

線形基底関数モデルにおける実行可能型 一般化リッジ回帰推定量の正確モーメント

地 道 正 行

I はじめに

一般に、同時分布 $F(\mathbf{x}, y)$ に従う $(d+1)$ 変量ベクトル $[\mathbf{X}', Y]'$ において標本ベクトル \mathbf{X} の関数 $r(\mathbf{X})$ で標本 Y を推定する問題は統計的回帰問題 (statistical regression problem) といわれる：

$$Y \stackrel{\text{推定}}{\longleftarrow} r(\mathbf{X})$$

(たとえば、稲垣 (2003) 参照.) ここで、関数 $r(\mathbf{X})$ は回帰関数 (regression function) と呼ばれ¹⁾、標本 Y は応答変数 (response variable)、 $\mathbf{X}=[X_1, \dots, X_d]'$ は共変量 (covariates) ベクトルと呼ばれる。なお、共変量ベクトルの実現値のベクトルは $\mathbf{x}=[x_1, \dots, x_d]'$ と書かれる。ここで、' はベクトル、行列の転置を表すことに注意しよう。

統計的回帰問題において、回帰関数 r をどのように選べばよいかの問題となるけれども、推定精度を平均 2 乗誤差²⁾ (mean squared error)：

$$\rho_2(Y, r(\mathbf{X})) := \int_{\mathbb{R}^{d+1}} |y - r(\mathbf{x})|^2 dF(\mathbf{x}, y)$$

ではかることにし、これを最小にするような回帰関数 r を探すが通常は考えられる：

1) 通常、 $r \in L^p$ (L^p 空間) と仮定される。
2) L^2 危険 (L^2 -risk) ともいわれる。

$$\rho_2(Y, r(\mathbf{X})) \longrightarrow \min_{r \in L^2}$$

この最小化問題の解は、回帰関数 r を応答変数 Y の共変量ベクトル $\mathbf{X} = \mathbf{x}$ ($\in \mathbb{R}^d$) に関する条件付き平均

$$\mu(\mathbf{x}) := E(Y|\mathbf{x})$$

によって与えられることが知られている。(たとえば、Györfi ら (2002), 稲垣 (2003) を参照.)

以上の結果から、統計的回帰問題は、平均構造 (mean structure) として条件付き平均 $\mu(\mathbf{x})$ がわかればよいということになるが、一般には (同時) 分布 $F(\mathbf{x}, y)$ が明らかでない場合が多く、よって条件付き平均 $\mu(\mathbf{x})$ もわからない。そこで、条件付き平均 $\mu(\mathbf{x})$ が共変量ベクトルの実現値 (ベクトル) $\mathbf{x} = [x_1, \dots, x_d]'$ の線形関数³⁾

$$\eta := \beta_0 + \boldsymbol{\beta}'\mathbf{x} = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d$$

で近似できる場合、すなわち、

$$\mu(\mathbf{x}) \simeq \eta = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d$$

をまずは考え、応答変数 Y と線形関数 η の差を誤差 ε として表すことによって、以下の線形回帰モデル (linear regression model) を利用することが考えられてきた：

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d + \varepsilon \quad (1)$$

ここで、 $\beta_0, \beta_1, \dots, \beta_d$ は回帰係数 (regression coefficients) と呼ばれ、未知のパラメータである。

線形回帰モデル(1)は、回帰モデル (regression model) の中でもっとも基本的なものであり、非常に扱いやすく、理論・実証の両面からこれまで多くの研究がなされたきたものであるけれども、実際にこのモデルをデータに当てはめる際の決定的な欠点として、その簡素な構造ゆえの表現力の乏しさにある。(Bishop (2006) 参照.) この問題に対して、本稿では近年学習理論

3) 線形予測子 (linear predictor) とも呼ばれる。

(learning theory) に関する分野でしばしば利用される線形基底関数モデル (linear basis function model) を扱う。このモデルは線形回帰モデルを少し拡張することによって柔軟なデータへの当てはめを実現しており、線形回帰モデルはもとより、多項式回帰モデルや B-スプライン回帰モデルなどを含むより一般的なものであることを注意されたい。

一方、線形回帰モデル(1)における共変量間に 1 次従属、または、それに近い関係がある場合に、回帰係数 $\beta_0, \beta_1, \dots, \beta_d$ の最小自乗法による通常の推定量の分散が大きくなることから、推定精度が悪くなるという、いわゆる、多重共線性 (multicollinearity) の問題があることが指摘されてきたけれども⁴⁾、学習理論の分野でも、この問題と類似する過学習 (over-fitting) と呼ばれる問題が存在することが指摘されており、この問題に対してパラメータを正則化最小自乗法 (regularized least squares method) によって推定することが提案されている。(たとえば, Bishop (2006), Konishi and Kitagawa (2007) を参照.) 本稿でも, Hoerl and Kennard (1970) によって線形回帰モデルの場合に提案された一般化リッジ推定量 (generalized ridge regression estimator) を線形基底関数モデルへ応用することを考える。

ただし、一般化リッジ回帰推定量を利用する際には、リッジ係数を適切に選択する必要があるが、やむおえず標本に依存した形式のリッジ係数をプラグインした推定量、すなわち、実行可能型一般化リッジ回帰推定量 (feasible generalized ridge regression estimator) を扱う必要が生ずる。一般に、正則化最小自乗法から導かれる実行可能型推定量の正確なモーメントを求めることがむずかしく、直接議論されることが少ないけれども、ここでは、Dwivedi ら (1980), Jimichi (2008) などの結果を利用して実行可能型一般化リッジ回帰推定量の正確なモーメントを与え、さらに得られたこれらのモーメントから、この推定量の総平均 2 乗誤差などの評価基準を正確に求める。この結果は、通常の最小自乗推定量からの精度の改良に関する程度を定量的

4) たとえば, Gunst (1983) 参照。

に評価するために利用できるという点で重要である。

本稿の構成としては、まずⅡ節で、線形基底関数モデルとその重みに関する推定量について述べ、Ⅲ節ではリッジ係数の選択アルゴリズムを与える。また、Ⅳ節では、まず最小自乗推定量と一般化リッジ回帰推定量の分布について考察し、次に実行可能型一般化リッジ回帰推定量の正確なモーメントを与える。さらに、この結果を利用して、Ⅴ節で実行可能型一般化リッジ回帰推定量の平均2乗誤差基準を理論的・数値的に評価し、Ⅵ節では例を与える。最後にⅦ節において総括を行う。なお、付録Aには一般的な平均2乗誤差基準と、付録Bには一般化リッジ回帰推定量の幾何学的説明を与える。なお、本稿における数値計算等は、R (Version 2.15.1) と Maple (Version 15) を用いて行われている。

Ⅱ モデルと推定量

本稿では、応答変数 Y の共変量 $\mathbf{X}=\mathbf{x}(\in\mathbb{R}^d)$ に関する条件付き平均 $\mu(\mathbf{x}):=E(Y|\mathbf{x})$ が基底関数⁵⁾ (basis function) と呼ばれる既知の関数 $b_j(\mathbf{x})$ ($j=0, 1, \dots, p$) の線形結合で以下のように近似できる場合を考える：

$$\mu(\mathbf{x}) \simeq w_0 + w_1 b_1(\mathbf{x}) + \dots + w_p b_p(\mathbf{x}) = \mathbf{w}'\mathbf{b}(\mathbf{x})$$

ここで、 $\mathbf{w}:=[w_0, w_1, \dots, w_p]'$ は重みベクトルと呼ばれ、未知パラメータベクトルとする。また、 $\mathbf{b}(\mathbf{x}):=[b_0(\mathbf{x}), b_1(\mathbf{x}), \dots, b_p(\mathbf{x})]'$ $= [1, b_1(\mathbf{x}), \dots, b_p(\mathbf{x})]'$ は基底関数ベクトル (既知) と呼ばれる。

この平均構造と誤差の和によって応答変数が表される以下のモデルは線形基底関数 (Linear Basis Functions: LBF) モデルと呼ばれる (Bishop (2006) 参照.)：

$$Y = w_0 + w_1 b_1(\mathbf{x}) + \dots + w_p b_p(\mathbf{x}) + \varepsilon$$

このモデルにおいて、個体 i に関する応答変数を Y_i 、共変量の実現値ベク

5) 数学的には、考察の対象となる (関数) 空間における基底 (basis) となりうる関数であることからこのような名称がついているものと思われる。

トルを \mathbf{x}_i とすると,

$$Y_i = w_0 + w_1 b_1(\mathbf{x}_i) + \cdots + w_p b_p(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n$$

と書くことができ、さらに行列とベクトルを使って以下のように表現される:

$$\mathbf{Y} = \mathbf{B}\mathbf{w} + \boldsymbol{\varepsilon} \quad (2)$$

ここで,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_0(\mathbf{x}_1) & b_1(\mathbf{x}_1) & \cdots & b_p(\mathbf{x}_1) \\ \vdots & \vdots & & \vdots \\ b_0(\mathbf{x}_n) & b_1(\mathbf{x}_n) & \cdots & b_p(\mathbf{x}_n) \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

であり, $b_0(\mathbf{x}_i) := 1$ とおいた. また, $\text{rank}(\mathbf{B}) = p+1 (< n)$ と仮定され, 誤差に関しては,

$$\boldsymbol{\varepsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (3)$$

が仮定されることに注意しよう. ここで, $\mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ は平均ベクトル $\mathbf{0} := [0, \dots, 0]' \in \mathbb{R}^n$, 分散共分散行列 $\sigma^2 \mathbf{I}_n$ の n 次元正規分布であり, σ^2 は未知の誤差分散, \mathbf{I}_n は n 次単位行列である.

重みベクトル \mathbf{w} の通常の推定量として, 誤差平方和:

$$\Delta^2(\mathbf{w}) := \|\boldsymbol{\varepsilon}\|^2 = \|\mathbf{Y} - \mathbf{B}\mathbf{w}\|^2$$

を \mathbf{w} に関して微分することによって得られる方程式 (正規方程式):

$$\mathbf{B}'\mathbf{B}\mathbf{w} = \mathbf{B}'\mathbf{Y}$$

を \mathbf{w} に関して解くことによって得られる以下のような通常最小自乗 (Ordinary Least Squares: OLS) 推定量が利用される:

$$\hat{\mathbf{w}} := (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Y}$$

この推定量は, 標準的な推定量として利用されるけれども, 過学習 (overfitting) の問題に対して推定精度が悪くなることがある. (たとえば, Bishop (2006), Konishi and Kitagawa (2007) 参照.) この問題は, 線形回帰分析において説明変数をモデルに取り入れすぎた場合に起こる多重共線性の問題と

本質的に同等であることに注意しよう。線形回帰においては多重共線性に対して通常リッジ回帰 (Ordinary Ridge Regression: ORR) 推定量やその一般化である一般化リッジ回帰 (Generalized Ridge Regression: GRR) 推定量が Hoerl and Kennard (1970) によって提案されており、同様に LBF モデルにおいても正則化最小自乗法の観点からリッジ型の推定量が考察されることがある。(このことに関しても, Bishop (2006), Konishi and Kitagawa (2007) を参照せよ.)

本稿では, LBF モデルにおける重みベクトル w に対して以下のように GRR 推定量を定義して推定することを考える:

$$\hat{w}(\mathbf{K}) := (\mathbf{B}'\mathbf{B} + \mathbf{\Gamma}\mathbf{K}\mathbf{\Gamma}')^{-1}\mathbf{B}'\mathbf{Y}$$

ここで, $\mathbf{\Gamma}$ は行列 $\mathbf{B}'\mathbf{B}$ を対角化

$$\mathbf{\Gamma}'\mathbf{B}'\mathbf{B}\mathbf{\Gamma} = \mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_p)$$

するような直交行列であり, $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_p (> 0)$ は, 行列 $\mathbf{B}'\mathbf{B}$ の固有値である。また,

$$\mathbf{K} := \text{diag}(k_0, k_1, \dots, k_p)$$

はりッジ係数 (ridge coefficients) $k_i (\geq 0)$ を対角成分に持つ対角行列である。

次に, 通常の線形回帰モデルにおいて GRR 推定量を考察する際にしばしば利用される正準形式 (canonical form) を本稿でも扱おう。(Hoerl and Kennard (1970) も参照.)

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (4)$$

ここで,

$$\mathbf{A} := \mathbf{B}\mathbf{\Gamma}$$

であり, パラメータベクトルの間には,

$$\boldsymbol{\alpha} := \mathbf{\Gamma}'\mathbf{w} \quad (5)$$

なる関係があることに注意しよう。なお, このようなモデルに変換することは, 一般性を失うことなくいつでも行うことができることに注意しよう。

正準形式(4)における $\boldsymbol{\alpha}$ の OLS 推定量としては,

$$\hat{\boldsymbol{\alpha}} := (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y} = \boldsymbol{\Lambda}^{-1}\mathbf{A}'\mathbf{Y}$$

であり, GRR 推定量は,

$$\hat{\boldsymbol{\alpha}}(\mathbf{K}) := (\mathbf{A}'\mathbf{A} + \mathbf{K})^{-1}\mathbf{A}'\mathbf{Y} = (\boldsymbol{\Lambda} + \mathbf{K})^{-1}\mathbf{A}'\mathbf{Y} \quad (6)$$

で定義される. なお, 正準形式に対する GRR 推定量の幾何学的な説明を付録Bに与えるので参照されたい.

注意II.1 GRR 推定量 $\hat{\boldsymbol{w}}(\mathbf{K})$ はリッジ係数行列 \mathbf{K} を適当にとることによって以下のような推定量を得ることができることに注意しよう:

$$\hat{\boldsymbol{w}}(\mathbf{K}) = \begin{cases} \hat{\boldsymbol{w}}, & \mathbf{K} = \mathbf{O} \text{ のとき,} \\ \hat{\boldsymbol{w}}(k), & \mathbf{K} = k\mathbf{I}_{p+1} \text{ のとき} \end{cases}$$

ここで, \mathbf{O} は零行列, \mathbf{I}_{p+1} は単位行列であり,

$$\hat{\boldsymbol{w}}(k) := (\mathbf{B}'\mathbf{B} + k\mathbf{I}_{p+1})^{-1}\mathbf{B}'\mathbf{Y}; \text{ (ORR 推定量)}$$

である.

注意II.2 パラメータ間の関係(5)が推定量間にも成り立つことに注意しよう:

$$\hat{\boldsymbol{w}} = \boldsymbol{\Gamma}\hat{\boldsymbol{\alpha}}, \quad \hat{\boldsymbol{w}}(\mathbf{K}) = \boldsymbol{\Gamma}\hat{\boldsymbol{\alpha}}(\mathbf{K})$$

よって, 正準形式において考察した結果は, これらの関係を利用することによって, オリジナルの LBF モデルの場合に引き戻すことが可能であることに注意しよう.

III 一般化リッジ回帰推定量に対するリッジ係数の選択アルゴリズム

GRR 推定量 $\hat{\boldsymbol{\alpha}}(\mathbf{K})$ におけるリッジ係数 k_i , ($i=0, 1, \dots, p$) を選択するためのアイデアとしては, その総平均2乗誤差 (Total Mean Squared Error: TMSE) を最小にするようなものを選択することである⁶⁾. (Hoerl and Kennard (1970)などを参照.)

この方法は,

6) TMSE を含む推定量の評価基準に関しては付録Aを参照されたい.

$$\begin{aligned}
 \rho(k_0, k_1, \dots, k_p) &:= \text{TMSE}(\hat{\boldsymbol{\alpha}}(\mathbf{K})) \\
 &= \text{trace } \mathbf{V}(\hat{\boldsymbol{\alpha}}(\mathbf{K})) + \|\text{bias}(\hat{\boldsymbol{\alpha}}(\mathbf{K}))\|^2 \\
 &= \sigma^2 \sum_{i=0}^p \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=0}^p \frac{\alpha_i^2 k_i^2}{(\lambda_i + k_i)^2} \\
 &=: \sum_{i=0}^p \rho_i(k_i) \longrightarrow \min_{\substack{k_i > 0, \\ i=0, 1, \dots, p}}
 \end{aligned}$$

と定式化され、ここで、

$$\rho_i(k_i) := \text{MSE}(\hat{\alpha}(k_i)) = \frac{\sigma^2 \lambda_i}{(\lambda_i + k_i)^2} + \frac{\alpha_i^2 k_i^2}{(\lambda_i + k_i)^2}$$

は、 α_i に対する GRR 推定量 $\hat{\alpha}(k_i)$ の MSE である。TMSE $\rho(k_0, k_1, \dots, k_p)$ を最小にする $k_i (i=0, 1, \dots, p)$ は、方程式

$$\frac{\partial \rho(k_0, k_1, \dots, k_p)}{\partial k_i} = \frac{\partial \rho_i(k_i)}{\partial k_i} = 2 \frac{\lambda_i (\alpha_i^2 k_i - \sigma^2)}{(\lambda_i + k_i)^3} = 0, \quad i=0, 1, \dots, p \quad (7)$$

の解として以下のように陽に与えられる：

$$k_i = \frac{\sigma^2}{\alpha_i^2} =: k_i^*, \quad i=1, \dots, p \quad (8)$$

すなわち、

$$\begin{aligned}
 (k_0^*, k_1^*, \dots, k_p^*) &= \arg \min_{\substack{k_i > 0, \\ i=0, 1, \dots, p}} \rho(k_0, k_1, \dots, k_p) \\
 &= \arg \min_{\substack{k_i > 0, \\ i=0, 1, \dots, p}} \text{TMSE}(\hat{\boldsymbol{\alpha}}(\mathbf{K}))
 \end{aligned}$$

が成り立つ。ただし、(8)式で与えられた k_i^* は、未知の α_i と σ^2 に依存するので、それらの推定量である $\hat{\alpha}_i$ と $\hat{\sigma}^2 := \|\mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\alpha}}\|^2 / (n-p-1)$ を使って、

$$\hat{k}_i^* = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}, \quad i=0, 1, \dots, p \quad (9)$$

によって選択する。

この \hat{k}_i^* を正準形式(4)における GRR 推定量(6)におけるリッジ係数 k_i に挿入 (plug-in) することによって得られる推定量

$$\hat{\boldsymbol{\alpha}}^* := \hat{\boldsymbol{\alpha}}(\hat{\mathbf{K}}^*) = (\mathbf{A}'\mathbf{A} + \hat{\mathbf{K}}^*)^{-1} \mathbf{A}'\mathbf{Y} = (\mathbf{A} + \hat{\mathbf{K}}^*)^{-1} \mathbf{A}'\mathbf{Y} \quad (10)$$

は、正準形式に対する実行可能型一般化リッジ回帰 (Feasible GRR: FGRR) 推定量と呼ばれる。

この結果とパラメータ間の関係(5)を使うと、LBP モデルに対する FGRR 推定量は以下のように与えられる：

$$\begin{aligned}\hat{\boldsymbol{w}}^* &= \boldsymbol{\Gamma} \hat{\boldsymbol{\alpha}}^* = \boldsymbol{\Gamma} (\boldsymbol{\Lambda} + \hat{\mathbf{K}})^{-1} \mathbf{A}' \mathbf{Y} \\ &= (\mathbf{B}' \mathbf{B} + \boldsymbol{\Gamma} \hat{\mathbf{K}}^* \boldsymbol{\Gamma}')^{-1} \mathbf{B}' \mathbf{Y}\end{aligned}\quad (11)$$

注意 III.1 ここでは、GRR 推定量 $\hat{\boldsymbol{\alpha}}(\mathbf{K})$ の TMSE を考えたけれども、 $\boldsymbol{\Gamma}' \boldsymbol{\Gamma} = \mathbf{I}_{p+1}$ より、

$$\begin{aligned}\text{TMSE}(\hat{\boldsymbol{w}}(\mathbf{K})) &= \text{trace } \mathbf{V}(\hat{\boldsymbol{w}}(\mathbf{K})) + \|\text{bias}(\hat{\boldsymbol{w}}(\mathbf{K}))\|^2 \\ &= \text{trace } \mathbf{V}(\boldsymbol{\Gamma} \hat{\boldsymbol{\alpha}}(\mathbf{K})) + \|\text{bias}(\boldsymbol{\Gamma} \hat{\boldsymbol{\alpha}}(\mathbf{K}))\|^2 \\ &= \text{trace } \boldsymbol{\Gamma} \mathbf{V}(\hat{\boldsymbol{\alpha}}(\mathbf{K})) \boldsymbol{\Gamma}' + \|\boldsymbol{\Gamma} \text{bias}(\hat{\boldsymbol{\alpha}}(\mathbf{K}))\|^2 \\ &= \text{TMSE}(\hat{\boldsymbol{\alpha}}(\mathbf{K})) = \rho(k_0, k_1, \dots, k_p)\end{aligned}$$

となり、TMSE で推定精度を評価する限り、 $\hat{\boldsymbol{w}}(\mathbf{K})$ の TMSE と一致することに注意しよう。なお、この結果は、GRR 推定量に関する議論に際して相關形式が利用されること理由でもある。また、同様の結果が、FGRR 推定量に関しても以下のように成り立つことに注意しよう：

$$\text{TMSE}(\hat{\boldsymbol{w}}^*) = \text{TMSE}(\hat{\boldsymbol{\alpha}}^*)$$

IV 実行可能型一般化リッジ回帰推定量のモーメント

誤差の分布に関する仮定(3)のもとで、正準形式(4)におけるパラメータベクトル $\boldsymbol{\alpha}$ に関する OLS 推定量の分布は以下のように与えられる：

$$\hat{\boldsymbol{\alpha}} \sim \mathbf{N}(\boldsymbol{\alpha}, \sigma^2 \boldsymbol{\Lambda}^{-1})$$

これより、OLS 推定量ベクトルの各成分は、

$$\hat{\alpha}_i \underset{\text{ind.}}{\sim} \mathbf{N}\left(\alpha_i, \frac{\sigma^2}{\lambda_i}\right), \quad i=0, 1, \dots, p$$

となることに注意しよう。ただし、 $X_i \underset{\text{ind.}}{\sim} F_i$ は確率変数 X_i が分布 F_i に独立に従うことを表す記号である。

一方、GRR 推定量の分布は以下のように与えられる：

$$\hat{\boldsymbol{\alpha}}(\mathbf{K}) \sim \mathbf{N}((\boldsymbol{\Lambda} + \mathbf{K})^{-1} \boldsymbol{\Lambda} \boldsymbol{\alpha}, \boldsymbol{\sigma}^2 (\boldsymbol{\Lambda} + \mathbf{K})^{-1} \boldsymbol{\Lambda} (\boldsymbol{\Lambda} + \mathbf{K})^{-1})$$

よって、OLS 推定量と同様に、GRR 推定量の成分は、

$$\hat{\alpha}_i(k_i) \stackrel{\text{ind.}}{\sim} \mathbf{N}\left(\frac{\lambda_i}{(\lambda_i + k_i)} \alpha_i, \frac{\lambda_i}{(\lambda_i + k_i)^2} \sigma^2\right), \quad i=0, 1, \dots, p$$

となる。

以上のことから、OLS 推定量 $\hat{\alpha}_i$ と GRR 推定量 $\hat{\alpha}_i(k_i)$ は独立に正規分布に従うことがわかったけれども、残念ながら FGRR 推定量(10)の分布は正規分布に従わないし、さらに互いに独立でもなく、正確な分布を求めることは難しい。そこで、本稿では線形回帰モデルに関するこれまでの結果を援用することによって、FGRR 推定量の推定精度の評価に欠かせないモーメントを以下に考察する。

まず、Dwivedi ら (1980) では、FGRR 推定量 $\hat{\alpha}_i^*$ ($i=0, 1, \dots, p$) の 1, 2 次モーメントが以下のように与えられている：

$$E(\hat{\alpha}_i^*) = \alpha_i \sum_{\ell=0}^{\infty} \left(\frac{\nu-1}{\nu}\right)^{\ell} \sum_{m=0}^{\infty} \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{5}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i), \quad (12)$$

$$E(\hat{\alpha}_i^{*2}) = \alpha_i^2 \frac{1}{2\delta_i} \sum_{\ell=0}^{\infty} (\ell+1) \left(\frac{\nu-1}{\nu}\right)^{\ell} \times \sum_{m=0}^{\infty} (2m+1) \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{7}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i) \quad (13)$$

ここで、 $\nu := n - p - 1$ は自由度であり、 $\delta_i := \alpha_i^2 / (\sigma^2 / \lambda_i) / 2$ は非心度 (non-centrality parameter) である。また、 $B(\cdot, \cdot)$ はベータ関数 (beta function)：

$$B(x, y) := \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

である。さらに、 $p(m; \delta_i)$ はポアソン分布 $\mathbf{Po}(\delta_i)$ の確率関数である：

$$p(m; \delta_i) := \frac{\delta_i^m}{m!} e^{-\delta_i}$$

次に, FGRR 推定量 $\hat{\alpha}_i^*$, $\hat{\alpha}_j^*$ ($i \neq j$) のクロスモーメントは, Jimichi (2008) で以下のように与えられている:

$$\begin{aligned} E(\hat{\alpha}_i^* \hat{\alpha}_j^*) &= \alpha_i \alpha_j \sum_{\ell_1=0}^{\infty} \sum_{\ell_2=0}^{\infty} \left(\frac{\nu-1}{\nu} \right)^{\ell_1+\ell_2} \\ &\quad \times \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \frac{\mathcal{B}_{\ell_1+1 \ell_2+1} \left(m_1 + \frac{5}{2}, m_2 + \frac{5}{2}, \ell_1 + \ell_2 + \frac{\nu}{2} \right)}{B \left(m_1 + \frac{3}{2}, m_2 + \frac{3}{2}, \frac{\nu}{2} \right)} \\ &\quad \times p(m_1; \delta_i) p(m_2; \delta_j) \end{aligned} \quad (14)$$

ここで, $\Gamma(x)$ をガンマ関数とすると,

$$B \left(m_1 + \frac{3}{2}, m_2 + \frac{3}{2}, \frac{\nu}{2} \right) = \frac{\Gamma \left(m_1 + \frac{3}{2} \right) \Gamma \left(m_2 + \frac{3}{2} \right) \Gamma \left(\frac{\nu}{2} \right)}{\Gamma \left(m_1 + m_2 + 3 + \frac{\nu}{2} \right)}$$

であり,

$$\begin{aligned} &\mathcal{B}_{\ell_1+1 \ell_2+1} \left(m_1 + \frac{5}{2}, m_2 + \frac{5}{2}, \ell_1 + \ell_2 + \frac{\nu}{2} \right) \\ &= \iint_{\mathcal{D}} \frac{1}{(1-t_1)^{\ell_2+1}} \frac{1}{(1-t_2)^{\ell_1+1}} t_1^{m_1+\frac{5}{2}-1} t_2^{m_2+\frac{5}{2}-1} (1-t_1-t_2)^{\ell_1+\ell_2+\frac{\nu}{2}-1} dt_1 dt_2 \end{aligned}$$

である. なお, 領域 $\mathcal{D} := \{(t_1, t_2) \mid t_1, t_2 > 0, t_1 + t_2 < 1\}$ は単体 (simplex) であることに注意しよう.

これらの結果を使うと, 正準形式に対する FGRR 推定量 (ベクトル) $\hat{\alpha}^*$ の平均ベクトルと分散共分散行列は, それぞれ, 以下のように与えられる:

$$E(\hat{\alpha}^*) = \mathbf{A}\boldsymbol{\mu}_1, \quad V(\hat{\alpha}^*) = \mathbf{A}(\mathbf{M}_2 - \boldsymbol{\mu}_1 \boldsymbol{\mu}_1') \mathbf{A}$$

ここで, $\mathbf{A} := \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_p)$ であり,

$$\boldsymbol{\mu}_1 := \begin{bmatrix} \frac{\mu_{10}}{\alpha_1} \\ \mu_{11} \\ \vdots \\ \mu_{1p} \end{bmatrix}, \quad \mathbf{M}_2 := \begin{bmatrix} \mu_{20} & \mu_{1101} & \cdots & \mu_{110p} \\ \mu_{1110} & \mu_{21} & \cdots & \mu_{111p} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{11p0} & \mu_{11p1} & \cdots & \mu_{2p} \end{bmatrix}$$

とおいた。なお,

$$\mu_{1i} := \mu_1(\delta_i, \nu) = \sum_{\ell=0}^{\infty} \left(\frac{\nu-1}{\nu} \right)^{\ell} \sum_{m=0}^{\infty} \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{5}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i), \quad (15)$$

$$\begin{aligned} \mu_{2i} := \mu_2(\delta_i, \nu) &= \frac{1}{2\delta_i} \sum_{\ell=0}^{\infty} (\ell+1) \left(\frac{\nu-1}{\nu} \right)^{\ell} \\ &\times \sum_{m=0}^{\infty} (2m+1) \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{7}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i), \end{aligned} \quad (16)$$

$$\begin{aligned} \mu_{11ij} := \mu_{11}(\delta_i, \delta_j, \nu) &= \sum_{\ell_1=0}^{\infty} \sum_{\ell_2=0}^{\infty} \left(\frac{\nu-1}{\nu} \right)^{\ell_1+\ell_2} \\ &\times \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \frac{\mathcal{B}_{\ell_1+1, \ell_2+1}\left(m_1 + \frac{5}{2}, m_2 + \frac{5}{2}, \ell_1 + \ell_2 + \frac{\nu}{2}\right)}{B\left(m_1 + \frac{3}{2}, m_2 + \frac{3}{2}, \frac{\nu}{2}\right)} \\ &\times p(m_1; \delta_i) p(m_2; \delta_j) \end{aligned} \quad (17)$$

である。

さらに、この結果から、LBF モデルに対する FGRR 推定量 (ベクトル) $\hat{\boldsymbol{w}}^*$ の平均ベクトルと分散共分散行列は、それぞれ、以下のように与えられる:

$$E(\hat{\boldsymbol{w}}^*) = \Gamma \mathbf{A} \boldsymbol{\mu}_1, \quad V(\hat{\boldsymbol{w}}^*) = \Gamma \mathbf{A} (\mathbf{M}_2 - \boldsymbol{\mu}_1 \boldsymbol{\mu}_1') \mathbf{A} \Gamma'$$

注意IV.1 ($\alpha_i=0$ の場合の FGRR 推定量のモーメント) 正準形式における回帰係数について $\alpha_i=0$ となる場合は、GRR 推定量のリッジ係数に関する最良値 $k_i^* = \sigma^2/\alpha_i^2$ は定義できないけれども⁷⁾、以下に示すように FGRR 推定量のモーメントは矛盾なく定まる (well-defined) ことに注意しよう。

まず,

7) この場合、MSE の導関数が k_i の単調減少関数となることから、しいて書くとすれば $k_i^* = \infty$ となる。

$$\alpha_i = 0 \implies \delta_i = \frac{1}{2} \frac{\alpha_i^2}{\sigma^2/\lambda_i} = 0$$

であることと、ポアソン分布の確率関数 $p(m; \delta_i)$ について

$$p(m; 0) = \begin{cases} 1, & m=0, \\ 0, & m \neq 0 \end{cases}$$

が成り立つことに注意すると、1次モーメント

$$E(\hat{\alpha}_i^*) = \alpha_i \mu_{1i}(\delta_i, \nu)$$

において、 $\mu_{1i}(\delta_i, \nu)$ が任意の $\delta_i \geq 0$ と任意の $\nu \geq 1$ に対して有限の値で定まるので、 $\alpha_i = 0$ に対して、

$$E(\hat{\alpha}_i^*) = 0 \times \mu_{1i}(0, \nu) = 0$$

となる。この結果は、 $\hat{\alpha}_i^*$ が $\alpha_i (=0)$ の不偏推定量であることを示していることに注意しよう。

一方、2次モーメントに関しては、

$$E(\hat{\alpha}_i^{*2}) = \alpha_i^2 \mu_{2i}(\delta_i, \nu)$$

$$= \frac{\sigma^2}{\lambda_i} \sum_{\ell=0}^{\infty} (\ell+1) \left(\frac{\nu-1}{\nu} \right)^\ell \sum_{m=0}^{\infty} (2m+1) \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{7}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i)$$

が成り立つことに注意すると⁸⁾、 $\alpha_i = 0$ ($\implies \delta_i = 0$) に対して、

$$E(\hat{\alpha}_i^{*2}) = \frac{\sigma^2}{\lambda_i} \sum_{\ell=0}^{\infty} (\ell+1) \left(\frac{\nu-1}{\nu} \right)^\ell \frac{B\left(\ell + \frac{\nu}{2}, \frac{7}{2}\right)}{B\left(\frac{\nu}{2}, \frac{3}{2}\right)}$$

となる。ここで、

$$\bar{\mu}_2(\nu) := \sum_{\ell=0}^{\infty} (\ell+1) \left(\frac{\nu-1}{\nu} \right)^\ell \frac{B\left(\ell + \frac{\nu}{2}, \frac{7}{2}\right)}{B\left(\frac{\nu}{2}, \frac{3}{2}\right)} \quad (18)$$

8) とくに、 α_i が全体の係数には残らないことに注意しよう。

は、 ν に関して単調減少であることに注意しよう。(図1も参照.)

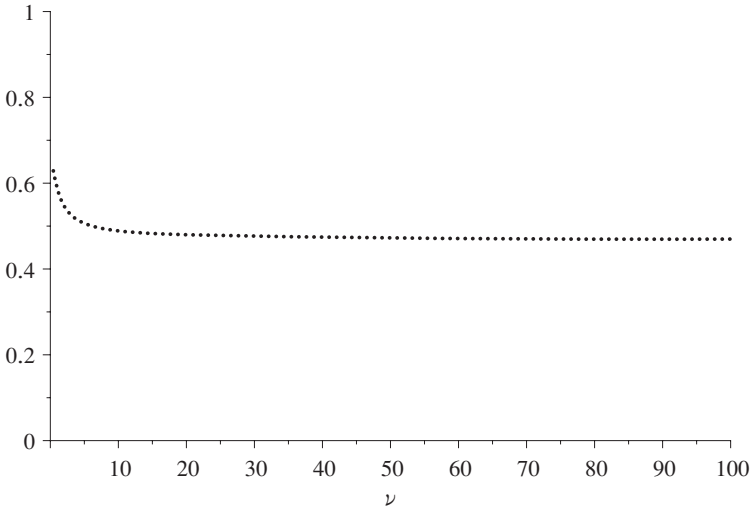


図1： $\tilde{\mu}_2(\nu)$ のプロット

1, 2次モーメントに関するこれらの結果から, $\alpha_i=0$ のときの FGRR 推定量の分散と平均2乗誤差が一致し,

$$V(\hat{\alpha}_i^*) = \text{MSE}(\hat{\alpha}_i^*) = \frac{\sigma^2}{\lambda_i} \tilde{\mu}_2(\nu)$$

で与えられることに注意しよう.

さらに, クロスモーメントに関しては,

$$E(\hat{\alpha}_i^* \hat{\alpha}_j^*) = \alpha_i \alpha_j \mu_{11}(\delta_i, \delta_j, \nu)$$

において, $\mu_{11}(\delta_i, \delta_j, \nu)$ が任意の $\delta_i, \delta_j \geq 0$ と任意の $\nu \geq 1$ に対して有限の値であることから, $\alpha_i=0$ または $\alpha_j=0$ に対して,

$$E(\hat{\alpha}_i^* \hat{\alpha}_j^*) = 0$$

となる.

V 平均 2 乗誤差基準

ここでは 前節で与えられた FGRR 推定量のモーメントに関する結果を使って、各種の平均 2 乗誤差 (Mean Squared Error: MSE) 基準と相対効率 (Relative Efficiency: REff) に関する数値評価を与える。

V.1 各種の平均 2 乗誤差基準

MSE 行列 (MSE matrix) は以下で与えられる：

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\alpha}}^*) &= \text{V}(\hat{\boldsymbol{\alpha}}^*) + \text{bias}(\hat{\boldsymbol{\alpha}}^*)\text{bias}(\hat{\boldsymbol{\alpha}}^*)' \\ &= \mathbf{A}(\mathbf{M}_2 - \boldsymbol{\mu}_1\boldsymbol{\mu}_1')\mathbf{A} + (\mathbf{M}_1 - \mathbf{I}_{p+1})\boldsymbol{\alpha}\boldsymbol{\alpha}'(\mathbf{M}_1 - \mathbf{I}_{p+1}) \end{aligned} \quad (19)$$

ここで、 $\mathbf{M}_i = \text{diag}(\mu_{10}, \mu_{11}, \dots, \mu_{1p})$ である。

MSE 行列の (i, j) 成分は、

$$[\text{MSE}(\hat{\boldsymbol{\alpha}}^*)]_{ij} = \begin{cases} \alpha_i^2(\mu_{2i} - 2\mu_{1i} + 1) = \text{MSE}(\hat{\alpha}_i^*) & ; \text{ if } i=j, \\ \alpha_i\alpha_j\{\mu_{11ij} - (\mu_{1i} + \mu_{1j}) + 1\} = \text{MCE}(\hat{\alpha}_i^*, \hat{\alpha}_j^*) & ; \text{ if } i \neq j \end{cases} \quad (20)$$

である。

さらに、OLS 推定量 $\hat{\alpha}_i$ の FGRR 推定量 $\hat{\alpha}_i^*$ に対する相対効率は以下のように定義される：

$$\text{REff}(\hat{\alpha}_i^*, \hat{\alpha}_i) := \frac{\text{MSE}(\hat{\alpha}_i^*)}{\text{MSE}(\hat{\alpha}_i)} = \frac{\text{MSE}(\hat{\alpha}_i^*)}{\text{V}(\hat{\alpha}_i)} = 2\delta_i(\mu_{2i} - 2\mu_{1i} + 1) \quad (21)$$

以下のことが成り立つことに注意しよう：

$$\text{REff}(\hat{\alpha}_i^*, \hat{\alpha}_i) < 1 \iff \text{MSE}(\hat{\alpha}_i^*) < \text{MSE}(\hat{\alpha}_i) \quad (22)$$

最後に、 $\hat{\boldsymbol{\alpha}}^*$ の総平均 2 乗誤差は MSE 行列のトレースで与えられることに注意しよう：

$$\begin{aligned} \text{TMSE}(\hat{\boldsymbol{\alpha}}^*) &:= \sum_{i=0}^p \text{MSE}(\hat{\alpha}_i(\hat{k}_i^*)) = \text{trace} \text{MSE}(\hat{\boldsymbol{\alpha}}^*) \\ &= \sum_{i=0}^p \alpha_i^2(\mu_{2i} - 2\mu_{1i} + 1) \end{aligned} \quad (23)$$

注意V.1 注意IV.1から、 $\alpha_i=0$ の場合の、MSE基準を考える。

まず、MSEは

$$\text{MSE}(\hat{\alpha}_i^*) = \frac{\sigma^2}{\lambda_i} \tilde{\mu}_2(\nu) (=V(\hat{\alpha}_i^*))$$

となり、続いて相対効率は、

$$\text{REff}(\hat{\alpha}_i^*, \hat{\alpha}_i) = \tilde{\mu}_2(\nu)$$

となることに注意しよう。

V.2 数値評価

ここでは、(21)式で定義されたOLS推定量 $\hat{\alpha}_i$ のFGRR推定量 $\hat{\alpha}_i^*$ に対する相対効率 $\text{REff}(\hat{\alpha}_i^*, \hat{\alpha}_i)$ を数値的に評価する。その際、1, 2次モーメントにおける級数 $\mu_{1i} := \mu_1(\delta_i, \nu)$, $\mu_{2i} := \mu_2(\delta_i, \nu)$ を数値的に評価するためには、以下の有限級数を利用する：

$$\mu_1(L, M, \delta_i, \nu) := \sum_{\ell=0}^L \left(\frac{\nu-1}{\nu} \right)^\ell \sum_{m=0}^M \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{5}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i), \quad (24)$$

$$\begin{aligned} \mu_2(L, M, \delta_i, \nu) := & \frac{1}{2\delta_i} \sum_{\ell=0}^L (\ell+1) \left(\frac{\nu-1}{\nu} \right)^\ell \\ & \times \sum_{m=0}^M (2m+1) \frac{B\left(\ell + \frac{\nu}{2}, m + \frac{7}{2}\right)}{B\left(\frac{\nu}{2}, m + \frac{3}{2}\right)} p(m; \delta_i) \end{aligned} \quad (25)$$

Jimichi (2008)でも同様の数値評価が $(L, M) = (100, 100)$ の場合に与えられているけれども、本稿では、さらに精度を高めるために $(L, M) = (700, 100)$ として再計算した結果を表1に与える⁹⁾。なお、図2, 3には相対効率のグラフを与えるのであわせて参照されたい。

9) $\nu \geq 20$ の場合に若干結果が異なる場合があることと、注意IV.1の結果をふまえて $\delta_i=0$ の場合も与えていることに注意しよう。

表 1 : OLS 推定量の FGRR 推定量に対する相対効率 (×100) :
(L, M) = (700, 100)

$\delta_i \backslash \nu$	1	2	5	10	15	20	30	40	50	100
0	62.500	56.194	50.975	48.919	48.198	47.831	47.459	47.272	47.159	46.932
0.01	63.060	56.847	51.701	49.674	48.962	48.600	48.233	48.048	47.937	47.713
0.05	65.253	59.398	54.542	52.626	51.952	51.609	51.262	51.086	50.981	50.768
0.10	67.892	62.465	57.955	56.171	55.542	55.222	54.898	54.735	54.636	54.437
0.20	72.843	68.209	64.341	62.802	62.259	61.982	61.701	61.559	61.473	61.301
0.50	85.388	82.690	80.393	79.460	79.127	78.955	78.781	78.693	78.640	78.533
0.70	92.120	90.404	88.907	88.283	88.058	87.941	87.823	87.762	87.726	87.652
0.90	97.778	96.844	95.984	95.609	95.471	95.398	95.324	95.286	95.263	95.217
1.00	100.257	99.649	99.056	98.785	98.683	98.630	98.574	98.546	98.529	98.494
2.00	115.883	116.927	117.683	117.947	118.034	118.077	118.120	118.141	118.154	118.179
5.00	123.249	123.500	123.612	123.636	123.642	123.644	123.646	123.646	123.647	123.647
10.00	117.501	116.323	115.391	115.028	114.901	114.835	114.769	114.736	114.716	114.675
20.00	110.531	109.169	108.218	107.873	107.755	107.695	107.635	107.605	107.587	107.550
50.00	104.676	103.875	103.363	103.187	103.127	103.098	103.068	103.053	103.044	103.026

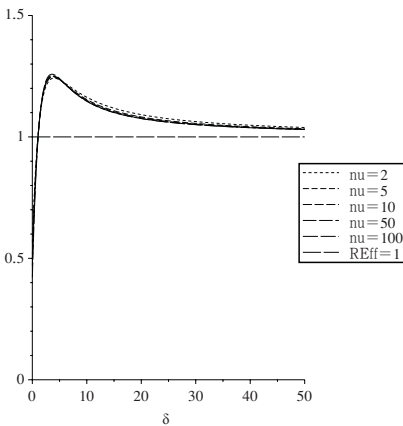


図 2 : OLS 推定量に対する FGRR 推定量の相対効率 : $0 \leq \delta_i \leq 50$

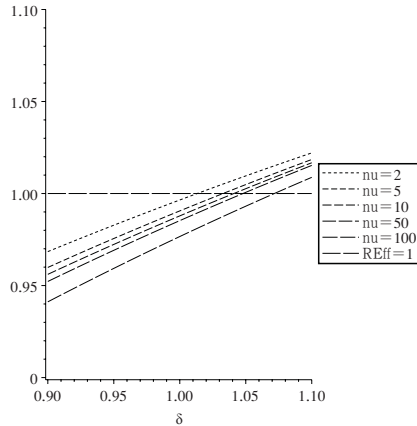


図 3 : OLS 推定量に対する FGRR 推定量の相対効率 : $0.9 \leq \delta_i \leq 1.1$

VI 例

この節では、LBF モデルの例¹⁰⁾を与える。ここでは、共変量 $X=x$ が与えられたときの応答変数 Y の条件付き平均が

$$E(Y|x) = \sin(x) \text{ (正弦関数)}$$

与えられている場合を考える。正弦関数は $x=0$ において 9 次の項までテイラー展開することによって、

$$\begin{aligned} \sin(x) &\simeq x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \frac{1}{5040}x^7 + \frac{1}{362880}x^9 \\ &= x - 0.1666666667x^3 + 0.008333333333x^5 - 0.0001984126984x^7 \\ &\quad + 0.000002755731922x^9 \end{aligned}$$

と多項式で (関数) 近似されるため¹¹⁾、LBF モデルとして、以下の多項式回帰モデル (polynomial regression model) を考えることが可能である。

$$y = \sum_{j=0}^p w_j b_j(x) + \epsilon = \sum_{j=0}^p w_j x^j + \epsilon$$

ここで、線形基底関数として、

$$b_j(x) = x^j; \text{ 巾関数}$$

をとった。

このモデルにもとづく以下のデータセットが得られたとする¹²⁾；

i	1	2	3	4	5	6	7	8	9	10	11
x_i	-3.14	-2.51	-1.88	-1.26	-0.63	0.00	0.63	1.26	1.88	2.51	3.14
y_i	-0.68	-0.64	-1.46	-1.09	-1.29	0.62	0.52	1.22	1.00	0.66	-0.45

このデータセットを使用し、OLS 推定値 \hat{w}_j によって 10 次の標本回帰曲線を求めると以下のように与えられる (図 4 も参照.)：

10) Bishop (2006) の第 1 章を参考にした。

11) 逆に、正弦関数を無限級数で定義する方法をとると、有限項で近似しているとみることがができる。

12) 誤差分散は $\sigma^2 = 0.5^2$ としている。

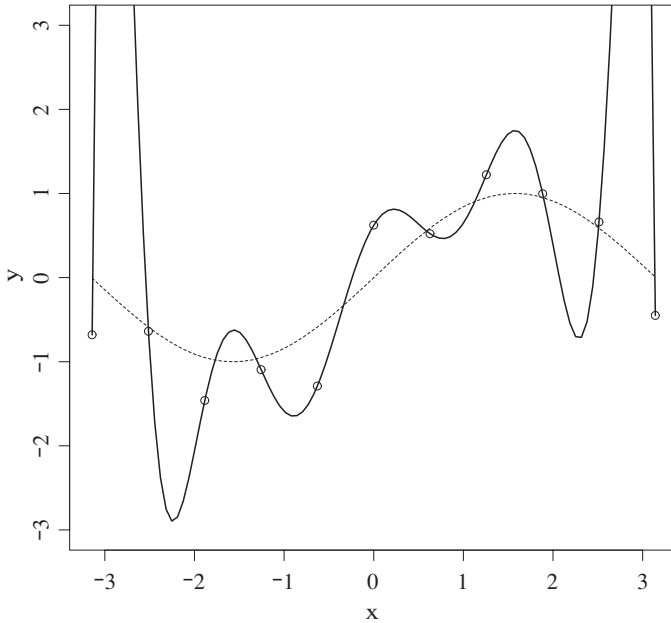


図4：正弦関数（点線）を条件付き平均（平均構造）に持つモデルの多項式回帰による当てはめ：標本回帰曲線（実線）が全てのデータ点を通っていることに注意

$$\begin{aligned}
 y &= \hat{w}_0 + \hat{w}_1 x + \hat{w}_2 x^2 + \cdots + \hat{w}_{10} x^{10} \\
 &= 0.624 + 1.736x - 3.996x^2 - 0.847x^3 + 4.229x^4 + 0.266x^5 - 1.535x^6 \\
 &\quad - 0.039x^7 + 0.217x^8 + 0.002x^9 - 0.010x^{10}
 \end{aligned}$$

この曲線は全てのデータ点を通るという意味ではデータに非常に良く当てはまっているけれども、大きく振動しており、条件付き平均（平均構造） $\sin(x)$ を説明するという意味では成功しているとはいえない。すなわち、過学習（過剰適合）（over-fitting）の典型的な例といえる。

ここで、次数を9として再度OLS推定値を求め、さらにFGRR推定量を求めることによって作った標本回帰曲線を以下に与える¹³⁾。

まず、OLS推定値にもとづく標本回帰曲線は以下のように与えられる：

$$\begin{aligned}
 y &= \hat{w}_0 + \hat{w}_1 x + \hat{w}_2 x^2 + \cdots + \hat{w}_9 x^9 \\
 &= 0.138 + 1.736x - 0.338x^2 - 0.847x^3 + 0.108x^4 + 0.266x^5 - 0.011x^6 \\
 &\quad - 0.039x^7 + 0.000x^8 + 0.002x^9
 \end{aligned}$$

この曲線をデータとともにプロットしたものが図5で与えられている。10次の場合と比べて幾分軽減しているけれども、やはり「揺らぎ」があることに注意しよう。

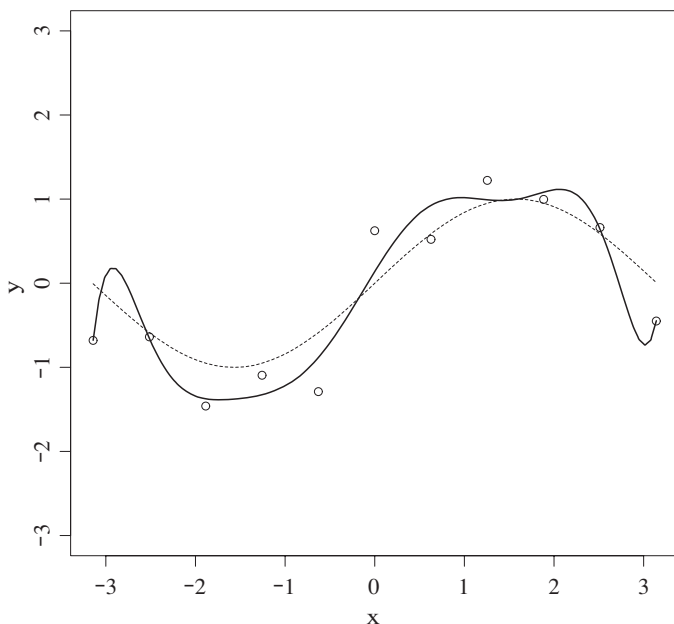


図5：OLS 推定値にもとづく標本回帰曲線（実線）

一方、FGRR 推定値にもとづく標本回帰曲線は以下のように与えられる：

$$\begin{aligned}
 y &= \hat{w}_0^* + \hat{w}_1^* x + \hat{w}_2^* x^2 + \cdots + \hat{w}_9^* x^9 \\
 &= 0.004 + 0.869x - 0.016x^2 - 0.041x^3 + 0.002x^4 + 0.061x^5 - 0.000x^6
 \end{aligned}$$

13) 0.000という表記がいくつかあるけれども、小数点以下3桁までに丸めた場合に零ということであり、正確な意味で零という意味でないことに注意しよう。

$$-0.008x^7 + 0.000x^8 + 0.000x^9$$

図5と同様に、この曲線をプロットしたものが図6で与えられている。OLS推定値による曲線よりも「揺らぎ」が軽減しており、さらに平均構造（正弦関数）をうまく説明したのものとなっていることに注意しよう。

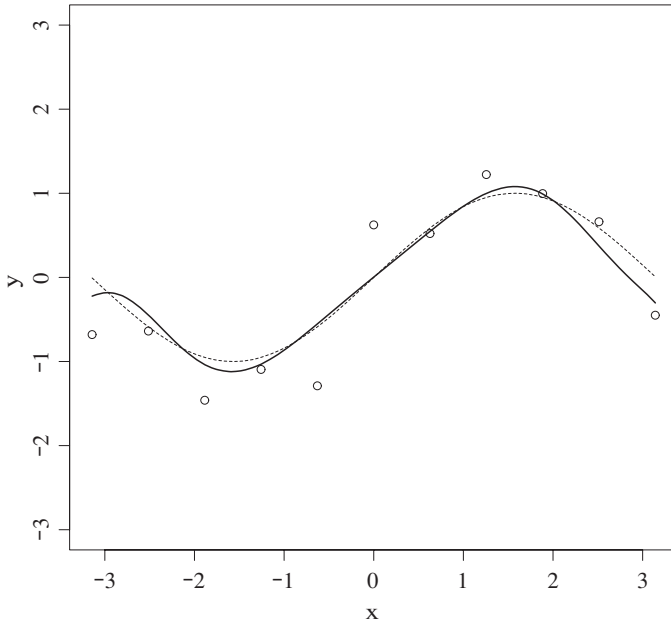


図6：FGRR推定値にもとづく標本回帰曲線（実線）

VII おわりに

本稿では、線形回帰モデルに関するこれまでの結果を援用することによって、線形基底関数モデルにおける重みベクトルに対する実行可能型一般化リッジ回帰推定量の正確なモーメントを与えるとともに、この結果を利用して各種の平均2乗誤差基準に関する結果も与えた。なお、数値評価として Jimichi (2008) で与えられた相対効率の結果を精度改良した結果も与えた。さらに、線形基底関数モデルの例として多項式回帰モデルを取り上げ、最小自乗推定

量を利用した場合よりも実行可能型一般化リッジ回帰推定量を利用したものが、真の平均構造をより忠実に再現できる場合を例示した。

今後の課題としては、B-スプライン関数、シグモイド関数（ニューラルネット）、動径基底関数などを用いた線形基底関数モデルの他のバリエーションを考察することと、これらのモデルを用いた実証分析への応用を検討することがあろう。

（筆者は関西学院大学商学部教授）

謝辞

本研究の一部は文部科学省科学研究費（基盤研究（C）課題番号：22540162, 研究代表者：北原和明）の助成により行った。ここに感謝の意を表するものである。

参考文献

- C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.
- T. D. Dwivedi, V. K. Srivastava, and R. L. Hall. Finite sample properties of ridge estimators. *Technometrics*, Vol. 22, pp. 205-212, 1980.
- R. F. Gunst. Regression analysis with multicollinear predictor variables: Definition, detection, and effects. *Communication in Statistics. Theory and Method*, Vol. 12, pp. 2217-2260, 1983.
- L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002.
- A. E. Hoerl and R. W. Kennard. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, Vol. 12, pp. 55-67, 1970.
- N. Inagaki. Two errors in statistical model fitting. *Annals of the Institute of Statistical Mathematics*, pp. 131-152, 1977.
- M. Jimichi. *Improvement of regression estimators by shrinkage under multicollinearity and its feasibility*. Ph. D. dissertation, Osaka University, 2005.
- M. Jimichi. Exact moments of feasible generalized ridge regression estimator and numerical evaluations. *Journal of the Japanese Society of Computational Statistics*, Vol. 21, pp. 1-20, 2008.
- S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling*. Springer-Verlag, 2007.
- V. K. Srivastava and A. Chaturvedi. Some properties of the distribution of an operational ridge estimator. *Metrika*, Vol. 30, pp. 227-237, 1983.
- 稲垣宣生. 数理統計学. 裳華房, 2003.

付録

A 推定量の評価基準

一般に、 θ をパラメータベクトルとし、その推定量 $\hat{\theta}$ の推定精度の評価基準として、平均 2 乗誤差 (Mean Squared Error: MSE) 行列：

$$\begin{aligned} \text{MSE}(\hat{\theta}) &:= E(\hat{\theta} - \theta)(\hat{\theta} - \theta)' \\ &= V(\hat{\theta}) + \text{bias}(\hat{\theta})\text{bias}(\hat{\theta})' \end{aligned}$$

が考えられる。ここで、

$$\begin{aligned} V(\hat{\theta}) &:= E(\hat{\theta} - E(\hat{\theta}))(\hat{\theta} - E(\hat{\theta}))'; \text{ 分散共分散行列,} \\ \text{bias}(\hat{\theta}) &:= E(\hat{\theta}) - \theta; \text{ バイアスベクトル} \end{aligned}$$

に注意しよう。MSE 行列の各成分は、

$$[\text{MSE}(\hat{\theta})]_{ij} = \begin{cases} E(\hat{\theta}_i - \theta_i)^2 =: \text{MSE}(\hat{\theta}_i), & i=j, \\ E(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) =: \text{MCE}(\hat{\theta}_i, \hat{\theta}_j), & i \neq j. \end{cases} \quad (26)$$

ここで、 $\text{MSE}(\hat{\theta}_i)$ は推定量 $\hat{\theta}_i$ の MSE であり、 $\text{MCE}(\hat{\theta}_i, \hat{\theta}_j)$ は推定量 $\hat{\theta}_i$ と $\hat{\theta}_j$ 間の平均交差誤差 (Mean Cross Error: MCE) と呼ばれる。

一般に、GRR 推定量などを含む縮小推定量 (shrinkage estimator) の推定精度を評価する基準としては、以下の総平均 2 乗誤差 (Total Mean Squared Error: TMSE) が利用される：

$$\text{TMSE}(\hat{\theta}) := \sum_i \text{MSE}(\hat{\theta}_i) \quad (27)$$

TMSE は以下のようにも表現できることに注意しよう：

$$\begin{aligned} \text{TMSE}(\hat{\theta}) &= \text{trace} \text{MSE}(\hat{\theta}) \\ &= E(\hat{\theta} - \theta)'(\hat{\theta} - \theta) = E\|\hat{\theta} - \theta\|^2 \\ &= \text{trace} V(\hat{\theta}) + \|\text{bias}(\hat{\theta})\|^2 \end{aligned} \quad (28)$$

$$(29)$$

B 一般化リッジ回帰推定量の導出と幾何学的説明

正準形式(4)における α の GRR 推定量は、以下のようにパラメータベクトルに関する制約を持つ正則化最小自乗推定 (regularized least squares estima-

tion) の枠組みから導出されることに注意しよう。(Inagaki (1977) も参照.)

まず、通常の最小自乗法は誤差平方和：

$$\Delta^2(\boldsymbol{\alpha}) := \|\boldsymbol{\varepsilon}\|^2 = \|\mathbf{Y} - \mathbf{A}\boldsymbol{\alpha}\|^2$$

を最小にする $\boldsymbol{\alpha}$ を求めるに対して、パラメータベクトル $\boldsymbol{\alpha}$ の最大値ノルム：

$$\|\boldsymbol{\alpha}\|_\infty := \max\{|\alpha_0|, |\alpha_1|, \dots, |\alpha_p|\}$$

に $c > 0$ として以下の制約を置くことを考える：

$$\begin{aligned} \|\boldsymbol{\alpha}\|_\infty \leq c &\iff |\alpha_0| \leq c \wedge |\alpha_1| \leq c \wedge \dots \wedge |\alpha_p| \leq c \\ &\iff \alpha_0^2 \leq c^2 \wedge \alpha_1^2 \leq c^2 \wedge \dots \wedge \alpha_p^2 \leq c^2 \end{aligned}$$

よって、正則化最小自乗法は以下のように定式化される：

$$\Delta^2(\boldsymbol{\alpha}) \longrightarrow \min_{\|\boldsymbol{\alpha}\|_\infty \leq c}$$

ラグランジュの未定乗数法によってこの問題を解く。まず、

$$\mathbf{g}(\boldsymbol{\alpha}) := [\alpha_0^2 - c^2, \alpha_1^2, \dots, \alpha_p^2 - c^2]'$$

とおき、 $\mathbf{k} := [k_0, k_1, \dots, k_p]'$ をラグランジュ乗数として、

$$L(\boldsymbol{\alpha}, \mathbf{k}) := \Delta^2(\boldsymbol{\alpha}) + \mathbf{k}'\mathbf{g}(\boldsymbol{\alpha})$$

と定義し、これを $\boldsymbol{\alpha}$ と \mathbf{k} でそれぞれ微分したものを $\mathbf{0}$ とおくと、

$$\begin{cases} \frac{\partial L(\boldsymbol{\alpha}, \mathbf{k})}{\partial \boldsymbol{\alpha}} = \frac{\partial \Delta^2(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{\partial \mathbf{g}'(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \mathbf{k} = \mathbf{0} \\ \frac{\partial L(\boldsymbol{\alpha}, \mathbf{k})}{\partial \mathbf{k}} = \frac{\partial \mathbf{k}'}{\partial \mathbf{k}} \mathbf{g}(\boldsymbol{\alpha}) = \mathbf{0} \end{cases}$$

$$\iff \begin{cases} -2\mathbf{A}'(\mathbf{Y} - \mathbf{A}\boldsymbol{\alpha}) + 2\mathbf{K}\boldsymbol{\alpha} = \mathbf{0} \\ \mathbf{g}(\boldsymbol{\alpha}) = \mathbf{0} \end{cases}$$

より、

$$(\mathbf{A}'\mathbf{A} + \mathbf{K})\boldsymbol{\alpha} = \mathbf{A}'\mathbf{Y}$$

を $\boldsymbol{\alpha}$ に関して解くと、

$$\boldsymbol{\alpha} = (\mathbf{A}'\mathbf{A} + \mathbf{K})^{-1}\mathbf{A}'\mathbf{Y} = (\boldsymbol{\Lambda} + \mathbf{K})^{-1}\mathbf{A}'\mathbf{Y} =: \hat{\boldsymbol{\alpha}}(\mathbf{K}) \quad (30)$$

となり GRR 推定量が導出できる。ここで、

$$\frac{\partial \Delta^2(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = -2\mathbf{A}'(\mathbf{Y} - \mathbf{A}\boldsymbol{\alpha}),$$

$$\frac{\partial \mathbf{g}'(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \mathbf{k} = 2 \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_p) \mathbf{k} = 2[k_0 \alpha_0, k_1 \alpha_1, \dots, k_p \alpha_p]' = 2\mathbf{K}\boldsymbol{\alpha},$$

$$\frac{\partial \mathbf{k}'}{\partial \mathbf{k}} = \mathbf{I}_{p+1}$$

を使った。OLS 推定量 $\hat{\boldsymbol{\alpha}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y} = \boldsymbol{\Lambda}^{-1}\mathbf{A}'\mathbf{Y}$ に注意すると、(30)式は、

$$\boldsymbol{\alpha} = (\mathbf{I} + \boldsymbol{\Lambda}^{-1}\mathbf{K})^{-1}\hat{\boldsymbol{\alpha}} \iff \alpha_i = \frac{\lambda_i}{\lambda_i + k_i} \hat{\alpha}_i, \quad i=0, 1, \dots, p$$

となる。一方、任意の $i=0, 1, \dots, p$ に対して、

$$\mathbf{g}(\boldsymbol{\alpha}) = \mathbf{0} \iff \alpha_i^2 = c^2 \iff |\alpha_i| = c (> 0)$$

が成り立つため、

$$\begin{aligned} |\alpha_i| = \frac{\lambda_i}{\lambda_i + k_i} |\hat{\alpha}_i| &\iff c = \frac{\lambda_i}{\lambda_i + k_i} |\hat{\alpha}_i| \\ &\iff (\lambda_i + k_i)c = \lambda_i |\hat{\alpha}_i| \\ &\iff k_i = \frac{\lambda_i}{c} (|\hat{\alpha}_i| - c) \end{aligned}$$

となる。ただし、 $k_i \geq 0$ に注意すると、 $c > 0$, $\lambda_i > 0$ より、

$$k_i = \begin{cases} \frac{\lambda_i}{c} (|\hat{\alpha}_i| - c), & |\hat{\alpha}_i| \geq c \text{ のとき,} \\ 0, & |\hat{\alpha}_i| < c \text{ のとき} \end{cases}$$

とおく必要がある。

さらに、この結果を GRR 推定量に代入すると、結局、

$$\hat{\alpha}_i(k) = \frac{\lambda_i}{\lambda_i + k_i} \hat{\alpha}_i = \begin{cases} c, & \hat{\alpha}_i > c \text{ のとき,} \\ \hat{\alpha}_i, & |\hat{\alpha}_i| \leq c \text{ のとき,} \\ -c, & \hat{\alpha}_i < -c \text{ のとき} \end{cases}$$

となる。

c を変化させたときに GRR 推定量がたどるパス（軌跡）を幾何学的に説明すると、誤差平方和を

$$\begin{aligned} \Delta^2(\boldsymbol{\alpha}) &= \|\mathbf{A}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\|^2 + \|\mathbf{Y} - \mathbf{A}\hat{\boldsymbol{\alpha}}\|^2 \\ &= (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})' \mathbf{A}'\mathbf{A}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) + \|\mathbf{e}\|^2 \end{aligned}$$

$$= \sum_{i=0}^p \frac{(\alpha_i - \hat{\alpha}_i)^2}{1/\lambda_i} + \sum_{i=1}^n e_i^2$$

のように OLS 推定量を中心とする楕円として表現したものと、パラメータの最大値ノルムによる制約 $\max\{|\alpha_0|, |\alpha_1|, \dots, |\alpha_p|\} \leq c$ の境界との共有点の軌跡によって与えられることに注意しよう．実際には、図7のように与えられる．

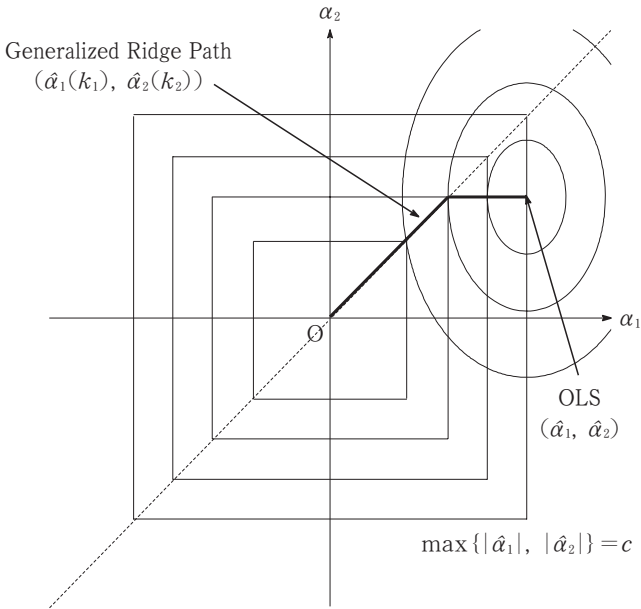


図7：GRR 推定量の幾何学的説明