

非正則分布族のフィンスラー幾何学

東大工学部 甘利俊一 (Shun-ichi Amari)

1. はじめに

本報告は、非正則な統計分布の族が幾何学的にはフィンスラー空間の構造を持つことと指摘し、非正則なモデルのもとでの統計的推論の漸近理論がフィンスラー幾何学を用いて建設できることを主張せんとするものである。これにはベクトル値率変数の安定分布が Minkowski 計量を持つことが関係している。本稿はこの問題に関する予備的考察であって、理論として定式化できているものではないことを断っておく。

2. Minkowski 計量と Finsler 空間

n 次元の線形空間 T と考えよう。 T の要素であるベクトル x は基底 $\{e_i\}$ ($i=1, \dots, n$) を用いて

$$x = x^i e_i$$

のように表わせる (ここで指標 i についての和をとる Ein-

steinの簡約和法を用いる)。このとき、 $x \in$ 成分 x^i ($i = 1, \dots, n$) で表わすことができる。線形空間 T に、ベクトル x の長さを表わす基本関数

$$\|x\| = F(x)$$

が与えられているものとしよう。 F は 1 次の齊次関数、すなわち

$$F(cx) = cF(x), \quad c > 0$$

を満足する凸関数であるとし、

$$F(x) \geq 0$$

で等号は $x=0$ のときこのときのみ成立するものとする。

さらに、ここでは $F(-x) = F(x)$ も仮定する。このようなノルムを定義する基本関数 F が与えられている線形空間を Minkowski 空間と呼ぶ。ユークリッド空間は F^2 が二次形式

$$F^2(x) = g_{ij} x^i x^j,$$

g_{ij} は正定の定数テンソル、で与えられる Minkowski 空間である。

T において

$$F(x) = 1$$

を満足する x の集合 (これは原点を囲む $n-1$ 次元曲面と見る) を *indicatrix* と呼ぶ。 $F(x) \leq 1$ で定義された領域は

凸である。ユークリッド空間とは, indicatrix が二次曲面 (楕円体) となる空間のことである。

多様体 S を考へ, S の局所座標系を $\theta = (\theta^1, \dots, \theta^n)$ としよう。一点 θ における S の接空間を T_θ とおく。各 T_θ に基本関数 $F(x, \theta)$ が与えられた多様体, すなわち接空間が Minkowski 空間であるような多様体を Finsler 空間という。(F の微分可能性を仮定する。) F^2 が x の二次関数となる Riemann 空間は Finsler 空間の特殊な場合である。Riemann 空間の場合と同様に, Finsler 空間の曲線に与えられた F の積分によつて, 曲線 $\theta(s)$ の長さを

$$\int F(\theta, \dot{\theta}(s)) ds$$

で定義できる ($\dot{\theta} = d\theta/ds$)。また, ここでは述べないが接空間の接続を導入することができる。

Minkowski 空間 T に話を戻そう。 F は 1 次の奇次関数だから, $F^2(x) = \{F(x)\}^2$ として

$$g_{ij}(x) = \frac{1}{2} \frac{\partial^2}{\partial x^i \partial x^j} F^2(x)$$

は 0 次の奇次関数と成分とする正定行列となる。すなわち

$$g_{ij}(Cx) = g_{ij}(x)$$

さらに

$$F^2(x) = g_{ij}(x) x^i x^j$$

が成立する。

例 L_p 空間

$$F(x) = \{|x^1|^p + |x^2|^p\}^{\frac{1}{p}}$$

のときは

$$g_{11}(x) = (2-p) |x^1|^{2p-1} \{|x^1|^p + |x^2|^p\}^{\frac{2}{p}-2} \\ + (p-1) |x^1|^{p-2} \{|x^1|^p + |x^2|^p\}^{\frac{2}{p}-1}$$

$$g_{12}(x) = |x^1|^{p-1} |x^2|^{p-1} \left(\frac{2}{p} - 1\right) p \{|x^1|^p + |x^2|^p\}^{\frac{2}{p}-2}$$

を得る。

二つのベクトル x, y に対して

$$g_{ij}(x) x^i y^j = 0$$

が成立するとき、 y は x に対して直交するといふ。このとき x は y に直交するとは限らぬ。

ベクトル x に対して双対ベクトル x^* を次式で定義しよう。

$$x = (x^i), \quad x^* = (x_i) \quad \text{として}$$

$$x_i = g_{ij}(x) x^j.$$

このとき

$$F^2(x) = x^i x_i$$

が成立する。

x と x^* は Legendre 変換で結ばれている。また、

$$u = x^* \quad (u_i = x_i)$$

として、関数 $H(u)$ が存在して

$$u_i = \frac{1}{2} \frac{\partial F^2(x)}{\partial x^i}, \quad x^i = \frac{1}{2} \frac{\partial H^2(u)}{\partial u_i}$$

の関係が成立し、

$$F^2(x) + H^2(u) - 2x^i u_i = 0$$

の関係で結ばれている。 $H(u)$ は 1 次 homogeneous 関数

$$H(cu) = |c| H(u)$$

で、

$$F(x) = H(u)$$

が成立する。 $H(u)$ は双対空間 $\Gamma^* = \{u\}$ に Minkowski 計量を定義する。

$$g^{ij}(u) = \frac{1}{2} \frac{\partial^2 H^2(u)}{\partial u_i \partial u_j}$$

とすると、

$$H^2(u) = g^{ij}(u) u_i u_j$$

であって、 $g_{ij}(x)$ と $g^{ij}(u)$ とは互に逆行列の関係で結ばれ

$$x^i = g^{ij}(u) u_j, \quad u_i = g_{ij}(x) x^j$$

である。

3. 安定分布と Minkowski 計量

いま T^* に値をとる確率変数 $Z = (Z_i) \quad (i=1, 2, \dots, n)$ を考よう。同一で独立な分布に従かう任意の良個の確率変数 Z_1, Z_2, \dots, Z_k に対してある a_k が存在して

$$\frac{S_k}{a_k} = \frac{1}{a_k} (Z_1 + \dots + Z_k)$$

が元の Z と同一の分布に従かうとき、この分布を安定分布という。分布の特性関数 $\varphi(t)$, $t = (t^i)$, $i=1, \dots, n$ と

$$\varphi(t) = E[\exp\{i t \cdot Z\}], \quad t \cdot Z = t^i Z_i$$

で定義すると、安定分布は特性関数を用いて

$$[\varphi(t)]^k = \varphi(a_k t)$$

で定義できる。 $a_k = k^{1/\alpha}$, $1 \leq \alpha \leq 2$ であることが知られていて、 $a_k = k^{1/\alpha}$ となる分布を指数 α の安定分布と

いう。キュムラント関数 $\psi(t)$

$$\psi(t) = \log \varphi(t), \quad \varphi(t) = e^{\psi(t)}$$

を用いると、 $\psi(t)$ は

$$\psi(t) = -A(t) \{1 + i G(t)\}$$

と書ける (Lévy-Khinchin 表示の一般化)。ここに、 $\psi(t)$

は α 次の齊次正関数

$$A(c\#) = |c^\alpha| A(\#),$$

$G(\#)$ は 0 次の齊関数で

$$G(-\#) = -G(\#)$$

である。ここでは簡単のため、 \mathcal{X} と $-\mathcal{X}$ の分布は等しいと仮定する。すると

$$\Psi(\#) = -A(\#), \quad \varphi(\#) = \exp\{-A(\#)\}$$

となる。

ここで $\#$ の空間 T に Minkowski 計量を

$$F(\#) = \{A(\#)\}^{1/\alpha}$$

により導入する。すると計量テンソルは

$$g_{ij}(\#) = \frac{1}{2} \frac{\partial^2 F^2}{\partial t^i \partial t^j}$$

で与えられる。かくして、安定分布に対して Minkowski 空間が一つ対応する。とくに $\alpha = 2$ の場合は安定分布は正規分布となり、 $g_{ij}(\#)$ は $\#$ によらなにから二次形式 g_{ij} による計量が一つ定義できる。明らかに \mathcal{X}_i は $N(0, g_{ij})$ に従う正規分布となる。一般の場合には、

$$\varphi(\#) = \exp\left\{-\left[g_{ij}(\#) t^i t^j\right]^{\frac{\alpha}{2}}\right\}$$

のように書ける。

議論を双対に進めると、確率変数 $z = (z_i)$ の空間に

$$\|z\| = H(z)$$

により、Minkowski ノルムを導入できる。H は F より Legendre 変換により、得られる。

$$g^{ij}(z) = \frac{1}{2} \frac{\partial^2 H(z)}{\partial z_i \partial z_j}$$

で、

$$\|z\|^2 = \frac{1}{2} g^{ij}(z) z_i z_j$$

である。 z と t と

$$t^i = g^{ij}(z) z_j, \quad z_i = g_{ij}(t) t^j$$

で結びつけると、 t を確率変数と見た空間 T に確率測度を導入できる。 t の分布を調べる必要がある。

いま、二つの指数 α の定常分布 z_1, z_2 があるとき、その直積 (直和) を $z = (z_1, z_2)$ とする。 z_1 と z_2 は確率変数として独立であるとする。このとき、 $t = (t_1, t_2)$ と対応する双対ベクトルとして、 z の特性関数は

$$\varphi(t) = E[e^{iz \cdot t}] = E[e^{iz_1 \cdot t_1}] E[e^{iz_2 \cdot t_2}]$$

となる。したがって t の基本関数 $F(t)$ は

$$F(t) = \left[\{F_1(t_1)\}^\alpha + \{F_2(t_2)\}^\alpha \right]^{\frac{1}{\alpha}}$$

ここに F_1, F_2 は μ_1, μ_2 の基本関数である。これは指数 α の安定分布の L_α 空間としての性格を明らかにする。このとき双対である $H(\mathbb{R})$ は L_α 空間 $\{(1/\alpha) + (1/\alpha') = 1\}$ の性質を持つと考えられる。

4. 非正則統計的モデルと Finsler 空間

n 次元空間の translation family として、分布密度関数

$$p(x, \theta) = f(x - \theta)$$

確率変数の分布族 $S = \{p(x, \theta)\}$ と考える。ここに $\theta = (\theta^1, \dots, \theta^n)$ は分布を指定するパラメータであり、関数 $f(x)$ の台は compact な凸領域で、その境界における立ち上がりは、境界からの距離 d に応じて

$$f(x) = c d^{\alpha-1}, \quad 1 \leq \alpha \leq 2$$

の形をとりうるものとする。1次元のときは、たとえば

$$f(x, \theta) = c_\alpha (x - \theta)^{\alpha-1} (1 - x + \theta)^{\alpha-1}$$

のようになる (x は区間 $[0, 1 + \theta]$ の外では 0)。

分布族 S は θ を座標系として n 次元の多様体となる。1点 θ における S の接空間を T_θ とする。 T_θ の自然基底を $\{\partial_i\}$ とし、 ∂_i の確率変数による表現を

$$\partial_i \sim \partial_i \rho(x, \theta), \quad \partial_i = \frac{\partial}{\partial \theta^i}$$

$$l(x, \theta) = \log p(x, \theta)$$

とする。 T_θ のベクトル

$$A = A^i \partial_i$$

は確率変数

$$A(x) = A^i \partial_i l(x, \theta)$$

で表現できる。

二つの確率分布 $p(x, \theta)$ と $p(x, \theta')$ の間の分離度と Hellinger 距離

$$D(\theta, \theta') = 4 \int (\sqrt{p(x, \theta)} - \sqrt{p(x, \theta')})^2 dx$$

または竹内を導入した

$$D(\theta, \theta') = -8 \log \int \sqrt{p(x, \theta) p(x, \theta')} dx$$

により定義しよう。このとき、 $d\theta$ を微小とすると

$$D(\theta, \theta + d\theta) = \begin{cases} O(|d\theta|^2), & \alpha > 2 \\ |d\theta|^2 \log |d\theta|, & \alpha = 2 \\ |d\theta|^\alpha, & 1 \leq \alpha < 2 \end{cases}$$

となることがわが、こゝる。複雑な議論をさけるため、とりあえず $1 < \alpha < 2$ および $\alpha > 2$ の正則の場合のこゝを考之よう。

T_θ における接ベクトル $h = h^i \partial_i$ ($h(x) = h^i \partial_i l$)

のノルムを基本関数

$$F(h, \theta) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \{D(\theta, \theta + \varepsilon h)\}^{\frac{1}{\alpha}}$$

$F(h, \theta)$ は h について 1 次斉次の正関数である。これより、計量テンソル (フィッシャー計量)

$$g_{ij}(h, \theta) = \frac{1}{2} \frac{\partial^2 F^2(h, \theta)}{\partial h^i \partial h^j}$$

が定義できるから、 S は Finsler 空間になる。

直積に分解できる場合、すなわち

$$P(x, \theta) = P_1(x_1, \theta_1) P_2(x_2, \theta_2),$$

$x = (x_1, x_2)$, $\theta = (\theta_1, \theta_2)$ と仮定する場合は、明らかに

$$F(h, \theta) = \left[\{F_1(h_1, \theta_1)\}^\alpha + \{F_2(h_2, \theta_2)\}^\alpha \right]^{1/\alpha}$$

となる。ただし $h = (h_1, h_2)$ で、 F_1, F_2 はそれぞれ成分系の基本関数である。このとき \rightarrow のベクトル

$$(h_1, 0), (0, h_2)$$

は互にどちらか一方みても直交していることがわかる。

同一独立分布の場合、すなわち

$$P(x_1, \dots, x_m; \theta) = \prod_{i=1}^m P(x_i, \theta)$$

の場合は、 m 個の観測に基づく基本関数 F_m は

$$F_m(h, \theta) = m^{\frac{1}{\alpha}} F(h, \theta)$$

である。

次に、確率変数 $z = (z_i)$,

$$z_i = a_i l(x, \theta)$$

に注目しよう。(点 θ は固定し, θ を書くのを省略することがある。) 明らかに

$$E[z_i] = 0$$

である。さらに, 独立な観測 x_1, \dots, x_k に対して

$$k^{-1/\alpha} \sum_{i=1}^k z_i$$

は指数 α の安定分布へ収束する。これをみるには, 1-次元の場合の ($\theta=0$ において)

$$z = l'(x) = \frac{c}{x}$$

の分布を調べて, z の分布 $f(z)$ は $z \rightarrow \infty$ のときに

$$f(z) dz = p(x) dx = z^{-\alpha-1} dz,$$

したがって

$$1 - P(z) = \int_z^{\infty} f(z) dz = z^{-\alpha},$$

これより (Feller, p. 303) l' は指数 α の安定分布の domain of attraction に入ることがわかる。同様の理由で行き $a_i a_j l(x, \theta)$ は指数 $\alpha/2$ の安定分布の d. of attraction に入る。

z の収束する安定分布の特性関数を $\varphi(t, \theta)$ とすると

$$\varphi(t, \theta) = E[\exp\{it' a_j l(x, \theta)\}]$$

を用いて求まる。この安定分布（ θ 自身は安定分布ではないが）が定義する Minkowski 計量 $\tilde{F}(\theta, \theta)$ としよう。 \tilde{F} と F との関係を知る必要がある。

雑な議論を進めると、

$$\begin{aligned}\varphi(\varepsilon\theta) &= E[\exp\{i\varepsilon t^2; l(x, \theta)\}] \\ &= E[\exp\{i\Delta_{\varepsilon\theta} l(x, \theta)\}] \\ &= E[1 - \frac{1}{2}(\Delta_{\varepsilon\theta} l)^2] \\ &= \exp\{-\frac{1}{2}E[(\Delta_{\varepsilon\theta} l)^2]\},\end{aligned}$$

ここで

$$\Delta_{\varepsilon\theta} l(x, \theta) = l(x, \theta + \varepsilon\theta) - l(x, \theta),$$

一方

$$\begin{aligned}D(\theta, \theta + \varepsilon\theta) &= -8 \log \int \{p + \frac{1}{2}\Delta_{\varepsilon\theta} p - \frac{1}{8}\frac{(\Delta p)^2}{p}\} dx \\ &= E[\{\Delta_{\varepsilon\theta} l(x, \theta)\}^2] = -E[\Delta_{\varepsilon\theta}^2 l(x, \theta)] \\ &= \varepsilon^\alpha F(\theta)^{\frac{\alpha}{2}}\end{aligned}$$

が成立している。ただし $\Delta_{\varepsilon\theta} l$ についての期待値をとるときは $p(x, \theta)$ と $p(x, \theta + \varepsilon\theta)$ の共通の台上でのみ積分をとるものとする。これより F と \tilde{F} とが同じようなものであることがわかる。

ここで、m.l.e. $\hat{\theta}$ について少し考へる。

$$\hat{\theta} = \theta_0 + d\theta$$

とおくと (θ_0 は真の値), N 個の測定 x_1, \dots, x_N に対し,

$$\sum_{k=1}^N \alpha_k \ell(x_k, \theta_0 + d\theta) = 0$$

これより

$$N^{-1/\alpha} \sum \alpha_k \ell(x_k, \theta_0) = -N^{-2/\alpha} \sum \alpha_k \ell(x_k, \theta_0) d\tilde{\theta}^j$$

$$d\tilde{\theta}^j = N^{1/\alpha} d\theta^j$$

左辺は確率変数 z_i に, 右辺の係数項は w_{ij} に収束するから

$$d\tilde{\theta}^j = w_{ij} z_i$$

と仮して $d\theta$ が $N^{-1/\alpha}$ のオーダーで収束することかわかる。確率変数 w_{ij} の性質を調べる必要がある。(正則な場合はこれは定数行列 g_{ij} に収束する。)

ここで $d\theta$ が双対ベクトルであり, かつ確率変数として扱われることに注意を要する。多分漸近的には

$$E[F(d\tilde{\theta})] \geq n \quad \text{or} \quad \{E[F(d\tilde{\theta})^\alpha]\}^{1/\alpha} \geq n$$

で, efficient な estimator に対しては等号が成立するといったものが Cramer-Rao の定理の非正則版として成立するのである。この際, 等号を漸近的に達成するのである。

は必ずしも m.l.e. ではなく、何らかの L_α , $L_{\alpha'}$ / L_μ を用いて得られる推定量であるかもしれない。

文献

H. Rund, The Differential Geometry of
Finsler Spaces, Springer-Verlag, Berlin-
Göttingen-Heidelberg, 1959.

竹内啓, Non-regular な統計的モデルの幾何学,
「統計的推論の漸近理論と幾何学的構造」研究会資料
1983年10月