



HOKKAIDO UNIVERSITY

Title	多群からなる変量に関する冗長性について
Author(s)	今井, 英幸; Imai, Hideyuki; 伊達, 惇 他
Citation	北海道大學工學部研究報告, 161, 23-27
Issue Date	1992-10-16
Doc URL	https://hdl.handle.net/2115/42320
Type	departmental bulletin paper
File Information	161_23-28.pdf



多群からなる変量に関する冗長性について

今 井 英 幸 伊 達 惇

(平成 4 年 6 月 30 日受理)

A Test for Redundancy of Some Sets of Variables

Hideyuki IMAI and Tsutomu DA-TE

(Received June 30, 1992)

Abstract

In canonical correlation analysis, Siotani(1957) and Fujikoshi(1982) obtained the likelihood ratio statistic and its asymptotic expansion under null hypothesis for redundancy of variables.

Horst(1961) extended ordinary canonical correlation analysis to those of two or more sets of variables.

In this paper, we obtain the likelihood ratio statistic and its asymptotic expansion under the null hypothesis for redundancy of several sets of variables.

1. は じ め に

2つの変数群の間の関係を調べる場合に用いられる手法に正準相関分析がある。正準相関分析において、変数群に含まれるいくつかの変数が群の相互関係に与える影響は無視できるほど小さいことがある。このような変数を冗長であるという。冗長な変数を選択する基準として、片方の変数群からだけ選択する基準が Siotani¹⁾により、両方の変数群から同時に選択する基準が Fujikoshi²⁾により得られており、それぞれの漸近分布も求められている。

一方、Horst³⁾は正準相関分析を拡張して3つ以上の変数群に対しても変数群相互の関連を分析する手法を提案した。本報告ではこのような多変数群に対する冗長性を定義し、冗長な変数を選択するための統計量とその漸近分布を求める。

2. 正準相関分析の多変数群への拡張

x_1, x_2 をそれぞれ p_1, p_2 次元確率変数で平均が $\mu_1, \mu_2, (x'_1, x'_2)'$ の分散共分散行列が

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

であるとき、制約条件 $a'\Sigma_{11}a = b'\Sigma_{22}b = 1$ のもとの $\rho = a'\Sigma_{12}b$ の最大値を第 1 正準相関係数、 ρ を最大にする p_1 次元ベクトル a と p_2 次元ベクトル b を用いた線形結合の組 $(a'x_1, b'x_2)$ を第 1 正準変量という。このような ρ は

$$(\sum_{12}\sum_{22}^{-1}\sum_{21}-\rho^2\sum_{11})a=0$$

$$(\sum_{21}\sum_{11}^{-1}\sum_{12}-\rho^2\sum_{22})b=0$$

すなわち一般固有値問題

$$|\sum_{12}\sum_{22}^{-1}\sum_{21}-\rho^2\sum_{11}|=0 \quad (1)$$

の解として得られる。いま(1)の解が大きさの順に

$$\rho_1^2 \geq \rho_2^2 \geq \dots \geq 0$$

であるとき ρ_i を第 i 正準相関係数という。 A_1, A_2 をそれぞれ p_1 次および p_2 次の正則行列として

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

のような変換をしたとき、正準相関係数 ρ_i^2 は不変である。この変換のことを内部正則線形変換という。

以下で、Horst³⁾による正準相関分析の多変数群への拡張について述べる。

$x_i, i=1, \dots, s$ をそれぞれ平均が μ_i であるような p_i 次元確率変数 ($p_i > 0$) で $(x_1', \dots, x_s)'$ の分散共分散行列が

$$\Sigma = \begin{bmatrix} \sum_{11} & \dots & \sum_{1s} \\ \vdots & & \vdots \\ \sum_{s1} & \dots & \sum_{ss} \end{bmatrix}$$

であるとする。このとき制約条件 $a_i' \sum_{ii} a_i = 1, i=1, \dots, s$ のもとで $Q = \sum_{ij} a_i' \sum_{ij} a_j = \sum_{i=1}^s \lambda_i$ (ただし

$\lambda_i = \sum_{j=1}^s a_i' \sum_{ij} a_j$) を最大にするような p_i 次元ベクトル $a_i, i=1, \dots, s$ と最大値 Q を求めることが多変数群への拡張である。このような $\lambda_i, a_i, i=1, \dots, s$ は

$$\begin{bmatrix} \sum_{11} & \dots & \sum_{1s} \\ \vdots & & \vdots \\ \sum_{s1} & \dots & \sum_{ss} \end{bmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_s \end{pmatrix} = \begin{bmatrix} \lambda_1 \sum_{11} & & \\ & \dots & \\ & & \lambda_s \sum_{ss} \end{bmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_s \end{pmatrix}$$

の解として得られる。この場合は正準相関分析とは異なり、固有値問題にはならないが $\lambda_i, i=1, \dots, s$ は内部正則線形変換により不変である。

3. 変数の冗長性

正準相関分析およびその多変数群への拡張した分析法では最大化基準（正準相関分析では相関係数）が内部正則線形変換によって不変であった。変数群が2つの場合、内部正則線形変換によって不変な量は、各々の変数群相互の正準相関係数にだけ依存する。したがって各変数群相互の関係に着目する場合には正準相関係数の変化だけを考慮すればよい。

変数の冗長性に対する帰無仮説を得るために u_i, v_i をそれぞれ r_i, q_i 次元確率変数 ($r_i, q_i > 0, r_i + q_i = p_i$) として $x_i = (u_i', v_i)'$ のように分割し、これらに対応して μ_i, \sum_{ij} を

$$\mu_i = \begin{pmatrix} \mu_{u_i} \\ \mu_{v_i} \end{pmatrix}, \quad \sum_{ij} = \begin{bmatrix} \sum_{u_i u_j} & \sum_{u_i v_j} \\ \sum_{v_i u_j} & \sum_{v_i v_j} \end{bmatrix} \quad (2)$$

とする。上で述べたことから、各 i, j について x_i と x_j との0でない正準相関係数と u_i と u_j との0でない正準相関係数がすべて等しいとき、 $(u_1', \dots, u_s)'$ は全変数 $(x_1', \dots, x_s)'$ のすべての情報を持っているといえる。あるいは $(v_1', \dots, v_s)'$ は $(u_1', \dots, u_s)'$ の関係について追加する情報を持っていないともいえる。 x_i と x_j , u_i と u_j の正準相関係数はそれぞれ $\sum_{ij} \sum_{jj}^{-1} \sum_{ji} \sum_{ii}^{-1}$, $\sum_{u_i u_i} \sum_{u_i u_i}^{-1} \sum_{u_i v_i} \sum_{u_i v_i}^{-1}$ の固有値であることから、 $(v_1', \dots, v_s)'$ が冗長であるという帰無仮説 H_0 は次

のようになる。

H_0 : すべての $1 \leq i < j \leq s$ について $\sum_{ij} \sum_{jj}^{-1} \sum_{ji} \sum_{ii}^{-1}$ と $\sum_{u_i u_j} \sum_{u_j u_i}^{-1} \sum_{u_i u_i} \sum_{u_i u_i}^{-1}$ の 0 でない全ての固有値が等しい

$\sum_{23,1} = \sum_{23} - \sum_{21} \sum_{11}^{-1} \sum_{13}$ のような記号を用いると Fujikoshi²⁾ から次の定理を得る。

定理 1: 帰無仮説 H_0 は次の 1, 2 と同値である。

$$1. \operatorname{tr} \sum_{ij} \sum_{jj}^{-1} \sum_{ji} \sum_{ii}^{-1} = \operatorname{tr} \sum_{u_i u_j} \sum_{u_j u_i}^{-1} \sum_{u_i u_i} \sum_{u_i u_i}^{-1}, 1 \leq i < j \leq s \quad (3)$$

$$2. [\sum_{u_i v_j \cdot u_j} \sum_{v_i v_j \cdot u_j}] = 0, i \neq j, 1 \leq i < j \leq s \quad (4)$$

4. 尤度比検定

以下では $P = \sum p_i$, (x_1', \dots, x_s') は P 変量正規分布に従うものとする。 (u_1', \dots, u_s') が与えられたときの (v_1', \dots, v_s') の条件付き分布の平均, 分散共分散行列はそれぞれ

$$E \left[\begin{pmatrix} v_1 \\ \vdots \\ v_s \end{pmatrix} \middle| \begin{pmatrix} u_1 \\ \vdots \\ u_s \end{pmatrix} \right] = \begin{pmatrix} \mu_{v_1} \\ \vdots \\ \mu_{v_s} \end{pmatrix} + \begin{bmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & & \vdots \\ B_{s1} & \cdots & B_{ss} \end{bmatrix} \begin{pmatrix} u_1 - \mu_{u_1} \\ \vdots \\ u_s - \mu_{u_s} \end{pmatrix}$$

$$V \left[\begin{pmatrix} v_1 \\ \vdots \\ v_s \end{pmatrix} \middle| \begin{pmatrix} u_1 \\ \vdots \\ u_s \end{pmatrix} \right] = \begin{bmatrix} \sum_{v_1 v_1 \cdot u_1 \cdots u_s} & \cdots & \sum_{v_1 v_s \cdot u_1 \cdots u_s} \\ \vdots & & \vdots \\ \sum_{v_s v_1 \cdot u_1 \cdots u_s} & \cdots & \sum_{v_s v_s \cdot u_1 \cdots u_s} \end{bmatrix}$$

ただし

$$\begin{bmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & & \vdots \\ B_{s1} & \cdots & B_{ss} \end{bmatrix} = \begin{bmatrix} \sum_{v_1 u_1} & \cdots & \sum_{v_1 u_s} \\ \vdots & & \vdots \\ \sum_{v_s u_1} & \cdots & \sum_{v_s u_s} \end{bmatrix} \begin{bmatrix} \sum_{u_1 u_1} & \cdots & \sum_{u_1 u_s} \\ \vdots & & \vdots \\ \sum_{u_s u_1} & \cdots & \sum_{u_s u_s} \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \sum_{v_1 v_1 \cdot u_1 \cdots u_s} & \cdots & \sum_{v_1 v_s \cdot u_1 \cdots u_s} \\ \vdots & & \vdots \\ \sum_{v_s v_1 \cdot u_1 \cdots u_s} & \cdots & \sum_{v_s v_s \cdot u_1 \cdots u_s} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{v_1 v_1} & \cdots & \sum_{v_1 v_s} \\ \vdots & & \vdots \\ \sum_{v_s v_1} & \cdots & \sum_{v_s v_s} \end{bmatrix} - \begin{bmatrix} \sum_{v_1 u_1} & \cdots & \sum_{v_1 u_s} \\ \vdots & & \vdots \\ \sum_{v_s u_1} & \cdots & \sum_{v_s u_s} \end{bmatrix} \begin{bmatrix} \sum_{u_1 u_1} & \cdots & \sum_{u_1 u_s} \\ \vdots & & \vdots \\ \sum_{u_s u_1} & \cdots & \sum_{u_s u_s} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{u_1 v_1} & \cdots & \sum_{u_1 v_s} \\ \vdots & & \vdots \\ \sum_{u_s v_1} & \cdots & \sum_{u_s v_s} \end{bmatrix}$$

である。これらを用いて帰無仮説は次のように言い換えることができる。

定理 2: (4) は

$$H_0: \sum_{v_i v_j \cdot u_1 \cdots u_s} = 0 \quad \text{かつ} \quad B_{ij} = 0, i \neq j, 1 \leq i, j \leq s$$

と同値である。

証明:

$$B_{1s} = \sum_{v_1 u_s \cdot u_1 \cdots u_{s-1}} \sum_{u_s u_s \cdot u_1 \cdots u_{s-1}}^{-1}$$

$$B_{s1} = \sum_{v_s u_1 \cdot u_2 \cdots u_s} \sum_{u_1 u_1 \cdot u_2 \cdots u_s}^{-1}$$

⋮

$$B_{s-1 s} = \sum_{v_{s-1} u_s \cdot u_1 \cdots u_{s-1}} \sum_{u_s u_s \cdot u_1 \cdots u_{s-1}}^{-1}$$

$$B_{s s-1} = \sum_{v_s u_{s-1} \cdot u_2 \cdots u_s} \sum_{u_{s-1} u_{s-1} \cdot u_2 \cdots u_s}^{-1}$$

⋮

$$B_{12} = \sum_{v_1 u_2 \cdot u_1} \sum_{u_2 u_2 \cdot u_1}^{-1}$$

$$B_{21} = \sum_{v_2 u_1 \cdot u_2} \sum_{u_1 u_1 \cdot u_2}^{-1}$$

$$\sum_{v_i v_j \cdot u_1 \cdots u_s} = \sum_{v_i v_j \cdot u_1 \cdots u_{s-1}} - \sum_{v_i u_s \cdot u_1 \cdots u_{s-1}} \sum_{u_s u_s \cdot u_1 \cdots u_{s-1}}^{-1} \sum_{u_s v_j \cdot u_1 \cdots u_{s-1}}$$

が成り立つことからこれらが(4)と同値であることがわかる。(証明終)

標本数を $N = n + 1 (> P)$ とする。 $n^{-1}S$ を標本分散共分散行列とすると (u_1', \dots, u_s') が与えられ

たときの (v_1', \dots, v_s') の条件付き分布をもちいて帰無仮説 H_0 に対する尤度比基準は

$$\begin{aligned} \lambda_N^2 = \Lambda &= \frac{\left| \begin{array}{ccc} S_{v_1 v_1 \cdot u_1 \dots u_s} & \dots & S_{v_1 v_s \cdot u_1 \dots u_s} \\ \vdots & & \vdots \\ S_{v_s v_1 \cdot u_1 \dots u_s} & \dots & S_{v_s v_s \cdot u_1 \dots u_s} \end{array} \right|}{|S_{v_1 v_1 \cdot u_1}| \dots |S_{v_s v_s \cdot u_s}|} \\ &= \frac{|S_{v_1 v_1 \cdot u_1 u_2 v_2 u_3 v_3 \dots u_s v_s}|}{|S_{v_1 v_1 \cdot u_1}|} \dots \frac{|S_{v_s v_s \cdot u_1 \dots u_s}|}{|S_{v_s v_s \cdot u_s}|} = \Lambda_1 \times \dots \times \Lambda_s \end{aligned}$$

となる。ただし、 S は Σ と同様の分割をしているものとする。

注意： Λ_1 の分子は $|S_{v_1 v_1 \cdot u_1 u_2 \dots u_{i+1} v_{i+1} \dots u_s v_s}|$ であり、 v_1 から $(u_1, u_2, \dots, u_i, u_{i+1}, v_{i+1}, \dots, u_s, v_s)$ の影響を除いた残差分散共分散行列の一般化分散、同様に分母は v_1 から u_1 の影響を除いた残差分散共分散行列の一般化分散である。

定理 3：帰無仮説 H_0 のもとで $\Lambda_1, \dots, \Lambda_s$ は互いに独立に分布し、

$$\Lambda_i \sim \Lambda(q_i, r_1 + \dots + r_{i-1} + p_{i+1} + \dots + p_s, n - (r_1 + \dots + r_i + p_{i+1} + \dots + p_s))$$

である。ここで A と B が互いに独立に Wishart 分布 $W_p(n, \Sigma), W_p(q, \Sigma)$ に従うとき $|A|/|A+B|$ 分布を $\Lambda(p, q, n)$ で表すものとする。

証明： X_1, \dots, X_n を互いに独立に $N_p(0, \Sigma)$ に従う確率変数、 $X = [X_1, \dots, X_n]'$ として $S = X'X$ と書くことができる。ここまでの記法では確率変数を列ベクトルで表してきたが、以下では記述統計的な多変量解析で用いられる記法に合わせて確率変数を行ベクトルで表すことにする。 X を (2) の分割に対応して $X = [U_1, V_1, \dots, U_s, V_s]$ とする。 U_i, V_i はそれぞれ $n \times r_i, n \times q_i$ 行列である。このように定義すると Λ_1 の分子は

$$|V_1' Q_1 V_1| = |V_1' (I_n - P_1) V_1|$$

と表すことができる。 A の一般逆行列を A^- とすると

$$P_1 = [U_1 U_2 V_2 \dots U_s V_s] ([U_1 U_2 V_2 \dots U_s V_s] [U_1 U_2 V_2 \dots U_s V_s]^- [U_1 U_2 V_2 \dots U_s V_s])'$$

であり、これは n 次元空間内で $U_1, U_2, V_2, \dots, U_s, V_s$ の列ベクトル全体で張られる部分空間への直交射影子である。したがって

$$\text{rank} P_1 = r_1 + p_2 + \dots + p_s$$

であり、

$$\text{rank} Q_1 = n - (r_1 + p_2 + \dots + p_s)$$

となる。分母も同じように

$$|V_1' Q_2 V_1| = |V_1' (I_n - P_2) V_1|, \quad P_2 = U_1 (U_1' U_1)^- U_1', \quad \text{rank} Q_2 = n - r_1$$

となる。直交射影子の性質から $Q_1^2 = Q_1, Q_2^2 = Q_2, U_1, U_2, V_2, \dots, U_s, V_s$ の列ベクトル全体で張られる部分空間は U_1 の列ベクトル全体で張られる部分空間を含むので $Q_1(Q_2 - Q_1) = Q_1 - Q_1 = 0$ 。

したがって $V_1'(Q_2 - Q_1)V_1$ と $V_1'Q_1V_1$ は互いに独立にそれぞれ自由度 $p_2 + \dots + p_s, n - (r_1 + p_2 + \dots + p_s)$ の Wishart 分布にしたがうので

$$\Lambda_1 = \frac{|V_1' Q_1 V_1|}{|V_1' Q_1 V_1 + V_1' (Q_2 - Q_1) V_1|} \sim \Lambda(q_1, p_2 + \dots + p_s, n - (r_1 + p_2 + \dots + p_s))$$

同様にして Λ_i についても

$$\Lambda_i \sim \Lambda(q_i, r_1 + \dots + r_{i-1} + p_{i+1} + \dots + p_s, n - (r_1 + \dots + r_i + p_{i+1} + \dots + p_s))$$

がいえる。帰無仮説 H_0 のもとでは V_i の分布は U_i だけに依存するので $\Lambda_1, \dots, \Lambda_s$ は互いに独立に分布する。(証明終)

Anderson¹⁾により尤度比 Λ の特性関数は Γ 関数によって表され、この特性関数を用いることにより Λ の漸近分布が次のようになることがわかる。

定理 4: 帰無仮説 H_0 のもとでの尤度比 Λ の漸近分布は

$$\Pr\{-m \log \Lambda \leq x\} = \Pr\{\chi_f^2 \leq x\} + \frac{\omega}{m^2} (\Pr\{\chi_{f+4}^2 \leq x\} - \Pr\{\chi_f^2 \leq x\}) + O(m^{-3})$$

ただし

$$\begin{aligned} f &= \sum_{j=1}^s (p_j p_j - r_j r_j), \quad m = n - \frac{1}{2}(P+1) + \frac{A}{f} \\ \omega &= -\frac{A^2}{4f} - \frac{1}{4}(P+1)A + \frac{2}{3}B - \frac{1}{16}(P+1)^2 - \frac{1}{12}f \\ A &= \frac{1}{2} \sum_{j=1}^s \alpha_j q_j \left(p_j + \sum_{k=1}^j q_k \right) \\ B &= \frac{1}{4} \left(\sum_{j=1}^s \alpha_j^2 r_j q_j + \frac{1}{2} \sum_{j=1}^s \alpha_j^2 q_j^2 + \frac{1}{2} \sum_{j=1}^s \alpha_j^2 q_j + \sum_{j=1}^s \alpha_j r_j^2 q_j + \sum_{j=1}^s \alpha_j r_j q_j^2 \right. \\ &\quad \left. + \sum_{j=1}^s \alpha_j r_j q_j + \frac{1}{3} \sum_{j=1}^s \alpha_j q_j^3 + \frac{1}{2} \sum_{j=1}^s \alpha_j q_j^2 + \frac{1}{6} \sum_{j=1}^s \alpha_j q_j \right) \\ \alpha_i &= r_1 + \dots + r_{i-1} + p_{i+1} + \dots + p_s \end{aligned}$$

定理 4 から変数の冗長性を検定する統計量の漸近的な性質がわかる。また、 $s=2$ としたときには Fujikoshi²⁾ の Theorem 3 と一致することから、定理 4 はこの結果を多群の場合に拡張したものであるといえる。

5. おわりに

多群からなる変数の冗長性を検定するための帰無仮説、帰無仮説を検定するための尤度比基準およびその漸近分布を求めた。この方法は内部正則線形変換で不変な量に注目する場合の変数選択として用いることができる。

この方法は全ての変数群から 1 個以上の変数を選択するためのものである。したがって、ある変数群には不要な変数がない場合には用いることができない。これは 2 群の場合(正準相関分析)でも同様であり、両側から選択する基準(Fujikoshi²⁾)は片側から選択する基準(Siotani⁴⁾)を含んでいない。この点を改良するためには一般化決定係数を選択基準として用いる方法などが考えられる。

参考文献

- 1) Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis, 2nd ed., Wiley
- 2) Fujikoshi, Y. (1982). A test for additional information in canonical correlation analysis, Ann. Inst. Statist. Math., 34, 523-540
- 3) Horst, P. (1961). Relation among m sets of variables, Psychometrica, 26, 129-149.
- 4) Siotani, M. (1957). Effect of the additional variables on the canonical correlation coefficients, Proc. Inst. Statist. Math., 5, 52-57.