Transport

Lecture 3

GOALS // o Define Transport Between Measures.

o Properties of Transport

o Variational Inference Via Transport

o Beyond KL Divergence.

WHAT IS TRANSPORT ? //

Definition / ρ pdf on $\mathbb{R}^d$; $g: \mathbb{R}^d \to \mathbb{R}^r$.
ρ_g pdf, on $\mathbb{R}^r$, of $g(Z)$, $Z \sim \rho$.
ρ_g is the pushforward of ρ under g, written
$\rho_g = g_{\#}\rho$.

Remark / $A \subseteq \mathbb{R}^r$ Borel set then

$$\mathbb{P}^{\rho_g}(A) = \mathbb{P}^{\rho}(g^{-1}(A))$$

Definition / Let $\rho_1, \rho_2 \in P(\mathbb{R}^d)$. A transport map $g: \mathbb{R}^d \to \mathbb{R}^d$ between ρ_1 & ρ_2 is map with property $\rho_2 = g_{\#}\rho_1$.

PROPERTIES OF TRANSPORT //

Important Remark / Recall the Wasserstein metric in Kantorwich form

$$W_p(\rho_1,\rho_2)^p = \inf_{\pi \in \Pi_{\rho_1,\rho_2}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|z-u\|^p_{\mathbb{R}^d} \, \pi(z,u)\, dz\, du$$

under certain smoothness assumptions this can be expressed in Monge form using transport:

$$W_p(\rho_1,\rho_2)^p = \inf_{g:\, g_\# \rho_1 = \rho_2} \int_{\mathbb{R}^d} \|z - g(z)\|^p_{\mathbb{R}^d} \, \rho_1(z)\, dz$$

because $\pi^*(z,u) = \delta(u - g^*(z))\, \rho_1(z)$

$\pi^*(z,u)$: infimizing π

$g^*(z)$: infimizing g

Theorem / Let $T \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ be invertible. Then

$$D_{KL}(\rho_1 \| \rho_2) = D_{KL}(T_\# \rho_1 \| T_\# \rho_2)$$

Lemma / $\pi \in \mathbb{R}^d$, $q \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ invertible. Then

$$\log q_\# \pi(u) = \log(\pi \circ q^{-1})(u) + \log \det D(q^{-1}(u))$$

Theorem / Let $T \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ be invertible. Then

$$D_{KL}(\rho_1 \| \rho_2) = D_{KL}(T_\# \rho_1 \| T_\# \rho_2)$$

Lemma / $\pi \in \mathbb{R}^d$, $q \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ invertible. Then

$$\log q_\# \pi(u) = \log(\pi \circ q^{-1})(u) + \log \det D(q^{-1}(u))$$

VARIATIONAL INFERENCE VIA TRANSPORT //

Recall Bayes Theorem:

$$\pi(u) = \frac{1}{Z} \ell(u) \rho(u), \quad Z = \mathbb{E}^{u \sim \rho}[\ell(u)].$$

We aim to find an approximate transport $T(\cdot;\theta): \mathbb{R}^d \to \mathbb{R}^d$ with property $T(\cdot;\theta)_\# \rho \simeq \pi$.

How should we choose θ? Define

$$F(\theta) = D_{KL}(T(\cdot;\theta)_\# \rho \| \pi).$$

By preceding Theorem

$$F(\theta) = D_{KL}(\rho \| T^{-1}(\cdot;\theta)_\# \pi)$$

By properties of D_{KL}

$$F(\theta) = -\mathbb{E}^{u\sim\rho} \log\left(T^{-1}(\cdot;\theta)_{\#}\pi\right) + \mathbb{E}^{u\sim\rho} \log \rho(u)$$

Minimizing $F(\cdot)$ is equivalent, by Lemma, to minimizing

$$-\mathbb{E}^{u\sim\rho}\left[\log \pi\circ T(u;\theta) + \log \det D_u T(u;\theta)\right]$$

But $\log \pi\circ T = \log \rho\circ T + \log \chi\circ T - \log Z$

& since Z is independent of T, and hence θ, our minimization problem is equivalent to

$$\theta^* \in \underset{\theta\in\Theta}{\operatorname{argmin}}\ J(\theta),$$

$$J(\theta) = -\mathbb{E}^{u\sim\rho}\left[\log \rho\circ T(u;\theta) + \log \chi\circ T(u;\theta) + \log \det D_u T(u;\theta)\right]$$

This is variational inference over set

$$Q = \left\{ q : q = T_{\#}\rho,\ T\in\mathcal{T}_\Theta \right\}$$

$\mathcal{T}$ is C^1 invertible maps:

$$\uparrow = \{ T(\cdot;\theta) \in C^1(\mathbb{R}^d, \mathbb{R}^d), \theta \in \Theta \mid \det D_u T(u;\theta) > 0 \ \forall (u,\theta) \in \mathbb{R}^d \times \Theta \}$$

BEYOND D_{KL} DIVERGENCE //

Preceding requires explicit knowledge of ρ. what if only samples from ρ are known, not ρ itself:

> **Data Assumption** / we are given
> $$\{ u^{(n)} \}_{n \in \{1,\dots,N\}} \overset{\text{i.i.d}}{\sim} \rho$$

Then define

$$\rho^N(u) = \frac{1}{N} \sum_{n=1}^{N} \delta_{u^{(n)}}(u)$$

$$\pi^N(u) = \frac{1}{Z} s(u) \rho^N(u), \quad Z^N = \mathbb{E}^{u \sim \rho^N}[s(u)].$$

It follows that

$$\pi^N(u) = \sum_{n=1}^{N} w^{(n)} \delta_{u^{(n)}}(u)$$

$$w^{(n)} = s^{(n)} / \sum_{m=1}^{N} s^{(m)}, \quad s^{(n)} = s(u^{(n)})$$

Find θ to minimize

$$J^N(\theta) = D_E\left(T(\cdot;\theta)_{\#}\rho^N,\ \pi^N\right).$$

This loss function can be evaluated using only samples from ρ^N & π^N.