振り返りと導入

前回は統計モデルを定義した。本稿では次のことを行う:

- 最尤推定量を定義する。
- Kullback-Leibler ダイバージェンスを定義する。

1 最尤推定

定義 1.1 (i.i.d. 拡張). $\mathcal{X}$ 上の統計モデル $(\Theta, \mathbf{p})$ に対し、$\mathcal{X}^k$ 上の統計モデル $(\Theta, \mathbf{p}^k)$ を $(\Theta, \mathbf{p})$ の **i.i.d. 拡張 (i.i.d. extension)** という。ただし $\mathbf{p}^k(\theta) \coloneqq \mathbf{p}(\theta) \times \cdots \times \mathbf{p}(\theta)$ (積測度) である。

命題-定義 1.2 (最尤推定量). 可測写像 $\hat{\theta} \colon \mathcal{X} \to \Theta$ に関する条件を考える。実現 (p, μ) に対し a.e.$x \in \mathcal{X}$, $\hat{\theta}(x) \in \operatorname{argmax}_{\theta \in \Theta} p_\theta(x)$ が成り立つとき、$\hat{\theta}$ を **(p, μ) に関する最尤推定量**と呼ぶことにする。このとき次は同値である:

(1) $\hat{\theta}$ はある実現 (p, μ) に関する最尤推定量である。
(2) $\hat{\theta}$ は任意の実現 (p, μ) に関する最尤推定量である。

そこで、上の条件のいずれかが成り立つとき、$\hat{\theta}$ を統計モデル $(\Theta, \mathbf{p})$ の**最尤推定量 (maximum likelihood estimator)** と呼ぶ。

証明 (2) ⇒ (1) は明らか。(1) ⇒ (2) は argmax の定義および $q_\theta = p_\theta \dfrac{d\mu}{d\nu}$ に注意すればわかる。 □

命題 1.3 (尤度方程式). (1) $\hat{\theta}$ が最尤推定量ならば、(2) a.e.x に対し $\hat{\theta}(x)$ は**尤度方程式 (likelihood equation)** $(dl_x)_\theta = 0$ の解である。さらに指数型分布族においては逆も成り立つ。

証明 (1) ⇒ (2) は l_x が θ に関し微分可能であることから従う。指数型分布族の場合、(2) ⇒ (1) は $l_x(\theta) = \langle \theta, T(x) \rangle - \psi(\theta)$ が上に凸であることから従う。 □

例 1.4 (最尤推定量の例).

- 正規分布族の場合
- Cauchy 分布族の場合

2 Kullback-Leibler ダイバージェンス

定義 2.1 (Kullback-Leibler ダイバージェンス). 関数 $D\colon \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to [0, \infty]$,

$$D(p\|q) := \begin{cases} E_q\left[\dfrac{dp}{dq}\log\dfrac{dp}{dq}\right] = E_p\left[\log\dfrac{dp}{dq}\right] & (p \ll q) \\ \infty & (p \not\ll q) \end{cases} \tag{2.1}$$

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

命題 2.2. $\mathcal{P}(\mathcal{X})$ 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 $p_n := \dfrac{1}{n}\delta^0 + \left(1 - \dfrac{1}{n}\right)\delta^1$, $q_n := e^{-n}\delta^0 + (1 - e^{-n})\delta^1$ が反例のひとつ。 □

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

$$D(p\|q) = \psi(\theta_q) + \psi^\vee(\eta_p) - \langle \theta_q, \eta_p \rangle \quad (\forall p, q \in \mathcal{P}) \tag{2.2}$$

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

$$\begin{aligned} \psi(\theta_q) + \psi^\vee(\eta_p) - \langle \theta_q, \eta_p \rangle &= \psi(\theta_q) - \psi(\theta_p) + \langle \theta_p, \eta_p \rangle - \langle \theta_q, \eta_p \rangle & (2.3)\\ &= E_p\left[\psi(\theta_q) - \psi(\theta_p) + \langle \theta_p, T \rangle - \langle \theta_q, T \rangle\right] & (2.4)\\ &= E_p\left[\log\frac{dp}{dq}\right] & (2.5)\\ &= D(p\|q) & (2.6) \end{aligned}$$

□

$\mathcal{X} = \{1, \ldots, n\}, n \in \mathbb{N}$ (カテゴリカル分布) 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 2.4 (経験分布). $x = (x_1, \ldots, x_k) \in \mathcal{X}^k$ に対し

$$\hat{p}_x := \frac{1}{k}\sum_{i=1}^{k} \delta^{x_i} \tag{2.7}$$

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

命題 2.5 (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。$x = (x_1, \ldots, x_k) \in \mathcal{X}^k$ 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:

$$\operatorname{argmin}_{\theta \in \Theta} D(\hat{p}_x \| \mathbf{p}^k(\theta)) = \operatorname{argmax}_{\theta \in \Theta} p^k_\theta(x) \tag{2.8}$$

証明 $\forall \theta \in \Theta$ に対し

$$D(\hat{p}_x \| \mathbf{p}^k(\theta)) = E_{\hat{p}_x}\left[\log \frac{d\hat{p}_x}{d(\mathbf{p}^k(\theta))}\right] \tag{2.9}$$

$$= E_{\hat{p}_x}\left[\log \frac{d\hat{p}_x}{di}\right] - E_{\hat{p}_x}\left[\log \frac{d(\mathbf{p}^k(\theta))}{di}\right] \tag{2.10}$$

$$= (\theta \text{ によらない項}) - \frac{1}{k}\log p_\theta^k(x) \tag{2.11}$$

ゆえに命題の主張が従う。 □

今後の予定

- 射影

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).