, b)$ is symmetric.

Also it is easy to see that

$$0 \leq W_p(a, a) \leq 0$$

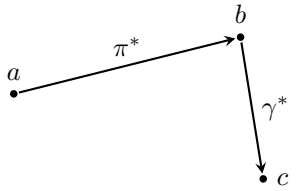
For the right inequality set $\pi_{ii} = a_i$ and the rest of the values to 0.

Claim: W_p satisfies the triangle inequality.

Proof:

Consider three points a, b, c

$$\sum a_i = 1 = \sum b_i = \sum c_i$$



Define a coupling

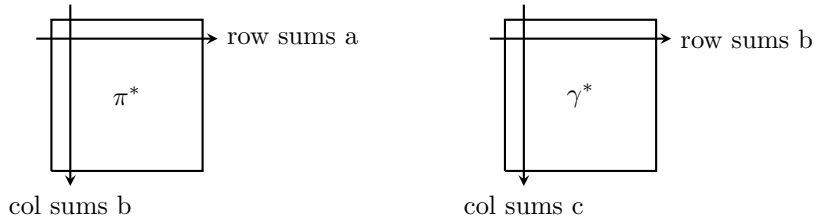
$$S = \pi^* diag(\frac{1}{\tilde{b}}) \gamma^*$$

Claim: S is a coupling of a and c

We can prove this as follows

$$S\mathbb{1} = \pi^* diag(\frac{1}{\tilde{b}}) \underbrace{\gamma^* \mathbb{1}}_{b} = \pi^* \mathbb{1} = a$$

Similarly, we can check that the column sums are c.



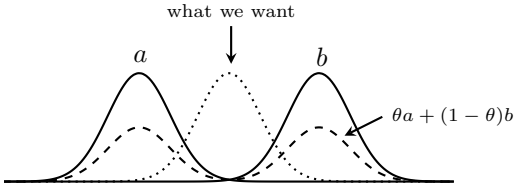
$$
\begin{aligned}
W_p(a,c) &= \Big(\min_{\pi} \sum_{ik} \pi_{ik} D_{ik}^p\Big)^{1/p} \\
&\le \Big(\sum_{ik} S_{ik} D_{ik}^p\Big)^{1/p} \\
&= \Big(\sum_{ik} D_{ik}^p \sum_j \pi_{ij}^* \frac{1}{\tilde{b_j}} \gamma_{jk}^*\Big)^{1/p} \\
&\le \Big(\sum_{ijk} (D_{ij} + D_{jk})^p \sum_j \pi_{ij}^* \frac{1}{\tilde{b_j}} \gamma_{jk}^*\Big)^{1/p} \quad \text{by the triangle inequality} \\
&\le \Big(\sum_{ijk} (D_{ij})^p \pi_{ij}^* \frac{1}{\tilde{b_j}} \gamma_{jk}^*\Big)^{1/p} + \Big(\sum_{ijk} (D_{jk})^p \pi_{ij}^* \frac{1}{\tilde{b_j}} \gamma_{jk}^*\Big)^{1/p} \\
&= \Big(\sum_{ij} (D_{ij})^p \pi_{ij}^* \underbrace{\sum_k \frac{\gamma_{jk}^*}{\tilde{b_j}}}_{1}\Big)^{1/p} + \Big(\sum_{jk} (D_{jk})^p \gamma_{jk}^* \underbrace{\sum_i \frac{\pi_{ij}^*}{\tilde{b_j}}}_{1}\Big)^{1/p} \\
&= W_p(a,b) + W_p(b,c)
\end{aligned}
$$

■

14.2 Curves

Aim: Find curve between distributions a and b.

When discussing convex functions we saw the formula $\theta x + (1-\theta)y$. This is not how we will find the curve because as we can see from the image below, $\theta a + (1-\theta)b$ looks like a mixture model.



The formula $\theta x + (1-\theta)y$ is a solution to an optimization problem. We are interested in the optimization problem.

Consider the problem in $\mathbb{R}^2$. Suppose we have $x, y \in \mathbb{R}^2$ and we want to find the midpoint $\frac{x+y}{2}$.

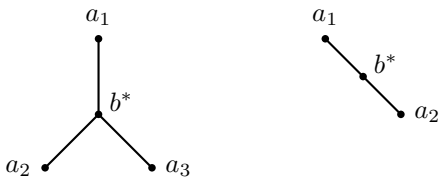
The problem that we want to solve is

$$z^* = \arg\min_z ||x-y||^2 + ||y-z||^2$$

The optimizer has $\frac{d}{dz} = 0 \to z(z-x) + z(z-y) = 0 \to z = \frac{x+y}{2}$.

Definition 14.3 *The barycenter (or center of mass) of two or more distributions $a_1, ..., a_n$ is*

$$b^* = \arg\min_b \sum_{i=1}^n W_p^p(b, a_i)\theta_i \qquad \sum \theta_i = 1$$



Definition 14.4 *Given a_0, a_1 a constant speed geodesic is a family of distributions $\{a_t\}_{t\in(0,1)}$ with*

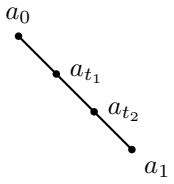
$$W_p(a_0, a_t) = tW_p(a_0, a_1)$$

We can construct a_t from an optimal coupling π^* of a_0, a_1.

Let

$$P_t : \chi \times \chi \to \chi$$
$$(x, y) \mapsto (1-t)x + ty$$

This is an averaging map.



Then define

$$\begin{aligned} a_t = \mathbb{E}_{\pi^*} \text{ of } & P_t(X, Y) \\ & X \sim a_o \qquad (X, Y) \sim \pi^* \\ & Y \sim a_1 \\ = & \sum_{ij=1}^n \pi_{ij} P_t(x_i, y_j) \end{aligned}$$

Claim: This is a constant speed geodesic, ie

$$W_p(a_0, a_t) = tW_p(a_0, a_1)$$

More generally,

$$W_p(a_{t_1}, a_{t_2}) = (t_2 - t_1)W_p(a_0, a_1)$$

Proof next lecture.