

推定関数とパラメトリックモデルの情報幾何

逸見 昌之 数理・推論研究系 准教授

【はじめに】

情報幾何学 ([1], [2]) においては、パラメトリックな統計モデル (確率密度関数の集合) を可微分多様体として扱い、その上の Fisher 計量と呼ばれる Riemann 計量と、指数接続、混合接続と呼ばれる2つの振れのない双対なアファイン接続が、統計的推論の構造を (微分) 幾何学的に議論する上で重要な役割を果たす。また、これらの幾何構造は、統計モデルの2点に対して定義される Kullback-Leibler ダイバージェンスから、ある操作によって自然に導かれることも知られている。本発表では、統計モデル上に与えられた任意の推定関数からも、類似の操作によってこのような双対微分幾何構造 (Riemann 計量とそれに関して双対な2つのアファイン接続) が誘導されることを紹介し、特に、推定関数がパラメータに関して可積分でない場合は、2つのアファイン接続の一方に振れが生じることについて述べる。

【Kullback-Leibler ダイバージェンスから誘導される幾何】

- ・正則な統計モデル $S = \{p(\mathbf{x}; \boldsymbol{\theta}) \mid \boldsymbol{\theta} = (\theta^1, \dots, \theta^d) \in \Theta \subset \mathbf{R}^d\}$
- ・ S の2点 $p(\mathbf{x}; \boldsymbol{\theta}_1)$ と $p(\mathbf{x}; \boldsymbol{\theta}_2)$ の Kullback-Leibler (KL) ダイバージェンス

$$\phi_{KL}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) := \int_{\Omega} p(\mathbf{x}; \boldsymbol{\theta}_2) \log \frac{p(\mathbf{x}; \boldsymbol{\theta}_2)}{p(\mathbf{x}; \boldsymbol{\theta}_1)} \nu(d\mathbf{x})$$

- ・ S の接ベクトル $(\partial_j)_{\boldsymbol{\theta}_1}$ と点 $p(\mathbf{x}; \boldsymbol{\theta}_2)$ の KL プレダイバージェンス (仮称)

$$\rho_{KL}((\partial_j)_{\boldsymbol{\theta}_1}, \boldsymbol{\theta}_2) := (\partial_j)_{\boldsymbol{\theta}_1} \phi_{KL}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = - \int_{\Omega} s^j(\mathbf{x}, \boldsymbol{\theta}_1) p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x})$$

- ・ ϕ_{KL} から誘導される Riemann 計量と双対接続

Fisher 計量 g^F

$$\begin{aligned} g_{jk}^F(\boldsymbol{\theta}) &:= g^F(\partial_j, \partial_k) \\ &= -(\partial_k)_{\boldsymbol{\theta}_2} \rho_{KL}((\partial_j)_{\boldsymbol{\theta}_1}, \boldsymbol{\theta}_2) |_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2=\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\{s^j(\mathbf{x}, \boldsymbol{\theta}) s^k(\mathbf{x}, \boldsymbol{\theta})\} \end{aligned}$$

指数接続 $\nabla^{(e)}$ と混合接続 $\nabla^{(m)}$

$$\begin{aligned} \Gamma_{ij,k}^{(e)}(\boldsymbol{\theta}) &:= g^F(\nabla_{\partial_i}^{(e)} \partial_j, \partial_k) \\ &= -(\partial_i)_{\boldsymbol{\theta}_1} (\partial_k)_{\boldsymbol{\theta}_2} \rho_{KL}((\partial_j)_{\boldsymbol{\theta}_1}, \boldsymbol{\theta}_2) |_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2=\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\{(\partial_i s^j(\mathbf{x}, \boldsymbol{\theta})) s^k(\mathbf{x}, \boldsymbol{\theta})\} \\ \Gamma_{ik,j}^{(m)}(\boldsymbol{\theta}) &:= g^F(\partial_j, \nabla_{\partial_i}^{(m)} \partial_k) \\ &= -(\partial_i)_{\boldsymbol{\theta}_2} (\partial_k)_{\boldsymbol{\theta}_2} \rho_{KL}((\partial_j)_{\boldsymbol{\theta}_1}, \boldsymbol{\theta}_2) |_{\boldsymbol{\theta}_1=\boldsymbol{\theta}_2=\boldsymbol{\theta}} = \int_{\Omega} s^j(\mathbf{x}, \boldsymbol{\theta}) \partial_i \partial_k p(\mathbf{x}; \boldsymbol{\theta}) \nu(d\mathbf{x}) \end{aligned}$$

但し、 $\partial_i := \frac{\partial}{\partial \theta^i}$, $s^i(\mathbf{x}; \boldsymbol{\theta}) = \partial_i \log p(\mathbf{x}; \boldsymbol{\theta})$ とし、 $E_{\boldsymbol{\theta}}$ は $p(\mathbf{x}; \boldsymbol{\theta})$ に関して期待値を取ることを意味するものとする。また、指数接続 $\nabla^{(e)}$ と混合接続 $\nabla^{(m)}$ が (Fisher 計量 g^F に関して) 双対接続であるとは、 S 上の任意のベクトル場 X, Y, Z に対して、

$$X g^F(Y, Z) = g^F(\nabla_X^{(e)} Y, Z) + g^F(Y, \nabla_X^{(m)} Z)$$

が成り立つことを意味する。さらに、指数・混合接続は共に振れのない (振率ゼロの) アファイン接続である。統計モデル S が特に指数型分布族の場合には、指数・混合接続は共に平坦 (振率だけでなく曲率もゼロ) となり、このとき、 S は双対平坦空間と呼ばれる。双対平坦空間では、測地線に関するピタゴラスの定理や正射影定理などが成り立ち、指数型分布族においては最尤推定量が混合接続に関する測地線の正射影によって得られることが知られている。

【推定関数から誘導される幾何】

- ・ S 上の推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) = (u^1(\mathbf{x}, \boldsymbol{\theta}), \dots, u^d(\mathbf{x}, \boldsymbol{\theta}))^T$
- ・推定方程式 $\sum_{i=1}^n \mathbf{u}(\mathbf{X}_i, \boldsymbol{\theta}) = \mathbf{0}$
(但し、 X_1, \dots, X_n は未知の分布 $p(\mathbf{x}; \boldsymbol{\theta}_0)$ からのランダムサンプル)
- ・M-推定量 $\hat{\boldsymbol{\theta}}$ (推定方程式の解) の一致性と漸近正規性 (適当な条件の下で)

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0, \quad \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \text{Avar}(\hat{\boldsymbol{\theta}})) \quad (n \rightarrow \infty)$$

- ・漸近分散共分散行列

$$\text{Avar}(\hat{\boldsymbol{\theta}}) = \{A(\boldsymbol{\theta}_0)\}^{-1} B(\boldsymbol{\theta}_0) \{A(\boldsymbol{\theta}_0)\}^{-1} = \{G(\boldsymbol{\theta}_0)\}^{-1}$$

但し、 $A(\boldsymbol{\theta}) := E_{\boldsymbol{\theta}}\{(\partial \mathbf{u} / \partial \boldsymbol{\theta})(\mathbf{x}, \boldsymbol{\theta})\}$, $B(\boldsymbol{\theta}) := E_{\boldsymbol{\theta}}\{\mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}(\mathbf{x}, \boldsymbol{\theta})^T\}$ であり、また $G(\boldsymbol{\theta})$ は、推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ の標準化 $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ (以下の式で定義) の分散共分散行列であり、Godambe 情報行列と呼ばれる (こともある)。

$$\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta}) := E_{\boldsymbol{\theta}}\{s(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}(\mathbf{x}, \boldsymbol{\theta})^T\} \left[E_{\boldsymbol{\theta}}\{\mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}(\mathbf{x}, \boldsymbol{\theta})^T\} \right]^{-1} \mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$$

但し、 $s(\mathbf{x}, \boldsymbol{\theta})$ は $\boldsymbol{\theta}$ に関するスコア関数であり、 $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ は、ある Hilbert 空間においてスコア関数を推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ に正射影したものであるという幾何学的な意味を持つ。

- ・推定関数 $\mathbf{u}(\mathbf{X}, \boldsymbol{\theta})$ から定義されるプレコントラスト関数

$$\rho_u((\partial_j)_{\boldsymbol{\theta}_1}, \boldsymbol{\theta}_2) := - \int_{\Omega} u_*^j(\mathbf{x}, \boldsymbol{\theta}_1) p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}) \quad (j = 1, \dots, d)$$

但し、 $u_*^j(\mathbf{x}, \boldsymbol{\theta})$ は標準化された推定関数 $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ の j 番目の成分である。
・ ρ_u から誘導される Riemann 計量と双対接続

$$\begin{aligned} g_{jk}(\boldsymbol{\theta}) &:= g(\partial_j, \partial_k) = E_{\boldsymbol{\theta}}\{u_*^j(\mathbf{x}, \boldsymbol{\theta}) u_*^k(\mathbf{x}, \boldsymbol{\theta})\} = G_{jk}(\boldsymbol{\theta}), \\ \Gamma_{ij,k}(\boldsymbol{\theta}) &:= g(\nabla_{\partial_i} \partial_j, \partial_k) = E_{\boldsymbol{\theta}}\{(\partial_i u_*^j(\mathbf{x}, \boldsymbol{\theta})) s^k(\mathbf{x}, \boldsymbol{\theta})\} \\ \Gamma_{ik,j}^*(\boldsymbol{\theta}) &:= g(\partial_j, \nabla_{\partial_i}^* \partial_k) = \int_{\Omega} u_*^j(\mathbf{x}, \boldsymbol{\theta}) \partial_i \partial_k p(\mathbf{x}; \boldsymbol{\theta}) \nu(d\mathbf{x}) \end{aligned}$$

但し、 $G_{jk}(\boldsymbol{\theta})$ は Godambe 情報行列 $G(\boldsymbol{\theta})$ の (j, k) 成分である。ここで、 $\Gamma_{ik,j}^* = \Gamma_{ki,j}^*$ が成り立っているので、 ∇^* は常に振れのないアファイン接続である。一方、 ∇ については、推定関数 $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ が $\boldsymbol{\theta}$ に関して可積分である (つまり $j = 1, \dots, d$ に対して $\partial_j \psi(\mathbf{x}, \boldsymbol{\theta}) = u_*^j(\mathbf{x}, \boldsymbol{\theta})$ を満たす関数 $\psi(\mathbf{x}, \boldsymbol{\theta})$ が存在する) ならば振れを持たないが、一般には振れが生じる。

【おわりに】

パラメトリックな統計モデル上に定義された Riemann 計量と振れのない双対アファイン接続という幾何構造は、「統計多様体」という名の下で数学的に抽象化され、(純粋) 数学の立場からも研究がなされている。その立場からすると、KL ダイバージェンスは「コントラスト関数」という可微分多様体上の2点関数の特別な場合であり、一般にコントラスト関数から上述のような手続きによって統計多様体の構造が得られることが知られている [3]。また、推定関数から定義されるプレコントラスト関数は、可微分多様体上で一般的に定式化された「プレコントラスト関数」[4] の特別な場合であり、そこから誘導される幾何構造は「振れを許す統計多様体」という名で数学的に一般化されている [5]。情報幾何学では、この幾何構造が量子統計モデルに現れることが以前から知られていたが [1]、本研究では、通常の統計モデルにも (可積分でない) 推定関数を通して現れることを示している。双対接続の一方に振れが生じるという状況では、双対平坦 (双対接続の両方が平坦) という事は起こり得ないが、片方の接続のみ平坦であるという事は起こり得る。ある具体的な統計的問題で用いられる推定関数がこのような状況を実現しているが、それについてはポスター発表の当日に述べる。本ポスターでは要点の概略のみしか述べられなかったが、詳細については参考文献 [6] に述べられている。

【参考文献】

- [1] Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*. Amer. Math. Soc., Providence, Oxford University Press, Oxford.
- [2] Amari, S. (2016). *Information Geometry and Its Applications*. Springer.
- [3] Eguchi, S. (1992). Geometry of minimum contrast. *Hiroshima Math. J.* **22**, 631-647.
- [4] 松添博 (2010). 振れを許す統計多様体とプレ・コントラスト関数. 大阪市立大学数学研究所情報幾何関連分野研究会2010講義録.
- [5] 黒瀬俊 (2007). Statistical manifolds admitting torsion. 2007年度福岡大学微分幾何研究会講演録.
- [6] Henmi, M. and Matsuzoe, H. (2018). Statistical Manifolds Admitting Torsion and Partially Flat Spaces. *Geometric Structures of Information* Springer, 37-50.