捩れをゆるす統計多様体とプレ・コントラスト関数

松添 博（名古屋工業大学）

（記：野田 知宣）

1. 序

多様体上のコントラスト函数から計量や接続など統計構造を導く事が出来る。数学的にコントラスト函数を拡張したとき、捩れのある統計構造が得られる。この構造を工学に応用出来ないだろうか？　それを知るためにプレ・コントラスト函数の布教を目的とし、この構造について解説を行う。

2. 統計多様体とコントラスト函数

先ずは確率分布全体の空間の有限次元部分多様体（パラメトリック・モデル）の幾何学的抽象化である統計多様体の概念の復習から始める。

M を多様体とする（以下、登場する対象は全て C^∞ 級とする）。M 上の計量を g または h で表す。計量に正定値性を仮定する場合、即ち Riemann 計量となっている場合には g を使用し、幾何学的に正定値性が必要なく擬 Riemann 計量で充分な場合は h を使用する。また ∇ を捩れのないアフィン接続とする。

定義 2.1 (統計多様体)**.** 多様体 M, 接続 ∇, 計量 h に対し

$$(\nabla_X h)(Y,Z) = (\nabla_Y h)(X,Z), \qquad {}^\forall X,Y,Z \in \mathfrak{X}(M),$$

が成立するとき、3 つ組 (M,∇,h) を**統計多様体**（statistic manifold, statistical manifold）と呼ぶ。

統計多様体は統計学で、特に情報量に焦点を合わせると自然に現れる幾何構造である。アフィン接続 ∇ に対し

$$Xh(Y,Z) = h(\nabla_X Y, Z) + h(Y, \nabla^*_X Z)$$

により ∇^* を定義する。これが捩れなしのとき (M,∇^*,h) は統計多様体となる。この ∇^* を **∇ の双対接続**と呼ぶ。

計量や接続など、これらの構造はコントラスト函数と呼ばれる M 上の距離の（自乗の）ような函数から導く事が出来る。それを説明する為に、先ずはコントラスト函数を定義したい。その準備として記号を一つ導入しておく。

ρ を $M\times M$ 上の函数とする（i.e., $\rho : M\times M \to \mathbb{R}$）。この ρ に対し M 上の函数を $r\in M$ に対し

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

で定める。ここで $X_1,\ldots,X_i$ と $Y_1,\ldots,Y_j$ は M 上のベクトル場であり、$(X_a)_p$ は X による $M\times M$ の第一成分に沿う微分を表し、$(Y_b)_q$ は第二成分に沿う微分を表

す[1]。また $|_{p=r,q=r}$ の部分で $M\times M$ の対角成分に制限し、これを M と同一視して M 上の函数とみている。この記号を用い、次のように定義する。

定義 2.2 (コントラスト函数)**.** $\rho: M\times M\to\mathbb{R}$ が次の 3 条件を満たすとき ***M* 上のコントラスト函数**（contrast function）と呼ぶ：
(1) 任意の $r\in M$ に対し $\rho(r,r)=0$;
(2) 任意の $X\in\mathfrak{X}(M)$ に対し $\rho[X|\cdot]=\rho[\cdot|X]=0$;
(3) 任意の $X,Y\in\mathfrak{X}(M)$ に対し $h(X,Y):=-\rho[X|Y]$ は M 上の擬 Riemann 計量を定める。

この定義において (2) が ρ の 1 階微分についての条件であり、(3) が 2 階微分についての条件である。(1)〜(3) の条件は全て対角成分に制限したときのものである事に注意する。対角成分では統計構造を指定するが、それ以外の部分での制約は特にない（用語の用法として、この辺りがダイバージェンスとの違い）。また h が正定値なら (2) の条件は自動的に満たされる。

コントラスト函数 ρ から計量 h が定まる事は定義に入っているが、接続が定まる事も直ぐに判る。

命題 2.3. コントラスト函数 ρ に対し

$$h(\nabla_X Y,Z)=-\rho[XY|Z],$$
$$h(\nabla_X Y,Z)=-\rho[XY|Z]$$

により ∇ と ∇^* を定めると、これらは M 上のアフィン接続を定める。更に ∇ と ∇^* は h に関し双対であり、(M,∇,h) と (M,∇^*,h) は共に統計多様体である。

これでアフィン接続も定まる事が判った。これには ρ の 3 階微分を使用している。では 4 階微分はどうなる？ と考えるのは自然な事である。4 階微分を考えると h の曲率の情報が現れてくる。

命題 2.4. M 上のコントラスト函数 ρ に対し (1,1)-テンソル B, B^* を

$$h(B(X,Y)Z,V)=-\rho[XYZ-\nabla_X\nabla_Y Z|V],\qquad X,Y,Z,V\in\mathfrak{X}(M),$$
$$h(Z,B^*(X,Y)V)=-\rho[Z|XYV-\nabla^*_X\nabla^*_Y V]$$

で定める。このとき

$$R^{\nabla}(X,Y)Z=B(X,Y)Z-B(Y,X)Z,$$
$$R^{\nabla^*}(X,Y)Z=B^*(X,Y)Z-B^*(Y,Z)X$$

が成立する。

[1] $(X_a)_p$ は X_a を $(X_a,0)$ により $M\times M$ 上のベクトル場と見做し、ρ を微分している。$(Y_b)_q$ も同様。

B を定義するところで $-\rho[XYZ|V]$ ではテンソルにならない。そこでテンソルになるように $-\nabla_X\nabla_Y Z$ の項が加わっている（引いてるけど）。B^* も同様。幾何学では曲率テンソル R は重要な量であるが、統計学では多様体の曲がり方によって統計量を補正する（少しずらす）ことに使いたいので、B の方が有用と思われる。コントラスト函数についてより詳しくは Eguchi [1], Matsuzoe [3] を参照の事。

3. 最尤法

前節でのコントラスト函数の一般化を考える動機付けの一つとして最尤法について復習をしておく。動機付けについてはこれ以外にも量子力学的な方法もある。

S を指数型分布族とする：

$$S = \{p\ ;\ p(x;\theta) = \exp\left[C(x) + \sum \theta^i F_i(x) - \psi(\theta)\right]\}.$$

S の n 次元部分多様体 M を曲指数型分布族と呼ぶ。$U \subset M$ を M の局所座標系とし、$x_1, \ldots, x_N$ を（データの）実現値とする。N は観測したデータの個数であり、M の次元 n とは別物である事に注意する。このとき尤度函数 $L(u)$ を

$$L(u) = p(x_1;\theta(u)) \cdots p(x_N;\theta(u))$$

と定める（$x_1, \ldots, x_N$ は独立に同一分布に従うとし、尤度は積の形とする）。この尤度を最大にする推定量を**最尤推定量**（Maximum Likelihood Estimator, mle）と呼び $\hat{u}$ で表す：

$$\hat{u} : \mathrm{mle} \overset{\mathrm{def}}{\Longleftrightarrow} \hat{u} := \underset{u\in U}{\mathrm{argmax}} L(u).$$

Kullback-Leibler ダイバージェンスを KL で表す：

$$\mathrm{KL}(p(\theta)\|p(\theta')) := \int_\Omega \log \frac{p(\theta)}{p(\theta')} p(\theta) dx.$$

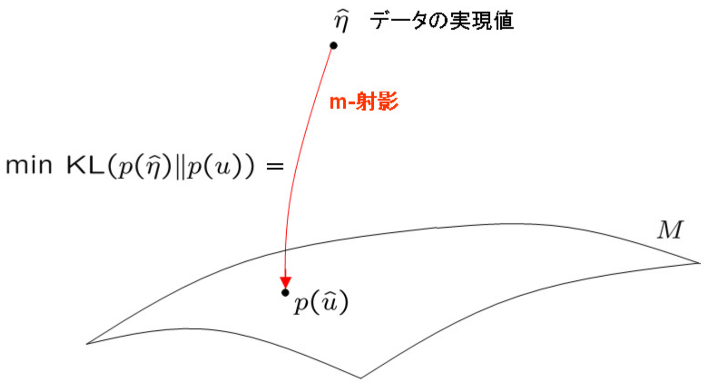
また標本平均値を $\hat{\eta}$ とする：

$$\hat{\eta}_i := \frac{1}{N}\sum_{j=1}^{N} F_i(x_j).$$

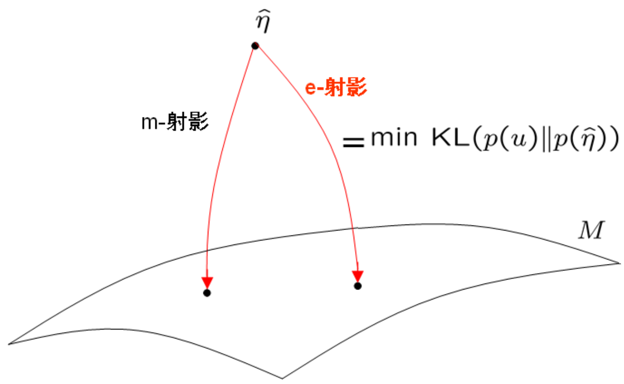
いま、指数型分布族を考えているので、これは双対座標の成分である。このとき最尤推定量 $\hat{u}$ に対し

$$\hat{u} : \mathrm{mle} \ \Leftrightarrow\ \hat{u} = \mathrm{argmin}\ \mathrm{KL}(p(\hat{\eta})\|p(u))$$

が成立する。Kullback-Leibler ダイバージェンスを距離のように思って、それが最小となっているところが最尤推定の $\hat{u}$ である。$p(\hat{\eta})$ から $p(\hat{u})$ への射影は直交であり、これを **m-射影**と呼ぶ。



KL ダイバージェンスは非対称であるから、入れ替えると別の推定量が出てくる。$\mathrm{KL}(p(u)\|p(\hat{\eta}))$ を最小とするような射影を **e-射影**と（か平均場近似と）呼ぶ。



推定量 $\hat{u}$ は尤度方程式の解だから、これを求めるには

$$d(\rho(,\cdot))(X) = 0, \qquad {}^{\forall}X \in T_{p(u)}M,$$

となる点を探す事が必要である。そこで、これを特徴付けるようなものを考える。

4. プレ・コントラスト函数

コントラスト函数を拡張してプレ・コントラスト函数を定義するのであるが、その前に記号を準備しておく。$\rho : TM \times M \to \mathbb{R}$ に対し

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

と定める。

定義 4.1. （プレ・コントラスト函数）$\rho : TM \times M \to \mathbb{R}$ が次の 3 条件を満たすとき**プレ・コントラスト函数**と呼ぶ：
(1) $\rho(f_1X_1 + f_2X_2, q) = f_1\rho(X_1, q) + f_2(X_2, q)$;
(2) $\rho[; X|] = 0$;
(3) $h(X, Y) := -\rho[; X|Y]$ は擬 Riemann 計量を定める。

ρ の条件 (1) は線型性（テンソル性）を表し、(2) は対角成分に制限すると 0 に当たる条件である。コントラスト函数の場合と同様に、プレ・コントラスト函数からも接続を定める事が出来る。

命題 4.2. プレ・コントラスト函数 ρ に対し

$$h(\nabla_X Y, Z) = -\rho[X; Y|Z],$$
$$h(\nabla_X Y, Z) = -\rho[; Y|XZ]$$

により ∇ と ∇^* を定めると、これらは M 上のアフィン接続を定める。更に ∇ と ∇^* は h に関して双対であり、∇^* は捩れなしである。

これらを踏まえ、表題にあるものを次のように定義する。

定義 4.3 (捩れをゆるす統計多様体)**.** 多様体 M, アフィン接続 ∇, 計量 h に対し

$$(\nabla_X h)(Y, Z) - (\nabla_Y h)(X, Z) = -h(T^\nabla(X, Y), Z), \qquad {}^\forall X, Y, Z \in \mathfrak{X}(M),$$

が成立するとき、３つ組を (M, ∇, h) を**捩れをゆるす統計多様体**と呼ぶ。

このとき次が成立する。

命題 4.4. ρ を多様体 M 上のプレ・コントラスト函数とし、h と ∇ はそれぞれ ρ から定まる計量とアフィン接続とする。このとき３つ組 (M, ∇, h) は捩れをゆるす統計多様体である。

§ 2 のコントラスト函数があれば、その微分によって自明なプレ・コントラスト函数が定まる。このとき誘導されるアファイン接続 ∇ は捩れを持たず、(M, ∇, h) は通常の統計多様体である。

最後に、プレ・ダイバージェンスも定義出来る事を注意しておく。ダイバージェンスの構成はいろいろな手法が知られているが、アファインはめ込みの幾何学を用いても構成する事が出来る（幾何学的ダイバージェンスと呼ぶ）。プレ・ダイバージェンスは、アフィン分布の幾何学を用いて構成出来る。捩れをゆるす統計多様体およびプレ・コントラスト函数については黒瀬 [2], Matsuzoe [4] を参照の事とする。

References

[1] S. Eguchi, Geometry of minimum contrast, Hiroshima Math. J., **22** (1992), 631–647.

[2] 黒瀬 俊, Statistical manifolds admitting torsion, 2007 年度福岡大学微分幾何研究会講演録, 2007.

[3] H. Matsuzoe, Geometry of contrast functions and conformal geometry, Hiroshima Math. J., **29** (1999), 175–191.

[4] H. Matsuzoe, Statistical manifolds and affine differential geometry, Advanced Studies in Pure Mathematics 57, Probabilistic Approach to Geometry, in press.