

推定関数と振れを許す統計多様体

統計数理研究所・データ科学研究系 逸見 昌之

Masayuki Henmi

Department of Data Science

The Institute of Statistical Mathematics

概要

パラメトリックな統計モデル S に対し、推定関数が 1 つ与えられると、プレコントラスト関数と呼ばれる $TS \times S$ 上の関数が (自然に) 定まり、そこから S に振れを許す統計多様体の構造が誘導される。特に、推定関数がパラメータに関して非可積分な場合には、誘導される双対アファイン接続の一方に、実際に振れが生じる ([1])。本稿では、この推定関数から誘導される統計モデルの微分幾何構造について、その背景となる基礎事項と合わせて解説を行う。

1 はじめに

情報幾何学 ([2]) においてこれまで主に扱われてきた幾何構造は、統計多様体 ([3], [4]) と呼ばれる、互いに双対な振れのない 2 つのアファイン接続を持つ Riemann 多様体の構造である。例えば、情報幾何学の主要なテーマの 1 つである統計的推論の微分幾何学 ([5]) においては、可微分多様体としてのパラメトリック統計モデル (確率密度関数の集合) に、Fisher 計量と呼ばれる Riemann 計量と α -接続と呼ばれるアファイン接続の 1 パラメータ族がある種の (統計的に自然な) 基準によって導入され、重要な役割を果たすが、 $(+\alpha)$ -接続と $(-\alpha)$ -接続は (Fisher 計量に関して) 互いに双対な振れのないアファイン接続である。また近年では、機械学習の分野や一般化エントロピーの問題などに関連して、これとは別の幾何構造が論じられているが、そこで扱われているものも基本的には統計多様体である。一方、量子統計モデル (量子状態を表す行列や作用素の集合) の幾何構造を論じる量子情報幾何学においては、以前から、双対接続の一方に振れが生じるような Riemann 多様体の構造が導入されていたが、近年、このような幾何構造を数学的な立場から研究するために、振れをゆるす統計多様体と呼ばれる概念が提唱された ([6])。これは、上記の量子統計モデルの幾何構造を数学的に抽象化したものであり、互いに双対な 2 つのアファイン接続を持つ Riemann 多様体という点では (通常の) 統計多様体と共通であるが、双対接続の一方にのみ振れがないことを要請し、もう一方には振れがあってもよいとする点で、統計多様体を含むより広い概念である。

さて、一般に可微分多様体上にコントラスト関数と呼ばれる二点関数が与えられると、そこから統計多様体の構造が誘導されることが知られている ([7], [8]). 例えば、パラメトリック統計モデル上の Kullback-Leibler ダイバージェンスは最も基本的なコントラスト関数の1つであるが、この場合、誘導される統計多様体の構造は Fisher 計量と ± 1 -接続 (e-接続と m-接続) である。これに対し、近年、可微分多様体上にプレコントラスト関数と呼ばれる関数が与えられると、そこから捩れを許す統計多様体の構造が誘導されることが示された ([9]). プレコントラスト関数とは、コントラスト関数の微分に相当するものを一般化した概念であるが、実は (以下の節で見るように) コントラスト関数から統計多様体の構造が誘導される過程において、少なくとも Riemann 計量と (それに関する) 双対接続を得るにはコントラスト関数の微分の情報さえあれば十分であり、そこでプレコントラスト関数からも同様の幾何構造が誘導される。但し、プレコントラスト関数から誘導される双対接続 (の一方) には捩れが生じる可能性がある。

本稿の主な目的は、パラメトリック統計モデルに推定関数 (未知パラメータの推定に用いられるデータとパラメータの関数) が与えられると、そこから自然にプレコントラスト関数が連想され、それを通して捩れを許す統計多様体の構造が誘導されること ([1]) の紹介および解説を行うことである。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、そのプレコントラスト関数から誘導される双対接続 (の一方) に実際に捩れが生じ得ることについて説明する。しかしながら、この幾何構造の統計的意味や役割についてはまだよく分かっていない部分が多く、最後に、そのことに関するいくつかの課題についても述べる。

2 統計多様体とコントラスト関数

本節ではまず、統計多様体とコントラスト関数について簡単に述べるが、その前に、本稿を通して重要な役割を果たす双対接続の概念について述べておく。なお、以下では多様体とその上の諸量はすべて滑らか (C^∞ 級) であるとする。また、本節と次節で述べる概念のほとんどは擬 Riemann 多様体に対しても定義可能であるが、本稿では Riemann 多様体しか扱わないので、その場合に限定して述べる。

(M, g) を Riemann 多様体、 ∇ を M 上のアファイン接続とする。このとき、Riemann 計量 g に関する ∇ の双対接続 ∇^* とは、以下の関係式を満たす M 上のアファイン接続のことである。

$\mathcal{X}(M)$ を M 上の (C^∞ -級) ベクトル場全体として、

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)).$$

ここで、任意のアフィン接続 ∇ に対して、その双対接続 ∇^* は Riemann 計量 g に応じて唯一つ定まり、また、 g の対称性から $(\nabla^*)^* = \nabla$ が成り立つ。

通常の微分幾何学でよく知られているように、 M 上のアフィン接続 ∇ に対し、その曲率テンソル場 R と捩率テンソル場 T はそれぞれ以下の式によって定義される。

$$\begin{aligned} R(X, Y)Z &:= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)), \\ T(X, Y) &:= \nabla_X Y - \nabla_Y X - [X, Y] \quad (\forall X, \forall Y \in \mathcal{X}(M)). \end{aligned}$$

捩率テンソル場 T が恒等的に 0 のとき、アフィン接続 ∇ は捩れないと言われるが、このときさらに $\nabla^* = \nabla$ であるとする、これは ∇ が (g に関する) Levi-Civita 接続であることを意味する (すなわち、双対接続は Levi-Civita 接続を特別な場合として含む、より広い概念である)。また、アフィン接続 ∇ の (g に関する) 双対接続 ∇^* の曲率テンソル場と捩率テンソル場をそれぞれ R^*, T^* とすると、一般に「 $R = 0 \iff R^* = 0$ 」が成り立つが、 $T = 0$ のとき必ずしも $T^* = 0$ は成り立たない。

2.1 統計多様体

統計多様体 (statistical manifold) は、その言葉が示す通り、パラメトリックな統計モデルの微分幾何学的な構造 (特に Fisher 計量と α -接続) を数学的に抽象化した概念で、もともと [3] によって導入されたものであるが、後にアフィン微分幾何学との関連から [4] により改めて定義がなされた。ここでは、すでに多くの場合で定着している [4] による定義を述べる。

定義 1 (統計多様体)

(M, g) を Riemann 多様体、 ∇ を M 上の捩れないアフィン接続とする。このとき、 (M, g, ∇) が統計多様体であるとは、 ∇g が対称な $(0, 3)$ -テンソル場である、すなわち

$$(\nabla_X g)(Y, Z) = (\nabla_Y g)(X, Z) \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)) \quad (1)$$

が成り立つことである。

M 上の捩れないアフィン接続 ∇ に対して、Riemann 計量 g に関するその双対接続を ∇^* とすると、条件 (1) は ∇^* が捩れないこと (*i.e.* $T^* = 0$) と同値である。よって、統計多様体 (M, g, ∇) が与えられることと、Riemann 多

様体 (M, g) に捩れない2つの双対接続 ∇, ∇^* が与えられることは幾何構造としては同等であるが, 統計多様体という概念は(ここでは)三つ組 (M, g, ∇) に対して定義されるものである. 統計多様体 (M, g, ∇) に対して, (M, g, ∇^*) もまた統計多様体となるが, これを双対統計多様体と呼ぶ. また, アファイン接続 ∇ が平坦であるとき (*i.e.* $R = 0, T = 0$ のとき), 統計多様体 (M, g, ∇) は特に Hesse 多様体 ([10]) と呼ばれ, さらにこのとき, 四つ組 (M, g, ∇, ∇^*) を双対平坦空間と呼ぶ(統計多様体では $T = T^* = 0$ なので, ∇ が平坦であれば自動的に ∇^* も平坦になることに注意).

2.2 コントラスト関数

コントラスト関数 (contrast function) は, 例えば統計モデル上の Kullback-Leibler ダイバージェンスのような, 多様体上の(必ずしも対称的でない)二点間の隔たり度を抽象化した概念で, [7] によって導入されたものであるが, まずその定義に必要な記法を準備する.

一般に, 多様体 M の直積 $M \times M$ 上の実数値関数 ϕ と M 上のベクトル場 $X_1, \dots, X_i, Y_1, \dots, Y_j$ に対して, M 上の関数 $\phi[X_1, \dots, X_i | Y_1, \dots, Y_j]$ を以下によって定める. $\forall r \in M$ に対して

$$\phi[X_1, \dots, X_i | Y_1, \dots, Y_j](r) := (X_1)_p \cdots (X_i)_p (Y_1)_q \cdots (Y_j)_q \phi(p, q)|_{p=r, q=r}.$$

但し, 右辺において (q に関して) Y_1, \dots, Y_j で微分を行わない場合は, 左辺を $\phi[X_1, \dots, X_i |]$ と記し, (p に関して) X_1, \dots, X_i で微分を行わない場合は, $\phi[| Y_1, \dots, Y_j]$ と記す.

定義 2 (コントラスト関数)

$M \times M$ 上の実数値関数 ϕ が以下の条件を満たすとき, ϕ を M 上のコントラスト関数という.

- (a) $\phi(p, p) = 0 \quad (\forall p \in M)$
- (b) $\phi[X |] = \phi[| X] = 0 \quad (\forall X \in \mathcal{X}(M))$
- (c) $g(X, Y) := -\phi[X | Y] \quad (\forall X, \forall Y \in \mathcal{X}(M))$ は M 上の Riemann 計量である.

ϕ が M 上のコントラスト関数であるとき, これらの条件が満たされることから, $M \times M$ の対角線集合 $\{(r, r) | r \in M\}$ のある近傍において

$$\phi(p, q) \geq 0, \quad \phi(p, q) = 0 \iff p = q \quad (2)$$

が成り立つ. コントラスト関数は一般に対称性 ($\phi(p, q) = \phi(q, p)$) を満たすとは限らないが, このように少なくとも (局所的には) 距離関数の第一公理を満たすことから, 何らかの意味で多様体 M 上の二点間の隔たりを測っているものと解釈される. ちなみに, もともと [7] では M 全体で (2) が成り立つことによってコントラスト関数を定義しているが, 以下に示すことから分かるように, コントラスト関数から統計多様体の構造を誘導する際には, $M \times M$ の対角線集合の近傍におけるコントラスト関数の挙動しか影響しないので, ここでは [8] に従い, 上記のような定義を行っている.

多様体 M 上にコントラスト関数 ϕ が与えられたとき, M 上のアフィン接続 ∇, ∇^* が以下の式によって定まる.

$$g(\nabla_X Y, Z) = -\phi[XY|Z], \quad g(Y, \nabla_X^* Z) = -\phi[Y|XZ].$$

ここで, g は ϕ から誘導される Riemann 計量 (定義 2 の (c) における g) であり, このとき ∇, ∇^* は g に関して互いに双対な捩れないアフィン接続となる. よって, $\nabla g, \nabla^* g$ はどちらも対称な $(0, 3)$ -テンソル場であり, (M, g, ∇) と (M, g, ∇^*) は互いに双対な統計多様体となるが, 特に (M, g, ∇) をコントラスト関数 ϕ から誘導された統計多様体と呼ぶ.

このように, コントラスト関数からは (1つの) 統計多様体の構造が誘導されるが, 逆に任意の統計多様体に対して, その構造を誘導するようなコントラスト関数が必ず存在することが知られている ([11]). 但し, 誘導される統計多様体の構造は, コントラスト関数の $M \times M$ の対角線集合の近傍における挙動のみに依存するので, その存在は一意的であるとは限らない. コントラスト関数の (局所的な) 構成法の 1 つとして, 双対接続に関する測地線を用いた方法がある ([12]).

2.3 パラメトリックモデルと Kullback-Leibler ダイバージェンス

ここでは, 統計多様体とその上のコントラスト関数の (最も基本的な) 例として, パラメトリック (統計) モデルと Kullback-Leibler ダイバージェンスについて述べておく.

統計的推論においては通常, データをある確率変数の実現値と見なし, 観測データからその確率変数の従う未知の確率分布に関する推測を行うが, その際, 問題に応じてしばしばその確率分布 (真の分布) があるクラス (集合) に属していると仮定する. そのクラスとしては, データ確率変数の取り得る値の集合 \mathcal{X} 上の, ある基準測度 ν (Lebesgue 測度や数え上げ測度など) に関する確率密度関数の集合を考えるのが通常であるが, 例えば最も基本的なパラ

メトリック推測においては, そのクラスとして

$$S = \{p(\mathbf{x}; \boldsymbol{\theta}) \mid \boldsymbol{\theta} = (\theta^1, \dots, \theta^d) \in \Theta \subset \mathbf{R}^d\}$$

なる形のを考える. ここで, $p(\mathbf{x}; \boldsymbol{\theta})$ は (任意の $\boldsymbol{\theta} \in \Theta$ に対して) 確率密度関数であるから,

$$\int_{\mathcal{X}} p(\mathbf{x}, \boldsymbol{\theta}) \nu(d\mathbf{x}) = 1 \quad (\forall \boldsymbol{\theta} \in \Theta), \quad p(\mathbf{x}, \boldsymbol{\theta}) \geq 0 \quad (\forall \boldsymbol{\theta} \in \Theta, \forall \mathbf{x} \in \mathcal{X})$$

が成り立ち, また Θ (パラメータ空間) は \mathbf{R}^d の開集合である.

集合 S はしばしばパラメトリック (統計) モデルと呼ばれ, 以下の正則条件の下で ($\boldsymbol{\theta}$ を 1 つの局所座標系とする) 可微分多様体と見なせる.

$$(i) \quad p(\mathbf{x}; \boldsymbol{\theta}_1) = p(\mathbf{x}; \boldsymbol{\theta}_2) \quad (\forall \mathbf{x} \in \mathcal{X}) \implies \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

$$(ii) \quad p(\mathbf{x}; \boldsymbol{\theta}) \text{ は } \boldsymbol{\theta} \text{ に関して } C^\infty \text{ 級 } \quad (\forall \mathbf{x} \in \mathcal{X})$$

$$(iii) \quad \det [E_{\boldsymbol{\theta}}\{\mathbf{s}(\mathbf{x}, \boldsymbol{\theta})\mathbf{s}(\mathbf{x}, \boldsymbol{\theta})^T\}] \neq 0 \quad (\forall \boldsymbol{\theta} \in \Theta).$$

但し, $\mathbf{s}(\mathbf{x}, \boldsymbol{\theta})$ は $\boldsymbol{\theta}$ に関するスコア関数, つまり

$$\begin{aligned} \mathbf{s}(\mathbf{x}, \boldsymbol{\theta}) &:= (s^1(\mathbf{x}, \boldsymbol{\theta}), \dots, s^d(\mathbf{x}, \boldsymbol{\theta}))^T, \\ s^i(\mathbf{x}, \boldsymbol{\theta}) &:= \frac{\partial}{\partial \theta^i} \log p(\mathbf{x}; \boldsymbol{\theta}) \quad (i = 1, \dots, d) \end{aligned}$$

であり, また $E_{\boldsymbol{\theta}}$ は $p(\mathbf{x}; \boldsymbol{\theta})$ に関して期待値を取ることを意味する. すなわち, 任意の (可積分) 関数 $f(\mathbf{x})$ に対して

$$E_{\boldsymbol{\theta}}\{f(\mathbf{x})\} := \int_{\mathcal{X}} f(\mathbf{x}) p(\mathbf{x}; \boldsymbol{\theta}) \nu(d\mathbf{x}).$$

さてよく知られているように, パラメトリックモデル S において, **Kullback-Leibler** ダイバージェンス ϕ_{KL} が以下のように定義される.

任意の $p_1(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_1)$, $p_2(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_2) \in S$ に対し,

$$\phi_{KL}(p_1, p_2) := \int_{\mathcal{X}} p_2(\mathbf{x}) \log \frac{p_2(\mathbf{x})}{p_1(\mathbf{x})} \nu(d\mathbf{x}). \quad (3)$$

条件 (i)-(iii) の下で, ϕ_{KL} はパラメトリックモデル S 上の (1 つの) コントラスト関数となるが, ここから S に以下のような統計多様体の構造が誘導される.

$$\text{Riemann 計量 } g_{jk}(\boldsymbol{\theta}) := g(\partial_j, \partial_k) = E_{\boldsymbol{\theta}}\{s^j(\mathbf{x}, \boldsymbol{\theta})s^k(\mathbf{x}, \boldsymbol{\theta})\} \quad (4)$$

$$\text{双対接続 } \begin{cases} \Gamma_{ij,k}(\boldsymbol{\theta}) := g(\nabla_{\partial_i} \partial_j, \partial_k) = E_{\boldsymbol{\theta}}\{[\partial_i s^j(\mathbf{x}, \boldsymbol{\theta})] s^k(\mathbf{x}, \boldsymbol{\theta})\} \\ \Gamma_{ik,j}^*(\boldsymbol{\theta}) := g(\partial_j, \nabla_{\partial_i}^* \partial_k) = \int_{\mathcal{X}} s^j(\mathbf{x}, \boldsymbol{\theta}) \partial_i \partial_k p(\mathbf{x}; \boldsymbol{\theta}) \nu(d\mathbf{x}) \end{cases} \quad (5)$$

但しここで, $\partial_i = \frac{\partial}{\partial \theta^i}$ ($i = 1, \dots, d$) とする. 上記の Riemann 計量 g と双対接続 ∇, ∇^* はそれぞれ Fisher 計量, 指数接続 (e -接続, $(+1)$ -接続), 混合接続 (m -接続, (-1) -接続) と呼ばれ, 統計的推論の微分幾何において最も基本的かつ重要な役割を果たす ([2], [5]). 例えば, パラメトリック推測において最も基本的な最尤推定量は, 指数型分布族において, m -接続に関する測地線 (m -測地線) の直交射影として捉えられる.

3 振れを許す統計多様体とプレコントラスト関数

3.1 振れを許す統計多様体

最初の節でも述べたように, 振れを許す統計多様体 (statistical manifold admitting torsion) は量子統計モデルに現れるような, 双対接続の一方に振れが生じる構造を数学的に抽象化した概念で, [6] によって導入されたものである. それは, 統計多様体の定義における (1) をより一般化することで, 次のように定義される.

定義 3 (振れを許す統計多様体)

(M, g) を Riemann 多様体, ∇ を M 上のアファイン接続とする. このとき, (M, g, ∇) が振れを許す統計多様体であるとは

$$(\nabla_X g)(Y, Z) - (\nabla_Y g)(X, Z) = -g(T(X, Y), Z) \quad (6)$$

$$(\forall X, \forall Y, \forall Z \in \mathcal{X}(M))$$

が成り立つことである. 但し, T は ∇ の振率テンソル場とする.

統計多様体の定義と異なる点は, アファイン接続 ∇ に振れが生じていてもよいということであり, (6) は, その下で ∇ の (g に関する) 双対接続 ∇^* が振れを持たないことと同値な条件である (∇ に振れがない場合は (1) に帰着することに注意). 但し, ∇^* が振れを持たないとは言っても, (M, g, ∇^*) は必ずしも統計多様体ではないことに注意する ($(\nabla^*)^* = \nabla$ に振れが生じている可能性がある). また上記の定義から, Riemann 多様体 (M, g) に振れのないアファイン接続 ∇ が与えられたとき, (M, g, ∇^*) は振れを許す統計多様体となる.

振れを許す統計多様体 (M, g, ∇) に対して, ∇ の曲率テンソル場 R が (恒等的に) 0 となるとき, ∇ は平坦とは限らないが, 双対接続 ∇^* は平坦となる ($R = 0$ より ∇^* の曲率テンソル場 R^* も 0 となり, またもともと ∇^* は振れがないため). このとき, (M, g, ∇) を遠隔平行性空間と呼ぶ(ことにする)が, 統

計多様体の場合と異なり, (M, g, ∇, ∇^*) は必ずしも双対平坦空間にはならないことに注意すべきである.

3.2 . プレコントラスト関数

2.2節でみたように, 多様体上にコントラスト関数が与えられるとそこから統計多様体の構造が誘導されるが, Riemann 計量と双対接続を定義する際には, コントラスト関数を(ベクトル場で)一回微分したものがあれば十分である. そこで, コントラスト関数の微分の持つ性質を公理化し, [9]により, プレコントラスト関数(pre-contrast function)と呼ばれる概念が導入されたが, まずその定義に必要な記法を準備する.

一般に, 多様体 M とその接束 TM の直積 $TM \times M$ 上の実数値関数 ρ と M 上のベクトル場 $X_1, \dots, X_i, Y_1, \dots, Y_j, Z$ に対して, M 上の関数 $\rho[X_1, \dots, X_i Z | Y_1, \dots, Y_j]$ を以下によって定める. $\forall r \in M$ に対して

$$\rho[X_1, \dots, X_i Z | Y_1, \dots, Y_j](r) := (X_1)_p \cdots (X_i)_p (Y_1)_q \cdots (Y_j)_q \rho(Z_p, q)|_{p=r, q=r}.$$

この記法は, 形式的にはコントラスト関数の場合と全く同じであるが, 右辺に含まれるベクトル場 Z の役割が異なることに注意されたい(その違いを強調するためにここでは Z という別の文字を用いている).

定義 4 (プレコントラスト関数)

$TM \times M$ 上の実数値関数 ρ が以下の条件を満たすとき, ρ を M 上のプレコントラスト関数という.

- (a) $\rho(f_1 X_1 + f_2 X_2, q) = 0$ ($\forall f_i \in C^\infty(M), \forall X_i \in \mathcal{X}(M), \forall q \in M$)
- (b) $\rho[X |] = 0$ ($\forall X \in \mathcal{X}(M)$) i.e. $\rho(X_p, p) = 0$ ($\forall p \in M$)
- (c) $g(X, Y) := -\rho[X | Y]$ ($\forall X, \forall Y \in \mathcal{X}(M)$) は M 上の Riemann 計量である.

多様体 M 上のコントラスト関数 ϕ に対し,

$$\rho(X_p, q) := X_p \phi(p, q) \quad (\forall p, \forall q \in M, \forall X_p \in T_p(M))$$

によって定義される $TM \times M$ 上の関数 ρ は, (当然のことながら) プレコントラスト関数となる. また, 多様体 M 上にプレコントラスト関数 ρ が与えられると, コントラスト関数の場合と同様に

$$g(\nabla_X Y, Z) = -\rho[XY | Z], \quad g(Y, \nabla_X^* Z) = -\rho[Y | XZ]$$

によって、 M 上の2つのアファイン接続 ∇, ∇^* が定まり、これらは ρ から誘導される Riemann 計量 g (定義4の(c)における g) に関して互いに双対的である。但しここで、 ∇^* は (常に) 捩れを持たないが、 ∇ の方には捩れが生じる可能性がある。したがってこのとき、 (M, g, ∇) は捩れを許す統計多様体となる。

3.3 遠隔平行性空間とプレコントラスト関数

統計多様体 (M, g, ∇) において、特に2つの双対接続 ∇, ∇^* が平坦となる双対平坦空間の場合には、もとの統計多様体の構造を誘導する標準的なコントラスト関数 (∇ -ダイバージェンス) が定まり、それに関してピタゴラスの定理や射影定理などが成り立つことはよく知られている ([2], [5])。一方、捩れを許す統計多様体 (M, g, ∇) においては、すでに述べたように ∇ に捩れが生じている可能性があるため、その曲率テンソルが消えていても双対平坦になるとは限らないが、少なくとも ∇ の双対接続 ∇^* は平坦となり、この場合、以下のような形で射影定理が成り立つ。

命題1 (標準的なプレコントラスト関数と一般化射影定理)

(M, g, ∇) を遠隔平行性空間 (*i.e.* ∇^* -平坦な捩れを許す統計多様体) とし、 (U, η_i) を ∇^* に関する任意のアファイン座標近傍とする。このとき、以下のことが成り立つ。

(a) U の任意の2点 p, q に対し、 $\gamma^* : [0, 1] \rightarrow U$ を $\gamma^*(0) = p, \gamma^*(1) = q$ となる ∇^* -測地線とし、 $\dot{\gamma}^*(0)$ を p における γ^* の接ベクトルとする。このとき、

$$\rho(Z_p, q) := -g_p(Z_p, \dot{\gamma}^*(0)) \quad (\forall p, \forall q \in U, \forall Z_p \in T_p(U)) \quad (7)$$

によって定まる $TU \times U$ 上の関数 ρ は、もとの捩れを許す統計多様体の構造 (の U への制限) (U, g, ∇) を誘導する U 上のプレコントラスト関数である。

(b) ρ を (7) で定まるプレコントラスト関数とすると、 U の任意の部分多様体 N に対して、以下のことが成り立つ。

$$\begin{aligned} & q \in U \text{ を起点とする } \nabla^* \text{-測地線が } p \in N \text{ において } N \text{ と直交する} \\ \iff & \forall Z_p \in T_p(N), \rho(Z_p, q) = 0 \end{aligned} \quad (8)$$

(M, g, ∇) が特に統計多様体の場合は, (7) のプレコントラスト関数は双対平坦空間における ∇ -ダイバージェンスを (Z_p) で微分したものに一致する ([2], [12]). そこで, (7) で定まる関数 ρ を, (遠隔平行性空間 (M, g, ∇) における) 標準的なプレコントラスト関数と呼ぶことにする. また, 双対平坦空間の場合には, (b) はダイバージェンスに関する射影定理に帰着する. そこで (b) を一般化射影定理と呼ぶことにする.

4 推定関数とその標準化

4.1 推定関数

推定関数 (estimating function) とは, データを生成している確率分布に関する未知パラメータを推定するために用いられる関数の一種であるが, 本稿では以下のように定義する.

定義 5 (パラメトリックモデルにおける推定関数)

$S = \{p(\mathbf{x}; \boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$ (Θ は \mathbf{R}^d の開集合) をデータ空間 $(\mathcal{X}, \mathcal{B}, \nu)$ 上の確率密度関数に対するパラメトリックモデルとする. このとき, S における (不偏な) 推定関数とは, 以下の条件を満たす $\mathcal{X} \times \Theta$ 上の \mathbf{R}^d 値関数 \mathbf{u} のことである.

- (i) $E_{\boldsymbol{\theta}}\{\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})\} = \mathbf{0} \quad (\forall \boldsymbol{\theta} \in \Theta)$
- (ii) $E_{\boldsymbol{\theta}}\{\|\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})\|^2\} < \infty \quad (\forall \boldsymbol{\theta} \in \Theta)$
- (iii) $\det \left[E_{\boldsymbol{\theta}} \left\{ \frac{\partial \mathbf{u}}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \right\} \right] \neq 0 \quad (\forall \boldsymbol{\theta} \in \Theta)$

これらの条件のうち, 特に (i) は推定関数の不偏性 (unbiasedness) と呼ばれ, (推定関数から得られる) 推定量の一致性を保証するための重要な条件である.

S に属する未知の確率分布 $p(\mathbf{x}; \boldsymbol{\theta}_0)$ からのランダムサンプル $\mathbf{X}_1, \dots, \mathbf{X}_n$ (互いに独立に同一の分布 $p(\mathbf{x}; \boldsymbol{\theta}_0)$ に従う n 個の確率変数) が与えられたとき, 推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ によって推定方程式

$$\sum_{i=1}^n \mathbf{u}(\mathbf{X}_i, \boldsymbol{\theta}) = \mathbf{0}$$

が立てられる. この解として得られる, 未知パラメータ $\boldsymbol{\theta}_0$ の推定量 $\hat{\boldsymbol{\theta}}$ を (しばしば) M-推定量というが, M-推定量は上記の条件を含むいくつかの正則条

件 (例えば [13] の第 5 章等を参照) の下で, 以下のような一致性と漸近正規性を持つ.

$$\begin{aligned} \text{一 致 性} \quad & \hat{\theta} \longrightarrow \theta_0 \quad (n \rightarrow \infty \text{ のとき確率収束}) \\ \text{漸 近 正 規 性} \quad & \sqrt{n}(\hat{\theta} - \theta_0) \longrightarrow N(0, \text{Avar}(\hat{\theta})) \quad (n \rightarrow \infty \text{ のとき分布収束}) \end{aligned}$$

ここで, $\text{Avar}(\hat{\theta})$ は $\hat{\theta}$ の漸近分散共分散行列と呼ばれ, 推定関数 $u(x, \theta)$ によって

$$\begin{aligned} \text{Avar}(\hat{\theta}) &= \{A(\theta_0)\}^{-1} B(\theta_0) \{A(\theta_0)\}^{-T}, \quad (9) \\ \text{但し } A(\theta) &:= E_{\theta} \left\{ \frac{\partial u}{\partial \theta}(x, \theta) \right\}, \quad B(\theta) := E_{\theta} \{u(x, \theta)u(x, \theta)^T\} \end{aligned}$$

と表されるが, これが (対称行列の順序関係の意味で) 小さいほど, $\hat{\theta}$ は (漸近的に) 精度の良い推定量ということになる.

2.3 節でも述べたように, パラメトリック推測 (パラメトリックモデル S に基づく統計的推測) における最も基本的なパラメータ推定法は最尤推定法 (最尤法) である. 最尤法ではスコア関数が推定関数の役割を果たし, またその意味では上記のような推定関数に基づくパラメータ推定法は最尤法の拡張とも見なせるが, よく知られているように, 最尤推定量の漸近分散共分散行列は Fisher 情報行列の逆行列で与えられ, M -推定量を含む漸近正規性を持った推定量のクラスにおいて最小の漸近分散共分散行列となる. したがって, パラメトリック推測においては多くの場合, 最尤法が用いられるが, 最尤法があまり適切ではないと考えられる場合にはそれ以外の方法が用いられることもある. 例えば, データに外れ値が混入している場合には最尤法はその影響を受けやすいので, (外れ値の影響を緩和するために) スコア関数をやや修正した形の推定関数を用いたり ([14], [15], [16]), 最尤法による計算が複雑あるいは困難な場合には, 計算がより単純となるような推定法 (推定関数) を用いたりすることがある ([17]).

また推定関数は, パラメータ θ が興味あるパラメータと局外パラメータ (興味のないパラメータ) から構成されている場合に, 興味あるパラメータだけを直接推定するためにも用いられるが ([18]), 特に局外パラメータが無限次元となるようなセミパラメトリックモデルにおいては, その局外パラメータの推定を介さずに有限次元の興味あるパラメータを直接推定する目的で, 応用上よく用いられる. セミパラメトリックモデルにおける推定関数については, [19] による情報幾何学の観点からの研究があるが, 本稿で扱う推定関数は定義 5 のようなパラメトリックモデルにおける推定関数で, しかも確率分布を規定するパラメータ全体を推定するためのものであり, そこから誘導される幾何構造として以下で議論するものは, [19] で扱われているものとは異なる.

4.2 推定関数の標準化

推定関数から誘導される幾何構造について述べる前に、そのために必要となる概念について述べておく。

パラメトリックモデル $S = \{p(\mathbf{x}; \boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta\}$ (Θ は \mathbf{R}^d の開集合) における2つの推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ と $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ に対し、パラメータ $\boldsymbol{\theta}$ のみに依存するある d 次正則行列 $M(\boldsymbol{\theta})$ が存在して

$$\mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) = M(\boldsymbol{\theta})\mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) \quad (\forall \boldsymbol{\theta} \in \Theta, \forall \mathbf{x} \in \mathcal{X})$$

と書けるとき、 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ と $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ は互いに同値であると言われる。同値と言われる所以は、これらの与える M-推定量が同じものになるということであるが、実際、上記の推定関数 $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ に対し、(S に属するある分布からの) ランダムサンプルに基づいて構成される推定方程式は、推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ からのもものに帰着する。

一般に (S における) 推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ が与えられたとき、それと同値な推定関数は無数にあるが、特に

$$\begin{aligned} \mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta}) &:= E_{\theta} \{ \mathbf{s}(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}(\mathbf{x}, \boldsymbol{\theta})^T \} [E_{\theta} \{ \mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}(\mathbf{x}, \boldsymbol{\theta})^T \}]^{-1} \mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) \\ &= -A(\boldsymbol{\theta}) \{ B(\boldsymbol{\theta}) \}^{-1} \mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) \end{aligned} \quad (10)$$

で与えられるものを推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ の標準化 (standardization), あるいは標準化された推定関数 (standardized estimating function) という ([20]). これは、幾何学的には Hilbert 空間

$$\mathcal{H}_{\theta} := \{a(\mathbf{x}) | E_{\theta} \{a(\mathbf{x})\} = 0, E_{\theta} \{a(\mathbf{x})^2\} < \infty\}$$

$$\text{内積 } \langle a(\mathbf{x}), b(\mathbf{x}) \rangle_{\theta} := E_{\theta} \{a(\mathbf{x})b(\mathbf{x})\} \quad (\forall a(\mathbf{x}), \forall b(\mathbf{x}) \in \mathcal{H}_{\theta})$$

において、 $\boldsymbol{\theta}$ に関するスコア関数 $\mathbf{s}(\mathbf{x}, \boldsymbol{\theta})$ の各成分を $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ の各成分が張る (d 次元) 線型部分空間に直交射影することによって得られるものであるが (すなわち $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ の第 i 成分はスコア関数の第 i 成分をこの部分空間へ直交射影したもの), この標準化を用いると、推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ から得られる M-推定量 $\hat{\boldsymbol{\theta}}$ の漸近分散共分散行列 (9) は以下のように表される。

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

$$\text{但し } G(\boldsymbol{\theta}) := E_{\theta} \{ \mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta}) \mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})^T \}.$$

推定関数 $u(\boldsymbol{x}, \boldsymbol{\theta})$ が特にスコア関数のときは、その標準化もまたスコア関数であり、行列 $G(\boldsymbol{\theta})$ は Fisher 情報行列に帰着するので、Fisher 情報行列のある種の一般化と見なされる。また、 $G(\boldsymbol{\theta}_0)$ が (対称行列の順序関係の意味で) 大きいほど推定量 $\hat{\boldsymbol{\theta}}$ の漸近分散共分散行列が小さくなる (漸近的に精度よく推定できる) という意味で、行列 $G(\boldsymbol{\theta})$ は Fisher 情報行列と同様の役割を果たし、Godambe 情報行列と呼ばれることもある。

推定関数 $u(\boldsymbol{x}, \boldsymbol{\theta})$ に対し、その標準化と言った場合、

$$\boldsymbol{u}^*(\boldsymbol{x}, \boldsymbol{\theta}) := \{A(\boldsymbol{\theta})\}^{-1} \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\theta})$$

で与えられるものを指すこともあるが ([21]), これは $u(\boldsymbol{x}, \boldsymbol{\theta})$ から得られる M-推定量 $\hat{\boldsymbol{\theta}}$ の影響関数 (influence function) と呼ばれるものの (-1) 倍に相当し、またこれを用いると $\hat{\boldsymbol{\theta}}$ の漸近分散共分散行列は以下のように表される。

$$\text{Avar}(\hat{\boldsymbol{\theta}}) = E_{\theta_0} \{ \boldsymbol{u}^*(\boldsymbol{x}, \boldsymbol{\theta}_0) \boldsymbol{u}^*(\boldsymbol{x}, \boldsymbol{\theta}_0)^T \}.$$

すなわち、この場合は (逆行列を取らずに) 標準化された推定関数の分散共分散行列そのもので与えられるが、これは $\boldsymbol{u}_*(\boldsymbol{x}, \boldsymbol{\theta})$ と $-\boldsymbol{u}^*(\boldsymbol{x}, \boldsymbol{\theta})$ が以下の意味で互いに双対な関係にあることによる。

$$\langle \boldsymbol{u}_{*i}(\boldsymbol{x}, \boldsymbol{\theta}), -\boldsymbol{u}^{*j}(\boldsymbol{x}, \boldsymbol{\theta}) \rangle_{\theta} = \delta_i^j,$$

但し $\boldsymbol{u}_{*i}(\boldsymbol{x}, \boldsymbol{\theta}), \boldsymbol{u}^{*i}(\boldsymbol{x}, \boldsymbol{\theta})$ はそれぞれ $\boldsymbol{u}_*(\boldsymbol{x}, \boldsymbol{\theta}), \boldsymbol{u}^*(\boldsymbol{x}, \boldsymbol{\theta})$ の第 i 成分 ($i = 1, \dots, d$) を表し、また δ_i^j はクロネッカーのデルタである。

このように、推定関数の標準化には 2 種類のもものが存在しているが、以下の議論で用いるのは前者、すなわち $\boldsymbol{u}_*(\boldsymbol{x}, \boldsymbol{\theta})$ の方であり、推定関数から誘導される幾何構造 (捩れを許す統計多様体) を論じる上で重要な役割を果たす。

5 推定関数から誘導される捩れを許す統計多様体

5.1 推定関数から連想されるプレコントラスト関数

2.3 節で見たように、パラメトリックモデル S 上の Kullback-Leibler ダイバージェンス (3) はコントラスト関数であるから、これを 1 回微分することにより、以下のような S 上のプレコントラスト関数 ρ_{KL} が得られる。

$$\begin{aligned} \rho_{KL}((\partial_j)_{p_1}, p_2) &:= (\partial_j)_{p_1} \phi_{KL}(p_1, p_2) \\ &= - \int_{\mathcal{X}} s^j(\boldsymbol{x}, \boldsymbol{\theta}_1) p(\boldsymbol{x}; \boldsymbol{\theta}_2) \nu(d\boldsymbol{x}) \end{aligned} \quad (11)$$

但し, $p_l(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_l)$ ($l = 1, 2$) は S の任意の元であり, また, $\partial_j = \frac{\partial}{\partial \theta_j}$ ($j = 1, \dots, d$) である. 3.2 節の手続きに従ってこのプレコントラスト関数から誘導される幾何構造は, コントラスト関数としての Kullback-Leibler ダイバージェンスから誘導される幾何構造 (Fisher 計量と e, m -接続) と同一のものであるが, この幾何構造は最尤推定量の幾何学的描像を与えるものである. 一方, プレコントラスト関数 (11) は最尤推定量を与える推定関数であるスコア関数 (のみ) によって記述されている. そこで, 一般の推定関数についても (11) と同様のものを考えれば, その M -推定量に関連する幾何構造を誘導するプレコントラスト関数が得られるであろうと期待される. しかしながら, (11) におけるスコア関数 (の第 j 成分) のところを単に一般の推定関数 (の第 j 成分) に置き換えただけでは, プレコントラスト関数が得られるとは限らない. ここで重要な役割を果たすのが, 推定関数の標準化である.

命題 2 (推定関数から連想されるプレコントラスト関数)

パラメトリックモデル S における推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ に対し,

$$\rho_u((\partial_j)_{p_1}, p_2) := - \int_{\mathcal{X}} u_*^j(\mathbf{x}, \boldsymbol{\theta}_1) p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}) \quad (j = 1, \dots, d) \quad (12)$$

によって, S 上のプレコントラスト関数 ρ_u が定まる. 但し, $u_*^j(\mathbf{x}, \boldsymbol{\theta})$ は (10) で与えられる, 推定関数 $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ の標準化 $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ の第 j 成分である. また, この定義は座標系 (パラメータ) の取り方に依らない.

プレコントラスト関数は, 3.2 節の定義 4 における (a) によって与えられるベクトル場 (あるいは接ベクトル) に関する線型性を持つので, (もしプレコントラスト関数であるならば) (12) のように S の任意の 2 点 p_1, p_2 と p_1 における接空間の基底に対して定義されていれば十分であるが, (12) の形から, この定義において推定関数の標準化 (の各成分) $\{u_*^1(\mathbf{x}, \boldsymbol{\theta}), \dots, u_*^d(\mathbf{x}, \boldsymbol{\theta})\}$ は S の接空間 $T_p(S)$ ($p(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta})$) の基底 $\{(\partial_1)_p, \dots, (\partial_d)_p\}$ の役割を担っているものと考えられる. 実際, 座標変換 (パラメータ変換) $\boldsymbol{\eta} = \Phi(\boldsymbol{\theta})$ に対して, 推定関数が $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ から $\mathbf{v}(\mathbf{x}, \boldsymbol{\eta}) = \mathbf{u}(\mathbf{x}, \Phi^{-1}(\boldsymbol{\eta}))$ へと変換されるものとする, これらの標準化の間には, 多様体の座標変換に伴う接空間の基底変換ルール

と同じものが成り立つ。すなわち、

$$\mathbf{v}_*(\mathbf{x}, \boldsymbol{\eta}) = \left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\eta}} \right)^T \mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta}).$$

またこのことから、(12) の定義が座標系の取り方に依らないとも言える。

式(12)によって定義される $TS \times S$ 上の関数 ρ_u がプレコントラスト関数の定義4における3つの条件を満たすことは以下のようにして分かる。まず、(a) のベクトル場(接ベクトル)に関する線型性については、(これを前提にして(12)のように定義しているわけであるが) (12) の右辺の積分の線型性による。次に(b)については、これは(ほとんど)推定関数の不偏性そのものである。すなわち、 $\forall p(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}) \in S, j = 1, \dots, d$ に対して

$$\rho_u[\partial_j |](p) = \rho_u((\partial_j)_p, p) = - \int_{\mathcal{X}} u_*^j(\mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}; \boldsymbol{\theta}) \nu(d\mathbf{x}) = 0$$

($\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$ の不偏性からその標準化 $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$ もまた不偏性を持つことに注意)。また(c)については、 $\forall p(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}) \in S, j = 1, \dots, d, k = 1, \dots, d$ に対して

$$\begin{aligned} g_{jk}(\boldsymbol{\theta}) &:= -\rho_u[\partial_j | \partial_k](p) \\ &= \frac{\partial}{\partial \theta_2^k} \Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}} \int_{\mathcal{X}} u_*^j(\mathbf{x}, \boldsymbol{\theta}_1) p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}) \\ &= E_{\boldsymbol{\theta}} \{ u_*^j(\mathbf{x}, \boldsymbol{\theta}) s^k(\mathbf{x}, \boldsymbol{\theta}) \} \\ &= E_{\boldsymbol{\theta}} \{ u_*^j(\mathbf{x}, \boldsymbol{\theta}) u_*^k(\mathbf{x}, \boldsymbol{\theta}) \} \quad (\text{推定関数の標準化の定義より}) \\ &= G(\boldsymbol{\theta})_{jk} \quad (\text{Godambe 情報行列の第 } (j, k) \text{ 成分}). \end{aligned}$$

すなわち、定義4の(c)における g は、今の場合、Godambe 情報行列を(座標系に関する)計量行列とするような Riemann 計量となる。

5.2 振れを許す統計多様体

3.2節で述べたように、一般にプレコントラスト関数からは振れを許す統計多様体の構造が誘導されるが、(12)によって与えられるプレコントラスト関数 ρ_u については、上記の Riemann 計量に加えて以下のような接続係数をも

つ双対接続 ∇, ∇^* が誘導される.

$$\begin{aligned}
\Gamma_{ij,k}(\boldsymbol{\theta}) &:= g(\nabla_{\partial_i} \partial_j, \partial_k) \\
&= -\rho_u[\partial_i \partial_j | \partial_k](p) \\
&= \frac{\partial}{\partial \theta_1^i} \Big|_{\theta_1 = \theta_2 = \boldsymbol{\theta}} E_{\boldsymbol{\theta}} \{ u_*^j(\mathbf{x}, \boldsymbol{\theta}_1) s^k(\mathbf{x}, \boldsymbol{\theta}_2) \} \\
&= E_{\boldsymbol{\theta}} [\{ \partial_i u_*^j(\mathbf{x}, \boldsymbol{\theta}) \} s^k(\mathbf{x}, \boldsymbol{\theta})], \\
\Gamma_{ik,j}^*(\boldsymbol{\theta}) &:= g(\partial_j, \nabla_{\partial_i}^* \partial_k) \\
&= -\rho_u[\partial_j | \partial_i \partial_k](p) \\
&= \frac{\partial}{\partial \theta_2^i} \Big|_{\theta_1 = \theta_2 = \boldsymbol{\theta}} \int_{\mathcal{X}} u_*^j(\mathbf{x}, \boldsymbol{\theta}_1) \partial_k p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}) \\
&= \int_{\mathcal{X}} u_*^j(\mathbf{x}, \boldsymbol{\theta}) \partial_i \partial_k p(\mathbf{x}; \boldsymbol{\theta}) \nu(d\mathbf{x}).
\end{aligned}$$

これらは形式的には、(Kullback-Leibler ダイバージェンスから誘導される) e -接続と m -接続の接続係数の表式 (5) において、スコア関数のところの一部を (一般の) 推定関数の標準化に置き換えただけであり、 ∇^* の方については、 m -接続の場合と同様に常に振れがないが (実際、 $\Gamma_{ik,j}^* = \Gamma_{ki,j}^*$ が成り立つ)、大きく異なるのは、 ∇ の方で振れが生じ得るということである。但し、もし推定関数 $u(\mathbf{x}, \boldsymbol{\theta})$ の標準化 $u_*(\mathbf{x}, \boldsymbol{\theta})$ がパラメータ $\boldsymbol{\theta}$ に関して可積分であるならば、つまり、ある実数値関数 $\psi(\mathbf{x}, \boldsymbol{\theta})$ が存在して $\partial_j \psi(\mathbf{x}, \boldsymbol{\theta}) = u_*^j(\mathbf{x}, \boldsymbol{\theta})$ ($j = 1, \dots, d$) が成り立つならば、 $\Gamma_{ij,k} = \Gamma_{ji,k}$ となり、 ∇ にも振れがなくなる。またこの場合は、上記の Riemann 計量 g と双対接続 ∇, ∇^* を誘導する S 上のコントラスト関数が、以下のように構成できる。

$\forall p_1(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_1), \forall p_2(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_2) \in S$ に対して、

$$\begin{aligned}
\phi_u(p_1, p_2) &:= \int_{\mathcal{X}} \psi(\mathbf{x}, \boldsymbol{\theta}_2) p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}) - \int_{\mathcal{X}} \psi(\mathbf{x}, \boldsymbol{\theta}_1) p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}) \\
&= \int_{\mathcal{X}} \{ \psi(\mathbf{x}, \boldsymbol{\theta}_2) - \psi(\mathbf{x}, \boldsymbol{\theta}_1) \} p(\mathbf{x}; \boldsymbol{\theta}_2) \nu(d\mathbf{x}).
\end{aligned}$$

しかしながら、 $u_*(\mathbf{x}, \boldsymbol{\theta})$ がパラメータ $\boldsymbol{\theta}$ に関して非可積分な場合は、アフライン接続 ∇ に振れが生じる可能性がある。

推定関数から ((12) のプレコントラスト関数を通じて) 誘導される、上記の振れを許す統計多様体の構造の具体的な例についてはここでは述べないが、それについては [1] を参照されたい。そこでは、非可積分な推定関数として [22] で取り上げられている例に対し、上記の Riemann 計量と双対接続の導出 (計算) を行っているが、この場合、アフライン接続 ∇ には実際に振れが生じ

ており、また ∇^* の方は特に平坦となっている。さらに、推定関数から連想されるプレコントラスト関数 (12) は、(遠隔平行性空間における) 標準的なプレコントラスト関数 (7) に一致している。

6 おわりに

本稿では、推定関数から (プレコントラスト関数を通じて) 誘導される振れを許す統計多様体の構造について、その背景となる基礎事項と共に解説を行った。推定関数から連想されるプレコントラスト関数は、形式的には Kullback-Leibler ダイバージェンスを微分して得られるプレコントラスト関数において、そこに現れるスコア関数の部分を推定関数の標準化に置き換えただけのものであるが、標準化の役割や導出される Riemann 計量が Godambe 情報行列で与えられることなどから、(少なくとも数学的には) 自然なプレコントラスト関数であると考えられる。しかしながら、そこから誘導される振れを許す統計多様体の統計的意味については、まだほとんどよく分かっていない。

例えば、第 3.3 節の一般化射影定理において、仮に多様体 M を有限離散分布全体、 N をその中におけるある統計モデル (パラメトリックモデル) とし、点 q をデータから得られる経験分布、 ρ を N におけるある推定関数から連想されるプレコントラスト関数とすると、(8) はまさに推定方程式そのものであり、推定方程式の解としての M -推定量は、経験分布から統計モデルへの ∇^* -射影 (∇^* -測地線による直交射影) で与えられることが期待される。しかしながら、プレコントラスト関数から直接誘導されるのは統計モデル N の幾何構造 (振れを許す統計多様体の構造) のみであり、それが M 全体にどのように (自然に) 拡張されるのかはあまり自明ではない。

また、最尤法の (外れ値に対する) ロバスト化を目的として導入された β -ダイバージェンス ([14], [23]) もまた、Kullback-Leibler ダイバージェンスと同様にパラメトリックな統計モデル上の 1 つのコントラスト関数となるが、それを微分して得られるプレコントラスト関数において現れる推定関数は、必ずしも標準化されていない。これは、推定関数から (プレコントラスト関数を通じて) 幾何構造を誘導する際、標準化を行うことが唯一の方法というわけではないことを示唆しているが、仮に β -ダイバージェンスから得られるプレコントラスト関数が特別な例だとしても、そこから誘導される幾何構造と標準化を行った場合の幾何構造が、(特に統計的推論との関わりにおいて) どのように異なるかもまだよく分かっていない。

他にも様々な観点から問題が考えられるが、本稿で論じた、推定関数から誘導される幾何構造の統計的意味の解明は今後の課題である。

参考文献

- [1] Henmi, M. and Matsuzoe, H. (2011). Geometry of pre-contrast functions and non-conservative estimating functions, *AIP Conference Proceedings* **1340**, 32-41.
- [2] Amari, S. and Nagaoka, H. (2000). *Method of Information Geometry*, Translations of Mathematical Monographs **191**, Oxford University Press.
- [3] Lauritzen, S.L. (1987). Statistical manifolds, In *Differential Geometry in Statistical Inferences*, IMS Lecture Notes Monograph Series **10**, Institute of Mathematical Statistics, 96-163.
- [4] Kurose, T. (1994). On the divergences of 1-conformally flat statistical manifolds, *Tohoku Math. J.* **46**, 427-433.
- [5] Amari, S. (1985). *Differential Geometrical Methods in Statistics*, Lecture Notes in Statistics **28**, Springer.
- [6] 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.
- [7] Eguchi, S. (1992). Geometry of minimum contrast, *Hiroshima Math. J.* **22**, 631-647.
- [8] Matsuzoe, H. (1999). Geometry of contrast functions and conformal geometry, *Hiroshima Math. J.* **29**, 175-191.
- [9] 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.
- [10] Shima, H. (2007). *The Geometry of Hessian Structures*, World Scientific.
- [11] Matsumoto, T. (1993). Any statistical manifold has a contrast function - On the C^3 -functions taking the minimum at the diagonal of the product manifold, *Hiroshima Math. J.* **23**, 327-332.
- [12] Henmi, M. and Kobayashi, R. (2000). Hooke's law in statistical manifolds and divergences, *Nagoya Math. J.* **159**, 1-24.
- [13] van der Vaart, A.W. (2000). *Asymptotic Statistics*, Cambridge University Press.
- [14] Basu, A., Harris, I.R., Hjort, N.L. and Jones, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence, *Biometrika* **85**, 549-559.
- [15] Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination, *J. Multivariate Anal.* **99**, 2053-2081.

- [16] Fujisawa, H. (2013). Normalized estimating equation for robust parameter estimation, *Electronic Journal of Statistics* **7**, 1587-1606.
- [17] Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods, *Statistica Sinica* **21**, 5-42
- [18] Liang, K.Y. and Zeger, S.L. (1995). Inference based on estimating functions in the presence of nuisance parameters, *Statist. Sci.* **10**, 158-173.
- [19] Amari, S. and Kawanabe, M. (1997). Information geometry of estimating functions in semi-parametric statistical models, *Bernoulli* **3**, 29-54.
- [20] Heyde, C.C. (1997). *Quasi-Likelihood and Its Application*, Springer.
- [21] Godambe, V.P. and Kale, B.K. (1991). Estimating functions: an overview, In *Estimating Functions*, Godambe, V.P. (ed.), Oxford University Press, 3-20.
- [22] McCullagh, P. and Nelder J.A. (1989). *Generalized Linear Models* (2nd ed.), Chapman and Hall.
- [23] Eguchi, S. and Kano, Y. (2001). Robustifying maximum likelihood estimation, *Research Memorandum of the Institute of Statistical Mathematics* No.802.