

重回帰分析の図的表示

—冗長性分析と主成分分析を用いて—

前川眞一

東京工業大学大学院 社会理工学研究科

961108,000830,31,0904

1 回帰分析の図的表示

$N \times p$ の基準変数 $\mathbf{Y} = [\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(p)}]$ と $N \times q$ の説明変数 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]'$ の間に、 $\boldsymbol{\beta} = [\beta_1, \beta_j, \beta_q]'$ を $p \times q$ の回帰係数行列として

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\mu}' + \mathbf{X}\boldsymbol{\beta}' + \boldsymbol{\varepsilon}$$

という回帰モデル（多変量重回帰モデル）、を考えた場合、説明変数の数 q が少ない場合には説明変数の各行（個体）を点として、各列（変数）をベクトルとしてバイプロット (biplot, Gower & Hand, 1996 に詳しいが、本稿では同時プロットと称する。前川, 1997 も参照。) を描くことができる。

一般に、 \mathbf{V} を $N \times p$ の基準変数、 \mathbf{U} を $N \times 2$ の説明変数 とした場合、

$$\mathbf{V} = \mathbf{1}\boldsymbol{\mu}' + \mathbf{U}\boldsymbol{\beta}' + \boldsymbol{\varepsilon}$$

バイプロットにおいては、 \mathbf{v}_j が \mathbf{u}_i をベクトル $\boldsymbol{\beta}_j$ へおろした垂線の足 $\mathbf{g}_i^{(j)}$ の原点からの方向付き距離と線形に関連することになる。図 1 に $q = 2$ の場合を図示したが、本稿の目的は $p = 1, q > 2$ の場合にこのような同時プロットを如何にして描くかということにあるが、そのためには、冗長性分析 ($\boldsymbol{\beta}$ の階数を $r, r < p, q$ に制約した回帰分析) を用いる方法、 \mathbf{X} の r 個の主成分に \mathbf{Y} を回帰する方法、および、 (\mathbf{Y}, \mathbf{X}) の r 個の主成分へ \mathbf{Y} を回帰する方法等がある。

2 冗長性分析モデル

回帰係数行列 $\boldsymbol{\beta}$ の階数を $r, r \leq q, p$ に制約するためには $p \times r$ の行列 \mathbf{A} と $q \times r$ の行列 \mathbf{B} を用いて $\boldsymbol{\beta} = \mathbf{A}\mathbf{B}'$ と置き

$$\text{RSS}_Y = \text{tr} \left[(\mathbf{Y} - \mathbf{X}\mathbf{A}\mathbf{B}')' (\mathbf{Y} - \mathbf{X}\mathbf{A}\mathbf{B}') \right]$$

を最小とするような \mathbf{A} および \mathbf{B} を求めればよい。このモデルは \mathbf{Y} を、 $N \times r$ の行列 $\mathbf{Z} = \mathbf{X}\mathbf{A}$ に回帰していることに他ならないから、 \mathbf{Z} と回帰係数行列 \mathbf{B} を用いて r 次元の同時プロットが描ける。これを \mathbf{Y} の \mathbf{X} への冗長性分析 (redundancy analysis, Van den Wollenberg, 1977) もしくは階数制約条件付きの回帰分析 (reduced rank regression, Izenman, 1975) と呼ぶ。また、 \mathbf{Z} が与えられた場合、

$$\mathbf{X} = \mathbf{Z}\mathbf{W}' + \boldsymbol{\varepsilon}_X$$

という形で \mathbf{X} を \mathbf{Z} に回帰すれば、同じ同時プロットの上に各説明変数 \mathbf{X} をもベクトルとして図示することが出来る。冗長性分析のひとつの解は、

$$\hat{\mathbf{A}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{B}})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{P}\boldsymbol{\Delta}^{-\frac{1}{2}} \quad \text{and} \quad \hat{\mathbf{B}} = \mathbf{P}\boldsymbol{\Delta}^{\frac{1}{2}}$$

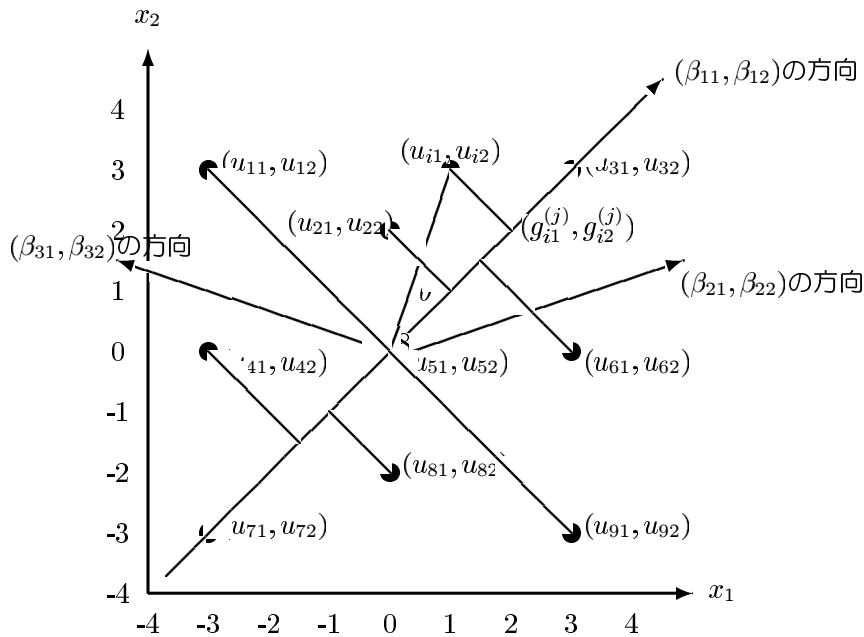


図 1: 回帰モデルの図的表示 (同時プロット)

ただし、 Δ および P は $\frac{1}{N}Y'JX(X'JX)^{-1}X'JY$ の最大 r 個の固有値および固有ベクトル行列である。また、 Z 行列と W 行列は

$$Z = X\hat{A} = X(X'X)^{-1}X'Y\hat{B}(\hat{B}'\hat{B})^{-1} = X(X'X)^{-1}X'YP\Delta^{-\frac{1}{2}}$$

$$\hat{W} = \frac{1}{N}X'Z = \frac{1}{N}X'YP\Delta^{-\frac{1}{2}} = \frac{1}{N}X'X\hat{A}$$

となる。なお、以下の比較のために、上記の解を A_1, B_1, Z_1, W_1 と呼び、これらを用いて残差の 2 乗和を評価したものを

$$RSS(Y, Z_1, B_1) = \text{tr}[(Y - Z_1B_1)'(Y - Z_1B_1)],$$

$$RSS(X, Z_1, W_1) = \text{tr}[(X - Z_1W_1)'(X - Z_1W_1)]$$

と置く。当然ではあるが、 $r \leq p, q$ の場合、 β に階数の制約のない場合と比べて、

$$RSS(Y, X, Y'X(X'X)^{-1}) \leq RSS(Y, Z_1, B_1) \quad \text{and} \quad RSS(X, X, I_q) \leq RSS(X, Z_1, W_1)$$

である。この方法の問題は、 $r > p$ の場合 (例えば、単純な重回帰分析 $p = 1, r = 2$, の場合) に r 次元の同時プロットが描けない点にある。この点を改善するために以下のような方法を提案する。

3 $r > p, r \leq q$ の場合の図的表示の方法

3.1 冗長性分析 改一 1 (X をうまく説明するように Z を補う方法)

$r > p$ の場合には、まず、 $r = p$ において Y の X への冗長性分析を行い、 $N \times p$ の行列 Z_2^* を求める。その後、求めた Z_2^* に直交するという条件の下で

$$\text{RSS}_X = \text{tr}[(X - FV')'(X - FV')]$$

を最小とするような $N \times (r - p)$ の F と $q \times (r - p)$ の V を求め、

$$Z_2 = [Z_2^*, F], \quad (N \times r)$$

とする。なお、このような F 行列は $X - Z_2^*(Z_2^{*'}Z_2^*)^{-1}Z_2^{*'}X$ 行列の最初の $r - p$ 個の主成分として与えられる。次に、 Y および X を Z_2 に回帰することにより

$$B_2 = Y'Z_2(Z_2'Z_2)^{-1}, \quad W_2 = X'Z_2(Z_2'Z_2)^{-1}$$

を求める。この場合、

$$\text{RSS}(Y, Z_2, B_2) = \text{RSS}(Y, Z_1, B_1) \quad \text{and} \quad \text{RSS}(X, Z_2, W_2) \leq \text{RSS}(X, Z_1, W_1)$$

である。これは Y の説明率を減少させることなく、しかも X をもある程度説明するような X の線形結合を求める方法である。

3.2 冗長性分析 改一 2 ((Y, X) の X への冗長性分析)

まず、 $U = (Y, X)$ とし、 U の X への冗長性分析を行い、 $N \times r$ 行列 Z_3 を求める。その後、 Y を Z_3 に回帰することにより $B_3 = Y'Z_3(Z_3'Z_3)^{-1}$ を、 X を Z_3 に回帰することにより $W_3 = X'Z_3(Z_3'Z_3)^{-1}$ を求める。この場合、

$$\text{RSS}(Y, Z_2, B_2) \leq \text{RSS}(Y, Z_3, B_3) \quad \text{and} \quad \text{RSS}(X, Z_3, W_3) \leq \text{RSS}(X, Z_2, W_2)$$

である。これは Y と X を同等によく説明するような X の線形結合を求める方法である。

3.3 主成分回帰 (Y の X の主成分への回帰)

X の主成分スコアを Z_4 とし、 Y を Z_4 に回帰することにより $B_4 = Y'Z_4(Z_4'Z_4)^{-1}$ を求める。この場合、

$$\text{RSS}(Y, Z_3, B_3) \leq \text{RSS}(Y, Z_4, B_4) \quad \text{and} \quad \text{RSS}(X, Z_4, W_4) \leq \text{RSS}(X, Z_3, W_3)$$

である。これは X を最もよく説明するような X の線形結合で Y を説明する方法である。

3.4 主成分分析 (Y の (Y, X) の主成分への回帰)

まず、 $U = (Y, X)$ とし、

$$\text{RSS}_U = \text{tr}[(U - Z_5V')'(U - Z_5V')]$$

を最小とするように $N \times r$ の Z_5 を求め、 Y と X を Z_5 に回帰することにより B_5 と W_5 を求める。なお、これらは V の行の分割として与えられる。この場合、

$$\text{RSS}(Y, Z_5, B_5) \quad ??? \quad \text{RSS}(Y, Z_4, B_4) \quad \text{and} \quad \text{RSS}(X, Z_5, W_5) \quad ??? \quad \text{RSS}(X, Z_3, W_3)$$

である。これは Y と X を同等によく説明するような (Y, X) の線形結合を求める方法であるが、この方法においてはもはや Y を X で予測するという回帰分析や冗長性分析特有の変数間の非対称性は失われてしまっている。

参考文献

Gower, J.C. & Hand, D.J. (1996) *Biplots*. London: Chapman and Hall.

Izenman, A. J. (1975) Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, **5** 248-264.

Van den Wollenberg, A. L. (1977) Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika*, **42** 207-219.

前川眞一 (1997) SASによる多変量データの解析 竹内啓監修 SASで学ぶ統計的データ解析
4. 東京大学出版会