土木学会第52回年次学術講演会（平成９年９月）

IV－251

パラメータ推定における多重共線性と適切化手法に関する一考察

東京大学大学院　学生会員　堤　盛人
大成建設株式会社　正会員　松葉保孝
東京大学大学院　正会員　清水英範

1. はじめに

線形地価モデルなどを実際のデータを用いてパラメータ推定する際には、しばしば多重共線性が問題となるため、これまでも、多重共線性の改善方法に関しては膨大な研究がなされてきた。本稿では、それらの既存研究の相互関係を簡潔にまとめて示す。

2. 多重共線性

多重共線性を説明するため、線形モデル $y=X\beta$ に関して $n(>k)$ 個のデータを得たとする。

$$\mathbf{y}=\mathbf{X\beta} \qquad (1)$$

ここで、$\mathbf{y}$ は $n\times 1$ の被説明変数ベクトル、$\mathbf{X}$ は説明変数ベクトルからなる $n\times k$ 行列、$\boldsymbol{\beta}$ は $k\times 1$ のパラメータベクトルである。 仮定より未知パラメータの数 k はデータ数 n より少ないため、一般には方程式(1)を満足する解は存在しない。このような問題は不適切問題と呼ばれる [1]。

(1)の近似解を求める最も一般的な方法は、次の最小二乗法である。

$$\min_{\boldsymbol{\beta}} \|\mathbf{y}-\mathbf{X\beta}\|^2 \qquad (2)$$

ただし、$\|\mathbf{x}\|$ は $\mathbf{x}$ のユークリッドノルムを表す。

（2）の一階条件から、正規方程式が導かれる。

$$\mathbf{X}^t\mathbf{X}\boldsymbol{\beta}=\mathbf{X}^t\mathbf{Y} \qquad (3)$$

ここで、t は行列の転置を表す。$\mathbf{X}^t\mathbf{X}$ が正則行列であれば(3)の解 $\hat{\boldsymbol{\beta}}$ は次のように得られる。

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \qquad (4)$$

$(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ は $\mathbf{X}$ の最小二乗型一般逆行列と呼ばれる [2]。

多重共線性とは、説明変数の間に高い相関がある状態、あるいは $\mathbf{X}^t\mathbf{X}$ がランク落ちに近い状態をいう。すなわち、方程式（4）が不安定な不適切問題となり、パラメータの分散が大きくなる。

3. 主成分回帰と Ridge 回帰

多重共線性が存在する場合に、従来から様々な解決法が提案されてきた。本稿ではなかでも特に有名な主成分回帰と Ridge 回帰について扱う。

主成分回帰では X に関して(5)のような主成分分析を行い、互いに無相関な主成分を導出する。

$$\mathbf{Z}=\mathbf{XP} \qquad (5)$$

ここで $\mathbf{P}$ は $\mathbf{X}^t\mathbf{X}$ の固有ベクトルから成る $k\times k$ の直交行列（$\mathbf{P}^{-1}=\mathbf{P}^t$）である。固有値が０に近い主成分をいくつか除いたあとの主成分行列を $\mathbf{Z}_L$（$l\times k$）と表すと、主成分回帰は $\mathbf{y}$ を $\mathbf{Z}_L$ に回帰する方法である。

$$\mathbf{y}=\mathbf{Z}_L\boldsymbol{\alpha} \qquad (6)$$

$\mathbf{Z}_L$ に対応した $\mathbf{P}$ を $\mathbf{P}_L$（$l\times k$）と書くと、主成分回帰によるパラメータ推定値は次のようになる。

$$\hat{\boldsymbol{\beta}}_S=\mathbf{P}_L\boldsymbol{\alpha} \quad ただし \quad \hat{\boldsymbol{\alpha}}=(\mathbf{Z}_L^t\mathbf{Z}_L)^{-1}\mathbf{Z}_l^t\mathbf{y} \qquad (7)$$

一方、Ridge 回帰では(4)の $\mathbf{X}^t\mathbf{X}$ を適当なスカラー c と単位行列 $\mathbf{I}$ を用いて $\mathbf{X}^t\mathbf{X}+c\,\mathbf{I}$ に置き換える。

$$(\mathbf{X}^t\mathbf{X}+c\,\mathbf{I})\boldsymbol{\beta}=\mathbf{X}^t\mathbf{y} \qquad (8)$$

これより、Ridge 推定量が次のように得られる。

$$\hat{\boldsymbol{\beta}}_R=(\mathbf{X}^t\mathbf{X}+c\,\mathbf{I})^{-1}\mathbf{X}^t\mathbf{y} \qquad (9)$$

ところで、任意の $m\times n$ 行列 $\mathbf{A}$ に対しては、次のような特異値分解が存在する。

$$\mathbf{A}=\mathbf{UDV}^t \qquad (10)$$

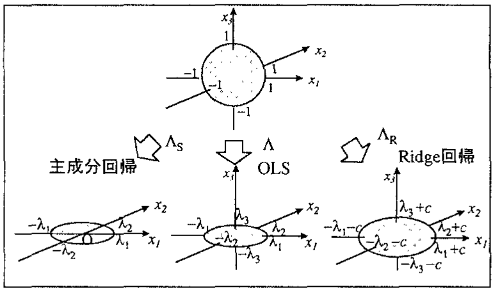
ただし、$\mathbf{U},\mathbf{V}$ はそれぞれ m,n 次の直交行列である。$\mathbf{D}$ は第（i,i）要素 σ_i 以外の要素がすべて０である $m\times n$ 行列で、σ_i はすべて非負の実数である。$\mathbf{U},\mathbf{V}$ を適当に選べば、

$$\sigma_1\geqq\sigma_2\cdots\geqq\sigma_r\geqq\ =\sigma_{r+1}=\cdots=\sigma_M=0 \qquad (11)$$

キーワード：多重共線性、主成分回帰、Ridge 回帰、一般逆行列、Tikhonov の適切化

連絡先　：　〒113　文京区本郷７－３－１　東京大学大学院工学系研究科　社会基盤工学専攻
TEL：03-3812-2111（代）　内線 6128　FAX：03-5689-7290

となるように順序づけ可能である。但し、r は rank **A** に等しく、M は m と n の小さい方の数である。直交行列はベクトルの回転・鏡映を表し、ベクトルの長さを変えない。一方、**A** の表す一次変換は単位球面 $\{\mathbf{x}:\|\mathbf{x}\|^2=1\}$ にひきおこす歪みの度合いを示す[3]。



図－1 対角行列の表す一次変換

正方行列の特異値分解は固有値分解である。

$$\mathbf{X}^t\mathbf{X}=\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^t=\sum_{i=1}^{k}\lambda_i\mathbf{P}_i\mathbf{P}_i{}^t \qquad (12)$$

ここで、$\mathbf{P}_i$ は固有値 λ_i に対応する固有ベクトル、**P** は固有ベクトルからなる行列（$\mathbf{P}\mathbf{P}^t=\mathbf{P}^t\mathbf{P}=\mathbf{I}$）、$\boldsymbol{\Lambda}$ は $\mathbf{X}^t\mathbf{X}$ の固有値（$\lambda_1\geqq\lambda_2\geqq\cdots\geqq\lambda_r\geqq\ \lambda_{r+1}\cdots\lambda_k\geqq 0$ とする）を対角要素とする対角行列である。多重共線性が存在すると、$\boldsymbol{\Lambda}$ の表す一次変換は単位球面を非常に歪んだ（縮んだ）楕円体面に移す変換となる。$\mathbf{X}^t\mathbf{X}$ が正則であれば次のようになる。

$$(\mathbf{X}^t\mathbf{X})^{-1}=\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^t=\sum_{i=1}^{k}\frac{1}{\lambda_i}\mathbf{P}_i\mathbf{P}_i{}^t \qquad (13)$$

固有値分解を用いると、主成分回帰や Ridge 回帰の相互の関係が明確に理解できる[4]。たとえば、λ_k が他の固有値に比べて非常に 0 に近く、これを除去する主成分回帰と Ridge 回帰は、それぞれつぎのような置き換えに等しい。

$$(X^tX)^{-1}\ \rightarrow\ PA_{(s)}P^t=\sum_{i=1}^{k-1}\frac{1}{\lambda_i}\mathbf{P}_i\mathbf{P}_i{}^t \qquad (14)$$

$$(X^tX)^{-1}\ \rightarrow\ PA_{(R)}P^t=\sum_{i=1}^{k}\frac{1}{\lambda_i+c}\mathbf{P}_i\mathbf{P}_i{}^t \qquad (15)$$

4. 一般逆行列と適切化手法

次に、X^tX が次のように完全にランク落ちした $\lambda_k=0$ の場合について考える。このとき X^tX の一般逆行列 $(X^tX)^+$ は

$$(X^tX)^+=\sum_{i=1}^{k-1}\frac{1}{\lambda_i}\mathbf{P}_i\mathbf{P}_i{}^t \qquad (16)$$

ところで、一般逆行列のうち Moore-Penrose 型一般逆行列 $\mathbf{X}^-_{MP}$ には、次の公式が知られている[5]。

$$\mathbf{X}^-_{MP}=\lim_{\delta\to 0}(\mathbf{X}^t\mathbf{X}+\delta\mathbf{I})^{-1}\mathbf{X}^t \qquad (17)$$

十分小さな δ に対し、$(\mathbf{X}^t\mathbf{X}+\delta\mathbf{I})^{-1}\mathbf{X}^t$ は **X** の Moore-Penrose 型一般逆行列の近似を与える。$\mathbf{X}^t\mathbf{X}$ が完全にランク落ちしている場合、Ridge 回帰は M-P 一般逆行列の近似を用いた推定と解釈できる。

さて、不適切問題の適切化手法の一つに Tikhonov の適切化[6]と呼ばれる手法があるが、それは方程式（4）が不適切である場合に、(2)の目的関数に安定化項 $F(\boldsymbol{\beta})$ を加えた次式を用いるものである。

$$M(\alpha)=\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2+\alpha F(\boldsymbol{\beta}) \qquad (18)$$

α は適切化パラメータと呼ばれる。ここで、$F(\boldsymbol{\beta})=\|\boldsymbol{\beta}\|^2$，$\alpha=c$ とおくと、(18)は

$$M(c)=\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2+c\|\boldsymbol{\beta}\|^2 \qquad (19)$$

となる。一階条件から容易に(8)式が導かれる。すなわち、Ridge 回帰は Tikhonov の適切化手法と解釈可能である。(2)の近似解法としての(19)と Ridge 回帰との関係については、Golub et al.[6]に詳しく示されている。

謝辞 Tikhonov の適切化については、東京大学地震研究所堀宗朗助教授からご指摘いただいた。記して感謝の意を表したい。

【参考文献】

(1) 久保司郎(1992)：計算力学と CAE シリーズ 10　逆問題，培風館，pp.8-9

(2) STRANG, G.(1976)：Linear Algebra and its Applications，山口昌哉監訳、井上昭訳(1978)：線形代数とその応用，産業図書，pp.152-153

(3) 田辺国士(1981)：数値的方法における特異値，数理科学，No.212，pp.46-50

(4) MARQUARDT, D. W. (1970)：Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, Technometrics, 12, No.3, pp.591-612

(5) 田島稔・小牧和雄(1986)：最小二乗法の理論とその応用，東洋書店，pp.349-350

(6) 前掲(1)，pp.45-47

(7) GOLUB, G. H., HEATH, M. and WAHBA, G.(1979)：Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter，Technometrics，21，No.2，pp.215-223