振り返りと導入

前回は最尤推定量と KL ダイバージェンスを定義した。本稿では次のことを行う:

- KL ダイバージェンスの性質を調べる。
- 双対平坦多様体への一般化を考える。

1 Kullback-Leibler ダイバージェンス

定義 1.1 (Kullback-Leibler ダイバージェンス). 関数 $D\colon \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to [0,\infty]$,

$$D(p\|q) := \begin{cases} E_q\left[\frac{dp}{dq}\log\frac{dp}{dq}\right] = E_p\left[\log\frac{dp}{dq}\right] & (p \ll q) \\ \infty & (p \not\ll q) \end{cases} \tag{1.1}$$

を $\mathcal{P}(\mathcal{X})$ 上の **Kullback-Leibler ダイバージェンス** と呼ぶ。

命題 1.2. $\mathcal{P}(\mathcal{X})$ に全変動で定まる位相を入れると、KL ダイバージェンスは連続とは限らない。

証明 $\mathcal{X} := \{0,1\}$ として $p_n := \frac{1}{n}\delta^0 + \left(1-\frac{1}{n}\right)\delta^1$, $q_n := \frac{1}{e^n}\delta^0 + \left(1-\frac{1}{e^n}\right)\delta^1$ が反例のひとつ。 □

$\mathcal{X} = \{1,\dots,n\}, n \in \mathbb{N}$ (カテゴリカル分布 [TODO]) の場合に最尤推定量と KL ダイバージェンスの関係を考える。

定義 1.3 (経験分布). $x = (x_1,\dots,x_k) \in \mathcal{X}^k$ に対し

$$\hat{p}_x := \frac{1}{k}\sum_{i=1}^{k}\delta^{x_i} \tag{1.2}$$

を x により定まる**経験分布 (empirical distribution)** という。

命題 1.4 (最尤推定量と KL ダイバージェンス). $(\Theta, \mathbf{p})$ を $\mathcal{X}$ 上の統計モデルとし、k 個の i.i.d. 拡張 $(\Theta, \mathbf{p}^k)$ を考える。$x = (x_1,\dots,x_k) \in \mathcal{X}^k$ とし、$\hat{p}_x$ を x により定まる経験分布とする。このとき、$\mathbf{p}^k(\Theta)$ が $\hat{p}_x$ を支配する確率測度を少なくともひとつ含むならば、次が成り立つ:

$$\operatorname{argmin}_{\theta\in\Theta} D(\hat{p}_x\|\mathbf{p}^k(\theta)) = \operatorname{argmax}_{\theta\in\Theta} p_\theta^k(x) \tag{1.3}$$

証明 [TODO] もう少し丁寧に $\forall\theta \in \Theta$ に対し

$$\begin{aligned} D(\hat{p}_x\|\mathbf{p}^k(\theta)) &= E_{\hat{p}_x}\left[\log\frac{d\hat{p}_x}{d(\mathbf{p}^k(\theta))}\right] && (1.4)\\ &= E_{\hat{p}_x}\left[\log\frac{d\hat{p}_x}{di}\right] - E_{\hat{p}_x}\left[\log\frac{d(\mathbf{p}^k(\theta))}{di}\right] && (1.5)\\ &= (\theta \text{ によらない項}) - \frac{1}{k}\log p_\theta^k(x) && (1.6) \end{aligned}$$

ゆえに命題の主張が従う。 □

2 双対平坦多様体とダイバージェンス

次のことを思い出しておく。

命題 2.1. $\mathcal{P}$ を可測空間 $\mathcal{X}$ 上の open な指数型分布族とし、$\mathcal{P}$ には自然な位相・多様体構造を入れる。このとき次が成り立つ:

(a) $\mathcal{P}$ の Fisher 計量を g、自然な平坦アファイン接続を ∇、g に関する ∇ の双対接続を ∇^* とおくと、(g, ∇, ∇^*) は $\mathcal{P}$ 上の双対平坦構造となる。

さらに次が成り立つ:

(1) (g, ∇, ∇^*) に関する双対アファイン座標 (θ, η) が存在する。
(2) ある関数 $\psi, \varphi \in C^\infty(\mathcal{P})$ であって $d\psi = \eta_i d\theta^i$, $d\varphi = \theta^i d\eta_i$ をみたすものが存在する。
(3) ψ は g のポテンシャルである。φ は g^{-1} のポテンシャルである。

証明 (a) 便覧.pdf 定理 3.7.3 より。(1), (2), (3) 便覧.pdf 定理 3.8.4 より。 □

定理 2.2 (双対アファイン座標と双対ポテンシャルの存在). M を多様体、(g, ∇, ∇^*) を M 上の双対平坦構造とする。このとき、次が成り立つ:

(1) (g, ∇, ∇^*) に関する双対アファイン座標 (θ, η) が存在する。
(2) ある関数 $\psi, \varphi \in C^\infty(M)$ であって $d\psi = \eta_i d\theta^i$, $d\varphi = \theta^i d\eta_i$ をみたすものが存在する。
(3) ψ は g のポテンシャルである。φ は g^{-1} のポテンシャルである。

証明 [TODO] □

命題 2.3 (指数型分布族と KL ダイバージェンス). $\mathcal{P}$ を指数型分布族とする。最小次元実現 (V, T, μ) に対し対数分配関数 ψ、自然パラメータ座標 θ、期待値パラメータ座標 η を考える。このとき

$$D(p\|q) = \psi(\theta_q) + \psi^\vee(\eta_p) - \langle \theta_q, \eta_p \rangle \quad (\forall p, q \in \mathcal{P}) \tag{2.1}$$

が成り立つ。ただし $\psi^\vee$ は ψ の Legendre 変換である。

証明 Legendre 変換の定義より $\psi(\theta_p) + \psi^\vee(\eta_p) = \langle \theta_p, \eta_p \rangle$ ゆえに

$$\begin{aligned}
\psi(\theta_q) + \psi^\vee(\eta_p) - \langle \theta_q, \eta_p \rangle &= \psi(\theta_q) - \psi(\theta_p) + \langle \theta_p, \eta_p \rangle - \langle \theta_q, \eta_p \rangle && (2.2)\\
&= E_p\left[\psi(\theta_q) - \psi(\theta_p) + \langle \theta_p, T \rangle - \langle \theta_q, T \rangle\right] && (2.3)\\
&= E_p\left[\log \frac{dp}{dq}\right] && (2.4)\\
&= D(p\|q) && (2.5)
\end{aligned}$$

□

命題-定義 2.4 (canonical ダイバージェンス).

$$\psi(\theta_q) - \varphi(\eta_p) - \langle \theta_q, \eta_p \rangle \tag{2.6}$$

の値は ψ, φ のとり方によらず定まる。これを $D(p\|q)$ とおき、双対平坦多様体 (M, g, ∇, ∇^*) の **canonical ダイバージェンス** という。

証明 [TODO] □

今後の予定

- 一般の多様体上のダイバージェンス
- ダイバージェンスから誘導される双対平坦構造
- ダイバージェンスから誘導されるシンプレクティック構造
- Bayes 更新

参考文献

[AJLS17] Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer, **Information Geometry**, Ergebnisse der Mathematik und ihrer Grenzgebiete 34, vol. 64, Springer International Publishing, Cham, 2017.

[Ama16] Shun-ichi Amari, **Information Geometry and Its Applications**, Applied Mathematical Sciences, vol. 194, Springer Japan, Tokyo, 2016 (en).

[AN07] Shun-ichi Amari and Hiroshi Nagaoka, **Methods of Information Geometry**, Translations of Mathematical Monographs, vol. 191, American Mathematical Society, Providence, Rhode Island, April 2007 (en).