シンプレクティック幾何的視点での BAYES の定理について

明治薬科大学・薬学教育研究センター　野田　知宣
TOMONORI NODA
GENERAL EDUCATION AND RESEARCH CENTER,
MEIJI PHARMACEUTICAL UNIVERSITY

Abstract. 本稿では多変量正規分布における母平均の推定の場合に Bayes の定理を正準変換で表せる事を述べる。

1. 序

シンプレクティック幾何学は解析力学の一般化・抽象化としての数学理論であり、相空間に対応する多様体 M^{2n} と力学構造を与える微分 2-形式 ω との組 (M^{2n}, ω) を研究する幾何学分野である。ここで ω は可積分条件 $d\omega = 0$ と非退化条件 $\omega^n \neq 0$ を満たす 2-形式である。ω に対する 2 つの条件から Hamilton 方程式の一般化である Hamilton ベクトル場 X_H やエネルギー保存則、Liouville 定理などが導かれる。また、局所的な性質である Darboux 定理が成立する。これらについては例えば Abraham-Marsden [1] などを参照とし、詳しい説明は略する。

シンプレクティック多様体の部分多様体は、シンプレクティック構造 ω が 2-形式だから、各点における部分多様体の接空間と ω の関係性により定められる。話を簡単にする為、正準シンプレクティック構造 ω_0 を備えたシンプレクティック・ベクトル空間 $\mathbb{R}^{2n}$ を考える。一般に、シンプレクティック幾何学において局所不変量は殆ど存在しない。$\mathbb{R}^{2n}$ の部分空間 W を考えると、（線型）正準変換の下で保たれるものとして先ず W の次元 $\dim W$, また ω_0 との関係として ω_0 を W に制限したものの階数 $\operatorname{rank} \omega_0|_W$ が考えられるが、本質的には不変量はこの 2 つしかない。これら 2 つの整数により部分空間（ひいては部分多様体）に名前が付くのであるが、それを簡潔に定義する為、部分空間 $W \subset \mathbb{R}^{2n}$ に対しその ω-直交部分空間を

$$W^{\perp} := \{v \in V \ ; \ \omega_0(w, v) = 0, \ {}^{\forall} w \in W\}$$

で定める。$W \subset W^{\perp}$ のとき W を isotropic ; 逆に $W^{\perp} \subset W$ のとき coisotropic[1] ; $W \subset W^{\perp}$ かつ $\dim W = \frac{1}{2} \dim \mathbb{R}^{2n} = n$ のとき Lagrangian ; $W \cap W^{\perp} = \{0\}$ のとき symplectic と

This work was supported by the Research Institute for Mathematical Sciences, a Joint Usage/Research Center located in Kyoto University.

[1]講演時に最も注意を払ったのがここである。

呼ばれる 4 種類の部分空間が、従って 4 種類の部分多様体が定義される。通常の幾何学においては外の空間の構造が遺伝する部分多様体、ここではシンプレクティック部分多様体が最も基礎的な研究対象となるのではないかと思われるが、シンプレクティック幾何学における部分多様体の中で最もよく研究されており、最も重要とされるのは Lagrange 部分多様体である。Weinstein [17] には

$$\boxed{\text{EVERYTHING IS A LAGRANGIAN SUBMANIFOLD}}$$

と書かれているくらいである。シンプレクティック構造との関係で云えば、Lagrange 部分空間は構造 ω_0 が消えるものの中で極大なものであり、非退化な構造 ω_0 から何も情報が取り出せないもの（で極大）という事になる。本稿でも Lagrange 部分多様体は Bayes 定理をシンプレクティック幾何学の言葉で表す上で非常に重要であるが、簡単なもののみ使用するだけであるので、Weinstein による上の言葉を簡潔に補足しておくに止める。

先ず、Lagrange 部分空間（多様体）は非退化な ω_0 を制限したら消える極大な部分空間であり、これだけでも興味深い。しかしながら、これだけに止まらず Lagrange 部分空間は空間（世界）、函数、シンプレクティック微分同相写像の一般化という側面を持つ。シンプレクティック多様体の典型例は $(T^*\mathbb{R}^n, \omega_0)$ であるが、このベクトル束の零切断である $\mathbb{R}^n$ は Lagrange 部分多様体である。$T^*\mathbb{R}^n$ を相空間と考えれば、零切断は配位空間、即ち位置座標を表す空間であり、これは世界である。次に、函数の一般化としての Lagrange 部分多様体について述べておく。函数概念を一般化する方法は、例えば $\mathbb{R}$-値ではなく $\mathbb{R}^m$-値としても一般化となるし、$\mathbb{R}$-値を直積束 $M \times \mathbb{R}$ の切断と見なせば、M 上の直線束の切断を函数の一般化と考える事も出来る[2]。Lagrange 部分多様体の場合、函数の定数項を無視する事で、函数 f の外微分 df を函数 f と実質同じものと見なす。すると、df は M から余接束 T^*M への写像、即ち T^*M の切断を定める。一般に 1-形式 $\alpha \in \Lambda^1$ を $\alpha : M \to T^*M$ と見なすと、像 $\alpha(M)$ が Lagrange 部分多様体である事と $d\alpha = 0$ である事は同値である。完全形式 df は閉形式だから、函数（の外微分）は Lagrange 部分多様体を定める。Lagrange 部分多様体の説明の最後として、シンプレクティック微分同相写像との関係を説明しておく。$\varphi : M \to M$ を微分同相写像とする。このとき φ がシンプレクティック構造を保つ為の必要充分条件は、$M \times M$ 内のグラフ $\{(x, \varphi(x)); x \in M\}$ が $(M \times M, \omega')$ の Lagrange 部分多様体となる事である。但し $\omega' = \mathrm{pr}_1^*\omega - \mathrm{pr}_2^*\omega$. このように、Lagrange 部分多様体は空間、函数、シンプレクティック微分同相の一般化であり、これらは『Everything』と云われる一端である。Lagrange 部分多様体の活躍場面はまだ多々存在するが、本稿では必要ないので略する（例えば Arnold 予想に関わる内容、特性類や Maslov 指数との関係など。これらを説明するのは著者の脳力を超える）。

[2]このような一般化については沢井氏（沼津工専）の方が詳しいので、こちらの話は彼に任せる。

次に Bayes の定理について簡単に復習をする。Bayes の定理は

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \qquad P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

である。前者が確率版、後者が密度函数版（のつもり）である。ここで A と B は事象、θ はパラメータ、y は標本を表す。この定理を簡単に述べると『事前分布と尤度から事後分布を導くもの』であり、対象についての事前情報にデータを合わせて再考し直す方法を与えている。推定したいパラメータを θ としたとき、古典的統計学では真値 θ_0 は確定しており（しかしそれを我々は知らない）、それをデータから、例えば信頼区間という形で推定する。信頼区間の各値の重みは一様であり、信頼区間に入るか否かで帰無仮説を容認するか棄却するかを決める方法が検定である。一方 Bayes 統計では θ_0 の値が定まっているかどうかは別にして、ピンポイントでの推定を諦める。その上で、θ を確率変数と考え、従う分布を推定する事を考える。データ収集前の事前分布にデータを考慮して事後分布を更新していく場合に Bayes 定理を使用する。

分布を更新するにあたり、事前分布と事後分布が同じ分布型であるものを共役分布と呼ぶ。本稿では共役分布での更新のみ考える。Bayes 更新は（古典論での推定でもそうであるが）、状況によって使用される共役分布族が変わる。例えば、データが正規分布 $N(\mu, \sigma^2)$ に従う場合でも母分散既知、母分散未知、母平均既知と場合分けされる。

本稿では事前分布から事後分布への Bayes 更新をシンプレクティック幾何学の言葉で表現する事を目的とする。具体的には正準変換を用いて表したい。このとき、考えられる方針が 4 通りある。θ の従う分布を考えるとして、パラメータ θ の空間に付随するシンプレクティック構造に関する正準変換を用いるか、共役分布族のパラメータ空間に付随するシンプレクティック構造に関するものを用いるかで 2 通り。付随するシンプレクティック構造としてパラメータ空間上の構造を用いるか、余接束上の構造を用いるかで 2 通り。合わせて（積だけど）4 通りである。本稿では θ の空間の余接束上のシンプレクティック構造に関する正準変換を対象とする。これを選択する理由を簡潔に述べておくと、本稿の目的は Bayes の定理を正準変換で表す事であるが、Bayes の定理だけでは不充分であり、Bayes 統計まで視野に入れるには事後分布を求めた後に事後分布による平均などの点推定を行う必要がある。そこで、密度函数が直接登場する方法を取る事にしている。

序の最後として、類似研究について言及しておく。シンプレクティック幾何学による Bayes 定理の記述として Mori [12], [13] がある。Mori は 1 変量正規分布における母平均に対する Bayes 更新として、共役分布族上のシンプレクティック構造を用いている。より詳しくは、

共役分布族である上半平面上に Poincaré 型のシンプレクティック構造を取り、上半平面の直積内の Lagrange 部分多様体上の flow を用いて Bayes 更新を記述しており、本稿のものとは異なる方法を用いている。

2. 確率分布の空間の微分幾何

本節では情報幾何学の基本的事項を復習する。詳しくは甘利・長岡 [2], [3] を参照下さい。

$(X, \mathfrak{B}, dx)$ を測度空間とする（とは云っても $\mathbb{R}^m$ と Borel 集合族、Lebesgue 測度で充分)。X 上の正値確率密度全体の空間を

$$\mathcal{P}(X) = \{p :\to \mathbb{R}\ ;\ p(x) > 0,\ \int_X p(x)\ dx = 1\}$$

とする。これは無限次元であり、多様体構造は Cena-Pistone [6] などで知られているが、やはり扱い難い。そこで、$\mathcal{P}(X)$ の有限次元の部分多様体を考える。

定義 2.1. $\Xi \subset \mathbb{R}^n$ を開集合とし、Ξ でパラメータ付けられた確率密度函数の集合

$$S = \{p_\xi = p(x;\xi) \in \mathcal{P}(X)\ ;\ \xi = (\xi^1, \ldots, \xi^n) \in \Xi\}$$

を n 次元**統計モデル**と呼ぶ。

簡単の為、$\xi \mapsto p_\xi$ は 1 対 1 C^∞ 級であり、ξ での微分と X 上の積分は可換などの正則性は仮定する。また、次に定義する指数型の場合は ξ の代わりに θ を用いる事も多い。

定義 2.2. $\Theta \subset \mathbb{R}^n$ を開集合とし、確率密度函数が

$$p_\theta = \exp\left[C(x) + \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)\right]$$

で与えられる統計モデル $S = \{p_\theta; \theta \in \Theta\}$ を**指数型分布族**と呼ぶ。但し $C(x)$, $F_i(x)$ は X 上の函数であり、$\psi(\theta)$ は θ の函数（で正規化要員[3]）。

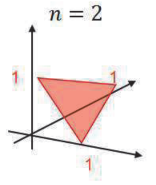
指数型分布族とはパラメータ θ が対数を取ると線型に入っている分布族である。θ を**自然パラメータ**（または**正準パラメータ**）と呼ぶ。

例 2.3 (有限集合上の確率分布)**.** $X = \{x_0, x_1, \ldots, x_n\}$ とし、X 上の確率分布全体を

$$\mathcal{P}(X) = \{(p_0, p_1, \ldots, p_n)\ ;\ p_0 + p_1 + \cdots + p_n = 1,\ p_i > 0\}$$

と置く。$n = 2$ の場合、$\mathcal{P}(X)$ は次図のようになる。

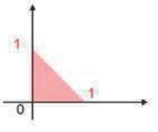
[3] しかしながら、正規化要員は例えば統計力学の分布函数を考えると非常に重要である事が判る。

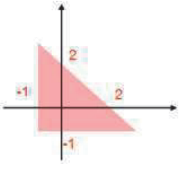


$\mathcal{P}(X)$ に多様体としての（局所）座標系 (ξ^1,ξ^2) を入れる事を考えると、例えば

- $\xi^i = p_i \ (1 \le i \le n)$,
- $\xi^i = 3p_i - 1 \ (1 \le i \le n)$

など幾らでも入れる事が出来る：





この場合、

- $C(x)=0,\ F_i(x)=\begin{cases}1 \ (x=x_i) \\ 0 \ (x \ne x_i)\end{cases}$, $\theta^i = \log \dfrac{p_i}{p_0},\ \psi(\theta) = \log p_0$

とすると指数型となる。

例 2.4 (正規分布族)**.** $X=\mathbb{R}$ 上の正規分布

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

の場合、

- $(\xi^1,\xi^2)=(\mu,\sigma^2)$ ：平均と分散,
- $(\xi^1,\xi^2)=(\mu,\sigma)$ ：平均と標準偏差,
- $(\xi^1,\xi^2)=(\mu,\mu^2+\sigma^2)$ ：モーメント

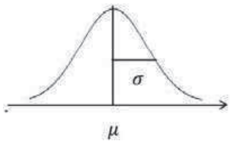
など色々座標系を入れられるが、

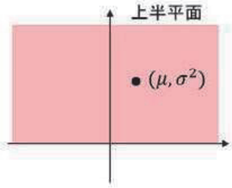
- $C(x)=0,\ F_1(x)=x,\ F_2(x)=x^2,\ \theta^1=\dfrac{\mu}{\sigma^2},\ \theta^2=-\dfrac{1}{2\sigma^2}$

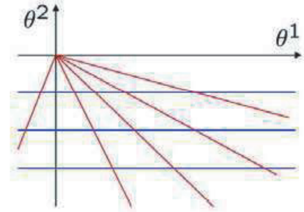
とすると指数型である。このとき

$$\psi(\theta) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^2}\right)$$

となる。







統計モデル S は集合としては $\mathbb{R}^n$ の開集合であるが、これが分布のパラメータ空間である事を考慮する事で幾何構造を入れる事が出来る。特に計量と接続が自然に定義される。次にそれを述べる。

n 次元統計モデル $S = \{p_\xi \in \mathcal{P}(X); \xi = (\xi^1, \ldots, \xi^n) \in \Xi\}$ に対し対数尤度を $l_\xi = \log p_\xi$ で表す。また座標函数 ξ^i $(1 \leq i \leq n)$ による偏微分を ∂_i で表す。

定義 2.5. $S = \{p_\xi; \xi \in \Xi\}$ に対し Fisher 情報行列 $g = [g_{ij}]$ を

$$g_{ij} = E_\xi[\partial_i l_\xi \ \partial_j l_\xi] = \int_X (\partial_i l_\xi)(\partial_j l_\xi) p_\xi dx = -E_\xi[\partial_i \partial_j l_\xi]$$

で定める。g が正定値であるとき **Fisher 計量**と呼ぶ。

本稿では統計モデル S に付随する g は非退化なもののみ考える。従って g は Fisher 計量である。Fisher 情報行列は不偏推定量の平均 2 乗誤差に関する Cramér-Rao の不等式において下限に登場するが、これを S 上の Riemann 計量と考えるべきである事を提唱したのは Rao であると記憶している。

次に接続を定めたい。S 上には Fisher 計量が定まっているので、微分幾何としては Levi-Civita 接続 ∇^0 が付随するのであるが、これがあまり統計学的に有難くない。推定を何かしらの意味でデータに最も近いパラメータの値を求める操作であると考えると、幾何学的には距離や接続が対応するのであるが、例えば指数型分布族の最尤推定は Fisher 計量の Levi-Civita 接続とは対応しない。そこで、各種の統計量（推定量）を含むように接続を定める必要がある。ここでは次のように定める。

定義 2.6. g を統計モデル S 上の Fisher 計量とし、その Levi-Civita 接続を ∇^0, その Christoffel 記号を $\Gamma^{(0)}_{ij,k}$ とする。S 上の $\alpha \in \mathbb{R}$ でパラメータ付けられた接続 $\nabla^{(\alpha)}$ の族を、

$$\Gamma^{(\alpha)}_{ij,k} := \Gamma^{(0)}_{ij,k} - \frac{\alpha}{2} T_{ijk}$$

を Christoffel 記号とするようなものとして定義する。ここで T_{ijk} は

$$T_{ijk} = E_\xi[\partial_i l_\xi \ \partial_j l_\xi \ \partial_k l_\xi]$$

で定められる対称 3-テンソル場。$\nabla^{(\alpha)}$ を $\boldsymbol{\alpha}$**-接続**と呼ぶ。

α 接続 $\nabla^{(\alpha)}$ の性質などを簡単に述べておく。先ず $\alpha = 0$, 即ち $\nabla^{(0)}$ は計量的 : $\nabla^{(0)} g = 0$ である（が、これは Levi-Civita 接続であるから明らか）。任意の $\alpha \in \mathbb{R}$ に対し $\nabla^{(\alpha)}$ は捩れ無し : $T^{\nabla^{(\alpha)}} = 0$ である。また $\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は双対（接続）である[4]。指数型分布族の場合、$\alpha = \pm 1$ が特に重要であり、実際、指数型分布族は $\nabla^{(\pm 1)}$ 平坦である（$\nabla^{(1)}$ を e-接続、$\nabla^{(-1)}$ を m-接続と呼ぶ）。更に自然パラメータ θ は $\nabla^{(1)}$ アフィン座標系である。$\nabla^{(-1)}$ アフィン座標系は $\eta_i = E_\theta[F_i]$ で与えられ、これを期待値座標系と呼ぶ。

例 2.7 (正規分布族の場合)**.** $X = \mathbb{R}$ 上の正規分布 $N(\mu, \sigma^2)$ の族において、Fisher 計量を例 2.4 で定めた自然パラメータ (θ^1, θ^2) で表すと

$$g_{11} = \frac{1}{2\theta^2}, \quad g_{12} = g_{21} = \frac{\theta^1}{2(\theta^1)^2}, \quad g_{12} = \frac{\theta^2 - (\theta^1)^2}{2(\theta^2)^2}$$

となる。この場合の期待値座標は $\eta_1 = \mu$, $\eta_2 = \mu^2 + \sigma^2$ となり、この座標系での Fisher 計量を $g = [g^{ij}]$ とすると

$$g^{11} = \frac{\eta_2 + \eta_1^2}{(\eta_2 - \eta_1^2)^2}, \quad g^{12} = g^{21} = -\frac{\eta_1}{(\eta_2 - \eta_1^2)^2}, \quad g^{22} = \frac{1}{2(\eta_2 - \eta_1^2)^2}$$

となる。因みに、座標系として平均と分散を用いた場合、即ち $(\xi^1, \xi^2) = (\mu, \sigma^2)$ とした場合

$$g = \begin{bmatrix} \sigma^{-1} & 0 \\ 0 & (2\sigma^4)^{-1} \end{bmatrix}$$

となる。

ここまでで統計モデル $S = \{p_\xi; \xi \in \Xi\}$ には Fisher 計量 g と α 接続 $\nabla^{(\alpha)}$ が付随する事が判った。そこで、これを抽象化して次のように定める。

定義 2.8. 多様体 S, Riemann 計量 g, アフィン接続 ∇ で

(i) ∇ は捩れ無し : $T^\nabla = 0$;

(ii) ∇ は Codazzi 方程式を満たす : $(\nabla_X g)(Y, Z) = (\nabla_Y g)(X, Z)$

を満たすものの 3 つ組 (S, g, ∇) を**統計多様体**、(g, ∇) を S 上の**統計構造**（または双対構造）と呼ぶ。

条件 (i), (ii) は $T^\nabla = 0 = T^{\nabla^*}$ と同値である事に注意しておく。統計モデルと Fisher 計量、α 接続の 3 つ組はこの定義を満たすから、統計モデルは統計多様体である（逆については Le [11] を見よ）。

先に述べたように、指数型分布族は $\nabla^{(\pm 1)}$ 平坦である。そこで、統計多様体の中で特に平坦なものに名前を付けておく。

---

[4]接続 ∇ に対し $Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y)$ を満たす ∇^* は一意的に存在し、∇^* を ∇ の双対接続と呼ぶ。また ∇ と ∇^* は双対であると云う。

定義 2.9. 統計多様体 (S, g, ∇) に対し、∇ が平坦、即ち曲率が $R^{\nabla} = 0$ を満たすとすると、双対接続 ∇^* についても $R^{\nabla^*} = 0$ が成立する。そこで、$R^{\nabla} = 0$（従って $R^{\nabla^*} = 0$）を満たす統計多様体を**双対平坦空間**と呼ぶ。これはまた Hesse 多様体とも呼ばれる。

双対平坦空間 (S, g, ∇) の（局所的）性質を纏めておく。

命題 2.10. (S^n, g, ∇) を双対平坦空間、∇^* を双対接続とする。

(1) ∇ アフィン座標系 $(\theta^1, \ldots, \theta^n)$ 及び ∇^* アフィン座標系 $(\eta_1, \ldots, \eta_n)$ で $g(\partial_i, \partial^j) = \delta_i^j$ を満たすものが存在する。但し $\partial_i = \partial/\partial\theta^i$, $\partial^j = \partial/\partial\eta_j$.

(2) S 上の函数 $\psi, \varphi : S \to \mathbb{R}$ で $d\psi = \eta_i d\theta^i$, $d\varphi = \theta^i d\eta_i$ を満たすものが存在する。このとき $\psi + \varphi = \theta^i \eta_i$ が成立。

(3) $\theta^i = \dfrac{\partial\varphi}{\partial\eta_i}$, $\eta_j = \dfrac{\partial\psi}{\partial\theta^j}$, $d\theta^i = g^{ij} d\eta_j$, $d\eta_j = g_{ij} d\theta^i$ が成立。また

$$g_{ij} = g(\partial_i, \partial_j) = \frac{\partial\eta_j}{\partial\theta^i} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}, \qquad g^{ij} = g(\partial^i, \partial^j) = \frac{\partial\theta^i}{\partial\eta_j} = \frac{\partial^2\varphi}{\partial\eta_i\partial\eta_j}$$

が成立。

次に分布の距離を測る函数について説明する。先ず、次のように定義する。

定義 2.11. 双対平坦空間 (S, g, ∇) に対し、命題 2.10 (2) の ψ, φ を用いて

$$D(p, q) = D(p\|q) = \varphi(p) + \psi(q) - \eta_i(p)\theta^i(q)$$

で定められる函数 $D : S \times S \to \mathbb{R}$ を（∇^* に関する）**canonical divergence** と呼ぶ。

この $D(p, q)$ は p, q に関して対称ではないが、距離の 2 乗のようなものである。実際、$^{\forall}p, q \in S$ に対し $D(p, q) \geq 0$ が成立し、等号は $p = q$ のときに限る。

例 2.12. 指数型分布族 $S = \{p_\theta; \theta \in \Theta\}$ に対し canonical divergence は

$$D(p, q) = \int_X p \log \frac{p}{q} dx$$

となり、これは相対エントロピーである。

いま、平坦統計構造 (g, ∇) から canonical divergence を構成したが、逆に D から統計構造を構成出来る。多様体 S^n の直積上の函数 $D : S \times S \to \mathbb{R}$ で $D(p\|q) \geq 0$ であり等号は $p = q$ に限るものを考える。この D から S^n 上の統計構造を構成する為、記号を準備する。

$S \times S$ の点を（局所）座標で $(\xi^1 \ldots, \xi^n, \xi'^1, \ldots, \xi'^n)$ と表す。また、$D(\partial_i\|\cdot) := \partial_i D(\cdot\|\cdot)$, $D(\cdot\|\partial'_j) := \partial'_j D(\cdot\|\cdot)$ と置く。更に、D に関する種々の微分などの対角成分 $\Delta \subset S \times S$ へ

の制限を $D[\cdot\|\cdot] := D(\cdot\|\cdot)|_\Delta$ で表す。このとき

$$D[\partial_i\|\cdot] = D[\cdot\|\partial'_j] = 0,$$
$$D[\partial_i\partial_j\|\cdot] = D[\cdot\|\partial'_i\partial'_j] = -D[\partial_i\|\partial'_j]$$

が成立する。いま $g^D = [g^D_{ij}] = [D[\partial_i\partial_j\|\cdot]]$ と置くと、これは半正定値である事が判る[5]。以下 g^D が正定値となる D のみ考える。このような D は divergence, contrast 函数、yoke などと呼ばれるが、ここでは divergence を使用する。D を divergence, X と Y を S 上のベクトル場としたとき、

$$g^D(X,Y) = -D[X\|Y]$$

により S 上の Riemann 計量が定まる。

次に $\Gamma^{(D)}_{ij,k} := -D[\partial_i\partial_j\|\partial'_k]$ を Christoffel 記号とする接続を ∇^D とすると $T^{\nabla^D} = 0$ かつ $g^D(\nabla^D_X Y, Z) = -D[XY\|Z]$ が成立する。更に $D^*(p\|q) = D(q\|p)$ とすると $g^{D^*} = g^D$ であり ∇^{D^*} は ∇^D の双対接続となる。これで divergence D から統計構造 (g^D, ∇^D) 及び双対接続 ∇^{D^*} が定まった。因みに、$(1,2)$-テンソル場 T を $D[T_X Y\|Z] = D[XY\|Z] - D[YZ\|X]$ で定めると $\nabla^{(\alpha)} := \nabla^{(0)} - \frac{\alpha}{2}T$ により α 接続も定める事が出来る。

3. 統計モデルとシンプレクティック構造

本節では統計構造とシンプレクティック構造について簡潔に述べる。詳しくは Furuhata [9], Noda [14], Boumuki-Noda [5] などを参照下さい。

前節で統計構造 (g,∇) と divergence D との関係、特に双対平坦空間 (S,g,∇) と canonical divergence D を考えた。先ず D からシンプレクティック構造を定め、その例などを述べる。

(S,g,∇) を統計多様体（双対平坦空間）、D を（canonical）divergence とする。$D : S\times S \to \mathbb{R}$ を第一変数に関して外微分する事で $d_1D : S\times S \to T^*S$, $d_1D = \dfrac{\partial D}{\partial \xi^i} d\xi^i$ を得る。但し $(\xi^1,\dots,\xi^n,\xi^{*1},\dots,\xi^{*n}) \in S\times S$ とし、$S\times S$ の第二成分の対象については全て $*$ を付す（前節では $\prime$ を使用していた）。この写像 d_1D によって T^*S 上の正準シンプレクティック構造 $-d\theta_0$ を引き戻す事で

$$\omega(\xi,\xi') = (d_1D)^*(-d\theta_0) = \frac{\partial D}{\partial\xi^i\partial\xi^{*j}} d\xi^i \wedge d\xi^{*j}$$

を得る。これは $S\times S$ 上のシンプレクティック構造を定める。即ち、divergence D からシンプレクティック構造 ω が定まった。最も簡単な例を挙げよう。$S = \mathbb{R}^n$ とし、$D : S\times D \to \mathbb{R}$ を $D(\xi,\xi') = \dfrac{1}{2}\|x - \xi^*\|^2$ とする。このとき得られる ω は同一視 $\mathbb{R}^n\times\mathbb{R}^n \cong T^*\mathbb{R}^n$ の下で正準構造 ω_0 となる。

双対平坦空間の canonical divergence から導かれるものは特に重要である。

---

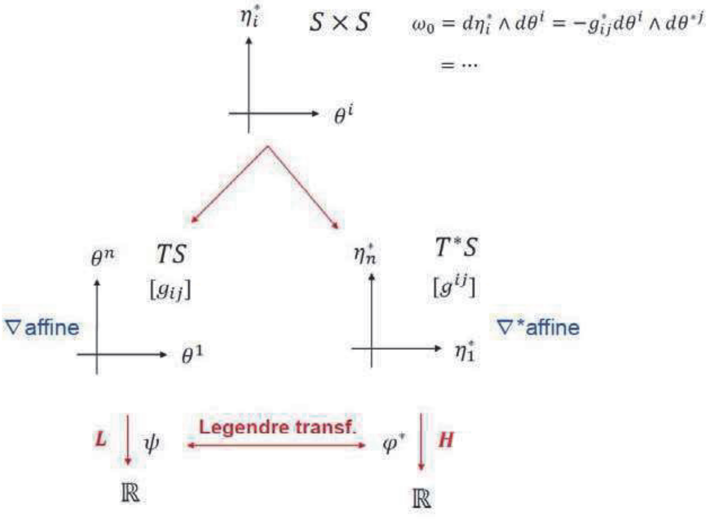
[5] 厳密には $S\times S$ 全体でとは限らず、Δ の開近傍上でのみである。

例 3.1. 双対平坦空間 (S, g, ∇) に対する（∇^* の）canonical divergence $D : S \times S \to \mathbb{R}$ の場合、

$$\omega = -g^*_{ij} d\theta^i \wedge d\theta^{*j} = d\eta^*_i \wedge d\theta^i = -g^*_{ij} g^{ik} d\eta_k \wedge d\theta^{*j} = -g^{ij} d\eta_i \wedge d\eta^*_j \tag{3.1}$$

となる。但し $\{\theta\}$ は ∇ アフィン座標系、$\{\eta\}$ は ∇^* アフィン座標系であり、g_{ij} などは Fisher 計量の対応する成分。

(3.1) において 2 つ目の表示が見易い。これは θ 座標系を接空間の座標、η 座標系を双対空間の座標と見なすと T^*S の正準構造と一致しており、ポテンシャル函数 ψ と φ が Legendre 変換の関係となっている事と整合する。



次に一般の多様体上で統計構造とシンプレクティック構造が両立する場合を簡単に述べておく。M^{2n} を多様体、g を Riemann 計量、J を概複素構造、ω をシンプレクティック構造、∇ をアフィン接続とする。(S, g, J, ω) が概 Kähler 多様体であり (M, g, ∇) が双対平坦空間、∇ がシンプレクティック接続（i.e., ω を保つ）とき、5 つ組 $(M, g, J, \omega, \nabla)$ を平坦 SS 多様体と呼びたい（Furuhata [9] では holomorphic statistical と呼んでいる）。このとき

$$\nabla^* = \nabla - J(\nabla J), \qquad (\nabla_X J)Y = (\nabla_Y J)X \tag{3.2}$$

が成立する（シンプレクティック構造と統計構造が両立する為の条件が大体 (3.2) である）。平坦 SS 多様体は局所的には双対平坦空間の余接束と見なす事が出来る。即ち、或る双対平坦空間 (S, g_S, ∇^S) が存在し $M \cong T^*S$ が成立。更に ∇ アフィン座標系 $\{\theta\}$, ∇^* アフィン座標系 $\{\eta\}$ で共に Darboux 座標系となっているものが存在する：

$$\omega = d\theta^1 \wedge d\theta^{n+1} + \cdots + d\theta^n \wedge d\theta^{2n} = d\eta_1 \wedge d\eta_{n+1} + \cdots + d\eta_n \wedge d\eta_{2n}.$$

但し、このとき J は $J\partial_i = \partial^{n+i}$ の形となる。

本節の最後に偶数次限の双対平坦空間の別の Darboux 型の標準形について述べておく。(M^{2n}, g, ∇) を偶数次限双対平坦空間とすると、或る双対平坦空間 (S^n, g_S, ∇^S) で局所的に $(M, \omega_1) \cong (T^*S, \omega_0)$ となるようなものが存在する。ここで

$$\omega_1 = \sum_{k=1}^{n} \frac{d\eta_{2k-1} \wedge d\eta_{2k}}{\eta_{2k-1}^2}$$

は M 上の Poincaré 型の（局所的）シンプレクティック構造。

4. 正準変換による Bayes 更新

序で述べたように、Bayes 更新を正準変換で表現するにはパラメータ空間上または共役分布族のパラメータ空間上の構造かで 2 通り、空間自体のシンプレクティック構造か余接束上のそれかで 2 通りの計 4 通りの可能性があった。ここでは推定したいパラメータの空間の余接束上の正準変換により確率密度函数を Hamilton 流で流す事により Bayes 更新を記述する事を考える。特に正規分布に従う変量の母平均の推定に話を限る。ここで用いる Bayes 更新は次のものである。詳しくは Bayes 統計の教科書、例えば照井 [15], 渡辺 [16] などを参照下さい。

事実 4.1 (多変量正規分布での母平均の分布の更新)**.** 標本の母集団は $N(\mu, \Sigma)$ の正規分布であるとする。

(1) 分散 Σ が既知の場合。μ の事前分布を $N(\mu_0, \Lambda_0)$ とする：

$$p(\mu) \propto \exp\left[-\frac{1}{2}(\mu - \mu_0)^T \Lambda_0^{-1}(\mu - \mu_0)\right].$$

標本 y が得られたときの事後分布は $\mu|y, \Sigma \sim N(\mu_n, \Lambda_n)$,

$$p(\mu|y, \Sigma) \propto \exp\left[-\frac{1}{2}(\mu - \mu_n)^T \Lambda_n^{-1}(\mu - \mu_n)\right].$$

但し

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + \Sigma^{-1}\bar{y}), \quad \Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}.$$

(2) 分散 Σ が未知の場合。μ の事前分布を $\mu|\Sigma \sim N(\mu_0, \Sigma/k_0)$, $\Sigma \sim IW(\nu_0, \Lambda_0)$ とすると事後分布は $\mu|y, \Sigma \sim N(\mu_n, \Sigma/k_n)$, $\Sigma|y \sim IW(\nu_n, \Lambda_n)$. 但し

$$\nu_n = \frac{k_0}{k_0 + n}\mu_0 + \frac{n}{k_0 + n}\bar{y}, \quad k_n = k_0 + n, \quad \nu_n = \nu_0 + n, \quad \Lambda_n = \Lambda_n + S + \frac{k_0 n}{k_0 + n}S_0.$$

正規分布に従う変量の母平均 $\mu \in \mathbb{R}^n$ に対する確率密度函数を ρ で表すと、ρ は多変量正規分布に従う。この場合、アフィン正準変換、即ち平行移動と線型正準変換により Bayes 更新を表す事が出来る。そこで先ず線型正準変換全体の群 $\mathrm{Sp}(2n, \mathbb{R})$ について復習しておく。

定義 4.2. $\mathbb{R}^{2n}$ 上の線型正準変換全体

$$\mathrm{Sp}(2n,\mathbb{R}) = \{S \in \mathrm{GL}(2n,\mathbb{R})\ ;\ S^T J S = J\}$$

を**シンプレクティック群**と呼ぶ。但し $J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$.

$\mathbb{R}^{2n}$ の点を z で表し、ω_0 を線型シンプレクティック空間 $\mathbb{R}^{2n}$ 上の正準シンプレクティック構造としたとき、

$$\omega_0(z,z') = \langle Jz, z' \rangle$$
$$\omega_0(Sz,Sz') = \langle JSz, Sz' \rangle = \langle S^T JSz, z' \rangle$$

の最右辺同士が一致する為の条件が $S^T JS = J$ である。任意の $S \in \mathrm{Sp}(2n,\mathbb{R})$ に対し $\det S = 1$ であり、$n = 1$ の場合は $\mathrm{Sp}(2n,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})$ が成立する（が、$n > 1$ では等号は成立しない）。一般に $\mathrm{Sp}(2n,\mathbb{R})$ は $n(2n+1)$ 次元の連結 Lie 群であり、Lie 環は $\mathfrak{sp}(n,\mathbb{R}) = \{M \in M(2n,\mathbb{R}); JM + M^T J = 0\}$ で与えられる。また、n 次ブロック行列を用いて $S = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{Sp}(2n,\mathbb{R})$ とすると

$$A^T C = C^T A, \quad B^T D = D^T B, \quad A^T D - C^T B = I_n$$

が成立する（逆も真)。これより S の逆行列が $S^{-1} = \begin{bmatrix} D^T & -B^T \\ -C^T & A^T \end{bmatrix}$ となる事が判る。これらについては Abraham-Marsden [1], de Gosson [10] などを参照下さい。

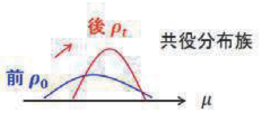
さて、$\rho_0 \sim N(0,\Sigma)$ である密度函数を Hamilton 流で流したい。正規分布に従う密度を線型 Hamilton 流で流すと正規分布である事が保たれる。実際、次が成立する。

補題 4.3. $\rho_0 \sim N(0,\Sigma)$ である密度函数を状態遷移行列 $\Phi_t \in \mathrm{Sp}(2n,\mathbb{R})$ の線型 Hamilton 系で発展させると $\rho_t \sim N(0,\Sigma(t))$ となる。ここで $\Sigma(t) = \Phi_t \Sigma \Phi_t^T$.

ここで状態遷移行列とは Hamilton 方程式

$$\begin{cases} \dot{q}_j = \dfrac{\partial H}{\partial p_j} \\ \dot{p}_j = -\dfrac{\partial H}{\partial q_j} \end{cases}$$

に対し $\Phi_t = \exp\left(t \begin{bmatrix} 0 & \frac{\partial H}{\partial p_j} \\ -\frac{\partial H}{\partial q_j} & 0 \end{bmatrix} \right)$ で与えられる行列であり、任意の t に対し $\Phi_t \in \mathrm{Sp}(2n,\mathbb{R})$ が成立する。



補題 4.3 は線型 Hamilton 流で正規分布を流しても正規分布のままであり、分散が状態遷移行列とその転置で挟んだものになる事を述べている[6]。証明自体は単純計算であり、$\Phi_{-t} = \Phi_t^{-1}$ と $\det \Phi_t = 1$ を用いる事で

$$\begin{aligned}\rho_t = \rho_0(\Phi_{-t}x) &= \left(\frac{1}{2\pi}\right)^n (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2}x^T \Phi_{-t}^T \Sigma^{-1} \Phi_{-t} x\right] \\ &= \left(\frac{1}{2\pi}\right)^n (\det(\Phi_t \Sigma \Phi_t^T))^{-1/2} \exp\left[-\frac{1}{2}x^T (\Phi_t \Sigma \Phi_t^T)^{-1} x\right]\end{aligned}$$

となる。

必要な準備が全て終了したので、本来の目的である Bayes 定理を正準変換で表す事について述べる。本稿では 母平均 μ の空間上のシンプレクティック構造ではなく、その余接束上の正準構造を使用する。その理由の一つを注意しておく。$\mu \in \mathbb{R}^{2n}$ とし、$\mathbb{R}^{2n}$ の（アフィン）正準変換で Bayes の定理が表現出来るためには、分散既知の場合なら $\Phi^T \Lambda_0^{-1} \Phi = \Lambda_0^{-1} + n\Sigma^{-1}$ を満たす $\mathrm{Sp}(2n, \mathbb{R})$ が見付かればよい（Φ, Σ, Λ は全て $2n$ 次である事に注意）。しかしながら、これは望み薄である。Bayse の定理では確率密度の切り口として函数型を決定し、その後に正規化する：

$$p(\theta|y) = \frac{l(\theta|y)p(\theta)}{p(y)} \propto l(\theta|y)p(\theta).$$

しかしながら、正準変換は体積保存であった。また、この場合では母平均ベクトル μ は $\mathbb{R}^{2n}$ の要素であり、偶数次元の場合しか考慮出来ない事になる。これらの弱点を補う方法が **Lagrange 部分多様体**の使用となる。μ の空間を適当な Lagrange 部分多様体とするようなシンプレクティック多様体を考え、その全空間での正準発展を構成する。このとき、μ の正規化のツケは Lagrange 部分多様体の補空間に押しつける。これが正準変換により Bayes 定理が記述出来るカラクリである。

では、具体的な記述方法を述べよう。先ずは分散 Σ が既知の場合から。$\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ 上の正準シンプレクティック構造を考える。いま、μ を第一成分（これが上記太字の Lagrange 部分多様体）の要素と考え、事前分布の分散に対応する行列として $\begin{bmatrix} \Lambda_0^{-1} & 0 \\ 0 & I_n \end{bmatrix}$ を取る。いま、

$$\Phi_1 = S^T = \begin{bmatrix} I_n & -(\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(n\Sigma^{-1})^{1/2} \\ (n\Sigma^{-1})^{1/2} & I_n - (n\Sigma^{-1})^{1/2}(\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(n\Sigma^{-1})^{1/2} \end{bmatrix} \in \mathrm{Sp}(2n, \mathbb{R})$$

[6] しかしながら、線型 Hamilton 系での発展で到達出来る正規分布には分散 $\Sigma(t)$ に制約が付く。具体的には $\det[\Sigma_{j,n+j}(t)] \geq (c(\Sigma)/\pi)^2$ が成立する。ここで $c(\Sigma)$ は Σ^{-1} の定める楕円体のシンプレクティック容量。

と置くと

$$S\begin{bmatrix}\Lambda_0^{-1} & 0\\ 0 & I_n\end{bmatrix}S^T=\begin{bmatrix}\Lambda_n^{-1} & 0\\ 0 & I_n-(n\Sigma^{-1})^{1/2}(\Lambda_0^{-1}+n\Sigma^{-1})^{-1}(n\Sigma^{-1})^{1/2}\end{bmatrix}$$

となる。従って、$\mathbb{R}^{2n}$ 内のベクトル (a,b) による平行移動を $T(a,b)$ で表すと

$$T(\mu_n,0)\circ\tilde{\Phi}_1\circ T(-\mu_0,0)$$

により Bayes 更新をアフィン正準変換で表す事が出来る。

次に分散未知の場合。この場合、事前分布に対応する行列として $\begin{bmatrix}k_0\Sigma^{-1} & 0\\ 0 & I_n\end{bmatrix}$ と取り、

$$\Phi_1=S^T=\begin{bmatrix}I_n & n^{1/2}\Sigma^{-1/2}\\ -(2k_0)^{-1}n^{1/2}\Sigma^{1/2} & (1-\frac{n}{2k_0})\Sigma\end{bmatrix}\in\mathrm{Sp}(2n,\mathbb{R})$$

と置き、

$$T(\mu_n,0)\circ\tilde{\Phi}_1\circ T(-\mu_0,0)$$

とすれば良い。

5. 補遺

ここでは講演では時間の関係などから充分に述べられなかった事柄を簡潔に纏めておくが、妄想に近い部分もある事を先に注意しておく。

前節で述べたように、多変量正規分布の母平均の場合については Bayes 更新を正準変換で表す事が出来た。分散未知の場合は分散の分布も推定する必要があるが、現時点ではこれには成功していない。簡単の為 1 変量の場合に制限すると、σ^2 の共役分布族は逆ガンマ分布である。この場合、分散が正値であるから、正準シンプレクティック構造ではなく Poincaré 型のシンプレクティック構造を取るべきであると考えられる。森 [13] では 1 変量の場合に分散未知の場合の Bayes 更新をシンプレクティック幾何学的に考察し、変動係数の逆数が（接触）Hamilton 函数であると述べているが、本稿でもそれに近い形で出来るのではないかと思われる。

Bayes 定理の正準変換での記述は Duistermaat-Heckman 定理 [7], [8] を統計モデルに付随するシンプレクティック構造に使用したら何かしら得られるのではないかとの思考による。Duistermaat-Heckman 定理とは調和積分の stationary phase approximation または Bott の留数定理の Hamilton 作用版という内容であるが、これを使用するには統計モデル（またはその余接束）上の積分に適用する事になり、統計モデル（またはその余接束）上の分布を考えるのは Bayes であるとの理由による。事後分布による点推定として例えば事後分布による平均を取る事を考えると、これには統計モデル上での積分を計算する事になるが、これと Duistermaat-Heckman 定理が何かしらの関係にあるのではないかと考えている（Mori

[12] では Lagrange 部分多様体（対応）を Fourier-like transformation と述べているが、それとも何かしらで整合するものと思われる）。しかしながら、現時点では（特定の場合の）Bayes 定理が正準変換で書けるのみであり、残念ながらこの定理の適用には至っていない。

References

[1] R. Abraham and J.E. Marsden, Foundations of Mechanics, Second Edition, Addison-Wesley Publishing Company, Inc., 1987.

[2] 甘利俊一, 長岡浩司, 情報幾何の方法, 岩波講座応用数学, 東京:岩波書店, 1993

[3] S. Amari and H. Nagaoka, Methods of information geometry, Translated from the 1993 Japanese original by Daishi Harada. Translations of Mathematical Monographs, 191. American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000. x+206 pp. ISBN 0-8218-0531-2

[4] O.E. Barndorff-Nielsen and P.E. Jupp, Statistics, yokes and symplectic geometry, Annales de la faculté des sciences de Toulouse Sèr. 6, 6 no. 3 (1997), p. 389–427.

[5] N. Boumuki and T. Noda, On gradient and Hamiltonian flows on even dimensional dually flat spaces, Fundamental J. Math. and Math. Sci. **6** (2016), 51–66.

[6] A. Cena and G. Pistone, Exponential statistical manifold, Annals of the Institute of Statistical Mathematics, **59** (1), 27–56 (2007).

[7] J. Duistermaat and G. Heckman, On the variation in cohomology of the symplectic forms of the reduced phase space, Invent. Math. **69**(1982), 259–268.

[8] J. Duistermaat and G. Heckman, Addendum to “On the variation in cohomology of the symplectic forms of the reduced phase space”, Invent. Math. **72**(1983), 153–158.

[9] H. Furuhata, Hypersurfaces in statistical manifolds, Differential Geom. Appl. 27(2009), 420–429.

[10] M.A. de Gosson, Symplectic Geometry and Quantum Mechanics. Birkhäuser, Basel, series ”Operator Theory: Advances and Applications” (2006).

[11] H.V. Le, Monotone Invariants and Embeddings of Statistical Manifolds, Advances in Deterministic and Stochastic Analysis, World Scientific (2007), 231–254.

[12] A. Mori, Information geometry in a global setting, OCAMI Preprint Series 2016, 16–12.

[13] 森 淳秀, 接触流，Anosov 流の観点から見た正規分布，Student-t 分布の空間について, ミニワークショップ 統計多様体の幾何学とその周辺 (9) [講演資料].

[14] T. Noda, Symplectic structures on Statistical manifolds, J. Aust. Math. Soc. 90 (2011), 371–384.

[15] 照井伸彦, R によるベイズ統計分析, 朝倉書店 (2010).

[16] 渡辺 澄夫, ベイズ統計の理論と方法, コロナ社 (2012).

[17] A. Weinstein, Symplectic geometry, Bull. Amer. Math. Soc. (N.S.) 5 (1981), no. 1, 1–13.

〒 204-8588 東京都清瀬市野塩 2-522-1 明治薬科大学 薬学教育研究センター 数理科学部門/数学

E-mail address: `noda@my-pharm.ac.jp`