

変換群を持つ統計モデルの幾何学

東大工学部 甘利 俊一 (Shun-ichi Amari)

東大工学部 倉田 耕治 (Koji Kurata)

§ 1 変換群モデルの例

R^n 上に, 2個のパラメータ a, b ($a > 0$) により, 2指定される変換 g を考える。

$$g(x) = g \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} g_1(x) \\ \vdots \\ g_n(x) \end{pmatrix} = \begin{pmatrix} ax_1 + b \\ \vdots \\ ax_n + b \end{pmatrix}$$

すると, このような変換 g の集合 $G = \{(a, b) \mid a > 0\}$ は, 変換の合成という演算に関して群を成す。

$$(a, b) \circ (c, d) = (ac, ad + b)$$

$$e = (1, 0) \quad (a, b)^{-1} = \left(\frac{1}{a}, -\frac{b}{a} \right)$$

ここで, R 上の密度関数をひとつ固定すると, G を母数空間とするモデルができる。すなわち, $p(x)$ から

$$p(x, e) = p(x_1) p(x_2) \cdots p(x_n),$$

$$\begin{aligned}
 p(x, g) &= p(g^{-1}x, e) \det \left(\frac{\partial g_i^{-1}(x)}{\partial x_j} \right) \\
 &= \frac{1}{a^n} p\left(\frac{x_1 - b}{a}\right) p\left(\frac{x_2 - b}{a}\right) \cdots p\left(\frac{x_n - b}{a}\right)
 \end{aligned}$$

のように作るわけである。

これはロケーション・スケール・モデルと呼ばれるものであるが、同じように、標本空間上の変換群と、ひとつの「種」となる分布によつて作られる分布には多次元正規分布、フィッシャーのサークルモデルなどがある。これらのモデルは、変換群モデル (transformation model) と呼ばれている。

変換群モデルに関しては、不変なロス、不変な推定量といった、自然な概念が定義される。

本稿では、変換群モデルに関して、近年筆者の一人が提唱している幾何学的な統計理論 (Amari 1985) を展開するため、変換群モデルの統計幾何学的諸量を求めた。

§2 変換群モデルの定式化

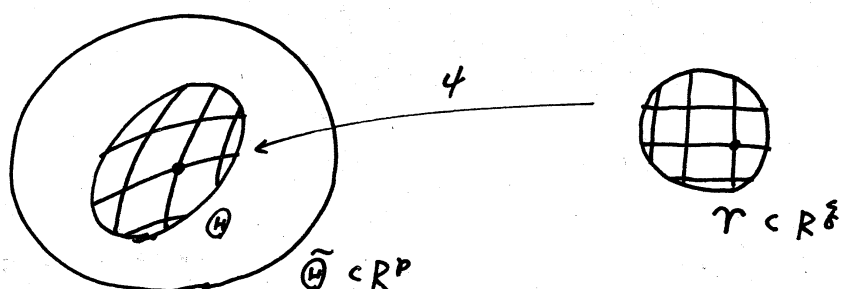
本節では Kariya (1983) に従つて、変換群モデルの定式化を行なう。

標本空間を X 、母数空間を Θ とし、 $X \subset \mathbb{R}^n$ 、 $\Theta \subset \mathbb{R}^p$ とする。 $\theta \in \Theta$ によつて指定される分布関数を P_θ と書く。 $\theta_1 \neq \theta_2$ ならば、 $P_{\theta_1} \neq P_{\theta_2}$ であるとする。

モデル $\Theta \subset \tilde{\Theta}$ を考え, Θ は, $\gamma \in R^b$ によって, 2 点間メトリクスを定めているとする:

$$\Theta = \{ \theta \in \tilde{\Theta} \mid \theta = \psi(\gamma) \ \gamma \in R \}.$$

$b < p$ であるとする.

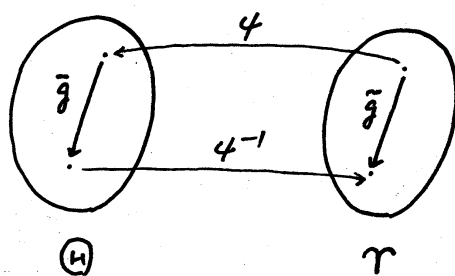


X 上に作用する群 G を考える. $gP_\theta = P_\theta g^{-1}$ と定めるとき, $gP(\tilde{\Theta}) = P(\tilde{\Theta})$ か, すべての $g \in G$ に関しても成り立つとする. このとき, G から, $\tilde{\Theta}$ 上に作用する群 $\bar{G}$ が induce される. つまり, $P_\theta g^{-1} = P_{\bar{g}\theta}$ とする. すると, G は $\bar{G}$ に準同型である.

Θ が $\bar{G}$ の orbit であるとは, すべての $\bar{g} \in \bar{G}$ に対し $\bar{g}\Theta = \Theta$ であり, かつ, すべての $\theta_1, \theta_2 \in \Theta$ に対し, ある $\bar{g} \in \bar{G}$ が存在して $\bar{g}\theta_1 = \theta_2$ となることである. 2 点を仮定する. なお, このふたつの条件のうち, 後者が成立するとき, $\bar{G}$ は Θ に transitive に作用するといふ.

すると, ψ により, γ 上に作用する変換群 $\tilde{G}$ が induce される:

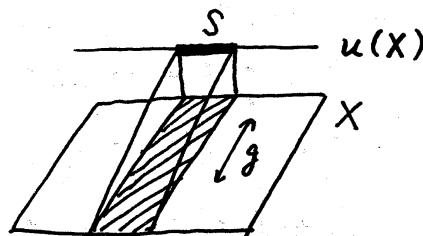
$$\tilde{G} = \{ \psi \bar{g} \psi^{-1} \mid \bar{g} \in G \}$$



$\tilde{H}$ は $\tilde{G}$ により, いくつかの orbit に分けられる。そのひとつが, $\tilde{H}$ である。 $\tilde{H}$ 上に定義された関数で, 各 orbit 内で, 一定値をとるものを invariant parameter と呼び, 特に, 異なる orbit 上では異なる値を持つものを maximal invariant parameter と呼ぶ。

$\lambda(\theta)$ を $\tilde{H}$ 上の maximal invariant parameter, $u(x)$ を X 上の maximal invariant parameter とすると, u の分布は, λ の ψ による。

なぜなら, θ_1 と θ_2 が同じ λ を与えたとすると, このふたつは, ある $\bar{g} \in \tilde{G}$ により, $\theta_1 = \bar{g} \theta_2$ と結びつけられる。



ずであり, $S \subset u(X)$ に対して,

$$\begin{aligned} P_{\theta_1}(u^{-1}(S)) &= P_{\tilde{g}\theta_2}(u^{-1}(S)) \\ &= P_{\theta_2}(g^{-1}u^{-1}(S)) \\ &= P_{\theta_2}(u^{-1}(S)) \end{aligned}$$

となるからである。

このより, もしモデルを④に限定し, $u(X)$ は, *ancillary* であることが分る。

真のハウメタ $\eta \in \mathcal{Y}$ により指定される分布から得られたデータを使って, η の推定値 $\hat{\eta}$ を得たとする。この時の損失を $L(\eta, \hat{\eta})$ とする。損失 $L(\eta, \hat{\eta})$ が不変であるとは, 任意の $\tilde{\eta} \in \hat{\mathcal{G}}$ に対し, $L(\eta, \hat{\eta}) = L(\tilde{\eta}, \hat{\eta})$ が成立することである。また, 推定量 $\hat{\eta}: X \rightarrow \mathcal{Y}$ が不変であるとは, 任意の $g \in G$, $x \in X$ に対し, $\hat{\eta}(gx) = \tilde{g}\hat{\eta}(x)$ が成立することである。

§3 不変なロスとフィッシャー情報量

以後, X と④が, G に対してひとつの orbit を成しており, この二者が互いに同型である場合を考える。群 G , $\tilde{G}$, $\hat{G}$ の区別はなく, すなわち, X , ④ と同一視される。

④中の1点, 例えば, 原点をある単位元と, 他の点との間に, ロスを定めると, このロスは, 不変性

を仮定すれば、すべての点とすべての点の間の間に2に拡張される。言い換えれば、1点の周りの12は、変換群によって、すべての点の周りの12に変換される。この変換はどのようなものなのかを、12の、各点の周りの2次の展開係数について調べてみよう。

$L(e, e+\xi)$ を ξ に関して展開すると、 $L(e, e) = 0, L \geq 0$ より ξ の1次の展開係数は0である。2次の係数を $L_{ij}(e)$ とする:

$$L(e, e+\xi) = \frac{1}{2} L_{ij} \xi^i \xi^j + \dots$$

ここで、右辺の1項は、 i, j に関して和をとることを略した記法である。すなわち、0の周りでは、

$$\begin{aligned} L(0, 0+\xi) &= L(e, \theta^{-1}(0+\xi)) \\ &= L(e, e + B_i^j \xi^j + \dots) \\ &= \frac{1}{2} L_{ij}(e) B_k^i B_l^j \xi^k \xi^l + \dots, \end{aligned}$$

$$2.2 \text{ に, } B_i^j(0) = \frac{\partial k^j(0, \eta)}{\partial \eta^i} \Big|_{\eta=0}, \quad k(0, \eta) = \theta^{-1}\eta.$$

$$\text{よって } L_{ij}(0) = L_{kl}(e) B_i^k(0) B_j^l(0) \text{ を得る。}$$

定理 1 フィッシャー-情報量を $g_{ij}(0)$ とすると、

$$g_{ij}(\theta) = g_{kl}(\eta) B_i^k B_j^l$$

これは, g_{ij} が L_{ij} と同じ変換に従うことを示している。よって 2 次の展開係数が, ファッシャー情報量に一致するような不変な η が存在するわけである。

証明 密度関数を $p(x, \theta)$ とおく。

$$p(gx, g\theta) d(gx) = p(x, \theta) dx$$

であるから,

$$\begin{aligned} p(x, \theta + \xi) dx &= p(\theta^{-1}x, \theta^{-1}(\theta + \xi)) d(\theta^{-1}x) \\ &= p(\eta, \eta + B_i^j(\theta) \xi^j) A(\theta, x) dx \end{aligned}$$

$$\text{22に, } \eta = \theta^{-1}x, \quad A(\theta, x) = \det \left(\frac{\partial \eta^i(\theta, x)}{\partial x^i} \right)$$

A は ξ によらない事を利用して, 上式の両辺の対数を取り, ξ で微分した後, $\xi = 0$ とおくと,

$$\begin{aligned} \partial_i \ell(x, \theta) &= \frac{\partial}{\partial \xi^i} \log p(x, \theta + \xi) \Big|_{\xi=0} \\ &= \frac{\partial}{\partial \xi^i} \log p(\eta, \eta + B_i^j \xi^j) \Big|_{\xi=0} \\ &= \partial_i \ell(\eta, \eta) B_i^j(\theta) \end{aligned}$$

2.2に ∂_i は $\frac{\partial}{\partial \theta^i}$ を意味する。

よって,

$$\begin{aligned} g_{ij}(\theta) &= E_\theta [\partial_i \ell(x, \theta) \partial_j \ell(x, \theta)] \\ &= E_\theta [\partial_k \ell(y, \theta) B_i^k(\theta) \partial_l \ell(y, \theta) B_j^l(\theta)] \\ &= B_i^k(\theta) B_j^l(\theta) g_{kl}(\theta) \end{aligned}$$

証明終

この定理を §1 であげたロケーション・スケール・モデルで確かめてみよう。まず B_j^i を求めると,

$$\begin{aligned} h((\theta^1, \theta^2), (\eta^1, \eta^2)) &= (\theta^1, \theta^2)^{-1} \cdot (\eta^1, \eta^2) \\ &= \left(\frac{\eta^1}{\theta^1}, \frac{\eta^2 - \theta^2}{\theta^1} \right) \end{aligned}$$

よって

$$B_j^i(\theta) = \begin{pmatrix} \frac{1}{\theta^1} & 0 \\ 0 & \frac{1}{\theta^1} \end{pmatrix}.$$

次に L_{ij} の変換は,

$$L(\theta, (1 + d\theta^1, d\theta^2)) = \frac{1}{2} L_{ij}(\theta) d\theta^i d\theta^j + \dots$$

$$L((\theta^1, \theta^2), (\theta^1 + d\theta^1, \theta^2 + d\theta^2))$$

$$= L(\theta, (1 + \frac{d\theta^1}{\theta^1}, \frac{d\theta^2}{\theta^1}))$$

$$= \frac{1}{2} L_{ij}(\theta) \left(\frac{1}{\theta^1}\right)^2 d\theta^i d\theta^j + \dots$$

一方, フィッシャー情報量の方は, $l(x) = \log p(x)$,

$y' = (x' - \theta^2) / \theta'$ とおくと,

$$\log p(x, (\theta', \theta^2)) = \log \left(\frac{1}{\theta'} p \left(\frac{x - \theta^2}{\theta'} \right) \right)$$

$$= -\log \theta' + l(y)$$

$$\therefore \partial_1 \log p(x, (\theta', \theta^2)) = -\frac{1}{\theta'} - \frac{1}{\theta'} y l'(y)$$

$$= \frac{1}{\theta'} \partial_1 \log p(y, \theta)$$

$$\partial_2 \log p(x, (\theta', \theta^2)) = -\frac{1}{\theta'} l'(y)$$

$$= \frac{1}{\theta'} \log p(y, \theta)$$

$$\therefore g_{ij}(\theta_1, \theta_2) = \left(\frac{1}{\theta'} \right)^2 g_{ij}(\theta)$$

例終

次に, 原点におけるフィッシャー情報量を具体的に求め, これに, 変換群と, 種になる分布が, どのように寄与しているかを見ることにする。この結果と, 定理1を合わせれば, すべての点でのフィッシャー情報量が得られる。

定理2 原点 $\theta = e$ におけるフィッシャー情報量は, 次の式で与えられる。

$$g_{ij}(e) = E_e \left[\left(r_i^k \frac{\partial}{\partial x^k} l(x) + \frac{\partial}{\partial x^k} r_i^k(x) \right) \right. \\ \left. \times \left(r_j^l \frac{\partial}{\partial x^l} l(x) + \frac{\partial}{\partial x^l} r_j^l(x) \right) \right]$$

2.2に $r_j^i(x) = \frac{\partial}{\partial \theta_j} k^i(e, x)$, 変換群のみで決まる量である。

証明 $l(x, \theta) = \log p(x, \theta)$, $l(x) = l(x, e)$,
 $p(x) = p(x, e)$ と書くことにする。

$$p(x, \theta) = p(k(\theta, x)) \left| \frac{\partial k}{\partial x} \right|$$

$$\therefore l(x, \theta) = l(k(\theta, x)) + \log \left| \frac{\partial k}{\partial x} \right|$$

これを θ_i で偏微分して

$$\partial_i l(x, \theta) = \frac{\partial l}{\partial k_j}(k(\theta, x)) \frac{\partial k_j}{\partial \theta_i} + \frac{\partial}{\partial \theta_i} \log \left| \frac{\partial k}{\partial x} \right|$$

上式右辺第1項は, $\theta = e$ とすれば, $r_i^k \frac{\partial}{\partial x^k} l$ となる。
 第2項を計算するために, $\left| \frac{\partial k}{\partial x} \right|$ を求める。

$$k^i(e + \theta, x) = x^i + \theta^j r_j^i(x) + O(\theta^2)$$

$$\therefore \frac{\partial k^i}{\partial x^j} = \delta_j^i + \theta^k \frac{\partial}{\partial x^j} r_k^i(x) + O(\theta^2)$$

$$\therefore \left| \frac{\partial k}{\partial x} \right| = 1 + \theta^k \frac{\partial}{\partial x^i} r_k^i(x) + O(\theta^2)$$

上式の右辺をもと、 $\tau \theta^k$ で偏微分すると、

$$\frac{\partial}{\partial \theta^k} \left| \frac{\partial h^i}{\partial x^j} \right|_{\theta=e} = \frac{\partial}{\partial x^i} r_k^i(x)$$

$$\therefore \partial_i l(x, e) = r_i^k \frac{\partial}{\partial x^k} l(x) + \frac{\partial}{\partial x^k} r_k^i(x)$$

これより定理の結果を得る。

証終

§4 α 接続について

推定と深い関係を持つ α 接続について、フィッシャー情報量と同じく、その受け子変換と、原点における値を求める事ができる。しかし、計算が煩雑なので、ここでは、 $\alpha=1$ の場合の変換について、結果のみ記す。

定理3 Γ_{ijk} の受け子変換は、次式で与えられる。

$$\hat{\Gamma}_{ijk}(\theta) = \hat{\Gamma}_{lmn}(e) B_i^l B_j^m B_k^n(\theta) + g_{ml}(e) C_{ij}^m(\theta) B_k^l(\theta)$$

$$\text{ここに } C_{ij}^m(\theta) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} k(\theta, \eta) \Big|_{\eta=\theta}$$

文献 Amari, S. (1985) Differential-geometrical methods in statistics. Lecture notes in statistics 28 Springer-Verlag.

Kariya, E. (1983) An invariance approach to estimation in a curved model. Discussion paper ser. No. 88 Hitotsubashi Univ.