数理解析研究所講究録
345巻 1979年 43-51

回帰行列に関する同時信頼推定

広大 綜合科学 塩谷 実

1. はじめに

同時信頼推定法は1950 年前後から始まつてきているようであるが，次に要約する事情に対処するためのものと思われる．(1) k個の母集団に関する多重比較を，同時に，与えられた一定の信頼係で推定すること。多変量の場合には，成分変量間の多重比較が母集団の間のものと同時に要求されている．よく知られているように，信頼領域(区間)推定は，よく仮説検定の役割をも荷なつている． (2) 検定しようとする仮説，推定しようとするパラメター関数は，本来，解明しようとしている目的に合わせて，予め設定されるべきものである．実験，観測によつて得られたデータを記述的に検討した後に設定されるのでは，検討中に入つてくる情報によつて，その確率空間は複雑に条件付けられて変形している筈で，このもとで構成された推論手法を使うことができなくなる．これは，実際的な立場から言えば誠に不便なことである．同時

信頼区間推定法は，近似的ではあるが，このような困難を克服する役割を果たす面を持っている．

本稿では，上記のことを説明するため，回帰行列に関する多重比較を考えそれらの同時信頼区間の構成を取り扱う．

2. k 個の回帰行列 $\underset{\sim}{\beta}^{(h)}$, $h=1,\cdots,k$ の多重比較

$\underset{\sim}{x}_i^{(h)}$ ($i=1,\cdots,N_h$; $h=1,\cdots,k$) を p 変量正規分布 $N_p(\underset{\sim}{\beta}^{(h)}\underset{\sim}{z}_i^{(h)},\underset{\sim}{\Sigma})$ からの観測ベクトルとする． $\underset{\sim}{\beta}^{(h)}: p\times q$, $\underset{\sim}{z}_i^{(h)}: q\times 1$, $\underset{\sim}{\Sigma}: p\times p$, >0 , $N_h \geq p+q$ とする．

$$\underset{\sim}{X}^{(h)}: p\times N_h = \{\underset{\sim}{x}_1^{(h)},\underset{\sim}{x}_2^{(h)},\cdots,\underset{\sim}{x}_{N_h}^{(h)}\}$$

$$\underset{\sim}{Z}^{(h)}: q\times N_h = \{\underset{\sim}{z}_1^{(h)},\underset{\sim}{z}_2^{(h)},\cdots,\underset{\sim}{z}_{N_h}^{(h)}\}$$

とおけば，

$$E(\underset{\sim}{X}^{(h)}) = \underset{\sim}{\beta}^{(h)}\underset{\sim}{Z}^{(h)}, \qquad h=1,\cdots,k \qquad (2.1)$$

である．簡単のために Full Rank すなわち $\mathrm{Rank}(\underset{\sim}{Z}^{(h)})=q$ の場合を考える．さらに

$$m = kq, \qquad N = N_1+N_2+\cdots+N_k$$

$$\underset{\sim}{X} : p\times N = \{\underset{\sim}{X}^{(1)},\underset{\sim}{X}^{(2)},\cdots,\underset{\sim}{X}^{(k)}\}$$

$$\underset{\sim}{\beta} : p\times m = \{\underset{\sim}{\beta}^{(1)},\underset{\sim}{\beta}^{(2)},\cdots,\underset{\sim}{\beta}^{(k)}\}$$

$$\underset{\sim}{Z} : m\times N = \begin{bmatrix} \underset{\sim}{Z}^{(1)} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{Z}^{(2)} & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{Z}^{(k)} \end{bmatrix}$$

とすれば

$$E(\underset{\sim}{X}) = \underset{\sim}{\beta}\,\underset{\sim}{Z} \tag{2.2}$$

とまとめて書くことができる．$\mathrm{Rank}(\underset{\sim}{Z}) = m$ で，$\underset{\sim}{\beta}$，$\underset{\sim}{\Sigma}$ の最尤推定量は

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}\underset{\sim}{Z}')(\underset{\sim}{Z}\underset{\sim}{Z}')^{-1} = \{\hat{\underset{\sim}{\beta}}^{(1)}, \hat{\underset{\sim}{\beta}}^{(2)}, \cdots, \hat{\underset{\sim}{\beta}}^{(k)}\}, \tag{2.3}$$

$$\hat{\underset{\sim}{\beta}}^{(h)} = (\underset{\sim}{X}^{(h)}\underset{\sim}{Z}^{(h)\prime})\underset{\sim}{A}_h^{-1}, \quad \underset{\sim}{A}_h = \underset{\sim}{Z}^{(h)}\underset{\sim}{Z}^{(h)\prime},$$

$$\begin{aligned} N\hat{\underset{\sim}{\Sigma}} &= (N-m)\underset{\sim}{S} = \underset{\sim}{X}\underset{\sim}{X}' - \hat{\underset{\sim}{\beta}}(\underset{\sim}{Z}\underset{\sim}{Z}')\hat{\underset{\sim}{\beta}}' \\ &= \sum_{h=1}^{k}\{\underset{\sim}{X}^{(h)}\underset{\sim}{X}^{(h)\prime} - \hat{\underset{\sim}{\beta}}^{(h)}\underset{\sim}{A}_h\hat{\underset{\sim}{\beta}}^{(h)\prime}\} = \sum_{h=1}^{k}(N_h - q)\underset{\sim}{S}^{(h)} \end{aligned} \tag{2.4}$$

で，これら最尤推定量の分布はよく知られているように，

(i) $\hat{\underset{\sim}{\beta}}^{(h)}$ の要素の同時分布は平均が $\underset{\sim}{\beta}^{(h)}$ の対応した要素である pq-変量正規分布で，その共分散行列は，$\Sigma\otimes\underset{\sim}{A}_h^{-1}$ である．

(ii) $(N_h - q)\underset{\sim}{S}^{(h)}$ は $\hat{\underset{\sim}{\beta}}^{(h)}$ と統計的に独立で，自由度 $(N_h - q)$ の Wishart 分布 $W_p(N_h - q, \underset{\sim}{\Sigma})$ にしたがう．

2.1 $\underset{\sim}{\beta}^{(h)}$, $h=1,\cdots,k$ の多重比較が事前に設定できない場合 ― すべての双一次結合のすべての比較 ―

実際に設定できる比較の個数は高々有限個であるのは確かであるが，これらを事前に決めることができない以上，信頼区間設定には予めすべての多重比較を考慮しておかねばなら

ないだろう．R_p でゼロ以外のすべての p-ベクトルの集合，C_k でゼロでない $c_1+\cdots+c_k=0$ を満たす k-ベクトル全体の集合を表わし，上記の意味で

$$\sum_{h=1}^{k} c_h \underline{a}' \beta^{(h)} \underline{b} = \underline{a}' \beta (\underline{b} \otimes I_k) \underline{c}, \quad {}^\forall \underline{a} \in R_p, {}^\forall \underline{b} \in R_q, {}^\forall \underline{c} \in C_k, \quad (2.5)$$

にたいする同時信頼係数 $1-\alpha$ の同時信頼区間の集合を構成することを試みることにしよう．この構成の方法は S. N. Roy (1953) の Union-Intersection Principle (U-I 原理) に基づいている．

まず $\underline{a}, \underline{b}, \underline{c}$ を止めて考えると，

$$\underline{a}'(\hat{\beta} - \beta)(\underline{b} \otimes I_k)\underline{c} = \sum_{h=1}^{k} c_h \underline{a}'(\hat{\beta}^{(h)} - \beta^{(h)})\underline{b} \qquad (2.6)$$

の分布は，平均 0，分散

$$(\underline{a}'\Sigma\underline{a})\left[\sum_{h=1}^{k} c_h^2 (\underline{b}' A_h^{-1} \underline{b})\right] = (\underline{a}'\Sigma\underline{a})\{\underline{c}'(\underline{b}\otimes I_k)'(ZZ')^{-1}(\underline{b}\otimes I_k)\underline{c}\}, \quad (2.7)$$

をもつ正規分布である．したがって

$$t_{\underline{a},\underline{b},\underline{c}} = \frac{\underline{a}'(\hat{\beta}-\beta)(\underline{b}\otimes I_k)\underline{c}}{(\underline{a}'S\underline{a})^{1/2}(\underline{c}'H_{\underline{b}}\underline{c})^{1/2}} \qquad (2.8)$$

は自由度 $n=N-m$ の Student の t 変量である．こゝに

$$H_{\underline{b}} : k \times k = (\underline{b}\otimes I_k)'(ZZ')^{-1}(\underline{b}\otimes I_k)$$

で，$\underline{b}'A_h^{-1}\underline{b}$，$h=1,\cdots,k$ を対角要素とする対角行列である．ゆえに $\underline{a}, \underline{b}, \underline{c}$ を止めたとき，自由度 n の t－分布に基き，信頼区間を作ることができる．これは most accurate unbiased なものである．すなわち信頼域は $[\,|t_{\underline{a},\underline{b},\underline{c}}| \leq g\,]$ あるいは $[\,t^2_{\underline{a},\underline{b},\underline{c}} \leq g^2\,]$ によつて表わされる．$g\,(>0)$ は当然 $\underline{a}, \underline{b}$,

$\underline{c}$に依存しておるべきであるのであるが，以後の操作が難しくなり複雑化して来て具体的に計算することができなくなるため，定数としている．S. N. Roy はこの場合を第1種と呼んでいる．$\underline{a}, \underline{b}, \underline{c}$ の値の組ごとにきめ細い信頼域を構成するのが望ましいのであるが．

さて求めるすべての $\underline{a}, \underline{b}, \underline{c}$ にたいする同時信頼域を求めることは，$[\, t^2_{\underline{a},\underline{b},\underline{c}} \leq g^2 \,]$ の $\underline{a}, \underline{b}, \underline{c}$ についての *intersection* をとることで可能である．容易に見られるように

$$\bigcap_{\underline{a}\in R_p,\, \underline{b}\in R_q,\, \underline{c}\in C_k} [\, t^2_{\underline{a},\underline{b},\underline{c}} \leq g^2 \,] \Leftrightarrow [\, \max_{\underline{a},\underline{b},\underline{c}} \{ t^2_{\underline{a},\underline{b},\underline{c}} \} \leq g^2 \,]$$
$$\Leftrightarrow [\, |t_{\underline{a},\underline{b},\underline{c}}| \leq g,\ \forall \underline{a}\in R_p,\ \forall \underline{b}\in R_q,\ \forall \underline{c}\in C_k \,]$$

なる同等関係があるから

$$\begin{aligned} &Pr\{ |t_{\underline{a},\underline{b},\underline{c}}| \leq g,\ \forall \underline{a}\in R_p,\ \forall \underline{b}\in R_q,\ \forall \underline{c}\in C_k \} \\ &\quad = Pr\{ \max_{\underline{a},\underline{b},\underline{c}} \{ t^2_{\underline{a},\underline{b},\underline{c}} \} \leq g^2 \} \end{aligned} \tag{2.9}$$

が得られる．ゆえに与えられた信頼係数 $1-\alpha$ にたいして，g^2 を計算することができれば，すなわち

$$Pr\{ \max(t^2_{\underline{a},\underline{b},\underline{c}}) \leq g^2 \} = 1-\alpha \tag{2.10}$$

により g^2 を決めれば，(2.9)の左辺の{ }の関係を区間の形に書きかえることにより，求める同時信頼区間の集合が得られる．しかしこのための演算は正確にはゆかず次のように不等式によって行われる：

$$\max_{\underset{\sim}{a},\underset{\sim}{b},\underset{\sim}{c}}(t^2_{a,b,c}) = \max_{\underset{\sim}{a},\underset{\sim}{b}}\left\{\sum_{k=1}^{k}\frac{\underset{\sim}{a}'(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})\underset{\sim}{b}\,\underset{\sim}{b}'(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})'\underset{\sim}{a}}{(\underset{\sim}{a}'\underset{\sim}{S}\underset{\sim}{a})(\underset{\sim}{b}'\underset{\sim}{A}_k^{-1}\underset{\sim}{b})}\right\}$$

$$\leq \sum_{k=1}^{k}\max_{\underset{\sim}{a},\underset{\sim}{b}}\left\{\frac{[\underset{\sim}{a}'(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})\underset{\sim}{b}]^2}{(\underset{\sim}{a}'\underset{\sim}{S}\underset{\sim}{a})(\underset{\sim}{b}'\underset{\sim}{A}_k^{-1}\underset{\sim}{b})}\right\} = n\sum_{k=1}^{k}\theta_{max\cdot k} \tag{2.11}$$

こゝに $\theta_{max\cdot k}$ は

$$|(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})A_k(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})' - \theta_k nS| = 0 \tag{2.12}$$

の最大根である。したがつて g^2 は

$$P_r\{|t_{\underset{\sim}{a},\underset{\sim}{b},\underset{\sim}{c}}|\leq g,\ {}^\forall\underset{\sim}{a}\in R_p,\ {}^\forall\underset{\sim}{b}\in R_q,\ {}^\forall\underset{\sim}{c}\in C_k\}$$
$$\geq P_r\{n\textstyle\sum_{k=1}^{k}\theta_{max\cdot k}\leq g^2\} = 1-\alpha \tag{2.13}$$

によつて求めなければならなくなる。これを満たす g^2 を $n\theta_{MS}(\alpha)$ と書けば，最終の同時信頼区間の集合は

$$\underset{\sim}{a}'\{\sum_{k=1}^{k}c_k\hat{\underset{\sim}{\beta}}^{(k)}\}\underset{\sim}{b} - \sqrt{E} \leq \underset{\sim}{a}'\{\sum_{k=1}^{k}c_k\underset{\sim}{\beta}^{(k)}\}\underset{\sim}{b} \leq \underset{\sim}{a}'\{\sum_{k=1}^{k}c_k\hat{\underset{\sim}{\beta}}^{(k)}\}\underset{\sim}{b} + \sqrt{E} \tag{2.14}$$

$$E = n\theta_{MS}(\alpha)(\underset{\sim}{a}'\underset{\sim}{S}\underset{\sim}{a})\{\underset{\sim}{b}'(\sum_{k=1}^{k}c_k^2\underset{\sim}{A}_k^{-1})\underset{\sim}{b}\} \tag{2.15}$$

$${}^\forall\underset{\sim}{a}\in R_p,\quad {}^\forall\underset{\sim}{b}\in R_q,\quad {}^\forall\underset{\sim}{c}\in C_k$$

となる。たゞし信頼係数は $\geq 1-\alpha$ となつている。

しかし実際には $\sum_{k=1}^{k}\theta_{max\cdot k}$ の分布は求まっていない。このため現段階では $\theta_{MS}(\alpha)$ を評価できないので実用にならないのであるが，さらに粗っぽくなることを覚悟すれば，逃げ道を考えることはできる。塩谷(1971)を参照。

2.2 多重比較が事前に与えられる場合.

おのおのの回帰行列にたいしてはすべての双一次結合が考えられ $\underset{\sim}{a}'\underset{\sim}{\beta}^{(k)}\underset{\sim}{b}$, $^\forall \underset{\sim}{a}\in R_p$, $^\forall \underset{\sim}{b}\in R_q$ とし，回帰間の比較が事前に m 個設定された場合を考えてみよう．すなわち

$$\sum_{k=1}^{k} c_{ik}\underset{\sim}{a}'\underset{\sim}{\beta}^{(k)}\underset{\sim}{b}\ , \quad ^\forall \underset{\sim}{a}\in R_p,\ ^\forall \underset{\sim}{b}\in R_q\ , \quad \underset{\sim}{c}_i'=(c_{i1},\cdots,c_{ik}),\ i=1,\cdots,m \tag{2.16}$$

にたいする同時信頼区間である．考え方としては前の場合と同様であるが同等関係が次のようになる。

$$\left[\ |t_{\underset{\sim}{a},\underset{\sim}{b},i}| = \frac{|\underset{\sim}{a}'\{\sum_{k=1}^{k} c_{ik}(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})\}\underset{\sim}{b}|}{(\underset{\sim}{a}'S\underset{\sim}{a})^{1/2}\{\underset{\sim}{b}'(\sum_{k=1}^{k} c_{ik}^2\underset{\sim}{A}_k^{-1})\underset{\sim}{b}\}^{1/2}} \le g,\ ^\forall \underset{\sim}{a}\in R_p,\ ^\forall \underset{\sim}{b}\in R_q,\ i=1,\cdots,m\right]$$

$$\begin{aligned} &\Longleftrightarrow \left[\ t^2_{\underset{\sim}{a},\underset{\sim}{b},i} \le g^2,\ ^\forall \underset{\sim}{a}\in R_p,\ ^\forall \underset{\sim}{b}\in R_q,\ i=1,\cdots,m\right] \\ &\Longleftrightarrow \left[\ \lambda_{max,i} \le g^2,\ i=1,\cdots,m\right] \\ &\Longleftrightarrow \left[\ \max_i\{\lambda_{max,i}\}\ \right] \end{aligned} \tag{2.17}$$

$\lambda_{max,i}$ は，i が固定されたときの，方程式

$$\left|\{\sum_{k=1}^{k} c_{ik}(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})\}\{\sum_{k=1}^{k} c_{ik}^2 A_k^{-1}\}^{-1}\{\sum_{k=1}^{k} c_{ik}(\hat{\underset{\sim}{\beta}}^{(k)}-\underset{\sim}{\beta}^{(k)})\}' - \lambda n S\right| = 0 \tag{2.18}$$

の最大根である．したがって

$$\lambda_{MM} = \max_i\{\lambda_{max\cdot i}\} \tag{2.19}$$

の分布の α 点を $\lambda_{MM}(\alpha)$ と表わせば，この場合の同時信頼区間の集合は

$$\underline{a}'\{\sum_{h=1}^{k} c_{ih}\hat{\underset{\sim}{\beta}}^{(h)}\}\underline{b} - \sqrt{F} \le \underline{a}'\{\sum_{h=1}^{k} c_{ih}\underset{\sim}{\beta}^{(h)}\} \le \underline{a}'\{\sum_{h=1}^{k} c_{ih}\hat{\underset{\sim}{\beta}}^{(h)}\}\underline{b} + \sqrt{F}$$

$$F = n\lambda_{MM}(\alpha)(\underline{a}'\underset{\sim}{S}\,\underline{a})\{\underline{b}'(\sum_{h=1}^{k} c_{ih}^2 A_h^{-1})\underline{b}\} \qquad (2.20)$$

$$\forall \underline{a} \in R_p, \quad \forall \underline{b} \in R_q, \quad i = 1, 2, \cdots, m$$

と書くことができる．

実際には $\lambda_{MM}(\alpha)$ の評価が大変難しく，(2.20)の代わりに，

$$P_r\{\lambda_{MM} \le g^2\} = P_r\{\max_i(\lambda_{max,i}) \le g^2\} \ge 1 - \sum_{i=1}^{m} P_r\{\lambda_{max,i} > g^2\} \qquad (2.21)$$

を利用することも考えられる．

3. λ_{MM} の分布について．

$$P\{\lambda_{MM} \le g^2\} = P_r\{\lambda_{max,1} \le g^2, \lambda_{max,2} \le g^2, \cdots, \lambda_{max,m} \le g^2\}$$
$$\equiv P(g^2, g^2, \cdots, g^2) \qquad (3.1)$$

とする．筆者はこの確率の評価について考えたことがある（Siotani & Geng (1974)）．漸近展開の形で

$$P(g^2, \cdots, g^2) = \prod_{i=1}^{k} P_i(g^2) - \frac{m^*}{4n}\Big[p(p+1)\prod_{i=1}^{k} P_i(g^2) + m^* H(2; g^2, \cdots, g^2)\Big]$$
$$+ \frac{m^*}{96n^2}\Big[\{-4p(2p^2+3p-1) + 3m^* p(p+1)(p^2+p+4)\}\prod_{i=1}^{k} P_i(g^2)$$
$$- 6m^{*2}(p^2+p+4)H(2; g^2, \cdots, g^2) + 16m^{*2}H(3; g^2, \cdots, g^2)$$
$$+ 3m^{*3}H(2,2; g^2, \cdots, g^2)\Big] + O(n^{-3}) \qquad (3.2)$$

こゝに $n=N-kq$, $m^*=kq$,

$$H(n;g^2,\cdots,g^2)=\left(\int_0^{qg^2 \underset{\sim}{I}_p}\right)^k tr\left(I_p-\frac{1}{m^*}\underset{\sim}{W}\right)^n \prod_{i=1}^{k} g_i(\underset{\sim}{U}_i)\prod_{i=1}^{k} d\underset{\sim}{U}_i \tag{3.3}$$

$$H(2,2;g^2,\cdots,g^2)=\left(\int_0^{qg^2 \underset{\sim}{I}_p}\right)^k tr\left(I_p-\frac{1}{m^*}\underset{\sim}{W}\right)^2 tr\left(I_p-\frac{1}{m^*}\underset{\sim}{W}\right)^2\cdot \cdot\prod_{i=1}^{k} g_i(\underset{\sim}{U}_i)\prod_{i=1}^{k} d\underset{\sim}{U}_i \tag{3.4}$$

$$g_i(\underset{\sim}{U})=\frac{1}{2^{qp/2}\Gamma_p(\frac{q}{2})}|\underset{\sim}{U}|^{\frac{1}{2}(q-p-1)}\exp\{-\frac{1}{2}tr\underset{\sim}{U}\}$$

$$\underset{\sim}{W}=\underset{\sim}{U}_1+\underset{\sim}{U}_2+\cdots+\underset{\sim}{U}_k$$

である。たゞし上の多重積分の数値計算をどうするか筆者には未解決である.