Lecture # 12 : Optimal Transport

- The $d=1$ case of optimal transport is unusually simple, in that the OT cost between $\mu,\nu \in \mathcal{P}(\mathbb{R})$ can be computed just based on integrating a suitable function of their cdfs.

- Recall that for any $\mu \in \mathcal{P}(\mathbb{R})$, the cumulative distribution function of μ is

$F_\mu : \mathbb{R} \to [0,1]$ $\quad F_\mu(t) := \mu((-\infty, t])$

$$= \int_{-\infty}^{t} d\mu$$

- Clearly $\lim_{t\to\infty} F(t) = 1$, $\lim_{t\to-\infty} F(t) = 0$, and one can show F_μ is càdlàg, i.e. continuous from the right.

- Moreover, F_μ is non-decreasing, albeit not strictly monotone increasing in general. Still, we can define a generalized inverse function $F_\mu^{-1} : [0,1] \to \mathbb{R}$

$$F_\mu^{-1}(s) = \inf_{t\in\mathbb{R}} \{F(t) \geq s\}$$

If F is in fact strictly increasing, then this coincides with the usual inverse.

In general, we have
- $F_\mu(F_\mu^{-1}(s)) \geq s \quad \forall s \in [0,1]$
- $F_\mu^{-1}(F_\mu(t)) \leq t \quad \forall t \in \mathbb{R}$

<u>Theorem</u> (OT on $\mathbb{R}$, Quadratic Cost): Let $\mu,\nu \in \mathcal{P}(\mathbb{R})$ and suppose we consider the quadratic cost function $c(x,y) = |x-y|^2$. Let F_μ, F_ν be the cdfs of μ,ν respectively and

F_μ^{-1}, F_ν^{-1} their generalized inverses. Let $\pi \in \mathcal{P}(\mathbb{R}^2)$ have the property that for any rectangle $R(x_0, y_0) = \{(x,y) \mid x \le x_0, y \le y_0\}$,

$$\int_{R(x_0,y_0)} d\pi = \min\{F_\mu(x_0), F_\nu(y_0)\} =: H(x_0,y_0)$$

Then (i) $\pi \in \Pi(\mu,\nu)$

(ii) π is optimal for the "primal" (KP)

(iii) $$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |x-y|^2 d\pi = \int_0^1 |F_\mu^{-1}(s) - F_\nu^{-1}(s)|^2 ds$$

Proof: We aim to invoke Knott-Smith optimality to establish (ii). To see (i), note that for any $\underset{(-\infty,a)}{A \overset{\shortparallel}{\subset}} \mathbb{R}$,

$$\pi[A\times\mathbb{R}] = \min\{F_\mu(a), F_\nu(+\infty)\} = F_\mu(a) = \mu(A)$$

and similarly $\pi(\mathbb{R}\times(-\infty,b)) = \nu(B)$. Intersections and unions of such intervals (and complements) generate the Borel sets, so this establishes $\pi \in \Pi(\mu,\nu)$.

Now let $F(x^-) = \lim_{x_0 \to x^-} F(x)$ denote the left-hand limit of a cdf F at x. This exists by monotonicity, and by right continuity $\lim_{x \to x_0^+} F_\nu(x) = F_\nu(x_0)$.

__Claim__: $\operatorname{supp}(\pi) \subset \{(x,y) \in \mathbb{R}^2 \mid F_\mu(x^-) \le F_\nu(y) \text{ and } F_\nu(y^-) \le F_\mu(x)\}$

Suppose $F_\mu(x^-) > F_\nu(y)$. By F_μ, F_ν ~~mon~~ non-decreasing and right continuity for (x', y') ~~wr~~ sufficiently close to (x,y), $F_\mu(x') > F_\nu(y')$ as well! ~~thing~~

$$\Rightarrow H(x',y') = \min\{F_\mu(x'), F_\nu(y')\}$$

$= F_\nu(y^-)$. So, on some small rectangle around (x, y), the value of $H(x, y)$ does not depend on x. But then integrating $d\pi$ over this rectangle cannot have non-zero value, because for any $R(x_0, y_0)$, ③

$$\int_{R(x_0,y_0)} d\pi = H(x_0, y_0)$$

$\Rightarrow (x,y) \notin \mathrm{supp}(\pi)$. Similarly, if $F_\mu(y^-) \geq F_\nu(x)$ then $(x,y) \notin \mathrm{supp}(\pi)$.

So, we must establish that $\{(x,y) \in \mathbb{R}^2 \mid F_\mu(x^-) \leq F_\nu(y) \text{ and } F_\mu(y^-) \leq F_\nu(x)\}$ (✪) is the graph of the subdifferential of some convex function. We can use the special geometry of $\mathbb{R}$ to make the following observations:

- gradients (derivatives) of convex functions are exactly the non-decreasing functions
- subgradients of convex functions are of the form $\Gamma \subset \mathbb{R}^2$ s.t.

$$(x_1, y_1), (x_2, y_2) \in \Gamma \Rightarrow \begin{cases} x_1 \leq x_2 \text{ and } y_1 \leq y_2 \\ \text{or} \\ x_1 \geq x_2 \text{ and } y_1 \geq y_2 \end{cases}$$

(✪✪) $\iff (x_1 - x_2)(y_1 - y_2) \geq 0$

So, we will show our set (✪) satisfies (✪✪). Suppose $x_1 > x_2$. Then we must show $y_1 \geq y_2$. Since $(x_1, y_1), (x_2, y_2) \in \Gamma$, we know $F_\nu(y_1) \geq F_\mu(x_1^-) \underset{F_\mu \text{ non-decreasing}}{\geq} F_\mu(x_2)$

$\geq F_\nu(y_2^-)$

So, we have $F_\nu(y_1) \geq F_\nu(y_2^-)$. If this is in fact a strict inequality, then $y_1 \geq y_2$ by left continuity, which gives the desired result. So, it suffices to consider the case

$F_\nu(y_1) = F_\nu(y_2^-)$. But this implies that $F_\nu(y_1) = F_\mu(x_1^-) = F_\mu(x_2) = F_\nu(y_2^-)$

Suppose that, for purposes of contradiction, that $y_2 > y_1$. Then since F_μ, F_ν are non-decreasing, we have that F_μ is constant on $[x_2, x_1)$ and F_ν on $[y_1, y_2)$.

<u>Exercise</u>: Show there exist $\varepsilon > 0$ s.t. the rectangle with vertices $(x_2-\varepsilon, y_2-\varepsilon)$ and $(x_2+\varepsilon, y_2+\varepsilon)$ has π-measure 0. (<u>Hint</u>: use that F_μ, F_ν are constant on $[x_2, x_1)$ and $[y_1, y_2)$, respectively.)

But $(x_2, y_2) \in \mathrm{supp}(\pi)$, which contradicts $\pi(\text{rectangle}) = 0 \Rightarrow$ contradiction

$\Rightarrow y_2 \leq y_1$, as needed. $\square$

So, by KS-optimality, (ii) holds.

To see (iii), let ν be the uniform distribution on $[0,1]$ (equivalently, the Lebesgue measure). It is enough to show $\pi = (F_\mu^{-1} \times F_\nu^{-1})_\# \nu$, since then for any measurable, non-negative $f: \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$,

$$\int_{\mathbb{R}^2} f(x,y)\, d\pi(x,y) = \int_0^1 f(F_\mu^{-1}(s), F_\nu^{-1}(s))\, ds$$

Taking $f(x,y) = |x-y|^2$ would give $\int_{\mathbb{R}^2} |x-y|^2\, d\pi(x,y) = \int_0^1 |F_\mu^{-1}(s) - F_\nu^{-1}(s)|^2\, ds$

To establish $\pi = (F_\mu^{-1} \times F_\nu^{-1})_{\#} \nu$, note that it is sufficient to show it (as before) on arbitrary rectangles. We leave this as an exercise. ∎