数学基础

邻域： $O_x^\varepsilon = \{ y : d(x,y) < \varepsilon \}$

开集： U ， $\forall x \in U$ ， $\exists \varepsilon$ ， $O_x^\varepsilon \subset U$

闭集： $U \Leftrightarrow X \setminus U$ 是开集

推前算子： $\int_{y \in A} T_\# \rho(y)\, dy = \int_{T(x) \in A} \rho(x)\, dx$

概率测度空间的度量

① 全变差距离

$\rho, \rho' \in C^\infty$

$$d_{TV}(\rho, \rho') = 0 \Leftrightarrow \|\rho - \rho'\|_{L_1} = 0 \Leftrightarrow \rho - \rho' = 0$$

$$d_{TV}(\rho, \rho') = d_{TV}(\rho', \rho')$$

$$d_{TV}(\rho_1, \rho_2) = \frac{1}{2}\int |\rho_1 - \rho_2|\, d\theta$$

$$= \frac{1}{2}\int |\rho_1 - \rho_3 + \rho_3 - \rho_2|\, d\theta$$

$$\le \frac{1}{2}\int |\rho_1 - \rho_3|\, d\theta + \frac{1}{2}\int |\rho_3 - \rho_2|\, d\theta$$

② Hellinger 距离

$$d_H(\rho_1, \rho_2)^2 = \frac{1}{2}\int |\sqrt{\rho_1} - \sqrt{\rho_2}|^2\, d\theta$$

$$= \frac{1}{2}\int |\sqrt{\rho_1} - \sqrt{\rho_3} + \sqrt{\rho_3} - \sqrt{\rho_2}|^2\, d\theta$$

$$= \frac{1}{2}\int |\sqrt{\rho_1} - \sqrt{\rho_3}|^2\, d\theta + \frac{1}{2}\int |\sqrt{\rho_3} - \sqrt{\rho_2}|^2\, d\theta$$

$$+ \int |\sqrt{\rho_1} - \sqrt{\rho_3}|\,|\sqrt{\rho_3} - \sqrt{\rho_2}|\, d\theta$$

$$= d_H(\rho_1, \rho_3)^2 + d_H(\rho_3, \rho_2)^2$$

$$+ \int \left|\sqrt{\rho_1} - \sqrt{\rho_3}\right| \left|\sqrt{\rho_3} - \sqrt{\rho_2}\right| d\theta$$

下证：

$$\int \left|\sqrt{\rho_1} - \sqrt{\rho_3}\right| \left|\sqrt{\rho_3} - \sqrt{\rho_2}\right| d\theta \le 2\, d_H(\rho_1\ \rho_3)\, d_H(\rho_3\ \rho_2)$$

使用 Cauchy 不等式

$$\int |f g| \le \sqrt{\int f^2 d\theta}\ \sqrt{\int g^2 d\theta}$$

良定性：

$$d_{TV}(\rho, \rho') = \frac{1}{2}\int |\rho - \rho'|\, d\theta \le \frac{1}{2}\underbrace{\left(\int |\rho| d\theta + \int |\rho'| d\theta\right)}_{=1}$$

$$d_H(\rho, \rho') = \left(\frac{1}{2}\int \left|\sqrt{\rho} - \sqrt{\rho'}\right|^2 d\theta\right)^{\frac{1}{2}}$$

$$= \left(\frac{1}{2}\int \rho + \rho' - 2\sqrt{\rho\rho'}\, d\theta\right)^{\frac{1}{2}}$$

$$\le 1$$

为什么不用 L_2 距离？

等价性：

$$\frac{1}{\sqrt{2}} d_{TV}(\rho, \rho') = \frac{1}{2\sqrt{2}} \int \left(\sqrt{\rho} + \sqrt{\rho'}\right) \left|\sqrt{\rho} - \sqrt{\rho'}\right| d\theta$$

$$\le \frac{1}{2\sqrt{2}} \sqrt{\int \left(\sqrt{\rho}+\sqrt{\rho'}\right)^2 d\theta \int \left(\sqrt{\rho}-\sqrt{\rho'}\right)^2 d\theta}$$

$$\le \frac{1}{2} \sqrt{4}\; d_H(\rho, \rho')$$

$$d_H(\rho,\rho')^2 = \frac{1}{2}\int \left|\sqrt{\rho}-\sqrt{\rho'}\right|^2 d\theta$$

$$d_{TV}(\rho,\rho') = \frac{1}{2}\int |\rho-\rho'|\, d\theta$$

由于 $\left|\sqrt{\rho}-\sqrt{\rho'}\right|^2 \le \left|\sqrt{\rho}-\sqrt{\rho'}\right| \left|\sqrt{\rho}+\sqrt{\rho'}\right|$

$$= |\rho-\rho'|$$

其它估计

首先证明

$$\frac{1}{2}\left|E_\rho f - E_{\rho'} f\right| \le d_{TV}(\rho,\rho')$$

$$\Longleftrightarrow \left|\int f(\rho-\rho')\, d\theta\right| \le \int |f|\,|\rho-\rho'|\, d\theta \qquad |f|_\infty \le 1$$

$$\le \int |\rho-\rho'|\, d\theta$$

再证明

$$\sup_{|f|_\infty\le 1} \frac{1}{2}\left|E_\rho f - E_{\rho'} f\right| \ge d_{TV}(\rho,\rho')$$

取 $f = \mathrm{sign}(\rho-\rho')$

$$|E_\rho f - E_{\rho'} f| = \left|\int f(\rho - \rho')\, d\theta\right|$$

$$\le \int |f|\, |\rho - \rho'|\, d\theta$$

$$\le 2|f|_\infty\, d_{TV}(\rho, \rho')$$

$$|E_\rho f - E_{\rho'} f| = \left|\int f(\sqrt{\rho} - \sqrt{\rho'})(\sqrt{\rho} + \sqrt{\rho'})\, d\theta\right|$$

$$\le \left(\int f^2 (\sqrt{\rho} + \sqrt{\rho'})^2\, d\theta \int (\sqrt{\rho} - \sqrt{\rho'})^2\, d\theta\right)^{\frac{1}{2}}$$

$$\le \left(\int f^2 2(\rho + \rho')\, d\theta\ 2\right)^{\frac{1}{2}} d_H(\rho, \rho')$$

最优传输问题

Kantorovich 问题，有 $\gamma(\theta_1, \theta_2)$ 的沙子从 θ_1 运到 θ_2。

Kantorovich 对偶，引入拉格朗日乘子，f，g

$$\inf_{\gamma} \sup_{f,g} \int \gamma\, c\, d\theta_1 d\theta_2 - \int \left(\int \gamma(\theta_1, \theta_2)\, d\theta_2 - \rho_A(\theta_1) \right) f(\theta_1)\, d\theta_1$$
$$- \int \left(\int \gamma(\theta_1, \theta_2)\, d\theta_1 - \rho_B(\theta_2) \right) g(\theta_2)\, d\theta_2$$

$$= \inf_{\gamma} \sup_{f,g} \int \gamma(\theta_1, \theta_2) \left[c(\theta_1, \theta_2) - f(\theta_1) - g(\theta_2) \right] d\theta_1 d\theta_2$$
$$+ \int \rho_A(\theta) f(\theta) + \rho_B(\theta) g(\theta)\, d\theta$$

$$= \sup_{f,g} \left[\int \rho_A(\theta) f(\theta) + \rho_B(\theta) g(\theta)\, d\theta + \inf_{\gamma} \int \gamma(\theta_1, \theta_2) \left(c(\theta_1, \theta_2) - f(\theta_1) - g(\theta_2) \right) \right]$$

$\gamma \geq 0$

$\vee\!|$ 可选 $f(\theta_1) + g(\theta_2) \leq c(\theta_1, \theta_2)$

$$\sup \int \rho_A(\theta) f(\theta) + \rho_B(\theta) g(\theta)\, d\theta$$

$$f(\theta_1) + g(\theta_2) \leq c(\theta_1, \theta_2)$$

$\wedge\!|$ 若 $c(\theta_1, \theta_2) > f(\theta_1) + g(\theta_2)$

那么 $\inf_{\gamma} = -\infty$

下面证明 Wasserstein-p 距离是距离

假设我们有 ρ_A ρ_B ρ_C

$$W_p(\rho_A, \rho_C) \leq W_p(\rho_A, \rho_B) + W_p(\rho_B, \rho_C)$$

$$\gamma^*_{AB} \qquad \gamma^*_{BC}$$

那么有 $\gamma(x_A\, x_B\, x_C)$ 的边缘分布为 γ^*_{AB} γ^*_{BC}

$x_B \sim \rho_B$ $x_A \sim \rho(\cdot \mid x_B)$ 根据 γ^*_{AB}

再生成 $x_C \sim \rho(\cdot \mid x_B)$ 根据 γ^*_{BC}

$$W_p(\rho_A, \rho_C) \leq \left(\int \|x_A - x_C\|_2^p \, \gamma(x_A\, x_B\, x_C)\, dx_A\, dx_B\, dx_C\right)^{\frac{1}{p}}$$

$$\leq \left(\int \{\|x_A - x_B\|_2 + \|x_B - x_C\|_2\}^p \gamma(x_A\, x_B\, x_C)\right)^{\frac{1}{p}}$$

$$\leq W_p(\rho_A, \rho_B) + W_p(\rho_B, \rho_C)$$

这里用了 $p \geq 1$

$$\left(\int (\|f\|_2 + \|g\|_2)^p \gamma \, d\theta\right)^{\frac{1}{p}}$$

$$\leq \left(\int \|f\|_2^p \gamma \, d\theta\right)^{\frac{1}{p}} + \left(\int \|g\|_2^p \gamma \, d\theta\right)^{\frac{1}{p}}$$

对 $c = \|\theta_1 - \theta_2\|_2$ ，下界的证明

$$\sup_h \int (\rho_A(\theta) - \rho_B(\theta))\, h(\theta)\, d\theta$$

$$= \sup_h \int \gamma(\theta_1, \theta_2)\big(h(\theta_2) - h(\theta_1)\big)\, d\theta_1 d\theta_2$$

$$\leq \sup_h \int \gamma(\theta_1, \theta_2)\, |h(\theta_2) - h(\theta_1)|\, d\theta_1 d\theta_2$$

$$\leq \sup_h \int \gamma(\theta_1, \theta_2)\, \|\theta_1 - \theta_2\|_2\, d\theta_1 d\theta_2$$

$$\leq W_1(\rho_A\ \rho_B)$$

上界的证明，目标

$$\sup_{f(\theta_1)+g(\theta_2)\leq\|\theta_1-\theta_2\|_2} E_{\rho_A} f + E_{\rho_B} g$$

$$\leq \sup_{|h(\theta_1)-h(\theta_2)|\leq\|\theta_1-\theta_2\|_2} E_{\rho_A} h - E_{\rho_B} h$$

构造

$$K(\theta) = \inf_u \left[\|\theta - u\|_2 - g(u)\right]$$ 由于 $f(\theta_1)+g(\theta_2)\leq\|\theta_1-\theta_2\|_2$

$$f(\theta_1) \leq \inf_{\theta_2} \|\theta_1 - \theta_2\|_2 - g(\theta_2) = K(\theta_1)$$

$$K(\theta_2) = \inf_{\theta_1} \|\theta_2 - \theta_1\|_2 - g(\theta_1)$$

$$\leq \|\theta_2 - \theta_2\|_2 - g(\theta_2) = -g(\theta_2)$$

我们有

$$\sup_{f(\theta_1)+g(\theta_2)\le \|\theta_1-\theta_2\|_2} E_{\rho_A} f + E_{\rho_B} g$$

$$\le E_{\rho_A} k - E_{\rho_B} k$$

下证 $|k(\theta_1) - k(\theta_2)| \le \|\theta_1-\theta_2\|_2$

$$k(\theta_1) = \inf_u \|\theta_1 - u\|_2 - g(u)$$

$$\le \inf_u \|\theta_1 - u\|_2 - g(u)$$

$$\le \inf_u \|\theta_1 - \theta_2\|_2 + \|\theta_2 - u\|_2 - g(u)$$

$$\le \|\theta_1-\theta_2\|_2 + k(\theta_2)$$

因此 $k(\theta_1) - k(\theta_2) \le \|\theta_1-\theta_2\|_2$

再由对称性可证。

Wasserstein 2 距离

测地线

$$T_0 : (\theta_1, \theta_2) \Rightarrow \theta_1, \quad T_0 \# \gamma^* (A) = \gamma^* (T_0^{-1}(A))$$

$$= \gamma^* (\theta_1 \in A) = \rho_A$$

同理 $T_1 \# \gamma^* = \rho_B$

下证 $W_2(\rho_s, \rho_t) \le (t-s) W_2(\rho_A, \rho_B) \ (s \le t)$

定义 $\gamma_{s,t}^* = (T_s, T_t) \# \gamma^*$

$$(\theta_1, \theta_2) \to ((1-s)\theta_1 + s\theta_2, (1-t)\theta_1 + t\theta_2)$$

$$\iint_{(\theta_1, \theta_2) \in (A, R^d)} \gamma_{s,t}^* (\theta_1, \theta_2) \, d\theta_2 = \int_{(\theta_1, \theta_2) \in (A, R^d)} \gamma_{s,t}^* (\theta_1, \theta_2) \, d\theta_1 d\theta_2$$

$$= \int_{\substack{[(1-s)x_1 + s x_2, \\ (1-t)x_1 + t x_2] \in (A, R^d)}} \gamma^* (x_1, x_2) \, dx_1 dx_2$$

$$= \int_{\substack{(1-s)x_1 + s x_2 \in A \\ (1-t)x_1 + t x_2 \in R^d}} \gamma^* (x_1, x_2) \, dx_1 dx_2$$

$$\int_{\theta_1 \in A} \rho_s(\theta_1) \, d\theta_1 = \int_{(1-s)\theta_1 + s\theta_2 \in A} \gamma^* (\theta_1, \theta_2) \, d\theta_2 d\theta_1$$

$$W_2(\rho_s, \rho_t)^2 \le \int \|\theta_1-\theta_2\|^2 \gamma^*_{st}(\theta_1, \theta_2)\, d\theta_1 d\theta_2$$

换元

$$= \int \|T_s(\theta_1, \theta_2) - T_t(\theta_1, \theta_2)\|^2 \gamma^*(\theta_1, \theta_2)\, d\theta_1 d\theta_2$$

$$= (s-t)^2 \int \|\theta_1-\theta_2\|^2 \gamma^*(\theta_1, \theta_2)\, d\theta_1 d\theta_2$$

$$= (s-t)^2 W_2(\rho_A, \rho_B)$$

因此

$$W_2(\rho_A, \rho_B) \le W_2(\rho_A, \rho_s) + W_2(\rho_s, \rho_t) + W_2(\rho_t, \rho_B)$$

$$\le s W_2(\rho_A, \rho_B) + (t-s) W_2(\rho_A, \rho_B) + (1-t) W_2(\rho_A, \rho_B)$$

$$= W_2(\rho_A, \rho_B)$$

所以全部取等

这对 W_p 距离也成立。

动力学观点

两点之间的距离

$X(s) = (1-s)X_0 + sX_1$ 满足

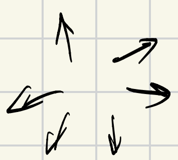
$$\inf_X \int_0^1 X'(s)^2 \, ds$$ （能量最小）

$$\int_0^1 X'(s)^2 \, ds \int_0^1 1 \, ds \geq \left(\int_0^1 \|X'(s)\|_2 \, ds\right)^2$$

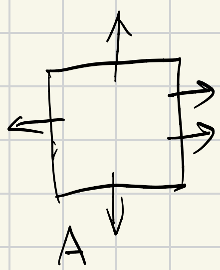
$$= \|X_1 - X_0\|_2^2$$

$$\int_0^1 (X'(s))^2 \, ds \geq \|X_1 - X_0\|_2^2$$

两个概率密度之间，有速度场 v_t



粒子随着 v_t 演化，ρ_t 也随之演化



A

$$\frac{\partial}{\partial t}\int_A \rho_t(\theta) \, d\theta = -\int_{\partial\Omega} \rho v \cdot n \, d\theta$$

$$= -\int_{\Omega} \nabla \rho_t v_t \, d\theta$$

$$\Rightarrow \partial_t \rho_t + \nabla_\theta(\rho_t v_t) = 0$$

定义 $T_t : R^d \to R^d$ 满足

$$\partial_t T_t(\theta) = v_t(T_t(\theta)), \quad T_0(\theta) = \theta$$

那么 $\rho_t = T_t \# \rho_0$，在时刻 t，粒子

从 $\theta \to T_t(\theta)$

给定 v_t，在时间 0 到 1，把 $\rho_0 = \rho_A$ 演化

到 $\rho_1 = \rho_B$，定义能量

$$A(\rho, v) = \int_0^1 \int \| v_t \|_2^2 \rho_t \, d\theta \, dt$$

① 我们有

$$= \int_0^1 \int \| v_t(T_t(\theta)) \|_2^2 \rho_t(T_t(\theta)) \, dT_t(\theta) \, dt$$

$$= \int_0^1 \int \| \frac{\partial}{\partial t} T_t(\theta) \|_2^2 \rho_0(\theta) \, d\theta \, dt$$

由于 $\int_0^1 \| \frac{\partial}{\partial t} T_t(\theta) \|_2^2 \, dt \quad \int_0^1 dt$

$$\geq \left(\int_0^1 \frac{\partial}{\partial t} T_t(\theta) \, dt \right)^2 = \| T_1(\theta) - T_0(\theta) \|^2$$

$$\geq \int \| T_1(\theta) - T_0(\theta) \|_2^2 \rho_0(\theta) \, d\theta$$

$$\geq W_2^2(\rho_A, \rho_B)$$

② 另一方面，如果有最优映射 T，定义

$$T_t(\theta) = (1-t)\theta + tT(\theta) \qquad (T = \nabla\varphi)$$

定义 $v_t = \frac{d}{dt} T_t(\theta) \circ T_t^{-1}(\theta)$

$$= (T - Id) \circ T_t^{-1}$$

我们有 $\frac{d}{dt} T_t(\theta) = v_t(T_t(\theta))$，$\rho_t$，$v_t$

满足连续性方程，且

$$A(\rho, v) = \int_0^1 \int \left\| \frac{\partial}{\partial t} T_t(\theta) \right\|_2^2 \rho_0(\theta)\, d\theta\, dt$$

$$= \int \| T(\theta) - \theta \|_2^2 \rho_0(\theta)\, d\theta$$

$$= W_2^2(\rho_A, \rho_B)$$

KL 散度 $\rho^* = \frac{1}{Z} e^{-\Phi_R(\theta)}$

$$KL[\rho \| \rho^*] = \int \rho \log \frac{\rho}{e^{-\Phi_R}} + \rho \log Z \, d\theta$$

$$= \int \rho \log \rho + \rho \log \Phi_R \, d\theta + \log Z$$

随机过程：

$$E\,dB_t = E\,B_{t+dt} - B_t = 0$$

$$E\,dB_t^2 = E(B_{t+dt} - B_t)(B_{t+dt} - B_t) \simeq dt$$

Itô 公式

$$dX_t = \partial_t f(t, \theta_t)dt + \nabla_\theta f(t, \theta_t)\, d\theta_t + \frac{1}{2}\nabla_\theta \nabla_\theta f(t, \theta_t)\, d\theta_t \cdot d\theta_t$$

$$= \partial_t f(t, \theta_t)dt + \nabla_\theta f(t, \theta_t)\left(b_t dt + \sigma_t dB_t\right) + \frac{1}{2}(\sigma_t dB_t)^T \nabla_\theta^2 f(t, \theta_t)\,(\sigma_t dB_t)$$

$$= \partial_t f(t, \theta_t)\,dt + \nabla_\theta f(t, \theta_t)\, b_t \cdot dt + \nabla_\theta f(t, \theta_t)\,\sigma_t dB_t + \frac{1}{2}\, dB_t^T\, \sigma_t^T\, \nabla_\theta^2 f(t, \theta_t)\, \sigma_t\, dB_t$$

$$\frac{1}{2}\nabla_\theta^2 f(t, \theta_t)\left(\underline{\sigma_t dB_t dB_t^T \sigma_t^T}\right)$$

$$dB_t \cdot dB_t^T \simeq dt\, I$$

Fokker Planck 方程

$$d f(\theta_t) = \nabla_\theta f(\theta_t) b_t dt + \frac{1}{2} \nabla_\theta^2 f(\theta_t) : \sigma_t \sigma_t^T dt$$
$$+ \nabla_\theta f(\theta_t)^T \sigma_t dB_t$$

$$\mathbb{E} f(\theta_t) = \int f(\theta) \rho_t(\theta) d\theta$$
$$= \int f(\theta_t) \rho_0(\theta) d\theta \quad \theta_t = \theta_t(\theta)$$

$$\partial_t \mathbb{E} f(\theta_t) = \mathbb{E}\left[\nabla_\theta f(\theta_t) b_t + \frac{1}{2} \nabla_\theta^2 f(\theta_t) : \sigma_t \sigma_t^T \right]$$
$$= \int \rho_t \left(\nabla_\theta f(\theta) b_t + \frac{1}{2} \nabla_\theta^2 f(\theta) : \sigma_t \sigma_t^T \right) d\theta$$
$$= -\int f(\theta) \nabla_\theta(\rho_t b_t) d\theta + \int \partial_{ij} f(\theta) \rho_t(\theta) D_{ij} d\theta$$
$$= -\int f(\theta) \nabla_\theta(\rho_t b_t) d\theta + \int f(\theta) \partial_{ij} (\rho_t(\theta) D_{ij}) d\theta$$

$$\Rightarrow \frac{\partial}{\partial t} \rho_t(\theta) = -\nabla_\theta(\rho_t b_t) + \sum \partial_{ij} (\rho_t D_{ij})$$