数理解析研究所講究録
第623巻 1987年 67-78

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個のパラメタ a, b $(a>0)$ によって指定される変換 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。

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

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

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

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

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

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」となる分布によっ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1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

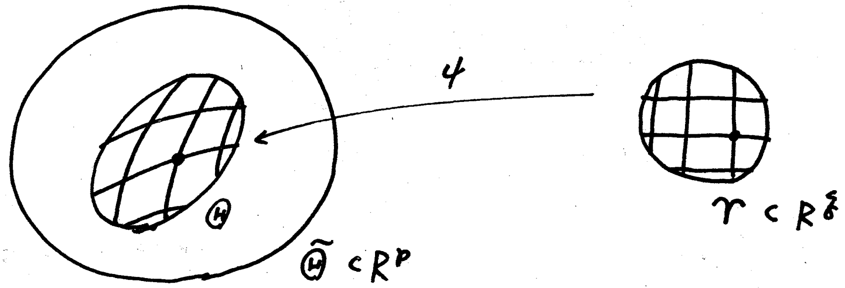
本節ではKariya (1983) に従って，変換群モデルの定式化を行なう。

標本空間を X，母数空間を Θ とし，$X \subset R^n$，$\Theta \subset R^p$ とする。$\theta \in \Theta$ によっ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，$\Theta$は，$\Upsilon \subset R^q$によってパラメトライズされているとする：

$$\Theta = \{\theta \in \tilde{\Theta} \mid \theta = \psi(\eta)\ \ \eta \in \Upsilon\}。$$

$q < p$であるとする。

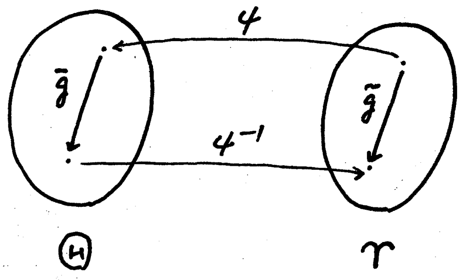


Xに作用する群Gを考える。$gP_\theta = P_\theta g^{-1}$と定めるとき，$gP(\tilde{\Theta}) = P(\tilde{\Theta})$が，すべての$g \in G$に関して成り立つとする。このとき，Gから，$\tilde{\Theta}$上に作用する群$\bar{G}$がinduceされる。つまり，$P_\theta g^{-1} = P_{\bar{g}\theta}$とする。すると，Gは$\bar{G}$に準同型である。

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。これを仮定する。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に作用するという。

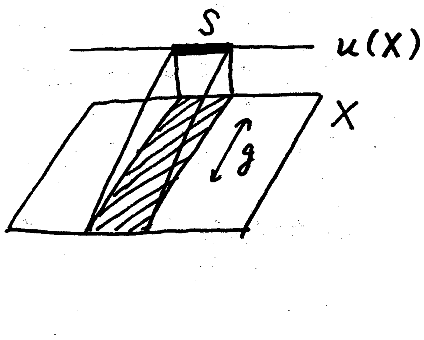
すると、ψにより、Υ上に作用する変換群$\tilde{G}$がinduceされる:

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



$\bar{\Theta}$は$\bar{G}$によって、いくつかのorbitに分けられる。そのひとつが、Θである。$\bar{\Theta}$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$\bar{\Theta}$上のmaximal invariant parameter、$u(x)$をX上のmaximal invariant parameterとすると、uの分布は、λのみによる。



なぜなら、θ_1とθ_2が同じλをあたえたとすると、このふたつは、ある$\bar{g} \in \bar{G}$によって、$\theta_1 = \bar{g}\theta_2$と結ばれているは

ずであり，$S \subset u(X)$ に対して，

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

となるからである。

これより，もしモデルをΘに限れば，$u(X)$は，ancillary であることが分る。

真のパラメタ$\eta \in \Upsilon$によって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$\tilde{g} \in \tilde{G}$に対し，$L(\eta, \hat{\eta}) = L(\tilde{g}\eta, \tilde{g}\hat{\eta})$が成立つことである。また，推定量$\hat{\eta}: X \to \Upsilon$が不変であるとは，任意の$g \in G$，$x \in X$に対し，$\hat{\eta}(gx) = \tilde{g}\hat{\eta}(x)$が成立つことである。

§3 不変なロスとフィッシャー情報量

以後，XとΘが，Gに対してひとつのorbitを成しており，この3者が互いに同型である場合を考える。群G，$\bar{G}$，$\tilde{G}$の区別はなく，すべて，X，Θと同一視される。

Θ中の1点，例えば，原点すなわち単位元と，他の点との間に，ロスL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ロスに拡張される。言い換えれば，1点の回りのロスは，変換群によって，すべての点の回りのロスに変換される。この変換はどのようなものなのかを，ロスの，各点の回りでの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：

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

ここで，右辺第1項は，i, j に関して和をとることを略した記法である。すると，θ の回りでは，

$$\begin{aligned} L(\theta, \theta+\xi) &= L(e, \theta^{-1}(\theta+\xi)) \\ &= L(e, e + B^i_j \xi^j + \cdots) \\ &= \frac{1}{2} L_{ij}(e) B^i_k B^j_l \xi^k \xi^l + \cdots, \end{aligned}$$

ここに，$B^j_i(\theta) = \left.\dfrac{\partial k^j(\theta, \eta)}{\partial \eta^i}\right|_{\eta=0}$, $k(\theta, \eta) = \theta^{-1}\eta$。

よって $L_{ij}(\theta) = L_{kl}(e) B^k_i(\theta) B^l_j(\theta)$ を得る。

定理 1　フィッシャー情報量を $g_{ij}(\theta)$ とすると，

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

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，フィッシャー情報量に一致するような不変なロス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，

$$p(x, \theta+\xi)\, dx = p(\theta^{-1}x, \theta^{-1}(\theta+\xi))\, d(\theta^{-1}x)$$
$$= p(y, e + B_j^i(\theta)\xi^j) A(\theta, x)\, dx$$

ここに，　$y = \theta^{-1}x$，　$A(\theta, x) = \det\left(\dfrac{\partial h^i(\theta, x)}{\partial x^j}\right)$

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

$$\partial_i l(x, \theta) = \frac{\partial}{\partial \xi^i} \log p(x, \theta+\xi) \Big|_{\xi=0}$$

$$= \frac{\partial}{\partial \xi^i} \log p(y, e + B_i \xi^i) \Big|_{\xi=0}$$

$$= \partial_j l(y, e) B_i^j(\theta)$$

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

よって,

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

証終

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。先ず 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,

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

$$
\begin{aligned}
&L((\theta^1,\theta^2),(\theta^1+d\theta^1,\theta^2+d\theta^2)) \\
&\quad = L\left(e,\left(1+\frac{d\theta^1}{\theta^1},\ \frac{d\theta^2}{\theta^1}\right)\right) \\
&\quad = \frac{1}{2}L_{ij}(e)\left(\frac{1}{\theta^1}\right)^2 d\theta^i d\theta^j + \cdots
\end{aligned}
$$

一方，フィッシャー情報量の方は，$l(x)=\log p(x)$，$y^i=(x^i-\theta^2)/\theta^1$ とおくと，

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

$$= -\log\theta^1 + l(y)$$

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

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

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

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

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

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。

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

ここに　$r_j^i(x) = \frac{\partial}{\partial \theta_j} k^i(e, x)$，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

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

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

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

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

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

上式の対数をとってθ^kで偏微分すると，

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

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

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，それの受ける変換と，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} のうける変換は，次式で与えられる。

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

ここに　$C_{ij}^m(\theta) = \dfrac{\partial^2}{\partial\eta_i\partial\eta_j} k(\theta,\eta)\Big|_{\eta=0}$

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.