推定関数から誘導される幾何構造

逸見　昌之　数理・推論研究系　准教授

【はじめに】

情報幾何学においては, 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つの捩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捩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構造】

・正則な統計モデル $S=\{p(\boldsymbol{x};\boldsymbol{\theta})\mid \boldsymbol{\theta}=(\theta^1,...,\theta^d)\in\Theta\subset \boldsymbol{R}^d\}$

・Sの2点$p(\boldsymbol{x};\boldsymbol{\theta}_1)$と$p(\boldsymbol{x};\boldsymbol{\theta}_2)$の**Kullback-Leibler(KL)ダイバージェンス**

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

・Sの接ベクトル$(\partial_j)_{\boldsymbol{\theta}_1}$と点$p(\boldsymbol{x};\boldsymbol{\theta}_2)$の**KLプレコントラスト関数**(仮称)

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

・ϕ_{KL}から誘導されるRiemann計量と双対接続

Fisher計量g^F

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

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

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

但し, $\partial_i:=\frac{\partial}{\partial\theta^i}$, $s^i(\boldsymbol{x};\boldsymbol{\theta})=\partial_i\log p(\boldsymbol{x};\boldsymbol{\theta})$ とし, $E_{\boldsymbol{\theta}}$は$p(\boldsymbol{x};\boldsymbol{\theta})$に関して期待値を取ることを意味する. 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,

$$Xg^F(Y,Z)=g^F(\nabla^{(e)}_X Y,Z)+g^F(Y,\nabla^{(m)}_X Z)$$

が成り立つことである. さらに, 指数・混合接続は共に**捩れのない(捩率ゼロの)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(捩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構造】

・S上の推定関数$\boldsymbol{u}(\boldsymbol{x},\boldsymbol{\theta})=(u^1(\boldsymbol{x},\boldsymbol{\theta}),\ldots,u^d(\boldsymbol{x},\boldsymbol{\theta}))^T$

・推定方程式$\sum_{i=1}^n\boldsymbol{u}(\boldsymbol{X}_i,\boldsymbol{\theta})=\boldsymbol{0}$

(但し, $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$は未知の分布$p(\boldsymbol{x};\boldsymbol{\theta}_0)$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)

$$\hat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}_0,\quad \sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\xrightarrow{d}N\left(\boldsymbol{0},\mathrm{Avar}(\hat{\boldsymbol{\theta}})\right)\quad(n\to\infty)$$

・漸近分散共分散行列

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

但し, $A(\boldsymbol{\theta}):=E_{\boldsymbol{\theta}}\{(\partial\boldsymbol{u}/\partial\boldsymbol{\theta})(\boldsymbol{x},\boldsymbol{\theta})\}$, $B(\boldsymbol{\theta}):=E_{\boldsymbol{\theta}}\left\{\boldsymbol{u}(\boldsymbol{x},\boldsymbol{\theta})\boldsymbol{u}(\boldsymbol{x},\boldsymbol{\theta})^T\right\}$. また, $G(\boldsymbol{\theta})$は推定関数$\boldsymbol{u}(\boldsymbol{x},\boldsymbol{\theta})$の**標準化**$\boldsymbol{u}_*(\boldsymbol{x},\boldsymbol{\theta})$ (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).

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

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

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

$$\rho_u((\partial_j)_{\boldsymbol{\theta}_1},\boldsymbol{\theta}_2):=-\int_{\boldsymbol{R}^p}u^j_*(\boldsymbol{x},\boldsymbol{\theta}_1)p(\boldsymbol{x};\boldsymbol{\theta}_2)d\boldsymbol{x}\quad(j=1,\ldots,d)$$

但し, $u^j_*(\boldsymbol{x},\boldsymbol{\theta})$は標準化された推定関数$\boldsymbol{u}_*(\boldsymbol{x},\boldsymbol{\theta})$ のj番目の成分である.

・ρ_u**から誘導されるRiemann計量**g**と双対接続**∇,∇^*

$$g_{jk}(\boldsymbol{\theta}):=g(\partial_j,\partial_k)=E_{\boldsymbol{\theta}}\{u^j_*(\boldsymbol{x},\boldsymbol{\theta})u^k_*(\boldsymbol{x},\boldsymbol{\theta})\}=G_{jk}(\boldsymbol{\theta}),$$

$$\begin{cases}\Gamma_{ij,k}(\boldsymbol{\theta}):=g(\nabla_{\partial_i}\partial_j,\partial_k)=E_{\boldsymbol{\theta}}[\{\partial_i u^j_*(\boldsymbol{x},\boldsymbol{\theta})\}s^k(\boldsymbol{x},\boldsymbol{\theta})]\\ \Gamma^*_{ik,j}(\boldsymbol{\theta}):=g(\partial_j,\nabla^*_{\partial_i}\partial_k)=\int_{\boldsymbol{R}^p}u^j_*(\boldsymbol{x},\boldsymbol{\theta})\partial_i\partial_k p(\boldsymbol{x};\boldsymbol{\theta})d\boldsymbol{x}\end{cases}$$

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, ∇^* **は常に捩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$\boldsymbol{u}_*(\boldsymbol{x},\boldsymbol{\theta})$が$\boldsymbol{\theta}$に関して可積分である(つまり任意の$j$に対して$\partial_j\psi(\boldsymbol{x},\boldsymbol{\theta})=u^j_*(\boldsymbol{x},\boldsymbol{\theta})$ を満たす関数$\psi(\boldsymbol{x},\boldsymbol{\theta})$が存在する)ならば捩れを持たないが, **一般には捩れが生じる**.

【具体例】

連続する2回の選挙において2つの政党CとLが争っているとする. 以下の表はn番目の選挙区での投票数で, X_{1n},X_{2n},X_nは確率変数とする.

Party	C	L	Total
C	X_{1n}	$m_{1n}-X_{1n}$	m_{1n}
L	X_{2n}	$m_{2n}-X_{2n}$	m_{2n}
Total	X_n	m_n-X_n	m_n

θ^1,θ^2: 1回目の選挙で政党C,Lに投票した人が2回目にCに投票する確率

$\boldsymbol{X}=(X_1,\ldots,X_N)$: 観測データ($N$は選挙区の総数)

$\boldsymbol{\theta}=(\theta^1,\theta^2)$の推定関数(疑似スコア関数)

$$q^1(\boldsymbol{x},\boldsymbol{\theta})=\sum_{n=1}^N\frac{m_{1n}\{x_n-\mu_n(\boldsymbol{\theta})\}}{V_n(\boldsymbol{\theta})},\quad q^2(\boldsymbol{x},\boldsymbol{\theta})=\sum_{n=1}^N\frac{m_{2n}\{x_n-\mu_n(\boldsymbol{\theta})\}}{V_n(\boldsymbol{\theta})}.$$

但し, $\mu_n(\boldsymbol{\theta}),V_n(\boldsymbol{\theta})$はそれぞれ$X_n$の平均(期待値)と分散である.

推定関数から誘導されるリーマン計量gと双対接続∇,∇^*

$$g_{ij}(\boldsymbol{\theta})=\sum_{\boldsymbol{x}}q^i(\boldsymbol{x},\boldsymbol{\theta})q^j(\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x};\boldsymbol{\theta})=\sum_{n=1}^N\frac{1}{V_n(\boldsymbol{\theta})}m_{in}m_{jn}$$

$$\Gamma^*_{ij,k}(\boldsymbol{\theta})=\sum_{\boldsymbol{x}}\{\partial_i\partial_j p(\boldsymbol{x};\boldsymbol{\theta})\}q^k(\boldsymbol{x},\boldsymbol{\theta})=0$$

$$\Gamma_{ij,k}(\boldsymbol{\theta})=\partial_i g_{jk}(\boldsymbol{\theta})-\Gamma^*_{ik,j}(\boldsymbol{\theta})=\sum_{n=1}^N\frac{1-2\theta^i}{V_n(\boldsymbol{\theta})^2}m_{in}m_{jn}m_{kn}.$$

$\boldsymbol{X}=(X_1,\ldots,X_N)$の(同時)確率関数から成る統計モデル$S$は, ∇^*に関しては平坦であるが, ∇に関しては(捩れがあるので)平坦ではない.

【参考文献】Henmi, M. and Matsuzoe, H. (2018). Statistical Manifolds Admitting Torsion and Partially Flat Spaces. *Geometric Structures of Information*, Springer, 37-50.