数理解析研究所講究録
60巻 1968年 35-46

多変量解析での点予測の良さ

大阪市大 商　石 井 吾 郎

§1. adequate 統計量

この§では§2，3で必要とする範囲で *adequacy* の話を *Bahadur* (A. M. S. 1954) — *Skibinsky* (A. M. S. 1967) — 竹内 (経済学論叢) — 杉浦 (大阪市大修士論文) — 森本 (1968 数学会) より引用する。

Notations

$X \in \mathfrak{X}$	:	観測し得る確率変数
$Y \in \mathcal{Y}$	:	未来の確率変数
$Z = (X, Y) \in \mathcal{Z}$	:	
$(\mathcal{Z}, \mathfrak{A}, \lambda)$	:	$\mathfrak{A}$は$\mathcal{Z}$上のσ-*field*，λはσ-*finite*な測度
$\mathcal{P} = \{P_\theta ; \theta \in \Theta\}$	:	$\mathcal{Z}$上の確率分布族，$P_\theta(z)$はλに関する密度関数
$\mathfrak{B} \subset \mathfrak{A}$	:	$\mathfrak{A}$の部分σ-*field*でXにより導かれるもの．$\mathfrak{X}$上のσ-*field*を$\mathfrak{F}$

とするとき

$\mathfrak{B} = \{A\ ;\ A = X^{-1}(B) \in \mathfrak{A},\ B \in \mathfrak{F}\}$

$\mathfrak{C} \subset \mathfrak{A}$: $\mathfrak{A}$の部分σ-field で Yにより導かれるもの. $\mathcal{Y}$上のσ-field を$\mathfrak{G}$とすると

$\mathfrak{C} = \{A\ ;\ A = Y^{-1}(C) \in \mathfrak{A},\ C \in \mathfrak{G}\}$

$\mathfrak{B}_0 \subset \mathfrak{B}$: $\mathfrak{B}$の部分σ-field で Xの関数 T(x) により導かれるもの. T(x)の値域を$\mathcal{T}$, $\mathcal{T}$上のσ-field を$\mathfrak{D}$とし

$\mathfrak{B}_0 = \{A\ ;\ A = X^{-1}T^{-1}(D) \in \mathfrak{A},\ D \in \mathfrak{D}\}$

定義. $\mathfrak{B}_0$ が$\mathfrak{B}$に対し$\mathfrak{C}$, $\mathfrak{P}$に関して adequate である

$\Leftrightarrow$ ① sufficient

$\forall B \in \mathfrak{B}$, $\exists\, \varphi_B(z)$ ($\mathfrak{B}_0$-m.f.) $\forall \theta \in \Theta$ に対し

$\varphi_B(z) = E_\theta(I_B(z) \mid \mathfrak{B}_0)$, $[\mathfrak{A}, \mathfrak{P}]$

② transitive

$\forall C \in \mathfrak{C}$, $\forall \theta \in \Theta$ に対し

$E_\theta(I_C(z) \mid \mathfrak{B}) = E_\theta(I_C(z) \mid \mathfrak{B}_0)$, $[\mathfrak{A}, \mathfrak{P}]$

但し $I_A(z)$ は Aの指示関数, $[\mathfrak{A}, \mathfrak{P}]$ は a.e. の意味.

このとき $\mathfrak{B}_0$ adq ($\mathfrak{B}$; $\mathfrak{C}$, $\mathfrak{P}$)

と書き T(x) を adequate 統計量と呼ぶ.

このとき次の定理が成り立つ.

Ⅰ　$\mathcal{B} \perp\!\!\!\perp \mathcal{C}$ ならば　（$\perp\!\!\!\perp$ は独立の意味）

$\mathcal{B}_0\ adq(\mathcal{B}; \mathcal{C}, \mathcal{P}) \Leftrightarrow \mathcal{B}_0\ suff(\mathcal{B}, \mathcal{P})$

Ⅱ　$\mathcal{Z} = \mathfrak{X} \otimes \mathcal{Y}$, $\mathcal{A} = \{B' \otimes C'\}$, $\mathcal{B} = \{B' \otimes \mathcal{Y}\}$

$\mathcal{C} = \{\mathfrak{X} \otimes C'\}$

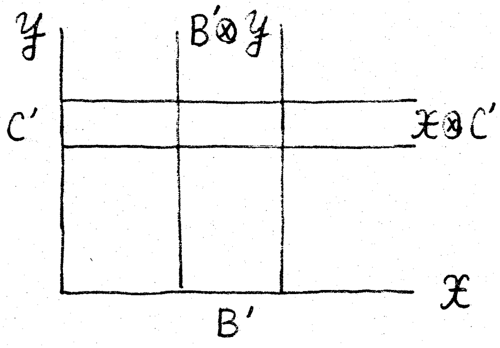
$\mathcal{P} \ll \lambda = \lambda_1 \otimes \lambda_2$

なるとき

$\mathcal{B}_0\ adq(\mathcal{B}; \mathcal{C}, \mathcal{P})$

$\Leftrightarrow$

$p_\theta(z) = g(z)\, h_\theta(z)$



但し g は $\mathcal{B}$ 可測，h は $\mathcal{B}_0 \vee \mathcal{C}$ 可測，すなわち

$p_\theta(z) = g(x)\, h_\theta(T(x), y)$

Ⅲ　Yに関する予測問題においてはXに基づく決定関数のリスクと同じかあるいはそれ以下のリスクを持つT(x)の関数である決定関数が存在する．

すなわちYに関する予測問題においては *adequate* 統計量 T(x) の関数になっている決定関数にかぎつてよいわけである．以下ではT(x) が観測されたと考えてT(x)を単にXと書く．又 $h_\theta(T(x), y)$ を $f(x, y; \theta)$ と書く．

定義．不偏予測量，$r(x)$ がYの不偏予測量である

$\Leftrightarrow$ $\forall \theta \in \Theta$ に対し $E_\theta^{\mathfrak{X}\mathcal{Y}}(Y - r(x)) = 0$

§2. 点予測における良さ

Notations

$$X = (X_1, X_2, \cdots, X_m) \in \mathcal{X} \subset R^m$$

$$Y = (Y_1, Y_2, \cdots, Y_n) \in \mathcal{Y} \subset R^n$$

$$\theta = (\theta_1, \theta_2, \cdots, \theta_k) \in \Theta \subset R^k$$

$$f(x, y; \theta) = p(x; \theta)\, g(y; \theta \mid x)$$

$$r(x) = (r_1(x), r_2(x), \cdots, r_n(x))$$

$$\mathbb{R} = \{ r(x) \,;\ E_\theta^{\mathcal{X}, \mathcal{Y}}(Y - r(x)) = 0, \quad \theta \in \Theta \}$$

: 不偏予測量全体

$$E_\theta^{\mathcal{Y}}(Y \mid X = x) = g(x, \theta) = (g_1(x, \theta), \cdots, g_n(x, \theta))$$

$$E_\theta^{\mathcal{Y}}(Y) = E_\theta^{\mathcal{X}}(g(x, \theta)) = h(\theta) = (h_1(\theta), \cdots, h_n(\theta))$$

$$V = V_\theta(Y, r(x)) = E_\theta^{\mathcal{X}\mathcal{Y}}(Y - r(x))'(Y - r(x))$$

$$V^1 = V_\theta^1(Y) = E_\theta^{\mathcal{X}\mathcal{Y}}(Y - g(x, \theta))'(Y - g(x, \theta))$$

$$V^2 = V_\theta^2(r(x)) = E_\theta^{\mathcal{X}}(g(x, \theta) - r(x))'(g(x, \theta) - r(x))$$

$$\phi_{i_1 \cdots i_k} = \phi_{i_1 \cdots i_k}(x, \theta) = \begin{cases} \dfrac{1}{p(x;\theta)} \dfrac{\partial^{i_1 + \cdots + i_k}}{\partial \theta_1^{i_1} \cdots \partial \theta_k^{i_k}}\, p(x; \theta) & \text{if } p > 0 \\ 0 & \text{if } p = 0 \end{cases}$$

$$i_1 + i_2 + \cdots + i_k \leqq \alpha$$

$$h_{i_1 \cdots i_k} = h_{i_1 \cdots i_k}(\theta) = \frac{\partial^{i_1 + \cdots + i_k}}{\partial \theta_1^{i_1} \cdots \partial \theta_k^{i_k}}\, h(\theta)' \quad \text{(列ベクトル)}$$

$$i_1 + i_2 + \cdots + i_k \leqq \alpha$$

$$\Phi = \Phi(x,\theta) = (\phi_{10\cdots0},\ \phi_{010\cdots0},\ \cdots,\ \phi_{0\cdots01},\ \phi_{20\cdots0},\ \cdots,\ \phi_{i_1\cdots i_k})$$

$i_1 + i_2 + \cdots + i_k \leqq \alpha$, Lーベクトル

$$H = H(\theta) = (h_{10\cdots0},\ h_{010\cdots0},\ \cdots,\ h_{0\cdots01},\ h_{20\cdots0},\ \cdots,\ h_{i_1\cdots i_k})$$

$i_1 + i_2 + \cdots + i_k \leqq \alpha$, $n \times L$ 行列

$$W = W(\theta) = E^{\mathfrak{X}}_{\theta}(\Phi(x,\theta)'\,\Phi(x,\theta)) \qquad L \times L \text{ 行列}$$

$$= \left(E^{\mathfrak{X}}_{\theta}(\phi_{i_1\cdots i_k}(x,\theta)\ \phi_{j_1\cdots j_k}(x,\theta))\right)$$

$$B = B(\theta) = E^{\mathfrak{X}}_{\theta}(\ g(x,\theta)'\,\Phi(x,\theta)) - H(\theta), \quad n \times L \text{行列}$$

Lemma 1. $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$: *sym. semi. pos. def.* $\exists\, M_{22}^{-1}$

$$\Rightarrow M_{11} \geqq M_{12} M_{22}^{-1} M_{21}$$

但し $A \geqq B$ は $A - B$ が *semi. pos. def.* の意味

Lemma 2 $M_{11} = M_{12} M_{22}^{-1} M_{21}$

$$\Leftrightarrow \operatorname{rank} M = \operatorname{rank} M_{22}$$

Lemma 3 確率変数 $X = (x_1, x_2, \cdots, x_m)'$ の共分散行列をCとする. $\operatorname{rank} C = \rho < m$

$$\Rightarrow \exists A : (m-\rho) \times m \text{ 行列},\ \operatorname{rank} A = m - \rho$$

$$A(X' - E(x)) = 0$$

仮定1. $p(x;\theta)$ は α 回連続的に微分可能

仮定2. θ の真の値を θ^0 とする. θ^0 の近傍 $\mathcal{V}(\theta^0) \subset \Theta$ と積分可能な関数 $G(x)$ が存在して, $\forall \theta \in \mathcal{V}(\theta^0)$

に対して $|f_{i_1 \cdots i_k}(x,\theta)\, p(x;\theta)| < G(x)$

仮定3. $W(\theta) > 0$: (pos. def.) $\forall \theta \in V(\theta^0)$

以上の準備の下で次の定理が成立する.

定理1. 仮定1,2,3が満たされているとき, すべての

$r(x) \in R$ に対して

(1) $$\nabla_\theta^2(r(x)) \geqq B(\theta) W^{-1}(\theta) B(\theta)' \ , \quad \forall \theta \in V(\theta^0)$$

$\geqq$ は実対称行列全体の空間の中で半正値行列全体をCとするときCは凸錐で, Cによる半順序の意味での不等式. (1)の右辺を予測問題における *Battacharya* の限界と呼ぶ.

定理2. 定理1で(1)の等号がθ^0で成り立つ

$$\Leftrightarrow \quad g(x,\theta^0)' - r(x)' = B(\theta^0) W^{-1}(\theta^0) \Phi(x,\theta^0)'$$

同時分布 $f(x, y; \theta)$ に従う確率変数Xを見てYを予測するとき Xの関数 $r(x)$ を作って

$$V(Y, r(x)) = E_\theta^{x y}(Y - r(x))'(Y - r(x))$$

が上の半順序の意味でなるべく小なる $r(x)$ を求めることを目標とする.

$$E_\theta^{xy}(Y - r(x))'(Y - r(x)) = E_\theta^{xy}(Y - g(x,\theta))'(Y - g(x,\theta)) + E_\theta^{x}(g(x,\theta) - r(x))'(g(x,\theta) - r(x))$$

$$V_\theta(Y, r(x)) = V_\theta'(Y) + V_\theta^2(r(x))$$

この式で ∇^2 はきまった行列で動かしようのない　　である。吾々の出来ることは $r(x)$ を色々と変えることにより ∇^2 を小にすることである。その ∇^2 を定理1で評価する。

定理1の証明

$$T(x) = (\underbrace{g(x,\theta) - r(x)}_{n},\ \underbrace{\Phi(x,\theta)}_{L})$$ とおく，$n+L$ ベクトル

$E_\theta^{\mathfrak{X}}\, T(x) = 0$ である。　∵仮定より $E_\theta^{\mathfrak{X}}(g(x,\theta) - r(x)) = 0$

$\int_{\mathfrak{X}} p(x;\theta)\, dx = 1$ 　と仮定2を用いて

$$\int_{\mathfrak{X}} \frac{\partial^{i_1+\cdots+i_k}}{\partial\theta_1^{i_1}\cdots\partial\theta_k^{i_k}}\, p(x;\theta)\, dx = \int_{\mathfrak{X}} \phi_{i_1\cdots i_k}(x,\theta)\, p(x;\theta)\, dx = 0$$

$$\mho(\theta) = E_\theta^{\mathfrak{X}}(T(x)'\,T(x)) \quad \text{とおく}$$

$$= \begin{array}{cc} \begin{bmatrix} E_\theta^{\mathfrak{X}}(g(x,\theta) - r(x))'(g(x,\theta) - r(x)), & E_\theta^{\mathfrak{X}}(g - r)'\Phi \\ E_\theta^{\mathfrak{X}}\Phi(x,\theta)'(g(x,\theta) - r(x)), & E_\theta^{\mathfrak{X}}\Phi'\Phi \end{bmatrix} & \begin{array}{c} n \\ L \end{array} \end{array}$$

$$= \begin{bmatrix} \nabla^2 & \overline{B} \\ \overline{B}' & \overline{W} \end{bmatrix} \geqq 0$$

Lemma 1 より

$$\nabla^2 \geqq \overline{B}\,\overline{W}^{-1}\,\overline{B}'$$

$$\int r(x)'\, p(x;\theta)\, dx = h(\theta)' \qquad (\text{列ベクトル})$$

の両辺をθにつき微分することにより

$$E_\theta^{\mathfrak{X}}\, r(x)'\, \phi_{i_1\cdots i_k}(x,\theta) = h_{i_1\cdots i_k}(\theta)$$

$$E_\theta^{\mathfrak{X}}\, r(x)'\, \Phi(x,\theta) = H(\theta)$$

以上により $\overline{B} = B$. よつて証明された.

定理2の証明.

定理1で等号が成立していると Lemma 2より

$\mathrm{rank}\ \Pi = \mathrm{rank}\ \overline{W} = L$, Lemma 3より

$\exists A$: $n \times (n+L)$ 行列, $\mathrm{rank}\ A = n$

$$A \begin{bmatrix} g(x,\theta)' - r(x)' \\ \Phi(x,\theta)' \end{bmatrix} = 0 \qquad (\because\ E\,T(x) = 0)$$

$A = [\underset{n}{A_1}, \underset{L}{A_2}]$ と分けて

$$A_1 [g(x,\theta)' - r(x)'] + A_2 \Phi(x,\theta)' = 0$$

$$A_1 E_\theta^{\mathfrak{X}} (g(x,\theta)' - r(x)') \Phi(x,\theta) + A_2 E_\theta^{\mathfrak{X}} \Phi(x,\theta)' \Phi(x,\theta) = 0$$

$$A_1 B + A_2 \overline{W} = 0 \quad , \quad A_2 = -A_1 B \overline{W}^{-1}$$

$$A_1 (g(x,\theta)' - r(x)' - B \overline{W}^{-1} \Phi(x,\theta)) = 0$$

ここで $\mathrm{rank}\ A_1 = n$ である. $\because$ $A = A_1 [I, -B\overline{W}^{-1}]$

$$n = \mathrm{rank}\ A \leqq \mathrm{rank}\ A_1 \leqq n$$

$$\therefore \quad g(x,\theta)' - r(x)' = B\,\overline{W}^{-1} \Phi(x,\theta)'$$

逆が成立することは明らか.

定理3 $\mathfrak{X} = [X^1, X^2]$ を adequate 統計量であるとする. $E_\theta^{y}(Y|\mathfrak{X}) = g(\mathfrak{X},\theta)$ が X^1 のみの関数であるとする, ($G(X^1, \theta)$ とおく). X^1を与えたときの X^2 の条件付分布が各 X^1 に対して Complete であり且つ $r(\mathfrak{X})$ は $G(X^1,\theta)$ の不偏推定量であるならば $r(\mathfrak{X})$ は Y の予測量として許容的である.

証明　$f(\mathbb{X},\theta)=f_1(X^1,\theta)\,f_2(X^2,\theta\,|\,X^1)$　とする．$f_2(X^2,\theta|X^1)$ に従う X^2 を観測して $G(X^1,\theta)$ を点推定する問題を考えると f_2 の complete 性よりそれは許容的である（Lehmann-Scheffé）

$$V^2 = E_\theta^{X^1} E_\theta^{X^2}\left[\left[(r(X^1,X^2)-G(X^1,\theta))\right]'\left[r(X^1,X^2)-G(X^1,\theta)\right]\,\middle|\,X^1\right]$$

より $r(\mathbb{X})$ が Y の予測量として許容的である．

§3.　正規分布に関する点予測

例1.　$X_i,\ i=1,2,\cdots,m$ ： $N(\mu,\sigma^2)$ に従う独立な m 個の観測し得る確率変数

$X_i,\ i=m+1,\cdots,m+n$ ：未来の確率変数

とするとき $X_i,\ i=1,2,\cdots,m$ を得て

$$Y_1=\frac{1}{m+n}\sum_{i=1}^{m+n}X_i$$

$$Y_2=\frac{1}{m+n-1}\sum_{i=1}^{m+n}(X_i-Y_1)^2$$

の点予測を行う．

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{m}X_i\ ,\qquad S=\sum_{i=1}^{m}(X_i-\bar{X})^2$$

$$\tilde{X}=\frac{1}{n}\sum_{i=m+1}^{m+n}X_i\ ,\qquad \tilde{S}=\sum_{i=m+1}^{m+n}(X_i-\tilde{X})^2$$

とおく．

adequate 統計量 $\mathbb{X}=(\bar{X},S)$　　（§1 の I より）

$$g_1(\mathbb{X},\theta)=E(Y_1\,|\,\mathbb{X})=\frac{1}{m+n}(m\bar{X}+n\mu)$$

$$g_2(\mathbb{X}, \theta) = E(Y_2 \mid \mathbb{X}) = \frac{1}{m+n-1}\left(S + (n-1)\sigma^2 + \frac{mn}{m+n}\left((\bar{x}-\mu)^2 + \frac{\sigma^2}{n}\right)\right)$$

$r_1(\mathbb{X}) = \bar{X}$, $r_2(\mathbb{X}) = \frac{1}{m-1} S$ を Y_1, Y_2 の予測量とする.

$$V = \begin{bmatrix} \frac{n}{(m+n)m}\sigma^2 & 0 \\ 0 & \frac{2n}{(m+n-1)^2(m-1)}\sigma^4 \end{bmatrix} = V^1 + V^2$$

$$V^1 = \begin{bmatrix} \frac{n}{(m+n)^2}\sigma^2 & 0 \\ 0 & \frac{2n((m+n)^2-n)}{(m+n-1)^2(m+n)^2}\sigma^4 \end{bmatrix}, \quad V^2 = \begin{bmatrix} \frac{n^2\sigma^2}{(m+n)^2 m} & 0 \\ 0 & \frac{2n^2((m+n)^2+m-1)\sigma^4}{(m+n-1)^2(m-1)\cdot(m+n)^2} \end{bmatrix}$$

$\Phi(x, \theta) = (\phi_{10}, \phi_{01}, \phi_{20})$ を考えることにより

$$W = \begin{bmatrix} \frac{m}{\sigma^2} & 0 & 0 \\ 0 & \frac{m}{2\sigma^4} & \frac{m}{\sigma^4} \\ 0 & \frac{m}{\sigma^4} & \frac{2m^2}{\sigma^4} \end{bmatrix}, \quad W^{-1} = \begin{bmatrix} \frac{\sigma^2}{m} & 0 & 0 \\ 0 & \frac{2\sigma^4}{m-1} & -\frac{\sigma^4}{m(m-1)} \\ 0 & -\frac{\sigma^4}{m(m-1)} & \frac{\sigma^4}{2m(m-1)} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{-n}{m+n} & 0 & 0 \\ 0 & \frac{-n}{m+n} & \frac{2mn}{(m+n-1)(m+n)} \end{bmatrix}$$

より $V^2 = B\,W^{-1}B'$ が成立することがわかる. 又

$$g(\mathbb{X}, \theta)' - r(\mathbb{X})' = \begin{bmatrix} \frac{n}{m+n}(\mu - \bar{x}) \\ \frac{1}{m+n-1}\left(\frac{mn}{m+n}(\bar{x}-\mu)^2 + (n-1+\frac{m}{m+n})\sigma^2 - \frac{n}{m-1} S\right) \end{bmatrix}$$

は $B\,W^{-1}\Phi'$ に一致する.

例2. $N(\mu,\ \sigma)$ での点予測

$$X = (x_1,\ x_2,\ \cdots,\ x_p,\ x_{p+1},\ \cdots,\ x_{p+q}) \in R^{p+q}$$
$$= (X^1,\ X^2)$$
$$X^1 = (x_1,\ x_2,\ \cdots,\ x_p)\ ,\quad X^2 = (x_{p+1},\ \cdots,\ x_{p+q})$$
$$X_i = (x_{i1},\ x_{i2},\ \cdots,\ x_{ip},\ x_{i,p+1},\ \cdots,\ x_{i,p+q})\ ,\ i=1,2,\cdots,m$$

$X^1_{m+1} = (x_{m+1,1},\ \cdots,\ x_{m+1,p})$ を X_{m+1} の初めの p 個の成分とする. X_i , $i=1,2,\cdots,m$ と X^1_{m+1} を観測して $Y=X^2_{m+1}$ を予測する問題を考える. このときの adequate 統計量は

$$\mathbb{X} = (\bar{X},\ S,\ X^1_{m+1})$$

である. 但し $\bar{X}$, S は X_i, $i=1,2,\cdots,m$ より作った平均及び $S = \sum X_i' X_i - m\bar{X}'\bar{X}$ である.

$$r(\mathbb{X}) = \bar{X}^2 + (X^1_{m+1} - \bar{X}^1)\, S_{11}^{-1} S_{12}$$

は不偏予測量であり, それが最適であることは定理3より示される.