

(5) 多変量統計解析の理論と応用

岩崎学 (防衛大学校) : カテゴリカルデータのスペクトル解析	151
菅野長武 (岡山県立津山東高等学校) : 関連性の順位尺度について	153
綿森葉子 (広島大学理学部) : Langevin 分布に関する最近の話題	155
松田真一 (名古屋大学工学部)、吉村功 (名古屋大学工学部) : 正規化した判別関数の性質について	157
椎名洋 (東京大学経済学部)、竹村彰道 (東京大学経済学部) : 正規分布の分散共分散行列の推定についての若干の考察—固有値の順序構造を維持しない直交不変推定量の非許容性—	159
百武弘登 (熊本大学工学部) : 多変量二段階法による同時信頼区間の構成について	160
狩野裕 (大阪府立大学工学部) : Asymptotic properties of statistical inference based on Fisher-Consistent estimators in the analysis of covariance structures	162
神田隆至 (広島工業大学) : 共分散構造をもつ多変量成長曲線モデルにおける信頼領域	164
長尾寿夫 (大阪府立大学工学部)、M. S. Srivastava (トロント大学) : 共分散行列に関する検定の仮説の近傍での漸近展開と Bootstrap 近似	166
栗木哲 (東京大学工学部) : 分散行列のランクに関する片側検定	168
藤越康祝 (広島大学理学部)、向井畑哲 (広島大学理学部) : ある種の Cornish-Fisher 展開に対する誤差評価	170
柳井晴夫 (大学入試センター)、高根芳雄 (マッギル大学) : Canonical correlation analysis with linear constraints	172
谷口正信 (大阪大学基礎工学部) : Statistical analysis of curved probability densities	174
繁榎算男 (東京工業大学工学部) : 正準相関分析における母数のベイズ的推定	176
有田清三郎 (川崎医科大学)、米田正也 (川崎医科大学) : 糖尿病判定のための新しい指標の試み—糖負荷試験データの解析—	178
豊田秀樹 (大学入試センター) : 交差妥当化を利用した統計モデルの選択法	180
藤越康祝 (広島大学理学部) : 経時測定データ解析への応用をもつ拡張 MANOVA モデルについて	182
丘本正 (追手門学院大学経済学部) : 多変量解析諸手法のモデル再現性に関する実験例	184

ここで言うスペクトル解析は、次のように定式化される。y をデータからなる n 次元列ベクトルとすると、y は n 次元ユークリッド空間 R_n の点と見なすことができる。そして、 $\{\ell_1, \dots, \ell_n\}$ を R_n におけるある正規直交基底とし、y を第 j 基底ベクトル ℓ_j に直交射影した射影ベクトルの長さを λ_j とすると、Parseval の等式（ピタゴラスの定理）

$$y' y = \sum_{j=1}^n \lambda_j^2$$

が成り立つ（プライム（'）は、行列またはベクトルの転置を表わす）。したがって、いくつかの λ_j の絶対値が他に比べて大きければ、それらの基底ベクトルでデータ y の変動の大部分が「説明」され、多次元データの縮約が情報の損失をもたらすことなく実行できることになる。以下では、問題を不変にする群を選択し、その群の作用の下で不変なベクトルを決定する、という形の議論を行なう。

k 個の 2 値変量があり、各変量の取る値は 0 および 1 とコード化されているとする。このとき、k 個の 0 および 1 の順列からなる異なるパターンは 2^k 通りあるが、それらを各行に持つ $2^k \times k$ 次の行列を X とする。ここで、各行は辞書式順序に並んだものとする。

データは、例えば、k 個の 2 値変数の値が $(0, 1, 1, \dots, 0)$ である度数、あるいは、そのような実験条件の下で得られた測定値のように、X の各行ごとに与えられる。ここでは、より一般的に、データを X の各行上に定義された関数として捉えることにし、それらを y_i ($i = 1, \dots, 2^k$) と書く。

k 個の 2 値データの解析においては、

①各変数における 0 と 1 とのコードを入れ替える。

②各変数間の順番を入れ替える。

の二つの作用のいずれかを施しても問題の本質が不変であることが多い。報告では上の両方の場合を扱ったが、ここでは後者に関する結果だけを簡単に述べる。

k 次対称群 S_k の元 σ に対応する $k \times k$ の置換行列を $Q(\sigma)$ とし、X の列を σ で置換した行列 $X(\sigma)$ を $X(\sigma) = XQ(\sigma)'$ とする。そして

$$P(\sigma)X = X(\sigma) = XQ(\sigma)'$$

で $P(\sigma)$ を定義する。この表現の既約分解，そしてそれらの表現行列による不変部分空間を求める。

まず， X の行ベクトルにおいて， 1 の個数が ℓ であるもののみからなる $\binom{k}{\ell} \times k$ 行列を X_ℓ とする。明らかに X_ℓ は X の列ベクトルの入れ替えの操作によって不変である。したがって，それら不変部分空間に対応する部分表現 P/X_ℓ を B_ℓ とすると，表現行列 P は

$$P = B_0 \oplus B_1 \oplus \cdots \oplus B_k$$

と直和分解される。当然 $\dim(B_\ell) = \binom{k}{\ell}$ である。以下では $\ell \leq k/2$ とする。

B_ℓ の既約分解は次の定理によって与えられる。

定理 1

自然数 k の分割 $[k-m, m]$ に対応する S_k の既約表現を $\rho[k-m, m]$ とする。そのとき，上で定義した部分表現 B は

$$B_\ell = \rho[k] \oplus \rho[k-1, 1] \oplus \cdots \oplus \rho[k-\ell, \ell]$$

と既約分解される。また， $\rho[k-m, m]$ の次元は， $\binom{k}{m} - \binom{k}{m-1}$ である。

定理 2

$0 \leq m \leq \ell$ である m に対して， T_m を $\binom{k}{\ell} \times \binom{k}{m}$ 行列で，次の性質を持つものとする。自然数の集合 $\{1, 2, \dots, k\}$ から m 個重複を許さずに選ぶ組み合わせに適当に $1, 2, \dots, \binom{k}{m}$ と順番を付ける。 X の第 i 行において，上で付けた順番の第 j 番目の組み合わせの数字 $\{j_1, \dots, j_m\}$ に対応する要素がすべて 1 であるときに T_m の第 (i, j) 要素 $= 1$ とし，そうでないとき 0 とする。このとき，

$$M_0 = \iota \iota' / \binom{k}{\ell},$$

ただし， $\iota = (1, \dots, 1)'$ ，および $m = 1, \dots, \ell$ に対し

$$M_m = T_m (T_m' T_m)^{-1} T_m' - T_m (T_m' T_m)^{-1} T_m'$$

と置くと， M_m は $\rho[k-m, m]$ の不変部分空間への直交射影行列となる。また， $\dim(M_m) = \binom{k}{m} - \binom{k}{m-1}$ である。

関連性の順位尺度について

津山東高 菅野長武

$(X_i, Y_i), i = 1, 2, \dots, N$ が互いに独立にそれぞれ同一の *c.d.f.* $F(x)$ 及び $G(y)$ をもつとき、 X, Y の関連度を測る順位統計量として、Spearman, Kendall の統計量以外にも種々の統計量が提案されている。本稿では、 Y_i について大きさの順に並べたときの X_i の順位を R_i とするとき、Sugano & Yamamoto (1983) が部分相関検定のために提案した Smoothness rank statistic $G = \sum_{i=1}^{N-1} |R_{i+1} - R_i|$ による統計量

$$G_N = N^{-1/2}(N-2)^{-1} \cdot G \quad (1)$$

を serial dependence のための統計量の一つとして取り上げ一次自己回帰モデル

$$X_j = \alpha X_{j-1} + \sigma E_j, 1 \leq j \leq N \quad (2)$$

を考え、 $H_1: \alpha > 0$ に対する $H_0: \alpha = 0$ の検定を $f(x) = F'(x)$ が適当な正則条件を満たしているとき、mean successive difference (Kamart(1953)) による統計量

$$T_N = (N-1)^{-1/2} \sum_{i=1}^{N-1} |X_{i+1} - X_i| \quad (3)$$

による検定に対して、その順位類似型となっている G_N による検定に対する Pitman 効率について、次式を得た。

$$e(G_N, T_N) = \frac{[\{\int_{-\infty}^{\infty} (\int_{-\infty}^x f^2(t) dt) x dF(x)\} \{1 + \int_{-\infty}^{\infty} (2 \int_{-\infty}^x F(t) dt - x)^2 dF(x)\}]^2}{[\int_{-\infty}^{\infty} F(x)(1-F(x)) dx]^2 / 360} \quad (4)$$

又、Wald & Wolfwitz (1943) の時系列での標本積率統計量の順位類似型である順位積率統計量 $W = \sum_{i=1}^{N-1} R_i R_{i+1}$ よりなる統計量

$$W_N = N^{-1/2}(N+1)^{-2} \cdot W \quad (5)$$

による H_0 の検定の自己相関係数 S_N による検定に対する Pitman 効率は Aiyar (1981) の結果から、 $f(x)$ が両側指数分布関数であるとき、 $e(W_N, S_N) \doteq 1.266$ を得る。この時、(4) の値は $e(G_N, T_N) \gg 1$ である。この様に、順位統計量は標本統計量による検定に対して良好な漸近相対効率をもっている。なお、 W の漸近正規性はいえるが、 W は可なり非線形な統計量である。しかし、その順列分布は数表化されておらず、そのため漸近分布への接近が、どの程度速いか知られていない。又、正確な分散については、定義に基づいて計算した結果

$$\text{Var}(W) = \frac{(N+1)(5N^4 + 16N^3 - 9N^2 - 64N - 44)}{720} \quad (6)$$

を得た。Sperman (1904) は、Foot-rule と呼ばれている統計量

$$D = \sum_{i=1}^N |i - R_i| \quad (7)$$

の平均、分散を証明なしで与えている。また、Diaconis & Graham (1977) は D の漸近正規性を証明し、Kendall の τ 等との比較に関する諸性質を求めている。しかし、正確な分布は求まっていない。そこで Jogdeo (1968) より D_N と漸近同値な統計量は

$$\sum_{i=1}^N |F(X_i) - G(Y_i)| \quad (8)$$

であるが、以下にこの分布関数を与えておく。 H_2 の下では $i \neq j$ にたいして $|F(X_i) - G(Y_i)|$ と $|F(X_j) - G(Y_j)|$ は互いに独立であるから (8) の *c.d.f.* を $H^{*N}(t)$ とおくと、

$$H^{*N}(t) = 2^N \sum_{m=0}^N \sum_{l=0}^{N-m} (-1)^{N+m+l} \binom{N}{m} \binom{N-m}{l} (t-m)^{2N-l} \frac{\varepsilon(t-m)}{(2N-l)!}$$

$$, \varepsilon(t-m) = \begin{cases} 1(t > m) \\ 0(t \leq m) \end{cases} \quad (9)$$

で与えられるから、(9) により近似分布を求めるとよい。

Langevin 分布に関する最近の話題

広大理

綿森 葉子

1 はじめに

一昨年、Directional statistics に関する最近までの研究の総合報告が Jupp and Mardia によってだされた。現在、様々な通常の統計的手法を標本空間が一般の球面の場合に適用されており、特に non-parametric の方面の関心も bootstrap の手法の導入などにより高まってきた。

p 変数 Langevin 分布は、密度が $\{a_p(\kappa)\}^{-1} \exp(\kappa \mu' x)$ で与えられる $(p-1)$ 次元球面上の最も重要な分布のひとつである。ここで

$$a_p(\kappa) = (2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa) \kappa^{-\frac{p}{2}+1},$$

$I_\nu(\kappa)$ は位数 ν の第一種変形 Bessel 関数である。以下、 p 変数 Langevin 分布を $M_p(\mu, \kappa)$ であらわす。その密度の形から exact theory は困難であるが漸近理論は比較的容易である。そしてその漸近理論は Watson によるところが大きく 1980 年代前半までに多くの研究がなされた。最近はその精密化である高次漸近理論が展開されてきた。

以下、特に断らない限り $x \sim M_p(\mu, \kappa)$ また、 x_1, \dots, x_n を $M_p(\mu, \kappa)$ からの独立標本とする。 $\bar{x} = \frac{1}{n} \sum x_j$ とおき、 $R = n \|\bar{x}\|$ は標本合成長である。 $(p-1)$ 次元単位球面 S^{p-1} の s 次元部分空間 V の $p \times s$ 基底行列 B_0 を $B_0' B_0 = I$, なるようにとる。 x の共分散行列は

$$\Sigma = A_p'(\kappa) \mu \mu' + \frac{A_p(\kappa)}{\kappa} (I_p - \mu \mu')$$

で、 $A_p(\kappa)$ は $A_p(\kappa) = \frac{d}{d\kappa} \{\log a_p(\kappa)\}$ で定義される κ の関数。 Σ は正定値であることに注意する。さらに κ, μ の最尤推定量をそれぞれ $\hat{\kappa}, \hat{\mu}$ とすると、

$$A_p(\hat{\kappa}) = \|\bar{x}\|, \quad \hat{\mu} = \frac{\bar{x}}{\|\bar{x}\|}$$

である。以下、特に注意する必要がない限り $A_p(\kappa)$ を A とかく。

2 推定

$\hat{\kappa}$ は偏りのある推定量であることが Best and Fisher (1981) によって指摘。Schou (1978) は R の分布に基づく周辺尤度を最大にすることによって κ を推定することを提案した。これを $\tilde{\kappa}$ とかくと、

$$\begin{aligned} \tilde{\kappa} &= 0 \quad \cdots \quad R \leq n^{\frac{1}{2}} \\ n A_p(\tilde{\kappa}) &= R A_p(\tilde{\kappa} R) \quad \cdots \quad R > n^{\frac{1}{2}}. \end{aligned}$$

これらの期待値及び平均自乗誤差の n についての展開は、

$$\begin{aligned}
 E[\hat{\kappa}] &= \kappa + \frac{1}{n} \left(\frac{p-1}{2A'\kappa} - \frac{A^{(2)}}{2A'^2} \right) + O(n^{-2}), \\
 E[\tilde{\kappa}] &= \kappa - \frac{1}{n} \frac{A^{(2)}}{2A'^2} + O(n^{-2}), \\
 E[(\hat{\kappa} - \kappa)^2] &= \frac{1}{nA'} + \frac{1}{n^2} \left\{ \frac{(p-1)(p-3)}{4A'^2\kappa^2} - \frac{3A^{(2)}(p-1)}{2A'^3\kappa} \right. \\
 &\quad \left. + \frac{11A^{(2)2}}{4A'^4} + \frac{A^{(3)}}{4A'^3} \right\} + O(n^{-3}), \\
 E[(\tilde{\kappa} - \kappa)^2] &= \frac{1}{nA'} + \frac{1}{n^2} \left\{ \frac{p-1}{2A'^2\kappa^2} + \frac{11A^{(2)2}}{4A'^4} \right. \\
 &\quad \left. + \frac{A^{(3)}}{4A'^3} \right\} + O(n^{-3}),
 \end{aligned}$$

で与えられ、かなり似通っているが多少改良されているようだ。しかしながら、 κ が大きいときはその確率展開から、どちらも偏りがあることが示される。

3 検定

仮説 $H_{01} : \mu = \mu_0$ と $H_{02} : \mu \in V$ の検定に関して、2つの統計量 T_W と T_L の n と集中化母数 κ が大のときの漸近展開を導出した。ここで T_W は Watson によって提案された統計量、 T_L は尤度比統計量である。 H_{01} については、早川氏の研究がある。 H_{02} のかわりに $H_{02}^* : \mu = B_0\zeta, \|\zeta\| = 1$ を考察、尤度比統計量は何れの場合にも Bartlett 補正が可能であり Watson の統計量はそうでないことがわかった。さらにこれらの power を比較すると、2つの統計量が非常に近いことがいえる。

参考文献

- [1] JUPP, P. E. and MARDIA, K. V. (1989). A unified view of the theory of directional statistics, 1975-1988. *Internat. Statist. Rev.* 57, 261-294.
- [2] SCHOU, G. (1976). Estimation of the concentration parameter in von Mises-Fisher distributions. *Biometrika.* 65, 369-377.
- [3] WATSON, G. S. (1983). *Statistics on Spheres.* Wiley : New York.

正則化した判別関数の性質について

名古屋大学 工学部 松田 眞一
吉村 功

1 問題

$N_p(\mu_k, \Sigma_k)$ で規定される母集団 Π_k , $k = 1, 2, \dots, a$ と判別対象標本 Y_Q が存在する。 $Y_Q \in \Pi_k$ の事前確率 π_k と (μ_k, Σ_k) が既知ならば, Y_Q と各母集団との近さを判別関数

$$d(Y_Q, \Pi_k) = (Y_Q - \mu_k)' \Sigma_k^{-1} (Y_Q - \mu_k) + \log |\Sigma_k| - 2 \log \pi_k$$

で評価し, $d(Y_Q, \Pi_y) = \min_k d(Y_Q, \Pi_k)$ のとき $Y_Q \in \Pi_y$ と判定するのがよい。事前確率と母数が未知で, 所属母集団 Π_k が既知の参照標本

$$X = \{X_{ki}, i = 1, \dots, N_k, X_{ki} \in \Pi_k, k = 1, \dots, a, N_1 + \dots + N_a = n\}$$

(ただし, $(N_1, \dots, N_a) \sim MN(n, \pi_1, \dots, \pi_a)$) があるときは $\hat{\pi}_k = N_k/n$, $\hat{\mu}_k = \sum_{i=1}^{N_k} X_{ki}/N_k$ と適当な $\hat{\Sigma}_k$ を上の判別関数に代入して用いるのが自然である。

ここでは, $p > n$ のときも正則な $\hat{\Sigma}_k$ で性能のよいものを求めるという問題について検討する。

2 正則化の方針

以下当分の間, 添え字 k が不要なので省くことにする。

標本 $\{X_i, i = 1, 2, \dots, N, X_i \in N_p(\mu, \Sigma)\}$ に対する個別分散行列 $V = \sum_{i=1}^N (X_i - \hat{\mu})(X_i - \hat{\mu})' / N$ (ただし, $\hat{\mu} = \sum_{i=1}^N X_i / N$) の代わりに次式で与える正則化個別分散行列 V^* を考える。

$$V^* = LD^*L' = \gamma e^*I + (1 - \gamma)V$$

ただし, $D^* = \text{Diag}(d_1^*, d_2^*, \dots, d_p^*)$, $d_i^* = e_i + \gamma(e^* - e_i)$, e_i は V の第 i 固有値, $e^* = \sum_{i=1}^q e_i / p$, $q = \min(m, p)$, $m = N - 1$ で, L は対応する固有値の順に V の固有ベクトルを並べた行列, そして γ が正則化のパラメータ。

γ は円形性の仮説の検定 (sphericity test) の尤度比検定統計量を $N \leq p$ でも使えるように修正した

$$\text{LR}_S = \left\{ \prod_{i=1}^q e_i / \left(\sum_{i=1}^q e_i / q \right)^q \right\}^{N/2}$$

の単調増加関数で設定することが考えられる。(Siotani ら (1985) 参照)

3 LR_S の分布の近似式と γ の関数形

[定理 1] m または p が大きいとき, 円形性の仮説 $\Sigma = \sigma^2 I$ の下で統計量 $T = -2 \log \text{LR}_S$ に関して次の漸近展開が成り立つ。

$$\Pr\{T \leq x\} = G_f(c) + \frac{\gamma_2}{\nu^2} [G_{f+4}(c) - G_f(c)] + O(\nu^{-3})$$

ただし, $G_f(c) = \Pr\{\chi_f^2 \leq c\}$, $f = \frac{1}{2}q(q+1) - 1$, $c = mpx\rho/qN$,

$$\rho = 1 - \frac{1}{6pm}(2q^2 + q + 2), \quad \nu = \frac{mp}{q}\rho,$$

$$\gamma_2 = \frac{1}{288q^2}(q-2)(q-1)(q+2)(2q^3 + 6q^2 + 3q + 2)$$

[定理 2] 定理 1 の最後の項を無視できるとき次の近似式が成り立つ.

$$E\{T\} \approx \frac{fNq}{m\rho} \left\{ 1 + \frac{(q-2)(q^3 + 3q^2 + 1.5q + 1)}{18(m\rho)^2} \right\}$$

[系] 定理 2 で q の低次の項を省くと次式が得られる.

$$E\{T\} \approx \frac{fNq}{mp - q^2/3} \left\{ 1 + \frac{q^4}{18(mp - q^2/3)^2} \right\}$$

近似式は $N-1=p$ となる場合を除くと非常によいが, その場合でもよくなるようにパラメータ α_2 を導入する.

そして, 松田ら(1990)で用いた γ の関数形の代わりに定理 2 と系に基づいた次の 2 つの式を考える.

$$f_5(LR_S) = \frac{-2(m\rho)^3 \log LR_S}{fNq\{(m\rho)^2 + \alpha_2(q-2)(q^3 + 3q^2 + 1.5q + 1)/18\}}$$

$$f_6(LR_S) = \frac{-2(mp - q^2/3)^3 \log LR_S}{fNq\{(mp - q^2/3)^2 + \alpha_2 q^4/18\}}$$

4 正則化判別関数

最初の問題に戻って添え字 k を再び導入する.

母集団 Π_k からの参照標本に基づいて得た正則化個別分散行列を V_k^* と書く. これにより正則化併合分散行列 $V_0^* = \sum_{k=1}^a N_k V_k^*/n$ を定義する. これらを用いて

$$\hat{\Sigma}_k = \frac{\lambda n V_0^* + (1-\lambda)N_k V_k^*}{\lambda n + (1-\lambda)N_k}$$

として判別関数をつくる. λ は平均化のパラメータで松田ら(1990)と同じとする.

5 判別関数の性能

α_1, β_2 のいろいろな値に対して繰り返し数 100 回のモンテカルロ実験で平均誤判別率を調べてみる. ここで, 新たなパラメータ c を次のように導入した.

$$\gamma = \alpha_1 f(LR_S)^c$$

なお, シミュレーションの設定は松田ら(1990)と同じである.

6 結論

今回の改良によって関数形の一部が理論的に明確になった. パラメータ $\alpha = (.90, 2.0)$, $\beta = (.90, 20.00)$, $c = 2.5$ のときの判別結果は前回と同程度である.

正規分布の分散共分散行列の推定についての若干の考察—固有値の順序構造を維持しない直交不変推定量の非許容性—

東京大学 経済学部 椎名 洋, 竹村 彰通

正規分布の平均は既知の場合を考えるので, 十分統計量の点から \mathbf{W} が自由度 k の p 次元ウィッシュャート分布 $\mathbf{W}_p(k, \Sigma)$ に従う時に Σ を推定する話になる. 以下, 直交不変推定量のみを考える. \mathbf{W} の固有値を $0 \leq l_p \leq \dots \leq l_1$ として, \mathbf{W} を次の様にスペクトル分解すると,

$$\mathbf{W} = \mathbf{H}\mathbf{D}\mathbf{H}', \quad \mathbf{D} = \text{diag}(l_1, \dots, l_p)$$

直交不変推定量は次の様になる.

$$\widehat{\Sigma} = \mathbf{H}\Psi\mathbf{H}', \quad \Psi = \text{diag}(\psi_1(\mathbf{l}), \dots, \psi_p(\mathbf{l})), \quad \psi_i(\mathbf{l}) > 0 \quad (1 \leq i \leq p), \quad \mathbf{l} = (l_1, \dots, l_p).$$

損失関数としては, Stein の損失関数

$$L(\widehat{\Sigma}, \Sigma) = \text{tr}(\widehat{\Sigma}\Sigma^{-1}) - \log|\widehat{\Sigma}\Sigma^{-1}| - p$$

を使う. この損失関数は直交不変であるから, 母数は $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ ($\sigma_1^2 \geq \dots \geq \sigma_p^2$). としても, 一般性を失わない. 直交不変推定量としては, 不偏推定量 $\psi_i(\mathbf{l}) = l_i/k$ があるが, $E[l_1/k] \geq \sigma_1^2$, $E[l_p/k] \leq \sigma_p^2$ であることから, ψ_i をもっと中央に集めたほうがよいという考えのもとにいくつかの改良された推定量が現れた. 例えば Stein(1977), Dey&Srinivasan(1985) があるが, これらは ψ_i をお互いに近づけた結果, 固有根の順序構造 $0 \leq l_p \leq \dots \leq l_1$ に対応した推定量の順序, すなわち

$$\psi_1 \geq \psi_2 \geq \dots \geq \psi_p \quad \forall \mathbf{l}$$

を維持していない. 固有根の順序構造を維持しない推定量は, 必ず順序構造を維持する推定量に改善できることを示したのが, 今回の発表である. 我々は二通りの改良方法を証明したが, 一つは順序統計量を使うもので $\psi_i^o(\mathbf{l})$ を $(\psi_1(\mathbf{l}), \dots, \psi_p(\mathbf{l}))$ の中で i 番目に大きなものとすれば,

$$\widehat{\Sigma}^o = \mathbf{H}\Psi^o\mathbf{H}', \quad \Psi^o = \text{diag}(\psi_1^o(\mathbf{l}), \dots, \psi_p^o(\mathbf{l}))$$

であたえられる推定量 $\widehat{\Sigma}^o$ は, もとの推定量 $\widehat{\Sigma}$ をドミネイトする. もう一つの改良法は, Isotonic Regression を用いるもので, $\mathcal{F} = \{f = (f_1, \dots, f_p) \in \mathcal{R}^p \mid f_1 \geq \dots \geq f_p\}$ としたとき $\min_{f \in \mathcal{F}} \sum_{i=1}^p (f_i - \psi_i)^2 = \sum_{i=1}^p (\psi_i^m - \psi_i)^2$, で与えられる ψ_i^m ($i = 1, \dots, p$) を使った推定量

$$\widehat{\Sigma}^m = \mathbf{H}\Psi^m\mathbf{H}', \quad \Psi^m = \overset{\text{diag}}{(\psi_1^m, \dots, \psi_p^m)}$$

は, 同様に $\widehat{\Sigma}$ をドミネイトする. さらに, これらの系として $\psi_i = d_i l_i$ (但し d_i は定数) のかたちをとる推定量は, 必ず非許容的であることも示される.

多変量二段階法による同時信頼区間の構成について

熊本大学 工学部 百武弘登

2つの p 変量正規分布 $N_p(\mu_1, \Sigma_1), N_p(\mu_2, \Sigma_2)$ において, 平均ベクトルの差の任意の 1 次結合に関する長さ信頼係数を与えたときの同時信頼区間の構成について考える. つまり,

$$P\{a'y - d \leq a'\mu \leq a'y + d, \text{ for all } a, a'a = 1\} \geq 1 - \alpha \quad (1)$$

をみたすような方法を与えることである. ただし, $\mu = \mu_1 - \mu_2$, y は μ の推定量とし, $\alpha(0 < \alpha < 1)$ と $d > 0$ はあらかじめ与えられたものである. また, $\Sigma_1 > 0, \Sigma_2 > 0, \Sigma_1 \neq \Sigma_2$ とする. 共分散行列 Σ_1, Σ_2 が未知であるときは, 標本数を固定すると, これをみたす方法が存在しないことが知られていることから, ここでは, Stein (1945, AMS, 243-258) の二段階標本抽出法により解いていく.

一つの方法として, HYakutake and Siotani (1987, AJMMS, 89-111) は不偏共分散行列のトレースによる二段階法をもとにした Heteroscedastic 法により, この問題を解決している. ところが, この方法において必要とされる標本数があまりにも大きいことや, 統計量の計算が面倒であることなどの欠点がある. そこで, Healy (1956, AMS, 687-702) が提案したように固有値を用いることにより次のような二段階法を提案する.

まず, 各母集団から $n_0(n_0 > p)$ 個の標本をとり, 不偏共分散行列 S_1 と S_2 をそれぞれ計算する. ℓ_{12} を $S_1 + S_2$ の最大固有値, ℓ_i を $S_i, (i = 1, 2)$ の最大固有値とする. このとき,

$$\hat{N} = \max\{n_0, [c_2 \ell_{12}] + 1\}, \quad (2)$$

と定義する. ただし, $[b]$ は b を越えない最大整数で, c_2 は与えられた信頼係数 $1 - \alpha$ と区間の長さ d によって定まる正定数である. ここで, 各母集団から同じ大きさ $\hat{N} - n_0$ 個の観測値をとり, \hat{N} 個の標本による標本平均 \bar{x}_1, \bar{x}_2 を求め, $y = \bar{x}_1 - \bar{x}_2$ とする. (2)において c_2 が $(n_0 - 1)pf_\alpha / (n_0 - p)d^2$ より大きいなら (1)をみたすことが示せる. ただし, f_α は自由度 $(p, n_0 - p)$ の F 分布の上側 100α パーセント点である. これは, Ghosh (1975, JASA, 457-462) による方法を多変量の場合に拡張したものである. この方法は本質的には一標本の問題に帰着している.

また、別の方法として

$$N_i = \max\{n_0, [c_3 \sqrt{\ell_i}(\sqrt{\ell_1} + \sqrt{\ell_2})] + 1\}, i = 1, 2 \quad (3)$$

と定義してやれば、(1)をみたすことも示せる。ただし、 c_3 は h_α/d^2 より大きいもので、 h_α は $(y - \mu)'(S_1/N_1 + S_2/N_2)^{-1}(y - \mu)$ の分布の上側 100α パーセント点である。この分布は漸近的 ($n_0 \rightarrow \infty$) に自由度 p の χ^2 分布になることはわかるが、正確な分布またはその漸近展開を与えることは困難である。

(2) および (3) の二段階法は少なくとも標本数の観点では漸近的 ($d \rightarrow 0$) にある条件のもとで Heteroscedastic 法をかなり改善されていることが示せる。しかしながら、(2) と (3) の比較においては、標本数が共分散行列の構造に強く依存することから、その構造によって良さが変わってくる。

Asymptotic Properties of Statistical Inference Based on
Fisher-consistent Estimators in the Analysis of Covariance Structures

大阪府立大学工学部 狩野 裕

要 旨

共分散構造モデルにおいて最も一般的な推測法は、正規分布の仮定の下での最尤法やそれと漸近的に同値な一般化最小2乗法などであり、今まではそれらの漸近的性質やロバストネスの研究がなされてきた。ここでは、それらの結果を Fisher-consistent な推定量の族に拡張する。Fisher-consistent な推定量の族は有効推定量でない最小2乗推定量や非反復推定量を含んでおり、漸近的には有効でなくとも、推定の簡便性や不適解があまり出ないなどのよい性質を有する。

結 果

共分散構造モデル $\Sigma(\theta)$ において、 θ の推定量の族として次のものを考える：

$$\hat{\theta} = \hat{\theta}(S)$$

$$(1) \quad \hat{\theta}(\Sigma(\theta)) = \theta \text{ for all } \theta \in N(\theta_0)$$

ここで、 S は標本分散共分散行列、 $N(\theta_0)$ は θ_0 の近傍を表す (θ_0 は θ の真値である)。2番目の性質からこのような推定量を Fisher-consistent と呼ぶことにする。

モデル $\Sigma(\theta)$ の $\theta = \theta_0$ での微分(ヤコビアン)を

$$\Delta = \frac{\partial v(\Sigma(\theta_0))}{\partial \theta'}$$

と書くとき、(1)から

$$(2) \quad \frac{\partial \hat{\theta}}{\partial v(\Sigma)'} \Delta = I_n$$

が得られる。これは本論文での鍵となる関係式である。

いま、母集団分布に関して次の仮定をおく：

適当な $\theta_n \in \Theta$, $q \times q$ 行列 G , 定数 η が存在して

$$(3) \quad \sqrt{n}(v(S) - v(\Sigma(\theta_n))) \rightarrow N(0, \eta \Gamma_n + \Delta G \Delta')$$

ここで Γ_n は正規性の下での S の漸近分散である。(3)の仮定は $\eta = 1$, $\theta_n = \theta_0$, $G = 0$ とすれば正規性に帰着し, その他, 楕円分布を含めた種々の重要な非正規分布を含む. このことから, ある意味で, η と G は非正規性を表していると考えられる.

このとき, 次の結果が成り立つ:

$$\sqrt{n}(\hat{\theta} - \theta_n) \rightarrow N\left(0, \eta \frac{\partial \hat{\theta}}{\partial v(\Sigma)'} \Gamma_n \frac{\partial \hat{\theta}}{\partial v(\Sigma)} + G\right)$$

この結果はいろいろと示唆に富んでいる. つまり, 漸近分散において, 推定法の

違いは $\frac{\partial \hat{\theta}}{\partial v(\Sigma)'}$ にのみ反映し, 仮定する母集団分布の違いは η と G に反映す

る. このことから次にことがわかる:

$$(i) \quad \frac{\partial \hat{\theta}}{\partial v(\Sigma)'} \Gamma_n \frac{\partial \hat{\theta}}{\partial v(\Sigma)}$$

は正規分布の仮定の下での推定量の漸近分散であ

るから, 正規性の下での有効推定量は仮定(3)を満たすより広い分布族の下でも有効推定量になる (S に基づく推定量の中で).

(ii) 非正規分布の下では正規分布に基づく漸近分散の公式を η と G で修正する必要があるが, その修正は推定法によらず一定である.

Fisher-consistent 推定量を利用して, 適合度検定

$$H_0: \Sigma = \Sigma(\theta) \text{ v.s. } H_1: \Sigma > 0$$

を行うこともできる.

共分散構造をもつ多変量成長曲線モデルにおける信頼領域

広島工大

神田隆至

次の多変量成長曲線モデル： $Y_{N \times p} = A_{N \times k} \Xi_{k \times q} B_{q \times p} + \epsilon_{N \times p}$ を考える。ただし、 A, B は既知な計画行列で $\text{rank}(A) = k, \text{rank}(B) = q \leq p, \Xi$ は未知母数行列、誤差行列 $\epsilon = (\epsilon_1, \dots, \epsilon_N)'$ の各行は互いに独立で、それぞれ $N_p(0, \Sigma)$ に従うものとする。 $n = N - k, C$ を $m \times q$ の行列、 d を $k \times 1$ のベクトルとして、 $\eta = C \Xi' d$ とおくと、 η の信頼係数 $1 - \alpha + O(n^{-2})$ の信頼領域を分散行列 Σ が次の3つの共分散構造の場合についてしらべる。

- (i) Σ : 正定値行列
- (ii) $\Sigma = \sigma^2[(1 - \rho)I_p + \rho \mathbf{1}_p \mathbf{1}'_p]$: 一様共分散構造
($\sigma > 0, -(p-1)^{-1} < \rho < 1$ は未知)
- (iii) $\Sigma = \sigma^2(\rho^{|i-j|})$: 系列共分散構造($\sigma > 0, |\rho| < 1$ は未知)

また、簡単のために次のようにおく。 $\hat{\Xi}$ を Ξ の最尤推定量、 $V(\hat{\eta})$ を $\hat{\eta} = C \hat{\Xi}' d$ の分散、 $T = B \Sigma^{-1} B', \tilde{B} = B \Sigma^{-1/2}, P_A = A(A'A)^{-1} A', S = \frac{1}{n} Y'(I_p - P_A) Y, P_B = \tilde{B}' T^{-1} \tilde{B}, \tilde{P}_B = \tilde{B}' T^{-1} C(C' T^{-1} C)^{-1} C' T^{-1} \tilde{B}$.

Σ が正定値行列の場合には Grizzle and Allen (1969) により

$$V(\hat{\Xi}) = \frac{n-1}{n-p+q-1} (A'A)^{-1} \otimes (B \Sigma^{-1} B')^{-1}$$

であることが知られている。このとき、 $V(\hat{\eta})$ は exact に求まるが、 Σ が一様、系列共分散構造の場合には exact な表現は困難であり、以下において、これら3つの共分散構造に対して信頼領域の比較を同じ立場で行うために漸近展開を求めると

$$V(\hat{\eta}) = \text{tr} H \cdot \left[\left(1 + \frac{p-q}{n}\right) C T^{-1} C' + O(n^{-2}) \right]. \quad (1)$$

である。また、 Σ が一様、系列共分散構造の場合には

$$V(\hat{\eta}) = \text{tr} H \cdot C \left[T^{-1} + \frac{1}{n} \gamma_i \Psi(Z) + O(n^{-2}) \right] C' \quad (i = 1, 2) \quad (2)$$

である。ただし、 $H = d'(A'A)^{-1} d, \Psi(Z) = T^{-1} \tilde{B} Z (I_p - P_B) Z \tilde{B}' T^{-1}$ であり、一様共分散構造の場合は $\gamma_1 = 2/\{p(p-1)\rho^2 \sigma^4\}$ 、 $Z = \Sigma$ 系列共分散構造の場合には $\gamma_2 = p/\{(p-1)r\rho^2 \sigma^4\}$ 、 $Z = Q, r = p - (p-2)\rho^2, Q = \Sigma^{1/2}(I_p - \rho^2 D_1)\Sigma^{1/2}, D_1 = \text{diag}(0, 1, \dots, 1, 0)$ である。

$V(\hat{\eta})$ の推定量 $\hat{V}(\hat{\eta})$ として (1) には S , (2) には $\hat{\Xi}$ と σ^2, ρ の MLE $\hat{\sigma}^2, \hat{\rho}$ を代入したものを考える。いま、

$$F_n = (\hat{\eta} - \eta)' \hat{V}(\hat{\eta})^{-1} (\hat{\eta} - \eta)$$

とおき、 F_n の特性関数を求めてこれを反転すると、

$$P(F_n \leq x) = P(\chi_{km}^2 \leq x) + \frac{1}{4n} \sum_{j=0}^2 h_j P(\chi_{km+2j}^2 \leq x) + O(n^{-2})$$

をうる。ただし、 $\sum_{j=0}^2 h_j = 0$ で Σ が正定値の場合には、 $h_0 = m\{2(p-q) - m\}$ 、 $h_2 = m(m+2)$ であり、 Σ が一様共分散構造の場合には

$$\begin{aligned} h_0 &= \frac{m}{p(p-1)} [m(p-3) + 2(p-1)\{k(p-q) - 3\} + 4q - 2], \\ h_2 &= \frac{m}{p(p-1)} [(m+2)(p-1) + 2(m-1)] \end{aligned}$$

である。 Σ が系列共分散構造の場合には $h_0 = \frac{1}{2}\lambda_1 - 2\lambda_2$ 、 $h_2 = \frac{1}{2}\lambda_1$ である。ただし、

$$\begin{aligned} \lambda_1 &= \frac{m(m+2)(1+\rho^2)}{(p-1)\rho^2} - \frac{2(m+2)}{(p-1)\rho^2\sigma^2} \text{tr} \tilde{P}_B Q + \gamma_2 [(\text{tr} \tilde{P}_B Q)^2 + 2\text{tr}(\tilde{P}_B Q)^2] \\ \lambda_2 &= \gamma_2 [\text{tr}(\tilde{P}_B Q)^2 - \text{tr} \tilde{P}_B Q P_B Q + \frac{k}{2} \text{tr} \tilde{P}_B Q \cdot \text{tr} P_B Q] \\ &\quad + \frac{m}{2(p-1)r\rho^2} [k pq - 2\{k(p^2 - p - q) - (p+2)\}\rho^2 \\ &\quad + (p-2)\{k(2p - q - 2) - 2\}\rho^4] \\ &\quad + \frac{1}{2(p-1)r\rho^2\sigma^2} [-k pq + (p-2)(kq + 4)\rho^2] \text{tr} \tilde{P}_B Q \\ &\quad - \frac{km}{2(p-1)\rho^2\sigma^2} \text{tr} P_B Q + \frac{p(1-\rho^2)}{(p-1)r\sigma^2} \text{tr} \tilde{P}_B \tilde{D}_1, \quad \tilde{D}_1 = \Sigma^{1/2} D_1 \Sigma^{1/2} \end{aligned}$$

である。

次に $P(F_n \leq x_\alpha) = 1 - \alpha$ となる自由度 km の χ^2 分布の上側 $\alpha\%$ 点を u_α とし、 $1/n$ の order で考える。 $x_\alpha = \tilde{x}_\alpha + O(n^{-2})$ とおくと \tilde{x}_α は Cornish-Fisher の定理をもちいると、それぞれの分散構造に従って

$$\begin{aligned} \Sigma > 0 \quad &: \quad \tilde{x}_\alpha = u_\alpha \left[1 + \frac{1}{n} \cdot \frac{1}{2m} (u_\alpha + m + 2p - 2q) \right], \\ \Sigma \text{ が一様または系列共分散構造} &: \\ \tilde{x}_\alpha &= u_\alpha \left[1 + \frac{1}{n} \cdot \frac{1}{2km(km+2)} \{h_2 u_\alpha - (km+2)h_0\} \right] \end{aligned}$$

である。よって信頼係数 $1 - \alpha + O(n^{-2})$ をもつ η に対する信頼領域は

$$A_\alpha(\eta) = \{ \eta : (\hat{\eta} - \eta)' \hat{V}(\hat{\eta})^{-1} (\hat{\eta} - \eta) \leq \tilde{x}_\alpha \}$$

である。

また、よく知られたデータを用いてこれら3つの共分散構造に対しての信頼領域の比較をした。

共分散行列に関する検定の仮説の近傍での漸近展開と Bootstrap 近似

大阪府立大学工学部 長尾 壽夫
トロント大 M.S. SRIVASTAVA

X_1, \dots, X_N を $p \times 1$ 確率ベクトルとし、平均 μ 、共分散行列 Σ からの標本とする。 Ω を正定値行列全体とし、 ω をある正定値行列の部分集合とする。仮説 $H_0: \Sigma \in \omega$ を対立仮説 $H_1: \Sigma \in \Omega - \omega$ に対して検定を行うとする。ただし ω の次元を $p(p+1)/2 - r(f)$ とする。このとき、次の検定統計量の棄却域とし、

$$\{S|nf(S) \geq c\}$$

を考える。ただし c_α は有意水準 α に対する棄却点である。 $S = (s_{ij}) = \sum_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})'/n$ とする。ただし $n = N - 1$ 。このとき仮説の近傍での漸近展開を考える。一般性を失うことなく、 $E(S) = \Sigma_n = I + \frac{1}{\sqrt{n}}\theta$ とおく。 θ は対称行列であり、その次元は $r(f)$ である。このとき $f(S)$ について、

$$f(I) = 0, \quad \frac{\partial f(I)}{\partial s_{i,j}} = 0,$$

$$\left(\frac{\partial^2 f(I)}{\partial s_{i,j} \partial s_{k,l}} \right)$$

は半正定値行列で rank は $r(f)$ とする。統計量 $nf(S)$ の分布の漸近展開を $\frac{1}{\sqrt{n}}$ まで求めた。極限分布は $r(f)$ 個の独立な自由度 1 の non-central χ^2 分布の重み付きの和であらわされる。またこの漸近展開の結果を用いて、正規分布の下で、次の対立仮説の下での検定力を求め、また、Bootstrap 近似として、繰り返し数 500 として power を求めた。その結果は次のようになる。(1) 仮説 $\Sigma = \sigma^2 I$ を対立仮説 $N_3(0, (1-\rho)I + \rho 11')$ に対して検定を行う。

表 1

$N = 80$				
ρ	LRT	B-LRT	Nagao	B-Nagao
0.01	0.051	0.063(0.020)	0.054	0.056(0.019)
0.03	0.060	0.072(0.023)	0.063	0.066(0.023)
0.05	0.078	0.089(0.023)	0.082	0.084(0.022)
0.08	0.129	0.130(0.030)	0.134	0.123(0.028)
0.10	0.186	0.181(0.030)	0.189	0.182(0.029)

表 2

N = 100				
ρ	LRT	B-LRT	Nagao	B-Nagao
0.01	0.051	0.061(0.018)	0.054	0.055(0.017)
0.03	0.062	0.067(0.018)	0.065	0.063(0.017)
0.05	0.085	0.091(0.022)	0.089	0.087(0.021)
0.08	0.152	0.152(0.029)	0.157	0.150(0.028)
0.10	0.221	0.215(0.030)	0.228	0.220(0.028)

bootstrap 検定力は、N個の標本に対してそれぞれ500回 bootstrap の繰り返しを行い カッコはN回試行の分散を表す。次は仮説 $H_0 : \Sigma = \sigma^2 I$ にたいして対立仮説として $N_3(0, \text{diag}(a_1, 1, a_2))$ を考える。

N = 80					
a_1	a_2	LRT	B-LRT	Nagao	B-Nagao
0.9	0.9	0.064	0.074(0.022)	0.067	0.068(0.020)
0.8	0.9	0.098	0.103(0.025)	0.102	0.095(0.023)
0.8	0.8	0.122	0.124(0.029)	0.127	0.119(0.028)
0.9	1.1	0.088	0.098(0.024)	0.092	0.092(0.022)
0.8	1.1	0.157	0.168(0.028)	0.163	0.149(0.028)
0.8	1.2	0.236	0.242(0.030)	0.243	0.224(0.030)

N = 100					
a_1	a_2	LRT	B-LRT	Nagao	B-Nagao
0.9	0.9	0.067	0.074(0.020)	0.070	0.069(0.020)
0.8	0.9	0.112	0.116(0.026)	0.115	0.108(0.025)
0.8	0.8	0.143	0.149(0.029)	0.148	0.145(0.028)
0.9	1.1	0.099	0.104(0.021)	0.102	0.095(0.019)
0.8	1.1	0.189	0.200(0.028)	0.194	0.180(0.028)
0.8	1.2	0.292	0.291(0.031)	0.300	0.277(0.031)

REFERENCES

1. Anderson, T.W., "An introduction to Multivariate Statistical Analysis 2nd ed.," Wiley, New York, 1984.
2. Beran, R., *Simulated power functions*, Ann. Statist. 14 (1986), 151-173.
3. Chan, Y. M. and Srivastava, M. S., *Comparison of powers for the sphericity test using both the asymptotic distribution and the bootstrap*, Commun. Statist. 17 (1988), 671-690.
4. John, S., *The distribution of a statistic used for testing sphericity of normal distributions*, Biometrika 59 (1972), 169-173.
5. Nagao, H., *On some test criteria for covariance matrix*, Ann. Statist. 1 (1973), 700-709.
6. Srivastava, M. S. and Khatri, C. G., "An Introduction to Multivariate Statistics," North-Holland, New York, 1979.

分散行列のランクに関する片側検定

東京大学工学部 栗木 哲

分散行列の 2 標本問題

$$H \sim W_p(M, \Phi), \quad G \sim W_p(N, \Psi)$$

で、次の階層的な仮説

$$H_0: \Phi = \Psi; \quad H_0^{(r)}: \Phi \geq \Psi, \text{rank}(\Phi - \Psi) \leq r; \quad H_1: \Phi \geq \Psi; \quad H_2: \Phi, \Psi \text{ は任意}$$

に対する尤度比検定

- (i) T_{01} : 帰無仮説を H_0 , 対立仮説を $H_1 - H_0$ とする LRT
- (i') $T_{01}^{(r)}$: 帰無仮説を $H_0^{(r)}$, 対立仮説を $H_1 - H_0^{(r)}$ とする LRT
- (ii) T_{12} : 帰無仮説を H_1 , 対立仮説を $H_2 - H_1$ とする LRT

を考える。ただし、 \geq は Löwner ordering (n.n.d. の意味での半順序) である。

次の多変量 1 元配置変量模型

$$X_{ij} = \mu + V_i + U_{ij}, \quad i = 1, \dots, g, \quad j = 1, \dots, n,$$

$$V_i \sim N_p(0, \Theta), \quad U_{ij} \sim N_p(0, \Psi)$$

の群の一様性検定は、群内、群間分散行列を考えることにより検定 T_{01} に帰着する。また主効果行列 Θ のランクに関する検定は、 $T_{01}^{(r)}$ に帰着する。検定 T_{12} はモデルの適合度検定である。

Anderson 他 (1986, AS) は検定 T_{01} , $T_{01}^{(r)}$ の尤度比検定統計量を構成し、その帰無分布が漸近的にも χ^2 分布に従わないことを指摘している。また Sakata (1987, Comm. Statist.), Anderson (1989, JMA) は $p = 2$ の場合の漸近分布を求めている。さらに Amemiya 他 (1990, Biometrika) は、一般の p について大規模シミュレーションによりこの漸近分布のパーセント点を求めている。

尤度比検定統計量は

$$T_{01}: \quad -2 \log \Lambda_{01} = (M + N) \sum_{l_i > 1} \{ \log(\rho l_i + 1 - \rho) - \rho \log l_i \}$$

$$T_{01}^{(r)}: \quad -2 \log \Lambda_{01}^{(r)} = (M + N) \sum_{l_i > 1, i \geq r+1} \{ \log(\rho l_i + 1 - \rho) - \rho \log l_i \}$$

$$T_{12}: \quad -2 \log \Lambda_{12} = (M + N) \sum_{l_i < 1} \{ \log(\rho l_i + 1 - \rho) - \rho \log l_i \}$$

で与えられる。ただし $\rho = M/(M + N)$ で、 l_i は $(H/M)(G/N)^{-1}$ の固有根である。

最初に Scott and Saw(1984,JMA) の議論より, $\Lambda_{01}^{(\tau)}$ の l.f.d. は, Λ_{01} の分布に帰着することを注意しておく.

l_i の同時密度関数に含まれる linkage factor $\prod_{i < j} (l_i - l_j)$ を Vandermonde 行列式で表現し, Laplace 展開することにより次の定理を得る.

定理 1 H_0 の下で, T_{01}, T_{12} の同時極限分布は,

$$\begin{aligned} & \lim_{N \rightarrow \infty} P_{H_0}(-2 \log \Lambda_{01} < y, -2 \log \Lambda_{12} < z) \\ &= d(p) \sum_{r=0}^p \sum_{p-1 \geq q_1 > \dots > q_r \geq 0} U_r(q_1, \dots, q_r) U_{p-r}(\bar{q}_1, \dots, \bar{q}_{p-r}) G_Q(y) G_{\bar{Q}}(z) \end{aligned}$$

で与えられる. ただし, $\bar{q}_1, \dots, \bar{q}_{p-r}$ は $\{0, \dots, p-1\} - \{q_1, \dots, q_r\}$ の要素を大きい順にならべた列, $Q = \sum_{k=1}^r (q_k + 1)$, $G_\nu(\cdot)$ は $\chi^2(\nu)$ の分布関数, $c(p) = 2^{-p/2} \pi^{p(p-1)/4} / \Gamma_p(p/2)$,

$$U_k(q_1, \dots, q_k) = \int \dots \int_{b_1 > \dots > b_k > 0} e^{-\frac{1}{2}(b_1^2 + \dots + b_k^2)} \det \begin{pmatrix} b_1^{q_1} & \dots & b_1^{q_k} \\ \vdots & & \vdots \\ b_k^{q_1} & \dots & b_k^{q_k} \end{pmatrix} db_1 \dots db_k .$$

Pillai(1956,AMS) と類似の方法で部分積分を行い, 以下の漸化式を得る.

定理 2 $q_1 \geq 1$ のとき

$$\begin{aligned} U_k(q_1, \dots, q_k) &= (-1)^{k-1} U_{k-1}(q_2, \dots, q_k) I(q_1 = 1) + (q_1 - 1) U_k(q_1 - 2, q_2, \dots, q_k) \\ &+ 2 \sum_{j=2}^k (-1)^j \frac{1}{2^{\frac{1}{2}(q_1 + q_j)}} U_1(q_1 + q_j - 1) U_{k-1}(q_2, \dots, q_{j-1}, q_{j+1}, \dots, q_k), \\ U_1(q_1) &= I(q_1 = 1) + (q_1 - 1) U_1(q_1 - 1). \end{aligned}$$

ただし $U_1(0) = \sqrt{\pi/2}$.

定理 3

$$\sum_{r=0}^p \sum_{p-1 \geq q_1 > \dots > q_r \geq 0} (-1)^Q U_r(q_1, \dots, q_r) U_{p-r}(\bar{q}_1, \dots, \bar{q}_{p-r}) = 0.$$

正規模型 $X \sim N_k(\mu, \Sigma)$, Σ : 既知 で, $H: \mu = 0, K: \mu \in C$ (polyhedral convex cone) の尤度比検定統計量の帰無分布は, χ^2 分布の mixture $\bar{\chi}^2 = \sum_{j=0}^k p(j; k) \chi^2(j)$ で与えられることは良く知られている. 定理 1 はこれに対応している. 定理 1, 2 により任意の p についての漸近分布が求められる. また任意の polyhedral convex cone C についての $\sum_{j=0}^k (-1)^j p(j; k) = 0$ の成立は, Gauss-Bonnet の定理の系として得られることが知られている (例えば McMullen, 1975, Math. Proc. Camb. Phil. Soc., Shapiro 教授より). 定理 3 はこれに対応している.

さらに, 定理 1 で与えた分布関数の漸近展開式を与えることができる. また, Mehta(1960, Nuclear Physics) を拡張する形で, 定理 3 の他にもいくつかの U_k の間の関係式を与えることができる.

ある種の Cornish-Fisher 展開に対する誤差評価

広島大・理 藤越 康祝
 広島大・理 向井 畑 哲

パラメータ n (n は必ずしも標本の大きさでなくても良い) に依存する確率変数 X_n の分布関数を $F_n(x)$ とする。 F_n を n が大のときの極限分布関数 $G(x)$ のまわりで展開したときの漸近展開の典型的な形は

$$F_n(x) = G(x) + g(x) \left\{ \frac{1}{n} a_1(x) + \frac{1}{n^2} a_2(x) + \dots \right\}, \quad (1)$$

または, (1) において n を n^k で置き換えたものである。ここに, $g(x)$ は G の密度関数で, $a_1(x), a_2(x)$ 等は適当な多項式である。 (1) の右辺の第 k 項までを用いた近似 $G_{k,n}(x) = G(x) + g(x) \sum_{j=1}^{k-1} a_j(x) n^{-j}$ に対し, 適当な正則条件のもとでその近似に対する誤差 $R_{k,n}(x) = F_n(x) - G_{k,n}(x)$ が $R_{k,n}(x) = O(n^{-k})$ となることが知られている。一方, 応用上

$$F_n(x_n(u)) = G(u) \quad (2)$$

をみたす $x_n(u)$ を求めることが重要となる。 (1) を用いると, (2) をみたす $x_n(u)$ は形式的に

$$x_n(u) = u + \frac{1}{n} b_1(u) + \frac{1}{n^2} b_2(u) + \dots \quad (3)$$

と展開される。この展開式は次のようにして求められる。まず,

$$F_n(x) = G(u_n(x)) \quad (4)$$

をみたす $u_n(x)$ の展開式

$$u_n(x) = x + \frac{1}{n} \tilde{a}_1(x) + \frac{1}{n^2} \tilde{a}_2(x) + \dots \quad (5)$$

を求める。つまり, 関係式 (5) を x について解けば

(3) が得られる。展開式 (1) は Edgeworth 展開, (3), (5) は Cornish-Fisher 展開と呼ばれる。 $F_m(x)$ の分位点 $x_m(u)$ は (3) の右辺の才 k 項までを用いた漸近展開式 $x_{k,m}(u) = u + \sum_{j=1}^{k-1} b_j(u) m^{-j}$ で近似されるが, この近似に対する誤差評価の研究はあまり進んでいない。一つの理想的な評価法として

$$\underline{Q}_m(u) \leq x_m(u) \leq \bar{Q}_m(u) \quad (6)$$

かつ,

$$\bar{Q}_m(u) - \underline{Q}_m(u) \leq D_k / m^k \quad (7)$$

をみたす上界 $\bar{Q}_m(u)$, 下界 $\underline{Q}_m(u)$, および定数 D_k が与えられればよい。実際, Student の t -分布に関しては, (5) をみたす $u_m(x)$ に対する (6), (7) 型の評価式 (ただし, $k=1/2$) とともに,

$$\underline{w}_m(x) \leq u_m(x) \leq \bar{w}_m(x) \quad (8)$$

かつ

$$\bar{w}_m(x) - \underline{w}_m(x) \leq \frac{D_1}{m} |x| \quad (9)$$

をみたす上界 $\bar{w}_m(x)$, 下界 $\underline{w}_m(x)$, および定数 D_1 が与えられている。本報告では, t -分布および F -分布における $x_m(u)$ に対する (6) 型の評価式, F -分布における $u_m(x)$ に対する (8), (9) 型の評価式などを与えている。

Canonical Correlation Analysis with Linear Constraints

大学入試センター：柳井晴夫　マッギル大学：高根芳雄

1:はじめに：線形回帰モデル $y = X\beta + \varepsilon$ における β の最小二乗推定量を $\Lambda\beta = 0$ の条件で解く試みは (Seber, 1977) 等にみられるが、正準相関分析の制約解を求める研究は見当らない。そこで、本稿では、正準相関分析における、二組変数群に対応する重みベクトルに線形制約を課して正準相関分析を行なう方法を提案する。

2:方法： p 、 q 個の変数の測定値に関する n 個のデータを各成分とする $n \times p$, $n \times q$ 型行列を、 X 、 Y とする (ただし、各変数の平均値はゼロと仮定) このとき、ある既知の行列 $A (p \times r_A; r_A \leq p)$ および $C (q \times r_C; r_C \leq q)$ 、について、

$$(1) \quad A' a = 0, \quad C' b = 0 \quad .$$

という制約条件のもとで合成変数 $f = X a$ 、 $g = Y b$ との相関係数

$$(2) \quad r_{fg} = (X a, Y b) / (\|X a\| \|Y b\|) \quad .$$

を最大にする問題を考察する。ここで、(1)式で与えられる行列 A 、 C に関して、次の条件

$$(3) \quad A' B = 0 \quad \text{and} \quad C' D = 0 \quad ,$$

$$(4a) \quad p = \text{rank}(A; B) = \text{rank}(A) + \text{rank}(B),$$

$$(4b) \quad q = \text{rank}(C; D) = \text{rank}(C) + \text{rank}(D) \quad .$$

を満たす行列を C 、 D とする。この時、次の結果が得られる。

定理 1：(1)の条件下で、(2)を最大にするベクトル a 、 b は、次の4つの式のいずれかによって与えられる。

$$(5a) \quad P_{XB} Y b = \mu X a, \quad P_{YD} X a = \mu Y b \quad ,$$

$$(5b) \quad (P_X - P_{X* A}) Y b = \mu X a, \quad (P_Y - P_{Y* C}) X a = \mu Y b$$

$$(5c) \quad (P_{XB} P_{YD}) X a = \mu^2 X a, \quad P_{YD} X a = \mu Y b \quad ,$$

$$(5d) \quad (P_X - P_{X* A}) (P_Y - P_{Y* C}) X a = \mu^2 X a,$$

$$(P_Y - P_{Y* C}) X a = \mu Y b,$$

ただし、 P_X , P_{XB} , P_{X^*A} は、行列 X , XB , X^*A (ただし、 $X^*=X(X'X)^{-1}$)の列ベクトルで生成される部分空間上への直交射影行列、 P_Y , P_{YD} , P_{Y^*C} も同様に定義される直交射影行列である。なお、 $X'X$ 、 $Y'Y$ が正則でない場合にも P_X , P_{XB} , P_{X^*A} は定義することができる。

系1: $X'X$ および $Y'Y$ が正則の場合、(5a), (5b)は、次のように表現される。

$$(5e) P_{B(N)} X' Y b = \mu X' X a, \quad (P_{D(N)})' (Y' Y)^{-1} Y' X a = \mu b,$$

$$(5f) (I - P_{A(M)}) X' Y b = \mu X' X a,$$

$$(I - P_{C(N)})' (Y' Y)^{-1} Y' X a = \mu b, \quad \text{ただし、}$$

$$P_{A(M)} = A (A' (X' X)^{-1} A)^{-1} A' (X' X)^{-1},$$

$$P_{B(N)} = (X' X) B (B' X' X B)^{-1} B',$$

$$P_{C(M)} = C (C' (Y' Y)^{-1} C)^{-1} C' (Y' Y)^{-1},$$

$$P_{D(N)} = (Y' Y) D (D' Y' Y D)^{-1} D',$$

定理2: $\rho_i(X, Y)$ および $\rho_i(XB, YD)$ を、 X と Y および XB と、 YD の i 番目の正準相関係数とする。この時、次の関係が成立する。

$$(i) \rho_i(X, Y) \geq \rho_i(XB, YD) \text{ for } i=1, r, \text{ ただし、 } r=\text{rank}(XB:YD)$$

$$(ii) \sum_{i=1}^r (\rho_i(XB, YD))^2 \leq \min(\text{rank}(X), \text{rank}(Y))$$

$$(iii) \text{rank}(XB) = \text{rank}(X), \text{rank}(YD) = \text{rank}(Y) \text{ が成立すれば、}$$

$$\rho_i(X, Y) = \rho_i(XB, YD), \quad i=1, \dots, r.$$

$$(iv) P_{XB} P_{YD} = P_{YD} P_{XB} \iff \rho_i(XB, YD) = 1 \text{ or } 0.$$

上記の定理1の特別な場合として、線形回帰モデルにおける制約解、Correspondence Analysisにおける制約解(Bockenholt, 1991)、および Canonical Correspondence Analysis (Takane, et. al, 1991)などが導かれる。

文献: (1) Bockenholt, U. & Bockenholt, I., Canonical analysis of contingency tables with linear constraints, Psychometrika, 55, 1990. (in press); (2) Seber, G.A.F., Linear Regression Analysis, Wiley, 1977; (3) Takane, Y, Yanai, H. & Mayekawa S., Relationships among several methods of linearly constrained correspondence analysis (Psychometrika in press).

Statistical Analysis of Curved Probability Densities

BY MASANOBU TANIGUCHI

Osaka University

Amari(1982,1985) developed differential geometry of statistical inference for a curved exponential family $M = [q(x;u) | q(x;u) = \exp \{ \theta^i(u)x_i - \psi(\theta(u)) \}]$ embedded in the exponential family $S = [p(x;\theta) | p(x;\theta) = \exp \{ \theta^i x_i - \psi(\theta) \}]$. In this paper we extend his analysis to the case where the n -consecutive observations are (i) not necessarily identically distributed, (ii) not necessarily independent. Suppose that the joint probability density function is given by $p_n(\cdot; \theta)$. Our main purpose is to estimate an unknown parameter u of a family of curved probability densities defined by $M = \{p_n(\cdot; \theta(u)), \dim u < \dim \theta\}$ embedded in $S = \{p_n(\cdot; \theta), \theta \in \Theta\}$, and to develop the higher-order asymptotic theory. Here we assumed that the observations come from M .

In Section 2 the third-order Edgeworth expansion for a class of estimators of u is derived. We investigate higher-order asymptotic properties of the maximum likelihood estimator in Section 3. Amari(1982, 1985) showed that the bias-adjusted maximum likelihood estimator is third-order asymptotically efficient for his i.i.d. curved exponential family. It can be shown that this result is still valid for our generalized case. However, our Edgeworth expansion contains two terms A_1^{ab} and A_3^{ab} which vanish in the case of curved exponential family. The term A_3^{ab} elucidates the followings. Suppose that a likelihood function ℓ depends on a vector parameter (u, s) , $u \in \Theta_1, s \in \Theta_2$. When s is known we can construct the following two estimators \hat{u} and \bar{u} of u , which are defined by

$$\begin{aligned} \max_{(u,s) \in \Theta_1 \times \Theta_2} \ell(u, s) &= \ell(\hat{u}, s), \\ \max_{u \in \Theta_1} \ell(u, s) &= \ell(\bar{u}, s), \end{aligned}$$

respectively. If $A_3^{ab} = 0$, it is shown that the third-order asymptotics of \hat{u} and \bar{u} are identical, otherwise this result does not hold. Our results are general enough to be applicable to various fields (e.g., regression analysis, multivariate analysis and time series analysis). In Section 4, we give a few examples (e.g., a family of curved ARMA models, a family of curved regression models). Especially the family of curved ARMA models seems interesting because $A_3^{ab} \neq 0$ for this family. Finally we briefly mention a development of differential-geometrical methods in time series analysis. Amari(1983) discussed differential geometry of spectrum estimation, and showed that AR model is 1-flat and MA model is -1-flat. Amari(1987) developed the α -geometry for a parametric family of invertible linear systems. Ravishanker et al. (1990) characterized ARMA models as members of the curved exponential family. They gave a framework for the determination of parameter transformations for which (i) the asymptotic bias vanishes, (ii) the Fisher information matrix is a constant matrix etc.

正準相関分析における母数のベイズ的推定
東京工業大学 繁樹算男

1. 問題

本報告の目的は、正準相関分析モデルにおいて、本質的ではない標本変動に対して安定的な推定手法をベイズ的観点にたって開発することである。この方法は重回帰分析モデルに対するリッジ回帰の手法に対応している。

2. モデルと事前分布

2つの変数群 x ($p \times 1$), y ($q \times 1$) と、それぞれの一次結合を $u = x' \alpha$, $v = y' \beta$ とする。 u と v が 2 元正現分布に従っていると仮定し、 n 個の観測値を得たとしてデータ X ($n \times p$)、 Y ($n \times q$) を所与とする尤度は次のようになる。

$$L(\sigma_u^2, \sigma_v^2 | X, Y, \underline{\alpha}, \underline{\beta}) = (2\pi\sigma_u\sigma_v\sqrt{1-\rho^2})^{-n} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{\underline{\alpha}' X' X \underline{\alpha}}{\sigma_u^2} - 2\rho \frac{\underline{\alpha}' X' Y \underline{\beta}}{\sigma_u\sigma_v} + \frac{\underline{\beta}' Y' Y \underline{\beta}}{\sigma_v^2} \right] \right\} \quad (1)$$

また、 α 、 β 、 σ_u^2 、 σ_v^2 、 ρ の事前分布を次のように設定する。

$$\underline{\alpha} \sim N_p \left(0, \frac{(1-\rho^2)\sigma_u^2}{k_x} I \right), \quad (2)$$

$$\underline{\beta} \sim N_q \left(0, \frac{(1-\rho^2)\sigma_v^2}{k_y} I \right). \quad (3)$$

$$p(\sigma_u^2, \sigma_v^2, \rho) = \sigma_u^{-2} \sigma_v^{-2} (1-\rho^2)^{-1/2} \quad (4)$$

ここでは、 α 、 β の各要素が相互に交換可能であることを想定している。このとき、 ρ の事後分布のモードは、

$$\rho^* = \frac{\underline{\alpha}' X' Y \underline{\beta}}{\underline{\alpha}' (X' X + k_x I) \underline{\alpha} \underline{\beta}' (Y' Y + k_y I) \underline{\beta}} \quad (5)$$

となる。このベイズ的推定値を最大化する α と β は次式によって与えられる。

$$\{ Y' X (X' X + k_x I)^{-1} X' Y + \gamma (Y' Y + k_y I) \} \underline{\beta}^* = 0. \quad (6)$$

(6) 式の解を α と β の推定値とする方法をリッジ正準相関分析と呼ぶことにする。このようにベイズ的枠組みによって正準相関分析を把える場合、定数 k_x 、 k_y は、 $k_x = (1-\rho^2)\sigma_u^2 / \sigma_\alpha^2$ 、 $k_y = (1-\rho^2)\sigma_v^2 / \sigma_\beta^2$ として与えられるので、有意な定数であり、適当な値を事前データ分析によって決めることができる。たと

例えば、事前に通常の正準分布を行って得た $\hat{\alpha}$, $\hat{\beta}$, $\hat{\rho}$ を使って $\hat{\sigma}_{\alpha}^2$, $\hat{\sigma}_{\beta}^2$ を計算し k_x , k_y を得る。

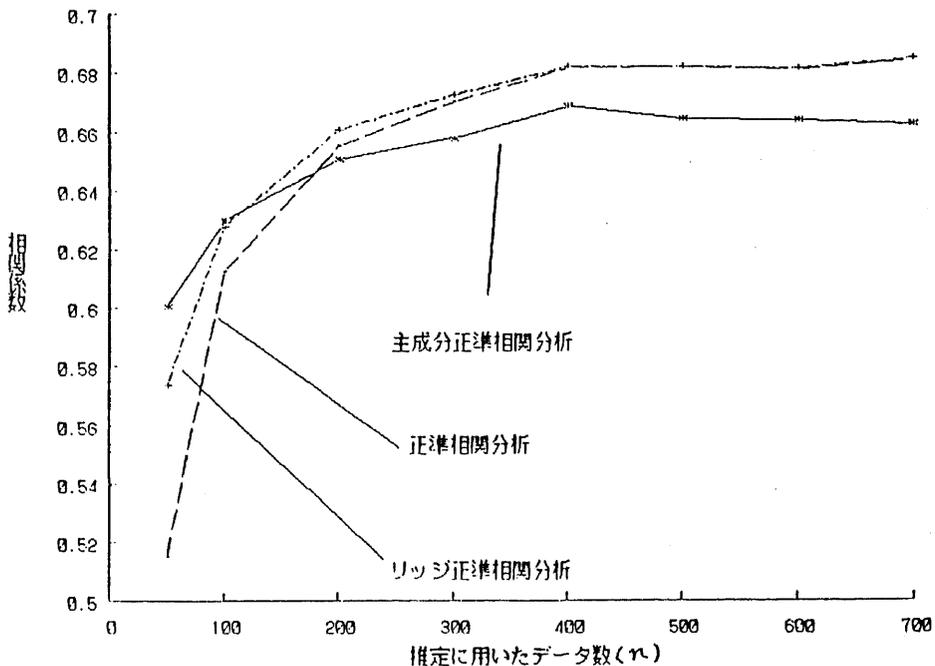
3. 主成分正準相関分析

標本変動に対し安定化する工夫として、上記のリッジ正準相関分析以外に主成分分析を使用する方法が考えられる。主成分回帰分析の考え方と同様に、小さい固有値に対応する成分は不安定であると考え主要な成分のみを用いることを考える。ただし、主成分正準相関分析の場合には、成分得点を計算するのではなく、正準方程式における共分散行列の推定値として主成分のみを用いる。

4. 実データへの適用

正準相関分析とリッジ正準相関分析主成分正準相関分析を実データに適用する。変数 x と変数 y は、 x (50m走、走り幅跳び、ハンドボール投げ、懸垂腕屈伸、1500m走)、 y は(身長、体重、胸囲、座高、握力、背筋力、伏臥上体反らし、立位体前屈、サイドステップ、垂直跳び、踏台昇降)である。

下図は、推定用データとして n 個のデータ ($n=50, 100, 200, \dots, 700$) を複数回とったときに、妥当性の検証用データ ($n=700$) における正準変数間の相関係数の平均を示すものである。推定用データの数 n が小さいときは、主成分正準相関分析やリッジ正準相関分析のほうが予測効率が低いことが示されている。ただし、 n が大になり、 S が Σ の推定値として有効であるときには、主成分正準相関分析の予測効率は低い。



糖尿病判定のための新しい指標の試み

－糖負荷試験データの解析－

川崎医科大学 数学教室 有田清三郎

内分泌内科 米田正也

人間集団の耐糖能は、連続的な変化量であると考えられている。しかも多様な経過を持って発症する糖尿病を反映し、単一の指標では評価しがたい多面性が存在する。特に75g糖負荷試験で境界型と判定される群では、ほぼ正常と言って良いようなものから、肥満、過食、その他のストレス等が加われば、糖尿病に移行する可能性を持った物まで、多種類の物が存在する。今回、WHO専門委員会の診断基準や、日本糖尿病学会の判定基準をふまえた上で、境界型群に対して糖尿病型の特徴をどれだけ強く有するかに基づいて、複数の指標によって連続的な評価値を与えることにより、患者への説明、経過観察、予後、予測等に有効となるのではないかと考え、糖負荷試験データの解析を行った。

1. はじめに

耐糖能評価で問題となるのは、糖負荷試験における耐糖能が十か一かといった明確な境界を持った変化量ではなく、非常に耐糖能が良いと思われる反応から非常に悪いと思われる反応まで連続的に変化して行くと考えられること、および形態学的診断における病理診断のような強力な検証手段がないことにある。このような対象に対して、WHO、および日本糖尿病学会の判定基準は妥当性のあるものであるが、ただ問題となるのは、実際的には糖尿病に近い多くの症例が2時間値で境界型と糖尿病型に分けられることになるが、たとえば198mg/dlが境界型となり204mg/dlが糖尿病型となるというように、境界値付近での限界分別力に問題がある点である。今回の我々の試みは、耐糖能の連続性、多用途性に対して、なんらかの尺度をつけようとするものである。

2. 新しい指標とデータ解析

数量化理論第II類の多変量解析を用いて糖尿病型、境界型、正常型の各3グループの判別要因をさぐった。説明変数としては、一応人工的な分別指標とし

での2時間値を除いた血糖ピーク値、血糖ピーク時間、空腹時血糖、 Δ IRI/ Δ BS30インスリンピーク値、インスリンピーク時間、およびBody Mass Indexの7項目を用い、糖尿病型グループと境界型グループ、正常型グループと境界型グループの2グループずつに分けた。解析の結果、正常型グループと境界型グループのグループ判別では①血糖のピーク値、②インスリンのピーク時間、③血糖のピーク時間が効いていること、糖尿病型と境界型のグループ判別には、①血糖のピーク値、②血糖のピーク時間、③インスリンのピーク値が効いていることがわかった。また、 Δ IRI/ Δ BS30が思ったほど影響していない。境界型と正常型の比較では境界型の正診率は95% (40/42)、境界型で正常型と診断されたものが5% (2/42)、正常型の正診率は100%であった。糖尿病型と境界型では、糖尿病型の正診率が87% (41/47)、境界型と診断されたものが13% (6/47)あった。また境界型の正診率は98% (43/44)で、糖尿病型と診断されたものが2% (1/44)あった。

文献

1. WHO Diabetes Mellitus. Report of a WHO study group. Technical Report Series 727, WHO. Geneva, 1985.
2. 伊藤知賀子:糖尿病に関連してみられる異常と診断のポイント. 治療, 70: 23, 1988.
3. 伊藤知賀子:新しいブドウ糖負荷試験判定法の評価. 内科, 53: 250, 1984.
4. 小坂樹徳、他:糖尿病の診断に関する委員会報告. 糖尿病, 25: 859, 1982.
5. Jarrett, R. J., Keen, H., Fuller, J. H., McCartney, M: Worsening to diabetes in men with impaired glucose tolerance ("borderline diabetes"). Diabetologia, 16: 25, 1979.
6. Arita, S., Yoneda, M. and Hori, Y.: Diagnostic System for Diabetes Mellitus based on the Response of Glucose Tolerance Test using a Fuzzy Inference. Proc. of the International Congress of Biomedical Fuzzy systems & the Third Annual Meeting of Biomedical Fuzzy Systems Assoc., 67-70, 1991.

豊田秀樹

目的

AICは、(1) 統計的仮説検定の「帰無仮説の採択」における消極的な判断とは異なり、候補に挙げられたモデルの中から最適なモデルを積極的に採用できる、(2) 標本数が増せば必ず棄却される(始めから偽であることが明白である) 帰無仮説を用いるという矛盾を犯さない、(3) 有意水準選択の任意性(あるいは5%と5.1%の間で判断が質的に変わるといふ不合理性)を免れる、(4) 一般モデルと特殊モデルを比較するとき必ずしも一般モデルの値が良いとは限らないなど、統計的検定理論の限界を克服した基準といえる(赤池, 1976)。

この様な多くの長所をもつAICも、対象となるモデルの型が(1) 最尤解を得る為の計算が非常に複雑であり、不適解や非収束が多い、(2) 最尤推定量の標準誤差が大きい、(3) 他に実用的な簡便解がある様な場合には、最尤解そのものが良い推定値(モデル)でない可能性が高いので役に立たないこともありうる。これ等はAICが最尤モデルしか評価できない為に生じた問題であり、最尤解以外の解をAICを用いたのと同じ様に評価できれば解決できる。本研究の目的は、交差妥当性の観点から統計モデルを評価することにより、候補に挙げられた複数のモデルの中から、AICと同じ様に、しかも最尤推定値に限らず最適なモデルを選択をすることが可能である統計量を導入することである。

方法

目的関数は尤度関数で表されているが、パラメータは最尤推定法で推定するとは限らない場合に以下の手続きに従って得られたモデルの良さを評価する。

[1] 同一母集団から独立に無作為に抽出した標本を2つ用意する。一方の標本をパラメータ推定の為の標本(標本x)とし、他方を妥当性評価の為の標本(標本y)とする。

[2] 評価されるべきモデルの型のパラメータを標本xを用いて推定する。推定法は任意である。モデルの型iの推定法jによるパラメータの推定値を $\hat{\theta}_{ij}$ とする。

[3] モデルの型iの $\hat{\theta}_{ij}$ によるyの対数確率

$$MSCV = \log P_i(y | \hat{\theta}_{ij}) \quad (1)$$

を「交差妥当化によるモデルセレクター (Model Selector by Cross Validation)」という意味で、MSCVと呼び、この値の最大になった $\hat{\theta}_{ij}$ を最適なモデルの推定値として採用する。Xによって得られた $\hat{\theta}_{ij}$ はYとは独立であるからXにとっては推定量であるがYにとってはモデルの型iの固定パラメータとなる。故に、どのような推定法によって得られた推定値であるかということには関係なく、(1)式は $\hat{\theta}_{ij}$ によって得られる分布の平均対数尤度の不偏推定量となる。

MSCVによるモデルの比較は、上述のままでは妥当性評価の為の標本をパラメータの推定に利用できない。そこで

[4] パラメータ推定の為の標本と妥当性評価の為の標本を合わせた標本に対してMSCVの値が

最大になったモデルの型と推定法を用いて再びパラメータを推定する。

以上の方法はモデルの型に依存しない一般的な方法であるが、本研究では特に共分散構造モデルに適用を試みる。共分散構造モデルは（１）最尤解を得る為の計算が複雑で、不適解や非収束が多く、（２）標本数が小さいときは解の標準誤差も大きいのでAICに適したモデルとはいえないが、（３）計算が容易で実用的な最小二乗解が存在するので本モデル選択法に適したモデルである。共分散構造モデルのモデル選択を行う為には、標本Xに標本共分散行列 S_x 、標本Yに標本共分散行列 S_y 、を用い、モデルセクターとしてはMSCVの差に影響しない部分を省略した

$$MSCV^* = \frac{-N}{2} \{ \log |\Sigma_x| + tr(\Sigma_x^{-1} S_y) \} \quad (2)$$

を用いる。ここで Σ_x は S_x による母共分散行列の推定値である。ただし相関構造分析を行う場合は（２）式の共分散行列を相関行列に置き換える。

適用例

本章で提案したモデル選択法を日本版WISC-Rの因子分析に適用する。分析の過程は以下の通りである。＜１＞パラメータ推定のための標本相関行列 R_x と妥当性評価のための標本相関行列 R_y は、この場合全年齢に渡る被験者を無作為に２等分し、それぞれ相関行列を計算した後に、年齢成分を除いた２つの標本偏相関行列を用いるのが正しい。しかし児玉-品川-茂木（１９８２）のデータを用いてそのような計算をすることができないので、正式な方法の代わりに、パラメータ推定のための標本相関行列 R_x には、年齢群別下位検査の６歳、８歳、１０歳、１２歳、１４歳、１６歳の内部相関の平均値を（合計被験者６００名）、妥当性評価のための標本相関行列 R_y には、年齢群別下位検査の７歳、９歳、１１歳、１３歳、１５歳の内部相関の平均値を用いる（合計被験者５００名）。ただし、この相関行列の算出法では R_x と R_y を計算した年齢群が等しくならないし、相関係数の平均値を用いることにも問題があるので、あくまでも間に合わせ的な方法である。＜２＞ R_y を用いて因子負荷行列を推定する。因子数は１因子から５因子まで求める。推定法は、最小残差法、相関係数の最大値によって共通性を推定する主因子法、そして通常は因子負荷行列の推定法としては用いないのであるが、比較するために主成分分析を加えた合計３つである。＜３＞ R_x を用いて（５－５）式によってMSCVを計算する。＜４＞相関係数の最大値によって共通性を推定した３因子の主因子解のMSCVが最大になったので、最適な解として採用する。

まず因子負荷行列の推定法についてMSCVの値を考察する。共通性を推定しない主成分分析の解は、因子分析の解としては他の２つの推定法より劣っていることが予想される。同一の因子数の解について比較すると主成分分析によって得られた解のMSCVの値が他の２つの推定法によって得られた解のMSCVの値よりも全て小さくしており予想と一致した。次に、因子数についてMSCVの値を考察する。上野（１９８２）は因子寄与の大きさを考察するという従来からの因子数の推定法の使用と、下位検査の性質という臨床的な見地から、３因子による解釈が妥当であることを主張しているが、MSCVによって最適と判断された因子数はこれと一致した。以上の結果は、MSCVを用いて構造モデルの選択をすることの実用可能性を示唆している。

経時測定データ解析への応用をもつ 拡張 MANOVA モデルについて

広島大学理学部 藤越康祝

変量 x が各個体についていくつかの時点（如
理条件）のもとで観測される場合のデータは経時
測定データとよばれる。観測が p 時点であって、
これらが N 個の個体について同一である場合のモ
デルは MANOVA モデルと密接に関係している。個
体 j の第 i 時点での測定を x_{ji} とし、

$$X = [x_1 \cdots x_N]', \quad x_j = (x_{j1} \cdots x_{jp})'$$

とする。このとき、MANOVA モデルの典型は

$$E(X) = A \Theta = A_1 \Theta_1 + A_2 \Theta_2$$

$$V(\text{vec}(X)) = \Sigma \otimes I_N$$

として与えられる。ここに、 $A: N \times q$ は個体間計
画行列、 $\Theta: R \times p$ は未知パラメータ行列である。
経時測定データの場合には、このモデルに対して、
さらに次の構造が仮定される。

(I) 平均構造

$$(I_1) \quad \Theta = \psi B$$

$B: r \times p$ は個体内計画行列

$\psi: R \times r$ は未知パラメータ行列

$$(I_2) \quad \textcircled{H}_1 = \textcircled{E}_1 B_1$$

$$(I_3) \quad \textcircled{H}_i = \textcircled{E}_i B_i, \quad i = 1, 2$$

(II) 分散構造

$$(II_1) \quad \Sigma > 0$$

(II₂) ランダム効果分散構造

$$\Sigma = B' \Delta B + \sigma^2 I_p, \quad \Delta \geq 0$$

(II₃) その他

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & & & \\ & \rho & & \\ & & \ddots & \\ \rho & & & 1 \end{bmatrix}, \quad \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{p-1} \\ \rho & \rho^2 & \dots & \rho^{p-1} \\ \vdots & \vdots & \ddots & \rho \\ \rho^{p-1} & \dots & \rho & 1 \end{bmatrix}$$

本報告においては、これらのモデルについての動機付け、推測に关する総合報告、並びに、平均構造(I₂)のもとでのランダム効果分散構造の適合性の検定法の提案を行った。

多変量解析諸手法の モデル再現性に関する実験例

追手門学院大学 丘本 正

1. 序 多変量解析の手法のうち、内部相関の解析を目的とし、いわゆる外的変量を持たないいくつかの手法が、与えられたモデルをどれほど再現するかを実験的に考察した。モデルとして平面図形を作り、これを計測することによって得られたデータに次の5つの手法を適用した：数量化法第3類、定性的正準相関分析、因子分析（セントロイド法と最小2乗法）、主成分分析。それぞれの手法によって得られた変数の散布図と個体の散布図が、もとの図形を再現する程度によって、手法の間の比較を行った。

2. 原図とデータ 原図は中心を共有する半径5cmの円と、半径10cmの4分円からなり、小円の周上に等間隔に16個の点をとって個体とみなし、大円の周上に等間隔に5つの点A, B, C, D, Eをとって観測点とする。各観測点から各個体に到る直線距離（単位mm）を変数 X_1, X_2, X_3, X_4, X_5 とすると、データは 5×16 行列となる。原図の対称性から、各変数の値は1個ずつずれて循環的であり、各変数について最小値=50, 最大値=150, 平均=106.50, 標準偏差=35.52である。

3. 解析 まず数量化法第3類を適用する。各変数 X を、不等式 $X < 90$, $90 \leq X < 130$, $130 \leq X$ によって3つのカテゴリー1, 2, 3に分けると、5つの変数によって15個のカテゴリーができる。解析の結果、カテゴリーの散布図を見ると、番号1, 4, 7, 10, 13はそれぞれ X_1, X_2, X_3, X_4, X_5 を最小にするカテゴリーであり、原図の5つの観測点の配置によく対応している。一方、個体の散布図では2つの個体7と15が中心に一致して、グラフはここで交差して奇妙にも8の字を描く。これでは原図を再現したとはいえない。

連続変数を離散化するときのカテゴリーの数（ k とする）の選択が結果に影響したのかも知れない。 $k=4$ として第3類を適用しても、個体の散布図はやはり8の字になった。しかし、 $k=2$ に減少し、 X の分点を100とすると、散布図はめで

たく円周状になった。

次に定性的正準相関分析を用いる。これは数量化法第3類を一般化して、ダミー変数 $\delta_i(j)$ の代わりに

e_{ij} : 個体 i がカテゴリー j に反応する程度

とする。解析結果について、変数の散布図は、たて軸方向に半分にちぢめたあと時計向きに 90° 回転すると原図に近い。一方、個体の散布図は逆向きに 90° 回転すると原図に似ないでもない。個体の16個の点の並び方が等間隔でないことが特徴的であるが、原図の再現性は第3類よりはよい。

3番目は因子分析（セントロイド法）である。個体の散布図は、原図に非常に近い。しかし、変数の散布図においては、第1軸の座標は妥当であるが、第2軸について観測点DとEの位置が不自然である。この原因は重心法の計算手順における独特な符号変換にあることがわかる。

第4の方法は因子分析（最小2乗法）であり、計算は部分ガウス・ニュートン法で行った。最尤法による解析もほぼ同じような結果になることと思われる。変数の散布図、個体の散布図ともに原図によく似ている。筆者は、部分ガウス・ニュートン法の反復を収束させず、1回だけの反復に止める1段推定量を推奨したことがあるが、この推定量に基づく結果は微小の変化にとどまる。

最後に扱った手法は主成分分析である。一般には、固有値解析を行う対象を標本相関行列とするか、標本分散行列とするかで答が違ってくるが、今の場合各変数の標本分散が同じ値をとるため、結果は一致する。変数の散布図は細長くなり、個体の散布図は平たい。縮尺を変えると、2つの散布図は原図に非常に近くなる。

4. 結論 　ここで取り上げた多変量解析の5つの手法を、円周データの再現性のよさという観点から見ると、よい方から順に

因子分析
(最小2乗法) } > 因子分析 > 定性的正準相関分析 > 数量化法第3類
主成分分析 } (重心法)

となる。この際、第3類は原データの情報を大幅に落としたために損をしていることを認める必要があり、その価値を一般に否定するものではない。また、上の結論は多分にデータに依存するものと思われ、今後の一層の検討が必要である。