Geometry of Entropic Optimal Transport.

Suppose π^ε is the solution of ε-entropic optimal transport problem between $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ w.r.t. the cost function c.

Consider $x_1, x_2 \in \mathrm{supp}(\mu)$ and $y_1, y_2 \in \mathrm{supp}(\nu)$

Then
$$\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_1,y_1)\,\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_2,y_2)$$
$$= e^{-\frac{1}{\varepsilon}\left(c(x_1,y_1)+c(x_2,y_2)\right)} \times e^{\frac{1}{\varepsilon}\left(\phi_\varepsilon(x_1)+\phi_\varepsilon(x_2)\right)} \times e^{\frac{1}{\varepsilon}\left(\psi_\varepsilon(y_1)+\psi_\varepsilon(y_2)\right)}$$

and,
$$\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_1,y_2)\,\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_2,y_1)$$
$$= e^{-\frac{1}{\varepsilon}\left(c(x_1,y_2)+c(x_2,y_1)\right)} \times e^{\frac{1}{\varepsilon}\left(\phi_\varepsilon(x_1)+\phi_\varepsilon(x_2)\right)} \times e^{\frac{1}{\varepsilon}\left(\psi_\varepsilon(y_1)+\psi_\varepsilon(y_2)\right)}$$

$$\Rightarrow \frac{\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_1,y_1)\ \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_2,y_2)}{\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_1,y_2)\ \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_2,y_1)}$$

$$= e^{-\frac{1}{\varepsilon}\left[c(x_1,y_1) + c(x_2,y_2) - c(x_1,y_2) - c(x_2,y_1)\right]}$$

Similarly if we take $x_1, \dots, x_n \in \operatorname{supp}(\mu)$
and $y_1, \dots, y_n \in \operatorname{supp}(\nu)$
and let $\sigma \in \text{Permutation}(\{1, \dots, n\})$

$$\prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_i,y_i) = e^{-\frac{1}{\varepsilon}\left[\sum_{i=1}^{n} c(x_i,y_i) - \sum_{i=1}^{n} c(x_i,y_{\sigma(i)})\right]} \times \prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_i, y_{\sigma(i)})$$

Definition: ((c,ε) - cyclical invariance)

If a probability measure $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ satisfies $\pi \ll \mu\otimes\nu$ and for any $x_1, \dots, x_n \in \operatorname{supp}(\mu)$, $y_1, \dots, y_n \in \operatorname{supp}(\nu)$ and $\sigma \in \text{Permutation}(\{1, \dots, n\})$

$$\prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_i,y_i) = e^{-\frac{1}{\varepsilon}\left[\sum_{i=1}^{n} \left(c(x_i,y_i) - c(x_i,y_{\sigma(i)})\right)\right]} \times \prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_i, y_{\sigma(i)})$$

Then π^ε is called (c,ε) - cyclically invariant measure and this property is called (c,ε) - cyclical invariance.

Theorem:

Let c be a lower semi-continuous function

1. There is atmost one (c,ε) - cyclically invariant coupling $\pi \in \Pi(\mu,\nu)$.

2. Suppose $\exists\, \pi \in \Pi(\mu,\nu)$ s.t. $\int c\, d\pi + KL(\pi \| \mu\otimes\nu) < \infty$,

then $\pi^\varepsilon \in \Pi(\mu,\nu)$ is (c,ε) - cyclically invariant if and only if

$$\pi^\varepsilon = \underset{\pi \in \Pi(\mu,\nu)}{\operatorname{argmin}} \left\{ \int c\, d\pi + KL(\pi \| \mu\otimes\nu) \right\}.$$

Application of (c,ε) - cyclically invariant

Notice for $x_1, \dots, x_n \in \operatorname{Supp}(\mu)$ and $y_1, \dots, y_n \in \operatorname{supp}(\nu)$

$$\prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_i,y_i) = \exp\left(-\frac{1}{\varepsilon}\left[\sum_{i=1}^{n} \left(c(x_i,y_i) - c(x_i,y_{\sigma(i)})\right)\right]\right) \times \prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x_i, y_{\sigma(i)})$$

Now for any set of n pairs of points $(x_1, y_1), \dots, (x_n, y_n)$, if there exists $\sigma \in \text{Perm}(\{1, \dots, n\})$ s.t.

$$\sum_{i=1}^{n} \left(c(x_i, y_i) - c(x_i, y_{\sigma(i)}) \right) > 0 \text{, then.}$$

$$\prod_{i=1}^{n} \frac{d\pi^\varepsilon}{d(\mu \otimes \nu)}(x_i, y_i) \longrightarrow 0 \quad \text{as } \varepsilon \to 0$$

So if $\pi^\varepsilon \xrightarrow{\text{weakly}} \pi^*$, then, $\pi^* \in \Pi(\mu, \nu)$ support (π^*) is c-cyclically monotone.

Hence $\pi^* \in \arg\inf\limits_{\pi \in \Pi(\mu,\nu)} \left\{ \int c \, d\pi \right\}$.

One can push the above argument to obtain a large deviation principle.

Theorem: Let $\Gamma = \text{supp}(\pi^*)$ where $\pi^* = \lim\limits_{\varepsilon \to 0} \pi^\varepsilon$

Define
$$I(x,y) := \sup_{k \geq 2} \sup_{(x_i, y_i)_{i=2}^{k}} \sup_{\sigma \in \text{Perm}(\{1, \dots, n\})} \sum_{i=1}^{n} \left(c(x_i, y_i) - c(x_i, y_{\sigma(i)}) \right)$$

$$= c(x,y) - \underbrace{\phi_*(x) - \psi_*(y)}_{\text{Kantorovich Potential}}$$

Equality holds when Kantorovich potential is unique.

Then

(a) For a closed bdd set $B \subseteq \mathcal{X} \times \mathcal{Y}$

$$\limsup_{\varepsilon \to 0} \varepsilon \log \pi^{\varepsilon}(B) \leq - \inf_{(x,y) \in B} I(x,y)$$

(b) For an open set $U \subseteq \mathcal{X} \times \mathcal{Y}$.

$$\liminf_{\varepsilon \to 0} \varepsilon \log \pi^{\varepsilon}(U) \geq - \inf_{(x,y) \in U} I(x,y)$$

Proof:

Proof of large deviation upper bound:

Suppose $(x,y) \in \mathcal{X} \times \mathcal{Y}$. and $(x_i, y_i)_{2 \leq i \leq K}$ $\subset \operatorname{supp}(\pi^*)$ with $K \geq 2$ such that

$$\delta_0 := \sum_{i=1}^{K} c(x_i, y_i) - \sum_{i=1}^{K} c(x_i, y_{i+1}) > 0$$

where $(x_1, y_1) := (x,y)$

Claim: Given $\delta < \delta_0$, there exists $\alpha, r, \varepsilon_0 > 0$ s.t.

$$\pi^{\varepsilon}(B_r(x,y)) \leq \alpha e^{-\delta/\varepsilon} \quad \text{for } \varepsilon < \varepsilon_0.$$

Proof:

Recall

$$\prod_{i=1}^{K} \frac{d\pi^{\varepsilon}}{d(\mu \otimes \nu)}(x_i, y_i) = e^{-\frac{1}{\varepsilon}\left[\sum_{i=1}^{K} c(x_i, y_i) - \sum_{i=1}^{K} c(x_i, y_{i+1})\right]}$$

$$\times \prod_{i=1} \frac{d\pi^{\varepsilon}}{d(\mu \otimes \nu)}(x_i, y_{i+1})$$

Integrating both side around small balls around $(x_i, y_i)_{i=1}^{K}$

$$\prod_{i=1}^{K} \pi^{\varepsilon}\left(B_r(x_i,y_i)\right) \lesssim e^{-\frac{1}{\varepsilon}\left[\sum_{i=1}^{K} c(x_i,y_i) - c(x_i,y_{i+1})\right]}$$

$$\Rightarrow \quad \pi^{\varepsilon}\left(B_r(x_1,y_1)\right) \lesssim \alpha_K e^{-\frac{1}{\varepsilon}\delta_0}$$

From above claim, we get

$$\limsup_{\varepsilon \to 0} \varepsilon \log \pi^{\varepsilon}\left(B_r(x_1,y_1)\right) \leq -\delta_0$$

Since $K, (x_i,y_i)_{i=2}^{K}$ are arbitrary, taking supremum over $K, (x_i,y_i)_{i=2}^{K} \in \Gamma$ and permutations of $(x_1,y_1), \ldots, (x_K,y_K)$ yields

$$\limsup_{\varepsilon \to 0} \varepsilon \log \pi^{\varepsilon}\left(B_r(x_1,y_1)\right) \leq - \sup_{K \in \mathbb{N}} \sup_{(x_i,y_i)_{i=2}^{K} \subset \Gamma} \sup_{\sigma \in \mathrm{Perm}\{1,\ldots,K\}} \left[\sum_{i=1}^{K} c(x_i,y_i) - \sum_{i=1}^{K} c(x_i, y_{\sigma(i)})\right]$$

Proof of lower bound:

Claim: Suppose Kantorovich potentials (ϕ^*, ψ^*) are unique (upto a constant), then

$$I(x,y) = c(x,y) - \phi^*(x) - \psi^*(y).$$

and as a consequence, for any $(x,y), (x',y')$ s.t. $(x,y'), (x',y) \in \Gamma$.

$$I(x,y) + I(x',y') = c(x,y) + c(x',y') - c(x,y') - c(x',y).$$

Proof: Solve in problem set.

We use the above claim to prove large deviation lower bound.

For any $(x,y) \in U$, choose $r>0$ small s.t. $B_r(x,y) \subseteq U$.

Choose (x',y') s.t. (x,y') and $(x',y) \in \Gamma$.

Then

$$\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x,y)\,\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x',y')$$
$$= e^{-\frac{1}{\varepsilon}\left[c(x,y) + c(x',y') - c(x,y') - c(x',y)\right]}\,\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x,y')\,\frac{d\pi^\varepsilon}{d(\mu\otimes\nu)}(x',y).$$

By continuity of c, we get

$$\liminf_{\varepsilon\to\infty} \varepsilon\log\left(\pi^\varepsilon\left(B_r(x,y)\right)\right) + \liminf_{\varepsilon\to\infty} \varepsilon\log\left(\pi^\varepsilon(B_r(x',y'))\right)$$
$$\geq -\left[c(x,y) + c(x',y') - c(x,y') - c(x',y)\right]$$
$$= I(x,y) + I(x',y')$$

Combining with upper bdd., we get

$$\liminf_{\varepsilon\to\infty} \varepsilon\log\left(\pi^\varepsilon\left(B_r(x,y)\right)\right) \geq -I(x,y)$$

Stability of Entropic Optimal Transport via cyclical invariance.

Suppose there is a sequence of quadruplets $(c_n, \mu_n, \nu_n, \pi_n^\varepsilon)$ s.t. $c_n \to c$ (cost function), $\mu_n \xrightarrow{d} \mu$, $\nu_n \xrightarrow{d} \nu$. Can we say π_n^ε also converge. If so, where?

Theorem: In the above setting suppose $c_n \to c$ locally uniformly and c is continuous. Then $\pi_n^\varepsilon \to \pi^\varepsilon$ where π^ε is the unique (c, ε)-cyclically invariant measure (s.t. $\pi^\varepsilon \ll \mu \otimes \nu$) and if $\int c \, d\pi + \varepsilon KL(\pi \| \mu \otimes \nu) < \infty$ for some $\pi \in \Pi(\mu, \nu)$, then

$$\pi^\varepsilon = \underset{\pi \in \Pi(\mu,\nu)}{\arg\inf} \int c \, d\pi + \varepsilon KL(\pi \| \mu \otimes \nu)$$

Proof Idea: Suppose π^ε is a sub-sequential limit of π_n^ε (Need tightness of π_n^ε — follows from weak convergence of μ_n, ν_n)

Recall

$$\prod_{i=1}^n \frac{d\pi_n^\varepsilon}{d(\mu_n \otimes \nu_n)}(x_i, y_i) = e^{-\frac{1}{\varepsilon}\left[\sum_{i=1}^n c_n(x_i, y_i) - \sum_{i=1}^n c_n(x_i, y_{\sigma(i)})\right]} \prod_{i=1}^n \frac{d\pi_n^\varepsilon}{d(\mu_n \otimes \nu_n)}(x_i, y_{\sigma(i)}).$$

If $\pi^{\varepsilon} \ll \mu\otimes\nu$, then taking limit on both sides of the above equation

(Note: one need to first integrate both sides on a small ball, then take $n\to\infty$ and then take radius of of those balls to 0)

yields — π^{ε} is (c,ε) cyclically invariant coupling of μ,ν.

We know there can be only one such coupling, hence the proof follows.

Suffices to show $\pi^{\varepsilon} \ll \mu\otimes\nu$.

Proof goes by contradiction.

Suppose $\pi^{\varepsilon} \not\ll \mu\otimes\nu$. Then

Claim:

$$\mu\otimes\mu\left(\left\{(x_1,x_2)\in\mathcal{X}^2 : \exists\, U_1,U_2 \in \mathcal{B}(\mathcal{Y}) \text{ with for } i=1,2,\ \mathbb{E}_{x_i}(\mathbb{1}(U_i))>0 \quad \mathbb{E}_{x_i}(\mathbb{1}(U_{i+1}))=0\right\}\right) > 0$$

Where $\mathbb{E}_x$ denotes conditional expectation under π^{ε}

In addition U_i are chosen to be disjoint, small diameter s.t. $\mathbb{E}_{x_i}(\partial U_i) = \mathbb{E}_{x_i}(\partial U_j) = 0$

Proof: Solve in the problem set.

We here assume the claim.

Fix (x_1,x_2) from the above set.

Consider

$$\frac{d\pi_n^{\varepsilon}}{d(\mu_n\otimes\nu_n)}(x_1,y_1)\ \frac{d\pi^{\varepsilon}}{d(\mu_n\otimes\nu_n)}(x_2,y_2)$$

$$= e^{-\frac{1}{\varepsilon}\left[c(x_1,y_1) + c(x_2,y_2) - c(x_1,y_2) - c(x_2,y_1)\right]}$$

$$\frac{d\pi_n^{\varepsilon}}{d(\mu_n \otimes \nu_n)}(x_1, y_2) \, \frac{d\pi^{\varepsilon}}{d(\mu_n \otimes \nu_n)}(x_2, y_1)$$

Now vary x_1, x_2 in a small ball and y_1 in U_1 and y_2 in U_2 and integrate both sided and let $n \to \infty$.

The right side will go to 0, but the left side does not. Contradiction.