

CIRJE-J-186

線形混合モデルの理論と応用 —特に小地域推定を巡って—

東京大学大学院経済学研究科
久保川達也

2007年10月

CIRJE ディスカッションペーパーの多くは
以下のサイトから無料で入手可能です。
http://www.e.u-tokyo.ac.jp/cirje/research/03research02dp_j.html

このディスカッション・ペーパーは、内部での討論に資するための未定稿の段階にある論文草稿である。著者の承諾なしに引用・複写することは差し控えられるたい。

Theory of Linear Mixed Models and its Applications to Small Area Estimation

Tatsuya Kubokawa

Linear mixed models (LMM) and the best linear unbiased predictor (BLUP) have received considerable attention in recent years from both theoretical and practical aspects. This article reviews the theory of LMM and illustrates how useful LMM and BLUP are through an example of the small area estimation. Linear mixed models for analyzing longitudinal data are also explained and an application to the posted land price data is given.

線形混合モデルの理論と応用 - 特に小地域推定を巡って -

久保川 達也* †

平成 19 年 10 月 10 日

概要

線形混合モデルの特徴は、観測値を共変量を用いて回帰するとき、個体や地域の違いを変量として組み入れ、それらの背後に共通な確率分布を想定して個体や地域の差異を推定している点である。全体の特性値だけでなく個体や地域ごとの特性値への関心が高まるにつれ、個々の差異を変量として捉えた線形混合モデルについての研究が盛んになり、このモデルの研究が始まった家畜育種学の分野はもとより医学・生物学分野から経済・教育など社会科学の分野、特に官庁統計分野での小地域推定において利用されている。本稿では、線形混合モデルとそこから導かれる経験最良線形不偏予測量について解説し、そのモデルがもっている予測精度を高めるための仕組みや経験最良線形不偏予測量の予測誤差の評価について小地域推定に焦点を当てて説明する。また経時測定データを解析するための線形混合モデルについても紹介し、地価公示価格データへの応用例を与える。

1 はじめに

線形混合モデル (Linear Mixed Model, LMM) と最良線形不偏予測量 (Best Linear Unbiased Predictor, BLUP) についての研究は C.R. Henderson の論文以来 50 年以上にわたって発展してきた。当初は、家畜育種学の分野で個体のもつ遺伝的能力などの推定を行うために研究されたが、次第に線形混合モデルの有用性が広く認識され、またベイズモデルとの関連においてベイズ推測の理論と計算方法についての顕著な発展に伴って、現在では実に広い分野で利用されている。LMM の離散分布への拡張である一般化線形混合モデル (Generalized Linear Mixed Model, GLMM) を含めれば、線形混合モデルに関する文献はかなりの量になっていることから、理論と応用の両面から関心が高いことがわかる。

LMM の応用例の一つに小地域推定の問題がある。これは標本調査に関連した問題で、通常は調査区全体の特性を調べるために標本調査が行われるが、そのデータを利用して地域ごとの特性値を推定したい状況がしばしば生ずる。例えば、得られたデータから各地域への予算配分の仕方を決めたり、政策を決定したりする場合がある。そのとき、狭い地域や人口が粗な地域に対しては十分なデータがとられていないため、その地域のデータだけでは特性値の十分な推測ができない。このような状況での推定問題を小地域推定という。この問題を解決する方法は、周辺地域のデータを組み込んで推定精度を高めることであり、どのような形でデータを取り込むかがポイントになる。そのために利用されるのが LMM であり、そのモデルから導かれる経験最良線形不偏予測量 (Empirical Best Linear Unbiased Predictor, EBLUP) が小地域の安定した推定値を与え

*113-0033 東京都文京区本郷 7-3-1 東京大学大学院経済学研究科

†E-Mail: tatsuya@e.u-tokyo.ac.jp

るのに役立つ。では、LMM がそのような性質をもつのはなぜであろうか。LMM は、基本的に共通母数に基づいて回帰する項と地域の差異を表す変量効果の項及び誤差項とから構成されている。すべての地域を通して回帰係数を共通に設定することによってすべてのデータをプールして安定した推定値を与えることができる。しかし、これだけでは地域の特徴や地域による差異を引き出すことができない。そこで地域の差異を変量効果としてモデルに取り込む。この効果を予測してやることにより、標本平均を縮小する作用が生ずることになる。LMM は、母数の共通化によるデータのプーリングと変量効果による標本平均の縮小作用を生み出すことのできるモデルであり、その結果生ずる予測量が EBLUP となる。したがって、EBLUP は、各々の地域の標本平均とプールされた回帰推定量との加重平均になっており、データ数が少ないときには標本平均をプールされた推定値の方向へ縮小することにより、推定精度の改善が図られている。

本稿では LMM の理論と小地域推定への応用について解説する。2 節では、LMM の紹介、混合モデル方程式と BLUP の説明、変量効果と共通母数の役割、分散成分を推定するための最尤法と制限最尤法についての解説を行う。3 節では、小地域推定の問題に焦点を絞り、予測精度を高めるために導出された EBLUP が実際どの程度推定誤差を改善しているのかについて、平均 2 乗誤差とその推定方法について説明する。また信頼区間の構成を行い、これらを用いた地価公示価格データへの応用を与える。4 節では、LMM の様々な応用や拡張について紹介する。経時測定データを解析するための LMM の紹介を行い、小地域推定のためのモデルの修正と上記の価格データへの適用結果について説明する。最後に GLMM への拡張、階層ベイズモデルへの拡張について若干の説明を与える。

なお、LMM や GLMM の解説書については、広津 (1992), McCulloch and Searle (2001), McCulloch (2003), 佐々木 (2007), 特に本格的なものとして Searle, Casella and McCulloch (1992), Demidenko (2004), 小地域推定に関するものとして Rao (2003) が挙げられるので参照してほしい。

2 線形混合モデルとその特徴

2.1 線形混合モデル

[1] 枝分かれ誤差回帰モデル. 線形混合モデルが使われている分野の一つに小地域推定がある。線形混合モデルをわかりやすく説明するために、その分野の啓蒙的な論文として知られる Battese, Harter and Fuller (1988) が取り上げた推定問題について紹介しよう。アイオワ州の k 個の郡について穀物(とうもろこし, 大豆等)の作付面積の調査がなされた。 k 個の郡それぞれをさらに約 250h の農作区画(segment)に細分し、その中から n_i 個の区画をランダムに抽出し、直接農家にインタビュー調査を行うことによってそれぞれの区画におけるトウモロコシの作付面積についてのデータをとった。 i 番目の郡における j 番目の区画に対するこのデータを y_{ij} で表すことにする。他方、人工衛星 LANDSAT からの観測により、約 0.45h のピクセル (picture element, 画像から識別する単位) に対していずれの穀物が作付けされているのかが識別され、 k 個の郡すべてにわたってこうした衛星データが補助情報として利用可能である。トウモロコシ, 大豆に関してピクセル単位での識別が可能であるとして、 i 番目の郡における j 番目の区画に対して、トウモロコシ, 大豆に識別されたピクセルの個数を x_{1ij}, x_{2ij} とすると、 y_i と (x_{1ij}, x_{2ij}) との間に線形関係が認められるため、

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$$

なるモデルが想定できる。ここで $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, $\mathbf{x}'_{ij}\boldsymbol{\beta} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2$ である。また u_{ij} は y_{ij} を \mathbf{x}_{ij} で説明したときの誤差項であるが、これが郡に依存する項 v_i と郡の差異に依らない項 e_{ij} に加法的に分解されて

$$(2.1) \quad u_{ij} = v_i + e_{ij}$$

と表されるとする。 v_i は地域の差異を表しているので地域効果と呼ばれる。これは未知母数もしくは確率変数として扱われ、それぞれ母数効果、変量効果と呼ばれる。どちらで扱うかは実際の問題に依存しており応用する際に検討する必要がある。地域の差異 v_i についてその背後に共通な分布が想定できるか否かに従って判断するのがよい。 v_i の背後に共通な分布が想定できる場合、すなわち v_i が変量効果として扱えるときには、 v_i, e_{ij} はすべて互いに独立な確率変数とし、それぞれ正規分布

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \quad e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$$

に従うとする。すると、線形モデルは

$$(2.2) \quad y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$$

と表される。 σ_v^2, σ_e^2 を一般に分散成分といい、特にこのモデルについては σ_v^2 を群間成分、 σ_e^2 を群内成分という。 $\boldsymbol{\beta}$ は回帰係数の未知母数ベクトルなので、このモデルの主要項は(母数効果) + (変量効果) の形に書かれており、(2.2) のようなモデルを一般に線形混合モデル (LMM) もしくは分散成分モデル (Variance Component Model) という。特に、(2.2) は誤差項が枝分かれ配置しているので、枝分かれ誤差回帰モデル (Nested Error Regression Model) ともいう。

[2] モデルの行列表現。モデルを行列を用いて表現しておくとう便利である。上の例では $\mathbf{x}_{ij}, \boldsymbol{\beta}$ は 3×1 のベクトルであるが、これ以降は $p \times 1$ のベクトルとして一般的な場合を扱うことにする。ベクトルと行列の大きさを適当に揃えて

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix}, \quad \mathbf{x}'_i = \begin{pmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{in_i} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_k \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

とし、 \mathbf{e} も \mathbf{y} と同様に $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})'$, $\mathbf{e} = (e'_1, \dots, e'_k)'$ と定義する。また、すべての成分が 1 の $n_i \times 1$ ベクトルを \mathbf{j}_{n_i} で表し、ブロック対角行列 $\text{block diag}(\cdot)$ を用いて $\mathbf{Z} = \text{block diag}(\mathbf{j}_{n_1}, \dots, \mathbf{j}_{n_k})$ とおき、 $\mathbf{v} = (v_1, \dots, v_k)'$ とおく。このときモデル (2.2) は、 $N = \sum_{i=1}^k n_i$ に対して

$$(2.3) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \\ \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_k), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}_N)$$

と表すことができる。 \mathbf{y}_i の共分散行列は $\text{Cov}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i(\sigma_e^2, \sigma_v^2) = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{J}_{n_i}$ と表される。ここで、 \mathbf{I}_{n_i} は $n_i \times n_i$ の単位行列、 $\mathbf{J}_{n_i} = \mathbf{j}_{n_i} \mathbf{j}'_{n_i}$ はすべての要素が 1 の $n_i \times n_i$ 行列である。従って、 \mathbf{y} の分散共分散行列は

$$\text{Cov}(\mathbf{y}) = \boldsymbol{\Sigma}(\sigma_e^2, \sigma_v^2) = \text{block diag}(\boldsymbol{\Sigma}_1(\sigma_e^2, \sigma_v^2), \dots, \boldsymbol{\Sigma}_k(\sigma_e^2, \sigma_v^2))$$

と書ける。

v_i を母数効果とした場合には $\text{Cov}(\mathbf{y}_i) = \sigma_e^2 \mathbf{I}_{n_i}$ となって \mathbf{y}_i の成分は互いに独立に分布するのに対して、変量効果とした場合 $\sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{J}_{n_i}$ という相関構造をもつことがわかる。一般に相関関係を利用してより推定精度の高い推定手法を導出することが可能になることは次の節で説明する。すなわち相関構造をどのように組み入れるかがより優れた推定を行う上で重要である。その際、誤差項の共分散行列に直接相関構造を埋め込んでしまえばよいわけであるが、変量効果を導入することによってモデルの意味が理解しやすくなるだけでなく、より豊かで複雑なモデルの構築も可能になる。さらに、変量効果の入っているモデルはベイズモデルの枠組みで捉えることができるため、たとえ複雑なモデルを作ったとしても、マルコフチェーン・モンテカルロ (MCMC) 法を利用して数値的な計算が可能になる。

[3] 一般的な線形混合モデル. (2.3) で記述されたモデルは、より一般的な線形混合モデル

$$(2.4) \quad \begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \\ \mathbf{v} &\sim \mathcal{N}_q(\mathbf{0}, \mathbf{G}), \quad \mathbf{e} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{R}) \end{aligned}$$

に拡張することができる。ここで、 \mathbf{y} は $N \times 1$ の観測データのベクトル、 \mathbf{X} は $N \times p$ の共変量からなる既知の行列、 \mathbf{Z} は $N \times q$ の既知の計画行列である。 \mathbf{y} の共分散行列は

$$(2.5) \quad \text{Cov}(\mathbf{y}) = \boldsymbol{\Sigma} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$$

と表される。共分散行列 \mathbf{G} , \mathbf{R} は一般に分散成分を含む母数を用いて表されるので、それらを $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$ として $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ と書く。

2.2 混合モデル方程式と最良線形不偏予測量 (BLUP)

[1] BLUP. 一般的な線形混合モデル (2.4) において観測できない母数効果 $\boldsymbol{\beta}$ と変量効果 \mathbf{v} の推定を考えよう。 \mathbf{v} は変量効果なので推定というよりは予測という方がふさわしい。共分散行列 \mathbf{G} , \mathbf{R} が既知の場合には \mathbf{v} の標準的な予測量は \mathbf{y} の線形関数として与えられる。線形で不偏な予測量の中で最良なものを最良線形不偏予測量 (BLUP) といい $\hat{\mathbf{v}}$ で表すことにする。また $\boldsymbol{\beta}$ の最良線形不偏推定量を $\hat{\boldsymbol{\beta}}$ で表すと、それらは次の連立方程式の解として与えられることが Henderson (1950) によって示された。

$$(2.6) \quad \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

これは、混合モデル方程式と呼ばれ、その解は

$$(2.7) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \quad \hat{\mathbf{v}} = \mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

で与えられる。ここで $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ は \mathbf{X} のランクが落ちている場合を考慮した $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ の一般化逆行列を表している。 $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の一般化最小 2 乗推定量 (GLS) であることがわかる。既知のベクトル $\mathbf{a} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^q$ に対して $\mu = \mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\mathbf{v}$ を推定したいときには、その BLUP は

$$(2.8) \quad \hat{\mu} = \mathbf{a}'\hat{\boldsymbol{\beta}} + \mathbf{b}'\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

で与えられることになる。

ここで，連立方程式 (2.6) の解が (2.7) で与えられることを確かめよう。まず，(2.6) の 2 番目の方程式 $Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{v} = Z'R^{-1}y$ より

$$(2.9) \quad \hat{v} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta)$$

と書ける。ここで， $(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}$ を変形すると，

$$\begin{aligned} & (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\ &= GZ'R^{-1} - G\{(Z'R^{-1}Z + G^{-1}) - G^{-1}\}(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\ &= GZ'R^{-1} - GZ'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \\ &= GZ'\{R^{-1} - R^{-1}Z(G^{-1} + Z'R^{-1}Z)^{-1}Z'R^{-1}\} \\ &= GZ'\Sigma^{-1} \end{aligned}$$

と書けることがわかる。最後の等式は，逆行列の計算でしばしば用いられる等式

$$(2.10) \quad \Sigma^{-1} = (ZGZ' + R)^{-1} = R^{-1} - R^{-1}Z(G^{-1} + Z'R^{-1}Z)^{-1}Z'R^{-1}$$

から従う。これを (2.9) に代入すると (2.7) の \hat{v} が得られることがわかる。次に，いま求めた \hat{v} を (2.6) の 1 番目の方程式 $X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{v} = X'R^{-1}y$ に代入して整理すると，

$$X'R^{-1}X\hat{\beta} + X'R^{-1}ZGZ'\Sigma^{-1}(y - X\hat{\beta}) = X'R^{-1}y$$

より，

$$X'R^{-1}(\Sigma - ZGZ')\Sigma^{-1}X\hat{\beta} = X'R^{-1}(\Sigma - ZGZ')\Sigma^{-1}y$$

となる。 $\Sigma = ZGZ' + R$ より， $R^{-1}(\Sigma - ZGZ') = I$ となるので，結局， $X'\Sigma^{-1}X\hat{\beta} = X'\Sigma^{-1}y$ となり，(2.7) の $\hat{\beta}$ が得られることが確かめられる。

[2] 混合モデル方程式の導出。混合モデル方程式 (2.6) の導出に関してはいくつかのアプローチが知られている。代表的なものに最尤法に基づいた方法と経験ベイズ法によるものがある。 y と v の同時密度関数は，基準化定数を除くと

$$|G|^{-1/2}|R|^{-1/2} \cdot \exp\left\{-\frac{1}{2}\begin{pmatrix} v \\ y - X\beta - Zv \end{pmatrix}' \begin{pmatrix} G^{-1} & 0 \\ 0 & R^{-1} \end{pmatrix} \begin{pmatrix} v \\ y - X\beta - Zv \end{pmatrix}\right\}$$

と書ける。 $\exp\{\cdot\}$ の中身を (-2) 倍したものを

$$h(\beta, v) = v'G^{-1}v + (y - X\beta - Zv)'R^{-1}(y - X\beta - Zv)$$

とおく。これを β と v に関して最小化するために β, v に関して偏微分すると

$$\begin{aligned} \frac{\partial h(\beta, v)}{\partial \beta} &= -2X'R^{-1}(y - X\beta - Zv), \\ \frac{\partial h(\beta, v)}{\partial v} &= 2G^{-1}v - 2Z'R^{-1}(y - X\beta - Zv), \end{aligned}$$

となり， $\partial h(\beta, v)/\partial \beta = 0$ ， $\partial h(\beta, v)/\partial v = 0$ の連立方程式を行列で表すと，(2.6) が得られる。これが最尤法に基づいた導出方法である。

もう1つの方法は、 y を与えたときの v の条件付き分布に基づいている。 (y, v) の共分散行列は

$$(2.11) \quad \text{Cov}(y, v) = \begin{pmatrix} \Sigma & ZG \\ GZ' & G \end{pmatrix}$$

で与えられるので、 y を与えたときの v の条件付き期待値は多変量正規分布の基本的な性質から

$$E[v|y] = GZ'\Sigma^{-1}(y - X\beta)$$

となる。この条件付き分布は、ベイズの枠組みでは、 y を与えたときの v の事後分布に相当しており、 $v|y \sim \mathcal{N}_q(GZ'\Sigma^{-1}(y - X\beta), G - GZ'\Sigma^{-1}ZG)$ で与えられる。また (2.10) を用いて y の周辺分布を計算すると、 $y \sim \mathcal{N}_N(X\beta, \Sigma)$ となることがわかる。周辺分布の密度は定数項を除いて

$$(2.12) \quad |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta)\right\}$$

と表されるので、周辺分布に基づいた β の最尤推定量は一般化最小2乗推定量 $\hat{\beta}$ に一致する。また v のベイズ推定量は事後分布の平均で与えられるので $GZ'\Sigma^{-1}(y - X\hat{\beta})$ がベイズ推定量になる。これに $\hat{\beta}$ を代入した $GZ'\Sigma^{-1}(y - X\hat{\beta})$ は経験ベイズ推定量と呼ばれるが、混合モデル方程式の解 \hat{v} に一致している。従って混合モデル方程式の解は経験ベイズ解として導出されることがわかる。

上の2つの方法の違いは、後者が事後分布の平均で推定するのに対して前者は事後分布のモードで推定している点である。正規分布の場合には両者が一致するので同じ解が得られたことになるが、一般には異なったものになり、前者はベイズ的最尤法と呼ばれる手法である。

観測できない変量を予測するためには (2.11) で与えられる相関関係が本質的であることを上で説明した。観測できなくても相関関係を利用して条件付き期待値で予測可能なわけである。このことは、広く用いられている考え方で、例えば、欠測値がある場合には条件付き期待値を用いることによって補完することができる。EM アルゴリズムや有限母集団の予測問題でも同様な方法が用いられている。

[3] 枝分かれ誤差回帰モデルにおける BLUP. 2.1 節で紹介したモデル (2.2) において、各郡におけるとうもろこしの平均的作付面積（農作区画単位）

$$\mu_i = \bar{x}'_i\beta + v_i$$

に対して BLUP を求めてみよう。ここで、 $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$ である。この場合、 $G(\sigma_v^2) = \sigma_v^2 \mathbf{I}_k$ 、 $\Sigma_i(\sigma_e^2, \sigma_v^2) = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_v^2 \mathbf{J}_{n_i}$ 、 $\Sigma(\sigma_e^2, \sigma_v^2) = \text{block diag}(\Sigma_1(\sigma_e^2, \sigma_v^2), \dots, \Sigma_k(\sigma_e^2, \sigma_v^2))$ となる。

$$\Sigma_i^{-1} = \frac{1}{\sigma_e^2} \left(\mathbf{I}_{n_i} - \frac{\sigma_v^2}{\sigma_e^2 + n_i \sigma_v^2} \mathbf{J}_{n_i} \right)$$

に注意し、 $\theta = \sigma_v^2/\sigma_e^2$ とおくと、 μ_i の BLUP $\hat{\mu}_i(\theta)$ は、(2.8) から

$$(2.13) \quad \hat{\mu}_i(\theta) = \bar{x}'_i \hat{\beta}(\theta) + \frac{\theta n_i}{1 + \theta n_i} \left\{ \bar{y}_i - \bar{x}'_i \hat{\beta}(\theta) \right\}$$

となる。ただし、 $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}$ であり、 β の GLS は次で与えられる。

$$\hat{\beta}(\theta) = \left\{ \sum_{i=1}^k \left(\mathbf{x}_i \mathbf{x}'_i - \frac{n_i^2 \theta}{1 + n_i \theta} \bar{\mathbf{x}}_i \bar{\mathbf{x}}'_i \right) \right\}^{-1} \sum_{i=1}^k \left(\mathbf{x}_i \mathbf{y}'_i - \frac{n_i \theta}{1 + n_i \theta} \bar{\mathbf{x}}_i \bar{y}_i \right)$$

2.3 変量効果と共通母数の役割

線形混合モデル (2.4) の特徴は、

$$(\text{データ}) = (\text{共通母数}) + (\text{変量効果}) + (\text{誤差項})$$

なる構造をしている点であり (共通母数) と (変量効果) がデータ解析においてどのように働くのか調べてみたい。モデル (2.2) において平均的作付面積 $\mu_i = \bar{x}'_i \beta + v_i$ の予測問題を取り上げてみると、各郡の標本平均 \bar{y}_i を用いて予測するのが基本であるが、 n_i が 1~5 程度であるため予測誤差が大きいという問題がある。これに対して、(2.13) で与えられる BLUP $\hat{\mu}_i(\theta)$ は \bar{y}_i と $\bar{x}'_i \hat{\beta}(\theta)$ との加重平均になっていることがわかる。 \bar{y}_i は直接取られたデータに基づいた平均値であるので、個々の郡の特徴を反映している。しかし n_i が小さいときには、 \bar{y}_i の予測誤差が問題となる。他方、 $\bar{x}'_i \hat{\beta}(\theta)$ は全データに基づいて構成されているので安定しているが、郡の特徴は \bar{y}_i ほど強くは現れないと考えられる。BLUP は、これらの点を考慮した方法であり、 n_i もしくは θ が小さければ \bar{y}_i を $\bar{x}'_i \hat{\beta}(\theta)$ の方向へ縮小することによって安定化を図っている。すなわち、 n_i が小さければ、データの不足を周辺もしくは全体のデータで補うことによって予測精度を高めていると解釈される。このようにして BLUP が小地域の予測問題に役立つことがわかる。言い換えれば、BLUP を生み出すところのモデルの形に、小地域の予測を効果的に行う仕組みが備わっていることになる。

[1] 変量効果と縮小推定. v_i を母数効果とし $\beta = 0$ としたときには、 μ_i の最適な推定量は \bar{y}_i となる。これに対して、 v_i を変量とすると、 (\bar{y}_i, v_i) の共分散行列が

$$\text{Cov}(\bar{y}_i, v_i) = \begin{pmatrix} \sigma_v^2 + \sigma_e^2/n_i & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2 \end{pmatrix}$$

となることからわかるように、 \bar{y}_i と v_i の間には相関が生ずる。このことから条件付き期待値が $E[v_i|\bar{y}_i] = \theta n_i(1 + \theta n_i)^{-1}(\bar{y}_i - \bar{x}'_i \beta)$ となり、 \bar{y}_i が $\bar{x}'_i \beta$ の方向へ縮小される。こうして、線形混合モデルにおいて変量効果が \bar{y}_i を縮小する作用を生むことがわかる。

[2] 共通母数によるデータのプーリング. モデル (2.2) からわかるように \bar{y}_i の期待値は $E[\bar{y}_i] = \bar{x}'_i \beta$ であり、これは i に依存しない共通な母数 β に基づいている。 β は全データ $\bar{y}_1, \dots, \bar{y}_k$ の加重平均 $\hat{\beta}(\theta)$ により推定される。すなわち、母数を共通にとることによってデータをプーリングする作用が働き、結果として安定した推定が可能になる。

以上述べてきたように、主要項の母数に等号制約や順序制約などの母数制約をいれることによってデータのプーリングがなされて安定した推定値が得られ、また変量効果を組み入れることによって \bar{y}_i をその安定化された推定値の方向へ縮小することができ、その結果、推定精度を高めることができる。この考え方は経験ベイズ法の枠組みで Efron and Morris (1975) の論文の中で示されたものであり、ベイズ的アプローチの現実的な有用性はこのような考え方に基づいている。

2.4 分散成分の推定と経験最良線形不偏予測量 (EBLUP)

[1] 最尤法 (ML) と制限最尤法 (REML) 線形混合モデル (2.4) において、一般に共分散行列 G , R は分散成分等の未知母数に依存している。それらの母数を $\alpha = (\alpha_1, \dots, \alpha_m)$ とし、 $\Sigma(\alpha) = R(\alpha) + ZG(\alpha)Z'$ と書くことにすると、BLUP (2.8) は $\hat{\mu}(\alpha) = a'\hat{\beta}(\alpha) + b'G(\alpha)Z'\{\Sigma(\alpha)\}^{-1}\{y - X\hat{\beta}(\alpha)\}$ と表される。 α をその推定量 $\hat{\alpha}$ で置き換えたもの $\hat{\mu}(\hat{\alpha})$ を経験最良線形不偏予測量 (EBLUP) という。

α を推定する代表的な方法に最尤法 (Maximum Likelihood, ML) と制限付き最尤法 (Restricted Maximum Likelihood, REML) がある。 \mathbf{y} の周辺分布は, (2.12) より $\mathbf{y} \sim \mathcal{N}_N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\alpha))$ で与えられるが, この尤度に基づいた推定法が ML である。 $\boldsymbol{\beta}$ を GLS $\widehat{\boldsymbol{\beta}}(\alpha)$ で推定すると, α の ML は, $\log |\boldsymbol{\Sigma}(\alpha)| + (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha))' \boldsymbol{\Sigma}(\alpha)^{-1} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha))$ を最小にする解として与えられる。一方, \mathbf{X} のランクを r とし \mathbf{K} を $\mathbf{K}'\mathbf{X} = \mathbf{0}$ なる $N \times (N-r)$ 行列とすると, $\mathbf{K}'\mathbf{y} \sim \mathcal{N}_{N-r}(\mathbf{0}, \mathbf{K}'\boldsymbol{\Sigma}(\alpha)\mathbf{K})$ となり, この分布に基づいた最尤法が REML である。従って, $\log |\mathbf{K}'\boldsymbol{\Sigma}(\alpha)\mathbf{K}| + \mathbf{y}'\mathbf{K}(\mathbf{K}'\boldsymbol{\Sigma}(\alpha)\mathbf{K})^{-1}\mathbf{K}'\mathbf{y}$ の最小値を達する解が REML になる。 $\mathbf{P}(\alpha) = \boldsymbol{\Sigma}(\alpha)^{-1} - \boldsymbol{\Sigma}(\alpha)^{-1}\mathbf{X}\{\mathbf{X}'\boldsymbol{\Sigma}(\alpha)^{-1}\mathbf{X}\}^{-1}\boldsymbol{\Sigma}(\alpha)^{-1}$ とおくと,

$$(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha))' \boldsymbol{\Sigma}(\alpha)^{-1} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha)) = \mathbf{y}'\mathbf{P}(\alpha)\mathbf{y}, \quad \mathbf{P}(\alpha) = \mathbf{K}(\mathbf{K}'\boldsymbol{\Sigma}(\alpha)\mathbf{K})^{-1}\mathbf{K}'$$

が成り立つことに注意する。また $(\partial/\partial\alpha_i) \log |\boldsymbol{\Sigma}| = \text{tr}(\boldsymbol{\Sigma}^{-1}\partial\boldsymbol{\Sigma}/\partial\alpha_i)$, $\partial\mathbf{P}/\partial\alpha_i = -\mathbf{P}(\partial\boldsymbol{\Sigma}/\partial\alpha_i)\mathbf{P}$, $(\partial/\partial\alpha_i) \log |\mathbf{K}'\boldsymbol{\Sigma}\mathbf{K}| = \text{tr}(\mathbf{P}\partial\boldsymbol{\Sigma}/\partial\alpha_i)$ が成り立つので, ML と REML は次の方程式の解として得られることがわかる。

$$(2.14) \quad [\text{ML}] \quad \text{tr} \left(\boldsymbol{\Sigma}(\alpha)^{-1} \frac{\partial\boldsymbol{\Sigma}(\alpha)}{\partial\alpha_i} \right) = \mathbf{y}'\mathbf{P}(\alpha) \frac{\partial\boldsymbol{\Sigma}(\alpha)}{\partial\alpha_i} \mathbf{P}(\alpha)\mathbf{y}$$

$$(2.15) \quad [\text{REML}] \quad \text{tr} \left(\mathbf{P}(\alpha) \frac{\partial\boldsymbol{\Sigma}(\alpha)}{\partial\alpha_i} \right) = \mathbf{y}'\mathbf{P}(\alpha) \frac{\partial\boldsymbol{\Sigma}(\alpha)}{\partial\alpha_i} \mathbf{P}(\alpha)\mathbf{y}$$

右辺は $\mathbf{y}'\mathbf{P}(\alpha)\{\partial\boldsymbol{\Sigma}(\alpha)/\partial\alpha_i\}\mathbf{P}(\alpha)\mathbf{y} = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha))' \boldsymbol{\Sigma}(\alpha)^{-1} \{\partial\boldsymbol{\Sigma}(\alpha)/\partial\alpha_i\} \boldsymbol{\Sigma}(\alpha)^{-1} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha)) = -(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha))' \{\partial\boldsymbol{\Sigma}(\alpha)^{-1}/\partial\alpha_i\} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha))$ と変形できるので, いずれか計算しやすいものを利用すればよい。ML と REML とではどちらがよいのかについて McCulloch and Searle (2001) の 6.10 節で議論されているが, REML は分散成分の推定が不偏に近くなるように自由度が調整されている点で優れていると思われる。ML と REML の具体的な例が 3.1 節で取り上げられているので両者の違いを見ることができる。

[2] 枝分かれ誤差回帰モデルにおける推定。一般には ML, REML を明示的に導くことができないため (2.14), (2.15) の方程式を数値的に解く必要がある。枝分かれ誤差回帰モデル (2.2) についても, n_1, \dots, n_k が異なる場合には明示的な解をもたない。そこで, ここでは分散成分の明示的な推定量を与えておこう。まず, σ_e^2 の不偏推定量は,

$$(2.16) \quad \hat{\sigma}_e^{2UB} = \frac{S_1}{N - k - p + \lambda}, \quad S_1 = \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ (y_{ij} - \bar{y}_i) - (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \widehat{\boldsymbol{\beta}}_1 \right\}^2$$

で与えられる。ただし, $p - \lambda$ は行列 $\sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$ のランクを表しており, 通常は線形モデル (2.2) が定数項をもつときには $\lambda = 1$, もたないときには $\lambda = 0$ となる。また $\widehat{\boldsymbol{\beta}}_1$ は $\sum_{i=1}^k \sum_{j=1}^{n_i} \{(y_{ij} - \bar{y}_i) - (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \boldsymbol{\beta}\}^2$ における $\boldsymbol{\beta}$ の最小 2 乗推定量を表している。一方, $\bar{y}_1, \dots, \bar{y}_k$ は $\hat{\sigma}_e^{2UB}$ と独立に分布し, $\bar{y}_i \sim \mathcal{N}(\bar{\mathbf{x}}_i' \boldsymbol{\beta}, \sigma_e^2/n_i + \sigma_v^2)$, $i = 1, \dots, k$, に従う。 $\widehat{\boldsymbol{\beta}}_2 = (\sum_{j=1}^k n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j')^{-1} \sum_{j=1}^k n_j \bar{\mathbf{x}}_j \bar{y}_j$ に対して, $S_2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{\mathbf{x}}_i' \widehat{\boldsymbol{\beta}}_2)^2$ とおき, $N = \sum_{i=1}^k n_i$, $N_* = N - \text{tr}(\sum_{i=1}^k n_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i') (\sum_{i=1}^k n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i')^{-1}$ とおくと, S_2 の期待値は $E[S_2] = (k-p)\sigma_e^2 + N_*\sigma_v^2$ と書ける。従って, σ_v^2 の 1 つの不偏推定量は

$$\hat{\sigma}_v^{2UB} = N_*^{-1} \{S_2 - (k-p)\hat{\sigma}_e^{2UB}\}$$

となる。分散成分 σ_v^2 の不偏推定量 $\hat{\sigma}_v^{2UB}$ の欠点は正の確率で負の値をとってしまうことである。Kubokawa (2000) は σ_e^2, σ_v^2 の不偏推定量に修正を加えた打ち切り推定量を提案し不偏推定量を

優越することを示した。ここでは分散成分の比 $\theta = \sigma_v^2/\sigma_e^2$ の打ち切り推定量として、

$$(2.17) \quad \hat{\theta} = \max\left\{\frac{1}{N_*}\left\{\frac{S_2}{\hat{\sigma}_e^2 UB} - (k-p)\right\}, \frac{1}{k^{2/3}}\right\}$$

を用いることにする。結局、この $\hat{\theta}$ を (2.13) に代入することにより EBLUP $\hat{\mu}_i(\hat{\theta})$ が得られる。

3 線形混合モデルを利用した小地域推定と誤差評価

2.3 節で説明されたように EBLUP は予測精度を高める手法として用いられ、特に官庁統計の分野では小地域推定のための実用的な方法として利用可能である。その際、予測誤差を見積もることが大切である。この節では、小地域推定に焦点をしばって EBLUP の平均 2 乗誤差の推定と EBLUP に基づいた信頼区間の構成を行い、実際の応用例に当てはめる。

3.1 地域レベルのモデル

官庁から発行される数値には集計データが多いことから、Fay and Herriot (1979) は、

$$(3.1) \quad \bar{y}_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, k,$$

なる形のモデルを扱った。ただし、誤差項 e_i は $e_i \sim \mathcal{N}(0, \sigma_e^2/n_i)$ に従う。個々のデータに基づいたモデル (2.2) が個体レベルのモデルであるのに対して、これは地域レベルのモデルとなっている。また、特に Fay-Herriot モデルとも呼ばれる。このモデルのもとでは誤差分散 σ_e^2 は既知として扱われるが、実際上は何らかの形で σ_e^2 を推定してやる必要がある。しかし解析上扱いやすいので、小地域推定の論文ではこのモデルを設定することが少なくない。モデル (2.2) との関係については記号上 $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$, $\bar{\mathbf{x}}_i = \sum_{j=1}^{n_i} \mathbf{x}_{ij}/n_i$ に対応しているが、観測されるデータは個々のデータではなく集計データ \bar{y}_i , $\bar{\mathbf{x}}_i$ である点に注意する。 $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_k)'$, $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_k)'$, $\mathbf{e} = (e_1, \dots, e_k)'$ として行列表現すると、

$$(3.2) \quad \begin{aligned} \bar{\mathbf{y}} &= \bar{\mathbf{X}} \boldsymbol{\beta} + \mathbf{v} + \mathbf{e}, \\ \mathbf{v} &\sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_k), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \end{aligned}$$

と表されることがわかる。ただし \mathbf{D} は対角行列 $\mathbf{D} = \text{diag}(\sigma_e^2/n_1, \dots, \sigma_e^2/n_k)$ である。これ以降は、このモデルを用いて説明する。モデル (3.1) において σ_v^2 及び分散成分比 $\theta = \sigma_v^2/\sigma_e^2$ は $(\bar{y}_1, \dots, \bar{y}_k)$ に基づいて推定される。 θ の推定量を $\hat{\theta} = \hat{\theta}(\bar{y}_1, \dots, \bar{y}_k)$ で表すことにすると、 $\mu_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + v_i$ の EBLUP は (2.13) よりこの $\hat{\theta}$ を用いて

$$(3.3) \quad \hat{\mu}_i(\hat{\theta}) = \bar{y}_i - \hat{\gamma}_i(\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\hat{\theta})), \quad \hat{\gamma}_i = \gamma_i(\hat{\theta}) = (1 + n_i \hat{\theta})^{-1}$$

なる形で表現できる。このモデルでは $\boldsymbol{\beta}$ の GLS は次のようになる。

$$(3.4) \quad \hat{\boldsymbol{\beta}}(\theta) = \left(\sum_{i=1}^k \frac{n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{1 + n_i \theta} \right)^{-1} \sum_{i=1}^k \frac{n_i \bar{\mathbf{x}}_i \bar{y}_i}{1 + n_i \theta}$$

θ の推定についてはいくつかの方法が知られている。ML, REML は (2.14), (2.15) より,

$$\begin{aligned} \text{[ML]} \quad \sigma_e^2 \sum_{i=1}^k \frac{n_i}{1+n_i\theta} &= \sum_{i=1}^k \frac{n_i^2 \{\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta)\}^2}{(1+n_i\theta)^2} \\ \text{[REML]} \quad \sigma_e^2 \sum_{i=1}^k \frac{n_i}{1+n_i\theta} - \sigma_e^2 \text{tr} \left(\sum_{i=1}^k \frac{n_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{1+n_i\theta} \right)^{-1} \sum_{i=1}^k \frac{n_i^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{(1+n_i\theta)^2} &= \sum_{i=1}^k \frac{n_i^2 \{\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta)\}^2}{(1+n_i\theta)^2} \end{aligned}$$

の非負の解もしくは 0 として与えられる。それらを $\hat{\theta}^{ML}$, $\hat{\theta}^{REML}$ で表す。また Fay and Herriot (1979) はモーメント推定量として次の方程式の解で推定することを提案した。

$$\text{[FH]} \quad \sigma_e^2(k-p) = \sum_{i=1}^k \frac{n_i \{\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta)\}^2}{1+n_i\theta}$$

これは, $\hat{\theta}^{FH}$ で表される。 $\hat{\boldsymbol{\beta}}_2 = (\sum_{j=1}^k n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j')^{-1} \sum_{j=1}^k n_j \bar{\mathbf{x}}_j \bar{y}_j$ に対して, $S_2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}_2)^2$ とおくと, (2.16) の下で与えられた議論と同様にして

$$\text{[TR]} \quad \hat{\theta}^{TR} = \max \left\{ \frac{1}{N_*} \left\{ \frac{S_2}{\sigma_e^2} - (k-p) \right\}, \frac{1}{k^{2/3}} \right\}$$

なる形の明示的な推定量が得られる。 $n_1 = \dots = n_k = n$ のときには, θ の ML は $n^{-1} \max\{n \sum_{i=1}^k \{\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta)\}^2 / (k\sigma_e^2) - 1, 0\}$, REML は $n^{-1} \max\{n \sum_{i=1}^k \{\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\theta)\}^2 / ((k-p)\sigma_e^2) - 1, 0\}$ となり, REML は不偏に近いことがわかる。この場合, FH は REML と同じであり, また $\hat{\theta}^{TR}$ も近い推定値を与えている。

3.2 平均 2 乗誤差の推定

さて EBLUP の予測誤差を見積もるために EBLUP の平均 2 乗誤差の推定量を求めよう。EBLUP の $\mu_i = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + v_i$ に対する推定誤差として,

$$(3.5) \quad M_i(\theta, \hat{\mu}_i(\hat{\theta})) = E[\{\hat{\mu}_i(\hat{\theta}) - \mu_i\}^2] / \sigma_e^2$$

を用いる。これは標準化平均 2 乗誤差と呼ぶべきものであるが, ここでは混乱がない限りこの誤差を平均 2 乗誤差 (Mean Squared Error, MSE) と呼ぶことにする。この MSE を推定することによって EBLUP の誤差を評価することができる。

まず, \bar{y}_i が $\mathcal{N}(\bar{\mathbf{x}}_i' \boldsymbol{\beta}, \sigma_e^2/n_i + \sigma_v^2)$ に従うことに注意し, 縮小推定の理論において知られている Stein の等式を用いると, EBLUP の MSE に対して正確な不偏推定量が得られる。詳しい形は久保川 (2007) で与えられているが, これは多くの項から構成されているのでそれだけ分散が大きくなってしまおうという弊害をもつことになる。実際, Datta, Kubokawa and Rao (2002) は正確な不偏推定量の推定誤差が大きいことを数値実験を通して検証している。データを当てはめてみるとほとんどの MSE 推定値が負の値をとってしまうこともおこる。そこで, MSE を漸近的に近似した推定量を求めることにしよう。

小地域推定の問題を扱っているので各郡の標本サイズ n_i が小さいため, 郡の個数 k が大きい場合を考えて k についての漸近近似を求める。 $\hat{\theta}$ のバイアスと分散を $Bias_{\theta}(\hat{\theta}) = E[\hat{\theta} - \theta]$,

$Var_{\theta}(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]$ で表す。 $Bias_{\theta}(\hat{\theta}) = O_p(k^{-1})$ となることを仮定する。 $\hat{\theta}$ が ML, REML, FH, TR で与えられるときにはこの条件がみたされる。

$$\begin{aligned} g_{1i}(\theta) &= n_i^{-1} - n_i^{-1} \gamma_i(\theta), \\ g_{2i}(\theta) &= \{\gamma_i(\theta)\}^2 \bar{\mathbf{x}}_i' \left\{ \sum_{j=1}^k \gamma_j n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right\}^{-1} \bar{\mathbf{x}}_i, \\ g_{3i}(\theta) &= n_i \{\gamma_i(\theta)\}^3 Var_{\theta}(\hat{\theta}), \end{aligned}$$

とおくと, EBLUP の MSE の k に関する 2 次近似は

$$M_i(\theta, \hat{\mu}_i(\hat{\theta})) = g_{1i}(\theta) + g_{2i}(\theta) + g_{3i}(\theta) + o(k^{-1})$$

となる。この近似式を用いると MSE の 2 次近似に基づいて MSE の 2 次漸近不偏推定量を構成することができる,

$$(3.6) \quad \widehat{M}_i^U(\hat{\theta}) = g_{1i}(\hat{\theta}) + g_{2i}(\hat{\theta}) + 2g_{3i}(\hat{\theta}) - Bias_{\hat{\theta}}(\hat{\theta}) \left\{ \gamma_i(\hat{\theta}) \right\}^2$$

で与えられる。実際, $E[\widehat{M}_i^U(\hat{\theta})] = M_i(\theta, \hat{\mu}_i(\hat{\theta})) + o(k^{-1})$ が成り立つ。

前節で与えられた θ の推定量 $\hat{\theta}^{ML}$, $\hat{\theta}^{REML}$, $\hat{\theta}^{FH}$, $\hat{\theta}^{TR}$ の分散については, $Var_{\theta}(\hat{\theta}^{ML}) = Var_{\theta}(\hat{\theta}^{REML}) = 2 / \sum_{i=1}^k (n_i \gamma_i)^2 + o(k^{-1})$, $Var_{\theta}(\hat{\theta}^{FH}) = 2k / (\sum_{i=1}^k n_i \gamma_i)^2 + o(k^{-1})$, $Var_{\theta}(\hat{\theta}^{TR}) = 2 \sum_{i=1}^k \gamma_i^{-2} / N^2 + o(k^{-1})$ となる。またバイアスについては, $Bias_{\theta}(\hat{\theta}^{REML}) = Bias_{\theta}(\hat{\theta}^{TR}) = o(k^{-1})$,

$$\begin{aligned} Bias_{\theta}(\hat{\theta}^{ML}) &= - \frac{\text{tr}(\sum_i n_i \gamma_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i')^{-1} \sum_i (n_i \gamma_i)^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'}{\sum_i (n_i \gamma_i)^2} + o(k^{-1}), \\ Bias_{\theta}(\hat{\theta}^{FH}) &= 2 \frac{k \sum_i (n_i \gamma_i)^2 - (\sum_i n_i \gamma_i)^2}{(\sum_i n_i \gamma_i)^3} + o(k^{-1}) \end{aligned}$$

となるので, これらを用いて MSE の 2 次漸近不偏推定値が計算できる。EBLUP の MSE $M_i(\theta, \hat{\mu}_i(\hat{\theta}))$ において $\hat{\theta}$ の影響は, $g_{3i}(\theta)$ の中の $Var_{\theta}(\hat{\theta})$ に現れるので, 分散の小さい推定量 $\hat{\theta}$ を用いれば MSE の 2 次近似が小さくなる。 $Var_{\theta}(\hat{\theta}^{ML}) = Var_{\theta}(\hat{\theta}^{REML}) \leq Var_{\theta}(\hat{\theta}^{FH})$ となることが示されるので, FH よりも ML, REML を用いた方が漸近的には MSE が小さくなることがわかる。また (3.6) の中の項の数が増えれば推定量 $\widehat{M}_i^U(\hat{\theta})$ のバラツキが大きくなるので, $Bias(\hat{\theta}) = 0$ となる推定量を用いた方がよい。その点から θ の推定量として REML を用いるのがよいと思われる。以上の内容は, Datta *et al.* (2002), Rao (2003), Datta, Rao and Smith (2005) で詳しく説明されている。

3.3 信頼区間の構成

EBLUP の誤差を評価するもう一つの方法は EBLUP に基づいた信頼区間を構成することで, その誤差の程度が信頼区間の幅として表されるので実用上有用である。前節と同様, i 番目の小地域の平均 $\mu_i = \bar{\mathbf{x}}' \boldsymbol{\beta} + v_i$ の推定に興味があるとし, σ_e^2 が既知として話を進める。まず, μ_i の標本平均に基づいた信頼区間は, μ_i を与えたときの \bar{y}_i の条件付分布が $\bar{y}_i | \mu_i \sim \mathcal{N}(\mu_i, \sigma_e^2 / n_i)$ であることに注意すると,

$$(3.7) \quad I_i^* : \bar{y}_i \pm z_{\alpha/2} \sqrt{\sigma_e^2 / n_i}$$

で与えられる。ただし、 $z_{\alpha/2}$ は標準正規分布の上側 $\alpha/2$ 点を表す。これは信頼係数 $1 - \alpha$ の信頼区間となるが、 n_i が小さいときには \bar{y}_i のバラツキが大きくなってしまふとともに信頼区間の長さが長くなってしまふ。

そこで、線形混合モデル (3.1) の下での信頼区間の構成について考えてみよう。そのモデルをベイズモデルの枠組みで捉えると、 μ_i に $\mu_i \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_k)$ なる事前分布を仮定していることになる。 μ_i のベイズ推定量は、 $\gamma_i = (1 + n_i \theta)^{-1}$ に対して $\hat{\mu}_i^B(\boldsymbol{\beta}, \theta) = \bar{\mathbf{x}}_i' \boldsymbol{\beta} + (1 - \gamma_i)(\bar{y}_i - \bar{\mathbf{x}}_i' \boldsymbol{\beta})$ で与えられるので、 \bar{y}_i を所与としたときの μ_i の事後分布は、

$$\mu_i | \bar{y}_i \sim \mathcal{N}(\hat{\mu}_i^B(\boldsymbol{\beta}, \theta), (\sigma_e^2/n_i)(1 - \gamma_i))$$

となり、 μ_i に対する信頼係数 $1 - \alpha$ のベイズ的信頼区間は、

$$I_i^B(\boldsymbol{\beta}, \theta) : \hat{\mu}_i^B(\boldsymbol{\beta}, \theta) \pm z_{\alpha/2} \sqrt{(\sigma_e^2/n_i)(1 - \gamma_i)}$$

と書ける。これは未知母数 $\boldsymbol{\beta}, \theta$ を含んでいるので、それらの推定量を代入した信頼区間が考えられる。一般に $\hat{\theta}$ を $\bar{y}_1, \dots, \bar{y}_k$ に基づいた θ の推定量とし、 $\boldsymbol{\beta}$ を (3.4) で与えられる一般化最小 2 乗推定量 $\hat{\boldsymbol{\beta}}(\hat{\theta})$ で推定すると、 μ_i の経験ベイズ推定量は

$$\hat{\mu}_i^{EB}(\hat{\theta}) = \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\hat{\theta}) + (1 - \hat{\gamma}_i) (\bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}(\hat{\theta})), \quad \hat{\gamma}_i = (1 + n_i \hat{\theta})^{-1}$$

と書けるので、経験ベイズ的信頼区間は $I_i^B(\boldsymbol{\beta}, \theta)$ から

$$I_i^{EB}(\hat{\theta}) : \hat{\mu}_i^{EB}(\hat{\theta}) \pm z_{\alpha/2} \sqrt{(\sigma_e^2/n_i)(1 - \hat{\gamma}_i)}$$

となる。経験ベイズ推定量 $\hat{\mu}_i^{EB}(\hat{\theta})$ は \bar{y}_i に比べて推定精度が高いだけでなく、信頼区間の長さも短くなっている。しかし、経験ベイズ信頼区間 $I_i^{EB}(\hat{\theta})$ の被覆確率が $1 - \alpha$ 以上になることは保証されていない。 $k = 50, p = 3, 1 - \alpha = 0.95$ とし $\boldsymbol{\beta}, n_i, \bar{\mathbf{x}}_i$ 等を適当に与えたときの被覆確率をシミュレーション実験によって求め、横軸に θ の値をとって描いたものが図 1 で示されている。この図からわかるように、 $I_i^{EB}(\hat{\theta})$ の被覆確率は 95% をかなり下回ってしまう。

そこで、3.2 節の議論を用いて、被覆確率が k に関して 2 次漸近的に $1 - \alpha$ で近似できるような信頼区間を構成する。このような信頼区間は、 $n_1 = \dots = n_k$ で σ_e^2 が既知のときには Basu, Ghosh and Mukerjee (2003) などによって求められ、 n_1, \dots, n_k が等しいことを仮定せず、 σ_e^2 が未知という一般的な設定の下では笹瀬-久保川 (2005) によって得られた。Basu *et al.* (2003) の論法に従って、 $z_{\alpha/2}$ の代わりに $z_{\alpha/2} \{1 + (2k)^{-1} h(\hat{\theta})\}$ を用いた信頼区間

$$(3.8) \quad I_i^{AEB} : \hat{\mu}_i^{EB}(\hat{\theta}) \pm z_{\alpha/2} \left[1 + (2k)^{-1} h(\hat{\theta}) \right] \sqrt{(\sigma_e^2/n_i)(1 - \hat{\gamma}_i)}$$

を考える。ここで、補正関数 $h(\hat{\theta})$ は、 $V_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2]$ を用いて

$$(3.9) \quad h(\hat{\theta}) = -\frac{kn_i \hat{\gamma}_i^2}{1 - \hat{\gamma}_i} \text{Bias}_{\hat{\theta}}(\hat{\theta}) + (1 + z_{\alpha/2}^2) \frac{kn_i^2 \hat{\gamma}_i^4}{4(1 - \hat{\gamma}_i)^2} \text{Var}_{\hat{\theta}}(\hat{\theta}) \\ + \frac{kn_i \hat{\gamma}_i^2}{1 - \hat{\gamma}_i} \left\{ \bar{\mathbf{x}}_i' \left(\sum_{j=1}^k \frac{n_j \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j'}{1 + n_j \hat{\theta}} \right)^{-1} \bar{\mathbf{x}}_i + 2n_i \hat{\gamma}_i \text{Var}_{\hat{\theta}}(\hat{\theta}) \right\}$$

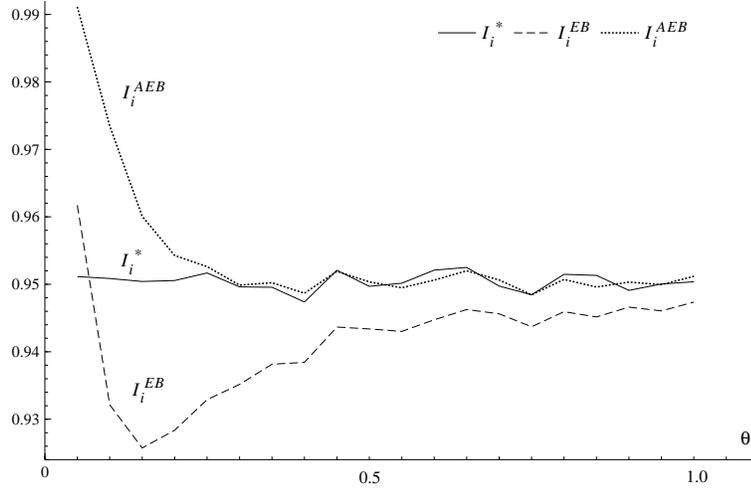


図 1: 標本平均に基づく信頼区間 I_i^* , 経験ベイズ信頼区間 I_i^{EB} , 補正後の信頼区間 I_i^{AEB} の被覆確率の比較

で与えられる。このとき, $Bias_{\theta}(\hat{\theta}) = O_p(k^{-1})$, $\partial\hat{\theta}/\partial\bar{y}_i = O_p(k^{-1})$ を仮定すると,

$$P[\mu_i \in I_i^{AEB}] = 1 - \alpha + o(k^{-1}), \quad (k \rightarrow \infty)$$

が成り立つことが示される。(3.1) で与えられた推定量 $\hat{\theta}^{TR}$ に対しては, $Bias_{\theta}(\hat{\theta}^{TR}) = o(k^{-1})$, $Var_{\theta}(\hat{\theta}^{TR}) = 2 \sum_{i=1}^k (1 + n_i \theta)^2 / N^2 + o(k^{-1})$ となるので, これを (3.9) に代入すれば補正項が得られる。特に $n_1 = \dots = n_k = n$ のときには次のようになる。

$$h(\hat{\theta}) = \frac{1 + z_{\alpha/2}^2}{2n^2 \hat{\theta}^2} + \frac{1}{n\hat{\theta}} \left\{ k \bar{x}_i' \left(\sum_{j=1}^k \bar{x}_j \bar{x}_j' \right)^{-1} \bar{x}_i + 4 \right\}$$

図 1 からわかるように, 2 次補正後の経験ベイズ信頼区間 I_i^{AEB} の被覆確率は 95% より大きく, $\theta > 0.25$ に対してはほぼ 95% になっている。図 2 は同じ設定のもとでの信頼区間の長さの平均値を描いたものであり, I_i^{AEB} の長さが I_i^* より短くなっていることがわかる。このことから, 補正した経験ベイズ信頼区間 I_i^{AEB} が被覆確率と区間の長さの意味で優れているといえよう。

3.4 応用例：地価公示価格の小地域推定

さて, 具体的な小地域に関するデータを用いて EBLUP の特徴, 平均 2 乗誤差推定及び EBLUP に基づいた信頼区間の挙動を調べてみよう。ここで用いるデータは, 京浜急行電鉄本線及び久里浜線沿いの宅地物件について 2001 年に公表された $1m^2$ 当たりの地価公示価格である。各駅を 1 つの小地域と考え, また i 番目の駅を最寄り駅とする物件のデータをその小地域からとられたデータと考えて, その個数を n_i で表す。小地域の総数は $k = 48$ であり, n_i は 1 から 12 まで不均一な値をとっているが平均 3.73 程度である。各地価公示価格を対数変換したものを y_{ij} とし, (2.2) に対応するモデル

$$y_{ij} = \beta_0 + x_{1i}\beta_1 + x_{2ij}\beta_2 + x_{3ij}\beta_3 + v_i + e_{ij}$$

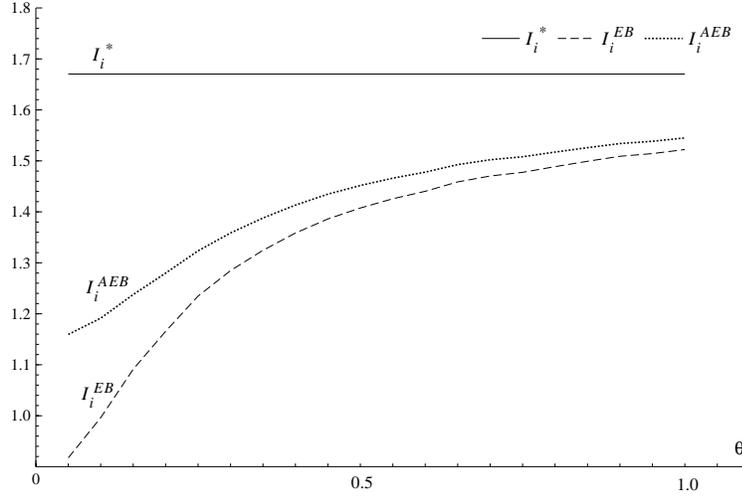


図 2: 標本平均に基づく信頼区間 I_i^* , 経験ベイズ信頼区間 I_i^{EB} , 補正後の信頼区間 I_i^{AEB} の長さの比較

を想定してみる。ここで、共変量 x_{1i} は i 番目の駅から品川駅に到着するのに要する時間、 x_{2ij} は物件 (i, j) から最寄り駅 (i) までの時間距離、 x_{3ij} は物件 (i, j) の容積率を表している。 $x_{ij} = (1, x_{1i}, x_{2ij}, x_{3ij})'$, $\bar{x}_i = (1, x_{1i}, \bar{x}_{2i}, \bar{x}_{3i})'$ とおくと、行列 $\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$ のランクは 2 となるので、(2.16) の $\hat{\sigma}_e^2$ における λ は $\lambda = 2$ となることに注意する。このモデルにおいて、駅 (小地域) ごとに平均的な地価公示価格

$$\mu_i = \beta_0 + x_{1i}\beta_1 + \bar{x}_{2i}\beta_2 + \bar{x}_{3i}\beta_3 + v_i$$

の予測に関して、これまで説明してきた手法の挙動を調べてみよう。

$\hat{\theta}$ として 3.1 節の打ち切り推定量 $\hat{\theta}^{TR}$ を用いて、データから推定値 $\hat{\theta}^{TR}$, $\hat{\beta}(\hat{\theta}^{TR})$, $\hat{\sigma}_e^2$ を計算すると、 $\hat{\theta}^{TR} = 0.551775$, $\hat{\beta}(\hat{\theta}^{TR}) = (12.927, -0.014251, 9.38 \times 10^{-6}, 0.001444)$, $\hat{\sigma}_e^2 = 0.020936$ となる。 β_1 の推定値が負の値であることから、品川駅から遠くなるにつれて地価公示価格は低くなる傾向にあり、合理的な符号を示している。表 1 は、京浜急行沿線の物件 $1m^2$ 当たりの駅ごとの平均価格の予測値と予測誤差を与えている。No.1 北品川から No.48 津久井浜までの 48 の駅が東京に近い順に番号付けられている。 n_i は利用可能なデータ数、 \bar{y}_i は標本平均値、 $EBLUP_i$ は (3.3) から計算される予測値、 $\hat{\beta}(\hat{\theta}^{TR})$ はプールされた推定値に基づいた回帰推定値を示している。これらの推定値を指数変換した値が、表 1 の「予測値」の欄で与えられている。この表から、 $EBLUP_i$ は \bar{y}_i を $\hat{\beta}(\hat{\theta}^{TR})$ の方向へ縮小しており、特に n_i が小さいときに縮小の度合いが大きくなり n_i が大きくなるにつれて縮小の程度は小さくなるのがわかる。表 1 の「予測誤差」の欄は、(3.5) で与えられた平均 2 乗誤差 (MSE) の推定値が与えられている。 $1/n_i$ は、 v_i を母数としたときの \bar{y}_i の MSE であり、 \widehat{M}_i^U は $EBLUP_i$ の MSE の 2 次漸近不偏推定値を表しており、(3.6) から計算されたものである。図 3 は \bar{y}_i の MSE と \widehat{M}_i^U と n_i の値を、駅を No.1 から No.48 まで横軸にとって描いた図である。これらの数値的な結果をながめてみると、 $EBLUP_i$ の予測誤差は \bar{y}_i よりもかなり小さく、特に n_i が小さいときには著しい改善がなされていることがわかる。 n_i が大きいところでは両者の差は小さいようである。表 1 では与えられていないが、MSE の正確な不偏推定量の値を計算したところ負の値になってしまった。全体的にかなり縮小がなされているた

表 1: 京浜急行線沿線の物件の $1m^2$ 当たりの駅ごとの平均価格の予測値と予測誤差
($EBLUP_i^*$ は時系列データを組み込んだ予測値 (4.6) による値)

No.	最寄り駅	n_i	\hat{v}_i^*	予測値			予測誤差		時系列予測値
				\bar{y}_i	$EBLUP_i$	$\hat{\beta}(\hat{\theta}^{TR})$	$1/n_i$	\widehat{M}_i^U	$EBLUP_i^*$
1	北品川	1	1.73	607000	470301	408538	1.000	0.428	465078
2	新馬場	3	0.29	536217	528604	516239	0.333	0.229	565306
3	青物横丁	1	-0.02	484000	485410	486190	1.000	0.402	535321
4	鮫洲	1	1.38	569000	463645	414112	1.000	0.405	470527
5	立会川	2	1.08	508199	469206	429639	0.500	0.291	482257
6	大森海岸	2	0.77	525882	496574	466124	0.500	0.293	528852
7	平和島	2	0.16	470500	464950	458901	0.500	0.289	502266
8	大森町	2	0.60	400999	383471	365017	0.500	0.287	400890
9	梅屋敷	1	0.32	382000	364090	354570	1.000	0.390	388314
10	京急蒲田	3	-0.05	484130	485518	487826	0.333	0.228	525930
11	雑色	4	0.60	389435	380913	362758	0.250	0.183	395263
12	六郷土手	2	-0.86	315899	336715	361281	0.500	0.295	372501
13	京急川崎	4	0.13	604995	601955	595299	0.250	0.197	650621
14	鶴見市場	7	-1.28	319928	328736	365095	0.143	0.119	353112
15	京急鶴見	2	-0.08	570920	574304	578061	0.500	0.330	648094
16	花月園前	2	-0.57	283069	295282	309371	0.500	0.286	322063
17	生麦	7	-0.56	293188	296724	310786	0.143	0.119	312264
18	京急新子安	2	-1.41	321134	356534	400146	0.500	0.286	407568
19	子安	1	1.70	525000	408045	355071	1.000	0.391	417252
20	神奈川新町	1	-0.56	270000	293576	307455	1.000	0.395	322851
21	神奈川	1	0.01	331000	330113	329624	1.000	0.391	357250
22	戸部	3	-0.26	375575	380448	388655	0.333	0.227	414867
23	日ノ出町	4	-0.32	322805	326724	335540	0.250	0.182	349921
24	黄金町	3	-0.49	305847	313364	326213	0.333	0.222	340868
25	南太田	2	0.01	371510	371021	370481	0.500	0.297	407216
26	井土ヶ谷	6	-0.12	271955	272818	275697	0.167	0.134	285193
27	弘明寺	6	0.02	243983	243815	243260	0.167	0.136	251979
28	上大岡	5	-0.84	245846	252081	270115	0.200	0.157	267897
29	屏風浦	4	-1.04	229245	238234	259343	0.250	0.183	257841
30	杉田	6	-0.59	226449	229798	241243	0.167	0.135	241506
31	京急富岡	8	0.25	228446	227393	222803	0.125	0.107	232265
32	能見台	12	1.42	244788	240536	214183	0.083	0.075	239566
33	金沢文庫	10	-0.29	236221	237245	242974	0.100	0.088	244637
34	金沢八景	2	-1.05	210000	226995	247350	0.500	0.282	253148
35	追浜	7	-1.92	189859	197708	231197	0.143	0.119	214788
36	京急田浦	3	-0.72	186865	193649	205425	0.333	0.226	210838
37	安針塚	3	-1.75	163998	178742	206117	0.333	0.233	202777
38	逸見	3	-0.57	184801	190123	199273	0.333	0.226	207148
39	汐入	2	-1.65	174379	196973	225318	0.500	0.284	226449
40	横須賀中央	3	0.54	258351	251518	240602	0.333	0.230	266109
41	県立大学	6	-0.03	208107	208260	208766	0.167	0.137	219065
42	堀ノ内	2	0.77	212941	201120	188836	0.500	0.306	208271
43	新大津	2	-0.36	189447	194667	200595	0.500	0.296	213176
44	北久里浜	5	-0.04	184702	184971	185715	0.200	0.157	194023
45	京急久里浜	6	2.31	240698	227358	188249	0.167	0.137	224582
46	Y R P野比	6	2.80	200674	187314	149112	0.167	0.138	181381
47	京急長沢	5	0.70	165187	161782	152748	0.200	0.162	166523
48	津久井浜	4	-0.12	158203	158957	160634	0.250	0.229	169784

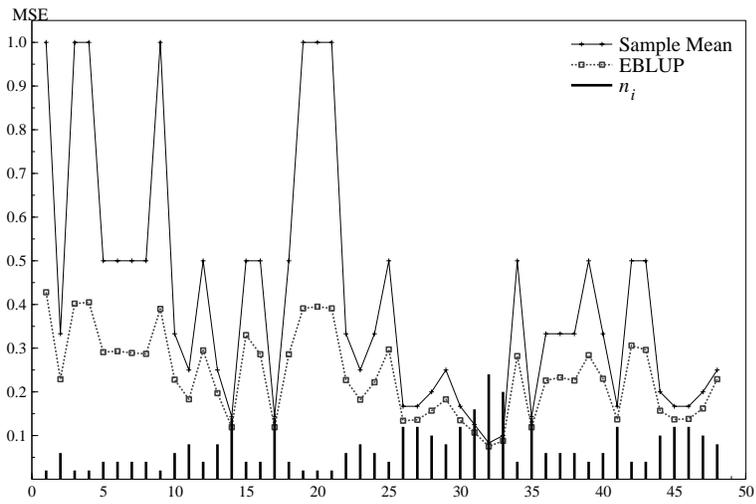


図 3: 標本平均 \bar{y}_i の MSE と EBLUP $_i$ の MSE 推定値 \widehat{M}_i^U の比較
(No.1 から No.48 までの各駅を横軸に並べている。)

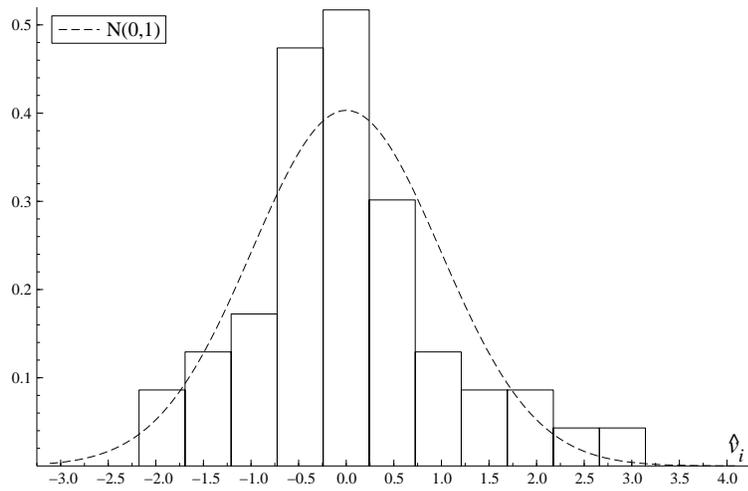


図 4: 変量効果の予測値 \hat{v}_i を標準化した値 \hat{v}_i^* の分布

め MSE の正確な不偏推定値が 0 を越えて負の値を取ってしまったのかもしれないが，正確な不偏推定値が好ましくないことを意味している。

表 1 中の \hat{v}_i^* は変量効果の予測値を標準化した値を示している。変量である地域効果の分布が駅毎に与えられており，共変量の影響を取り除いた後の地域的な差をみるのに役に立つ。例えば，能見台，京急久里浜，Y R P 野比はそれぞれ 1.42, 2.31, 2.80 と高く，追浜，汐入は -1.92, -1.65 と低いが，これらの要因を調べてみるのも興味深いであろう。図 4 は \hat{v}_i^* の分布のヒストグラムと標準正規分布のグラフを描いたものである。変量効果に正規分布を仮定したことは悪くなかったようであるが， t -分布のような裾の厚い分布やその拡張である正規混合分布などを考えてみるのもよいであろう。

図 5 は，2 次補正した経験ベイズ的信頼区間 I_i^{AEB} と \bar{y}_i に基づいた信頼区間 I_i^* の両端の値を，駅を No.1 から No.48 まで横軸にとって描いた図である。 I_i^{AEB} の値は (3.8) から計算できる。 I_i^* の信頼区間の動きが n_i が小さいときに不安定になるのに比べ， I_i^{AEB} は安定した信頼区間を与えている。 I_i^{AEB} の動きをながめてみると，「快特」「特急」電車が停車する駅では土地価格が高くなることを反映して駅ごとに変動しながら，全体として東京から遠くなるにつれて価格が減少するという合理的な傾向がみられる。図 6 は，信頼区間の長さとの関係を示したものである。 I_i^* の長さは， n_i が大きいときには I_i^{AEB} と同程度であるものの， n_i が小さいときには I_i^{AEB} に比べてかなり大きくなってしまふことがわかる。

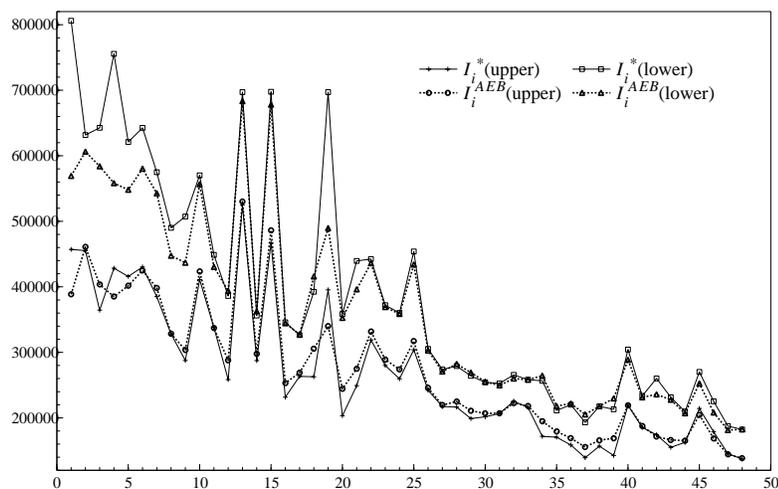


図 5: 補正後の信頼区間 I_i^{AEB} と 標本平均に基づく信頼区間 I_i^* の両端の値の比較 (No.1 から No.48 までの各駅を横軸に並べている。)

4 線形混合モデルの様々な応用

線形混合モデル (LMM) の理論と小地域推定への応用について紹介してきたが，線形混合モデルは長い歴史と幅広い応用分野があり，様々な拡張や変形，推測手法及び計算方法などが活発に研究されてきた。2.3 節で説明したように，LMM が小地域推定を行う上で優れた予測量を導くこ

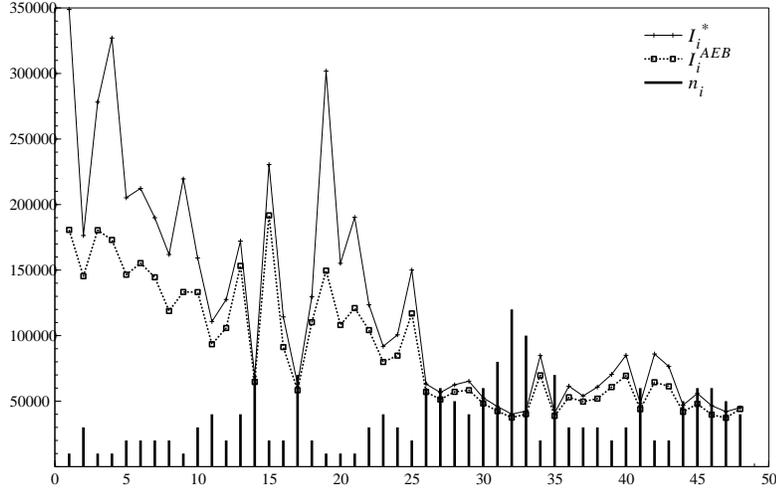


図 6: 補正後の信頼区間 I_i^{AEB} と 標本平均に基づく信頼区間 I_i^* の長さの比較とデータ数 n_i との関係 (n_i のスケールは縦軸 1000 が 1 個のデータを表している。No.1 から No.48 までの各駅を横軸に並べている。)

とのできる要因は、共通母数と変量効果を組み込んでいる点である。したがって、共通母数と変量効果を巧みに組み入れることによって、変量係数モデル、分散変動モデルなど応用例に即した様々なモデルを構築することができる。ここでは、経時測定データのモデル、一般化線形混合モデル、経験ベイズと階層ベイズモデルについて説明する。なお、久保川 (2007) でも詳しく解説されているので参照してほしい。

4.1 経時測定データのモデル

線形混合モデルの最近の中心的話題は、空間的かつ時系列的に取られたデータを扱うモデルであり、Laird and Ware (1982), Tsimikas and Ledolter (1997), Das, Jiang and Rao (2004) などの論文の他、Diggle, Liang and Zeger (1994), Verbeke and Molenberghs (2000), McCulloch and Searle (2001), Demidenko (2004), Fitzmaurice, Laird and Ware (2004), Molenberghs and Verbeke (2006) などの解説書でも章を割いて扱っている。またこのモデルは、計量経済学の分野ではパネル計量モデルと呼ばれ、経済データを解析するために用いられている。パネルデータの解析については、Hsiao (2003) で詳しく解説されている。

ここでは、地域レベルのモデル (3.1) に関連して地域レベルの集計データが T 個の時点で時系列的にとられている場合を考える。これは、経時測定データ (repeated measures data, longitudinal data) という。本来ならば標本サイズはデータがとられた時点と小地域により異なるのであるが、簡単のために同じ小地域ならば各時点で同じサイズの標本が得られているという設定のもとで考える。いま時点 $t = 1, \dots, T$ において $\bar{y}_{i1}, \dots, \bar{y}_{iT}$ なるデータが観測され、そのときの共変量を $\bar{x}_{i1}, \dots, \bar{x}_{iT}$ とする。 $\bar{y}_i = (\bar{y}_{i1}, \dots, \bar{y}_{iT})'$, $\bar{X}_i = (\bar{x}_{i1}, \dots, \bar{x}_{iT})$ とおき、 y_i は

$$(4.1) \quad \bar{y}_i = \bar{X}_i' \beta + j_T v_i + e_i$$

に従うとする。ただし, e_i と v_i は独立に分布し, $e_i \sim \mathcal{N}_T(\mathbf{0}, (\sigma_e^2/n_i)\mathbf{Q})$, $v_i \sim \mathcal{N}(0, \sigma_v^2)$ に従うとする。 $e_i = (e_{i1}, \dots, e_{iT})'$ として成分で表すと

$$\bar{y}_{it} = \bar{x}'_{it}\beta + v_i + e_{it}, \quad i = 1, \dots, k, \quad t = 1, \dots, T,$$

となる。このモデルは多くの文献で扱われており, e_{is} と e_{it} , $s \neq t$, の間に相関関係が入っている。その関係は共分散行列 \mathbf{Q} の中に時間的な相関構造として埋め込まれることになり, 通常は一様共分散構造もしくは自己共分散構造 AR(1) を仮定する。例えば, $T = 4$ の場合, それらは $|\rho| < 1$ に対して

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix} = (1 - \rho)\mathbf{I}_4 + \rho\mathbf{J}_4,$$

$$\mathbf{Q}_2 = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} = (\rho^{|i-j|})$$

と表される。いまこのようなモデルを k 個の地域について考えることにし,

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_k \end{pmatrix}, \quad \bar{\mathbf{X}} = \begin{pmatrix} \bar{\mathbf{X}}'_1 \\ \vdots \\ \bar{\mathbf{X}}'_k \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{j}_T & & \\ & \ddots & \\ & & \mathbf{j}_T \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k \end{pmatrix}$$

とおくと,

$$(4.2) \quad \bar{\mathbf{y}} = \bar{\mathbf{X}}\beta + \mathbf{Z}\mathbf{v} + \mathbf{e}$$

と表される。 $p \times q$ 行列 $\mathbf{A} = (a_{ij})$, $r \times s$ 行列 \mathbf{B} のクロネッカー積は $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$ で定義されるが, これを用いると \mathbf{Z} は $\mathbf{Z} = \mathbf{I}_k \otimes \mathbf{j}_T$ と書ける。 \mathbf{v} の共分散行列を $\text{Cov}(\mathbf{v}) = \mathbf{G}$ とし, $\mathbf{D} = \text{diag}(\sigma_e^2/n_1, \dots, \sigma_e^2/n_k)$ とおくと, $\bar{\mathbf{y}}$ の共分散行列は,

$$\Sigma = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{D} \otimes \mathbf{Q}$$

となる。さらに $\mathbf{Z}\mathbf{G}\mathbf{Z}' = (\mathbf{I}_k \otimes \mathbf{j}_T)(\mathbf{G} \otimes \mathbf{1})(\mathbf{I}_k \otimes \mathbf{j}'_T) = \mathbf{G} \otimes \mathbf{J}_T$ に注意すると, 結局

$$\Sigma = \mathbf{G} \otimes \mathbf{J}_T + \mathbf{D} \otimes \mathbf{Q}$$

と書ける。 $\mathbf{G} = \text{Cov}(\mathbf{v})$ には空間的相関構造を入れることができ, 例えば一様共分散構造

$$\mathbf{G} = \sigma_v^2 \{ (1 - \rho_v)\mathbf{I}_k + \rho_v\mathbf{J}_k \}$$

を仮定すると, 他の地域の情報を有効に利用する予測量が作れる。空間的な広がり方は2次元的なので, 局所的にこの相関構造を用いて隣接する地域の情報を利用する手法を構成することができるが, 広い範囲にわたってこの相関構造をおくことは好ましくない。 \mathbf{G} として空間的な自己回帰モデルを仮定するのも自然である。

モデル (2.2), (3.1) では空間的に無相関 $\text{Cov}(v) = \sigma_v^2 \mathbf{I}_k$ を仮定している。その代わり β を共通にとることによってデータの空間的プーリングがなされている。以降は v が無相関として議論する。このとき, $\Sigma = \text{Cov}(\bar{y})$ は

$$\begin{aligned}\Sigma &= \sigma_v^2 \mathbf{I}_k \otimes \mathbf{J}_T + \mathbf{D} \otimes \mathbf{Q} = \text{diag}(\Sigma_1, \dots, \Sigma_k), \\ \Sigma_i &= \sigma_v^2 \mathbf{J}_T + (\sigma_e^2/n_i) \mathbf{Q}, \quad i = 1, \dots, k\end{aligned}$$

と書ける。 $\Sigma^{-1} = \text{diag}(\Sigma_1^{-1}, \dots, \Sigma_k^{-1})$ であり,

$$\Sigma_i^{-1} = \frac{n_i}{\sigma_e^2} \left\{ \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{j}_T \mathbf{j}_T' \mathbf{Q}^{-1}}{\sigma_e^2/(n_i \sigma_v^2) + \mathbf{j}_T' \mathbf{Q}^{-1} \mathbf{j}_T} \right\}$$

となることから, $\hat{v} = (\hat{v}_1, \dots, \hat{v}_k)' = \mathbf{GZ}'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$ は, $\theta = \sigma_v^2/\sigma_e^2$ に対して

$$\begin{aligned}\hat{v}_i &= \sigma_v^2 \mathbf{j}_T' \Sigma_i^{-1} (\bar{y}_i - \bar{\mathbf{X}}_i' \hat{\beta}) \\ &= \frac{n_i \theta}{1 + n_i \theta \mathbf{j}_T' \mathbf{Q}^{-1} \mathbf{j}_T} \mathbf{j}_T' \mathbf{Q}^{-1} (\bar{y}_i - \bar{\mathbf{X}}_i' \hat{\beta})\end{aligned}$$

となる。また, β の GLS は

$$\hat{\beta} = \left(\sum_{i=1}^k \bar{\mathbf{X}}_i \Sigma_i^{-1} \bar{\mathbf{X}}_i' \right)^{-1} \sum_{i=1}^k \bar{\mathbf{X}}_i \Sigma_i^{-1} \bar{y}_i$$

となる。従って, 小地域の T 期間における平均値 $\mu_i = \mathbf{j}_T' \bar{\mathbf{X}}_i' \beta / T + v_i$ を推定したいときには, その推定量は $\hat{\mu}_i = \mathbf{j}_T' \bar{\mathbf{X}}_i' \hat{\beta} / T + \hat{v}_i$ で与えられる。

\mathbf{Q} に一様共分散構造 $\mathbf{Q}_1 = (1 - \rho) \mathbf{I}_T + \rho \mathbf{J}_T$ を仮定すると,

$$\mathbf{Q}_1^{-1} = \frac{1}{1 - \rho} \left\{ \mathbf{I}_T - \frac{\rho}{1 + (T - 1)\rho} \mathbf{J}_T \right\}$$

より, $\mathbf{j}_T' \mathbf{Q}_1^{-1} = \{1 + (T - 1)\rho\}^{-1} \mathbf{j}_T'$ となることに注意して,

$$\hat{v}_i = \frac{n_i \theta}{1 + (T - 1)\rho + n_i \theta T} \sum_{t=1}^T (\bar{y}_{it} - \bar{\mathbf{x}}_{it}' \hat{\beta})$$

と書ける。 \mathbf{Q} に自己共分散構造 $\mathbf{Q}_2 = (\rho^{|i-j|})$ を仮定すると, 例えば $T = 4$ の場合

$$\mathbf{Q}_2^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & -\rho & 1 \end{pmatrix}$$

と書けることから, $\mathbf{j}_T' \mathbf{Q}_2^{-1} = (1 - \rho)^{-1} (1, 1 - \rho, \dots, 1 - \rho, 1) = \mathbf{j}_T' + \rho(1 - \rho)^{-1} (1, 0, \dots, 0, 1)$, $\mathbf{j}_T' \mathbf{Q}_2^{-1} \mathbf{j}_T = T + 2\rho/(1 - \rho)$ となることに注意する。従って,

$$\hat{v}_i = \frac{n_i \theta}{1 + n_i \theta \{T + 2\rho/(1 - \rho)\}} \left\{ \sum_{t=1}^T (\bar{y}_{it} - \bar{\mathbf{x}}_{it}' \hat{\beta}) + \frac{\rho}{1 - \rho} (\bar{y}_{i1} - \bar{\mathbf{x}}_{i1}' \hat{\beta} + \bar{y}_{iT} - \bar{\mathbf{x}}_{iT}' \hat{\beta}) \right\}$$

となる。いずれの場合も, n_i を大きくし $\rho \rightarrow 0$ とすると全体の平均 $\sum_{t=1}^T (\bar{y}_{it} - \bar{\mathbf{x}}_{it}' \hat{\beta}) / T$ に近づくことになる。

4.2 モデルの修正と小地域推定への応用

以上の設定では、変量効果 v_i を全期間 $t = 1, \dots, T$ にわたって同一としているため v_i の予測量は T 個の時点にわたって合算した量 $\sum_{t=1}^T (\bar{y}_{it} - \bar{x}'_{it}\beta)$ に基づいている。しかし、1年ごとにデータが取られていて最新の予測値を知りたいときには、上述の予測量は明らかに好ましくない。第 T 期のデータにウェイトをかけた予測量を導くためにはモデル (4.1) の中の変量効果の部分 $j_T v_i$ を地域と時点に依存させて

$$(4.3) \quad \bar{y}_i = \bar{X}'_i \beta + v_i + e_i$$

と変えてみることにする。ここで $v_i = (v_{i1}, \dots, v_{iT})'$ は $\mathcal{N}_T(0, \sigma_v^2 I_T)$ に従う変量とする。すなわち、変量効果が時点に関して独立に分布すると仮定する。このとき、 \bar{y}_i の共分散行列は

$$\Sigma_i = \sigma_v^2 I_T + (\sigma_e^2/n_i)Q, \quad i = 1, \dots, k$$

となる。第 T 期の平均値 $\mu_{iT} = \bar{x}'_{iT}\beta + v_{iT}$ を予測したいときには、条件付き期待値が $E[v_{iT}|\bar{y}_i] = \sigma_v^2(0, \dots, 0, 1)\Sigma_i^{-1}(\bar{y}_i - \bar{x}'_i\beta)$ となることより、予測量は

$$\hat{\mu}_{iT} = \bar{x}'_{iT}\hat{\beta} + \sigma_v^2(0, \dots, 0, 1)\Sigma_i^{-1}(\bar{y}_i - \bar{X}'_i\hat{\beta})$$

で与えられる。 $\Sigma_i = (\sigma_e^2/n_i)(n_i\theta I_T + Q)$ と書けるので、

$$n_i\theta I_T + Q = A = \begin{pmatrix} A_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} A^{11} & a^{12} \\ a^{21} & a^{22} \end{pmatrix}$$

と表すことにする。ただし、 a_{22} はスカラーである。 $a_{22.1} = a_{22} - a_{21}A_{11}^{-1}a_{12}$ とおくと $a^{21} = -a_{21}A_{11}^{-1}/a_{22.1}$ 、 $a^{22} = 1/a_{22.1}$ と書けるので、 $\hat{\mu}_{iT}$ は

$$(4.4) \quad \hat{\mu}_{iT} = \bar{x}'_{iT}\hat{\beta} + \frac{n_i\theta}{a_{22.1}} \left\{ (\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta}) - (a_{21}A_{11}^{-1}, 0)(\bar{y}_i - \bar{X}'_i\hat{\beta}) \right\}$$

となる。これは、 $\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta}$ を過去の $T-1$ 個のデータに基づいた予測量の方向へ縮小した形をしていることがわかる。

Q に一様共分散構造 $Q_1 = (1-\rho)I_T + \rho J_T$ を仮定すると、 $a_{22.1} = (n_i\theta + 1 - \rho)(n_i\theta + 1 + (T-1)\rho)/(n_i\theta + 1 + (T-2)\rho)$ 、 $a_{21}A_{11}^{-1} = \{\rho/(n_i\theta + 1 + (T-2)\rho)\}j'_{T-1}$ となることから、

$$(4.5) \quad \hat{\mu}_{iT} = \bar{x}'_{iT}\hat{\beta} + \frac{n_i\theta(n_i\theta + 1 + (T-2)\rho)}{(n_i\theta + 1 - \rho)(n_i\theta + 1 + (T-1)\rho)} \left\{ (\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta}) - \frac{\rho}{n_i\theta + 1 + (T-2)\rho} \sum_{t=1}^{T-1} (\bar{y}_{it} - \bar{x}'_{it}\hat{\beta}) \right\}$$

と表される。 $n_i \rightarrow \infty$ とすると $\hat{\mu}_{iT} \rightarrow \bar{y}_{iT}$ 、 $\rho \rightarrow 0$ とすると $\hat{\mu}_{iT} \rightarrow \bar{x}'_{iT}\hat{\beta} + \{n_i\theta/(1+n_i\theta)\}(\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta})$ に近づくことがわかる。一方、 Q に自己共分散構造 $Q_2 = (\rho^{|i-j|})$ を仮定したときには一般の次元に対しては予測量 (4.4) を明示的に書き下すことはできないが、例えば $T=3$ の場合には、

$$(4.6) \quad \hat{\mu}_{iT} = \bar{x}'_{iT}\hat{\beta} + \frac{n_i\theta\{(n_i\theta + 1)^2 - \rho^2\}}{(n_i\theta + 1 - \rho^2)\{(n_i\theta + 1)^2 - (1 - n_i\theta)\rho^2\}} \left\{ (\bar{y}_{iT} - \bar{x}'_{iT}\hat{\beta}) - \frac{\rho}{(n_i\theta + 1)^2 - \rho^2} \left[n_i\theta\rho(\bar{y}_{i,T-2} - \bar{x}'_{i,T-2}\hat{\beta}) + (n_i\theta + 1 - \rho^2)(\bar{y}_{i,T-1} - \bar{x}'_{i,T-1}\hat{\beta}) \right] \right\}$$

と表される。(4.5) が過去のデータを同じ重みでプーリングしているのに対して, $n_i\theta\rho < n_i\theta+1-\rho^2$ となることからわかるように, (4.6) はより最近のデータをより重くして加重和をとっている。

θ, ρ を推定するには ML もしくは REML を数値的に求める必要がある。しかし次のような簡便な推定量も利用可能である。 $\hat{\beta}_t = (\sum_{j=1}^k n_j \bar{x}_{jt} \bar{x}'_{jt})^{-1} \sum_{j=1}^k n_j \bar{x}_{jt} \bar{y}_{jt}$ とおいて $\hat{e}_{it} = \bar{y}_{it} - \bar{x}'_{it} \hat{\beta}_t$ とする。 θ については (2.16) の下で与えられた議論と同様にして, $S_2 = \sum_{t=1}^T \sum_{i=1}^k n_i \hat{e}_{it}^2$, $N_* = \sum_{t=1}^T \{ \sum_{i=1}^k n_i - \text{tr}(\sum_{i=1}^k n_i^2 \bar{x}_{it} \bar{x}'_{it}) (\sum_{i=1}^k n_i \bar{x}_{it} \bar{x}'_{it})^{-1} \}$ に対して

$$\hat{\theta}^{TR} = \max \left\{ \frac{1}{N_*} \left(\frac{S_2}{\sigma_e^2} - T(k-p) \right), \frac{1}{k^{2/3}} \right\}$$

で推定できる。 ρ については, $\hat{\rho} = \sum_{i=1}^k n_i \hat{\rho}_i / N$ で推定できる。ここで, $\hat{\rho}_i$ は, Q_1 の場合, $\bar{e}_i = \sum_{t=1}^T \hat{e}_{it} / T$ に対し, $\hat{\rho}_i = 2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T (\hat{e}_{is} - \bar{e}_i)(\hat{e}_{it} - \bar{e}_i) / \{ (T-1) \sum_{t=1}^T (\hat{e}_{it} - \bar{e}_i)^2 \}$ であり, Q_2 の場合, $\hat{\rho}_i = \sum_{t=2}^T (\hat{e}_{it} - \bar{e}_i)(\hat{e}_{i,t-1} - \bar{e}_i) / \sum_{t=1}^T (\hat{e}_{it} - \bar{e}_i)^2$ である。いずれも $|\hat{\rho}| < 1$ であり一応これらを用いて推定できるが, それらを初期値として ML, REML を数値的に解くのがよい。

さて, 自己共分散構造 Q_2 を仮定するとき, (4.4) もしくは (4.6) によって予測される値を実際のデータに基づいて計算してみよう。実は, 3.4 節で取り上げた宅地物件の地価公示価格に関するデータは時系列的に得られており, ここでは 1997 年から 2001 年までの 5 年間 ($T=5$) のデータを利用することができる。 $\hat{\theta}^{TR} = 0.686732$, $\hat{\rho} = 0.244273$ となり, 駅毎の予測値が表 1 の EBLUP $_i^*$ の欄で与えられている。ほとんどの予測値が 2001 年だけのデータに基づいた予測値 EBLUP $_i$ を上方修正していることがわかる。これは, この 5 年間の宅地の価格が年々低くなる傾向にあることと, (4.4) が過去のデータに基づいた予測値の方向へ縮小する作用があることに起因している。図 7 は, 変量効果の予測値 \hat{v}_{it} を標準化した値 \hat{v}_{it}^* の経年的変化を, Nos.1, 3, 4, 13, 14, 33 の 6 つの駅について描いたものである。地価公示価格から共変量の影響を取り除いたものを縮小した値が \hat{v}_{it} であり, \hat{v}_{it}^* はその相対的な値を示していることになる。この 5 年間で全体的には価格が低下している状況の中にあって, 品川駅に近い駅 (No.1, 3, 4) の地域効果が相対的に上昇し, No.13, 14, 33 の地域効果が相対的に下降していることがわかる。地価の下落の影響を強く受けた地域とそれほど影響をうけなかった地域など, 小地域に対する経年的変化の様子を捉えることができる。

4.3 一般化線形混合モデル

データが死亡数など離散的に変動するときには, 2 項分布やポアソン分布などに基づいたモデルを考える必要がある。こうした離散分布に回帰項と変量効果を組み入れてモデルを構築することができ, それを統一的に扱ったモデルが一般化線形混合モデル (GLMM) である。いま, 全体で k 個の地域があり, i 番目の地域から n_i 個のデータ (もしくは集計データ) y_{i1}, \dots, y_{in_i} が取られており, v_i を所与としたときの y_{ij} の条件付分布が

$$f(y_{ij}|v_i) = \exp \{ [y_{ij}\theta_{ij} - b(\theta_{ij})] / \tau_{ij} + c(y_{ij}, \tau_{ij}) \},$$

$j = 1, \dots, n_i; i = 1, \dots, k$, で与えられるとする。これは一般化線形モデルと呼ばれる。この密度関数は自然母数 θ_{ij} と尺度母数 $\tau_{ij} (> 0)$ を用いて表現されており, τ_{ij} は既知と仮定される。 y_{ij} の条件付期待値を $E[y_{ij}|v_i] = \mu_{ij}$ と書くとき, μ_{ij} がリンク関数 $g(\cdot)$ を通して共変量 x_{ij} と関係づけることができることを仮定して

$$g(\mu_{ij}) = x'_{ij}\beta + v_i$$

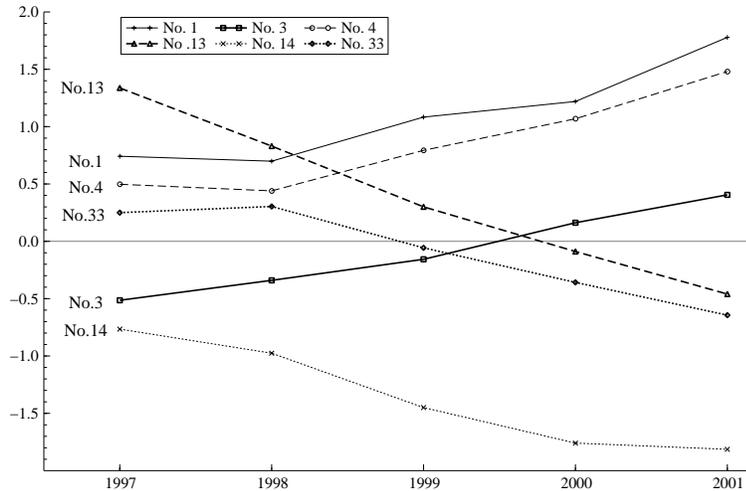


図 7: 変量効果の予測値 \hat{v}_{it} を標準化した値 \hat{v}_{it}^* の経年的変化

なる形で表現できるとする。ここで、 $v_i \sim \mathcal{N}(0, \sigma_v^2)$ であり、すべて互いに独立であると仮定する。これを GLMM という。GLMM の詳しい解説については、McCullagh and Nelder (1989, 14.5 節), Fahrmeir and Tutz (2001), McCulloch (2003) で与えられている。また、McCulloch and Searle (2000) はわかりやすく書かれた本であり、疾病地図など空間モデルを扱った本としては、Lawson, Browne and Vidal Rodeiro (2003), Lawson (2006) などがあるので参照してもらいたい。

共通母数と変量効果を組み入れた GLMM は、地域の特性値に対して精度の高い安定した推定値を与えることができる。その代表的な例が疾病地図の作成に用いられる死亡率及び死亡指標の推定である。死亡率の代表的な指標として用いられるのが標準化死亡率 (Standardized Mortality Rate, SMR) で、(観測死亡数)/(期待死亡数) で定義される。地域と年齢階級を細分していくと SMR のバラツキが大きくなってしまふ。この問題に対して GLMM が安定した推定値を与えるために有用であることが示されており肺ガンの死亡率地図が作成されている。丹後 (1988) はポアソン・ガンマモデルによる経験ベイズ推定値の導出方法について議論し安定した疾病地図の作成が可能であることを示した。

4.4 経験ベイズモデルと階層ベイズモデル

枝分かれ誤差回帰モデル (2.2) を $y_{ij} = \mu_{ij} + e_{ij}$, $\mu_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + v_i$ と分解すると、条件付分布は $y_{ij} | (\mu_{ij}, \sigma_e^2) \sim \mathcal{N}(\mu_{ij}, \sigma_e^2)$ となり、 μ_{ij} の事前分布が $\mu_{ij} \sim \mathcal{N}(\mathbf{x}_{ij}'\boldsymbol{\beta}, \sigma_v^2)$ で与えられるベイズモデルの形で捉えることができる。母数 $\boldsymbol{\beta}$, σ_v^2 , σ_e^2 が既知のときが主観的ベイズモデルであり、それらの事前の値に強く依存してしまう。そこで客観性を持たせるために 2 つの方法が知られている。1 つは $\boldsymbol{\beta}$, σ_v^2 , σ_e^2 を未知母数として扱う方法で経験ベイズモデルと呼ばれる。未知母数とすることで事前分布の恣意性を排除できる。経験ベイズ推定量は事前分布の情報を取り入れながらもその情報の誤りに対する実害が生じないという利点をもっている。いいかえると、事前情報に関して頑健であることを意味する。(2.2) などこの論文で扱ってきた線形混合モデルは経験ベイズモデルに対応している。

事前分布に客観性を持たせるもう1つの方法は、 $\beta, \sigma_v^2, \sigma_e^2$ を変量として扱う方法で階層ベイズモデルと呼ばれる。この場合、 (i, j) -成分に μ_{ij} をもつ行列を μ とおくと、 $(\mu, \sigma_e^2) | (\beta, \sigma_v^2) \sim \pi_1(\mu, \sigma_e^2 | \beta, \sigma_v^2)$ を第1段階事前分布、 $(\beta, \sigma_v^2) \sim \pi_2(\beta, \sigma_v^2)$ を第2段階事前分布という。客観的なベイズ推定を構成するためには、一般に、第1段階事前分布は正確に、第2段階事前分布は無情報的に設定するのがよいとされている。例えば、第1段階事前分布 $\pi_1(\mu, \sigma_e^2 | \beta, \sigma_v^2)$ として

$$\begin{aligned} \mu_{ij} | (\beta, \sigma_v^2) &\sim \mathcal{N}(x'_{ij}\beta, \sigma_v^2), \\ \sigma_e^2 &\sim \sigma_e^{-2} d\sigma_e^2 \end{aligned}$$

がとられる。 $\sigma_e^{-2} d\sigma_e^2$ は尺度変換に関して不変な測度で無情報事前分布を表している。また第2段階事前分布 $\pi_2(\beta, \sigma_v^2)$ としては σ_v^2 には無情報事前分布 $\sigma_v^{-2} d\sigma_v^2$ を想定し、 β に対しては、(1) 一様分布 $d\beta$ 、(2) $\beta | \sigma_v^2 \sim \mathcal{N}(\beta_0, \sigma_v^2 A)$ 、(3) $\beta | (\sigma_v^2, \lambda) \sim \mathcal{N}(\beta_0, \lambda \sigma_v^2 A)$ 、 $\lambda \sim \pi_3(\lambda)$ 、などの設定が考えられる。ここで β_0, A は既知の値とする。このような階層ベイズ推定量の理論的な性質については Kubokawa and Strawderman (2007) とその中の参考文献を参照してほしい。また階層ベイズを用いた空間データの解析については、Banerjee, Carlin and Gelfand (2004) に述べられている。

5 おわりに

線形混合モデルと経験最良線形不偏予測量について基本的な性質を解説し、小地域推定や経時測定データへの応用について主に説明してきた。最後に、線形混合モデルと関係のある推定問題について若干説明を加えて本稿を終えることにする。

まず、家畜育種学という分野で生まれ育ってきた線形混合モデルの手法が、数理統計学において有名な Stein 問題と関係している点を指摘しておこう。C. Stein が 1956 年に発見した理論は、「3次元以上の正規分布の平均を同時に推定する問題において標本平均よりも平均2乗誤差 (MSE) を一様に小さくする縮小推定量が存在する」ということであった。そこで登場した縮小推定量を $n_1 = \dots = n_k = n$ の地域モデル (3.1) に当てはめてみると、平均 $\mu_i = \bar{x}'_i \beta + v_i$ からなるベクトル $\mu = (\mu_1, \dots, \mu_k)'$ の縮小推定量は

$$\hat{\mu}^S = \bar{X}\hat{\beta} + \max\left\{1 - \frac{(k-p-2)\sigma_e^2}{n\|\bar{y} - \bar{X}\hat{\beta}\|^2}, 0\right\}(\bar{y} - \bar{X}\hat{\beta})$$

という形で表される。ここで $\hat{\beta} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}$ は β の OLS である。Stein の理論によると、 $k-p \geq 3$ ならば $\hat{\mu}^S$ の MSE は \bar{y} の MSE より一様に小さいことになる。Stein 問題については篠崎 (1991)、下平、伊藤、久保川、竹内 (2004) に詳しく説明されているが、この問題は理論家の興味の対象となり 20 世紀の数理統計学の一分野に発展した。一方、 $n_1 = \dots = n_k = n$ の地域モデル (3.1) に対して μ の EBLUP は、 θ を REML で推定すると (3.3) より、縮小推定量 $\hat{\mu}^S$ において $(k-p-2)$ を $(k-p)$ に置き換えたものに一致する。Henderson の BLUP についての論文が出されたのが 1950 年であり、その数年後に全く違う分野で Stein が EBLUP と同等な手法を考案していたことになる。しかも、Henderson は家畜育種という応用分野、Stein は統計的決定理論という理論分野であり、応用・理論の双方向から研究されてきたトピックであるといえよう。

次に、2.4 節で扱ったように、LMM の分散成分に関する推定は LMM の主要な研究テーマの一つであるが、この問題が分散の不等式制約に関係している点を指摘しておこう。例えば、 $n_1 = \dots = n_k = n$ の枝分かれ誤差回帰モデル (2.2) を取り上げてみると、(2.16) 周辺の議論から S_1

と S_2 が独立になり, $S_1/\sigma_1^2 \sim \chi_{m_1}^2$, $S_2/\sigma_2^2 \sim \chi_{m_2}^2$ に従うことがわかる。ここで m_1, m_2 は対応する自由度を表し, $\sigma_1^2 = \sigma_e^2$, $\sigma_2^2 = \sigma_e^2 + n\sigma_v^2$ である。従って分散 σ_1^2, σ_2^2 の間に $\sigma_1^2 < \sigma_2^2$ なる不等式制約が入っていることがわかる。一般に多元配置の变量モデルにおいては分散の間に複雑な順序関係が入ることになり, 母数制約のもとでの推定問題として定式化される。また多変量へ拡張したモデルを考えると, 共分散行列の間に不等式制約が入った推定問題になり, Srivastava and Kubokawa (1999), Kubokawa and Tsai (2006) などで議論されている。

本稿で取り上げてきた内容からわかるように, 筆者が線形混合モデルの魅力に惹かれたのは小地域推定という問題を通してであった。カナダのオタワにある Carleton 大学へ滞在していた 1989 年, 91 年当時, その数学・統計学科の教授 John N.K. Rao の研究テーマが LMM を利用した小地域推定であった。Rao はオタワ市内にあるカナダ統計局 (Statistica Canada) のスタッフとつねに連携を取り統計調査の現場において何が必要とされているのかに関心をもちながら研究を進めていた。当時の PhD 学生の論文テーマもその現場にモチベーションをもつ内容であり, PhD を取得した後にカナダ統計局へ就職して活躍している研究者もいる。Rao のこのような研究姿勢に, 応用と理論の両輪で発展する統計学の生きた姿をみた思いがして大変感銘したことが思い出される。

謝辞. 編集者及び査読者の方々から貴重なコメントを頂きました。ここに深く感謝申し上げます。本研究は, 科学研究費補助金 16500172 及び東京大学大学院経済学研究科 21 世紀 COE プログラムから研究助成を受けております。

参考文献

- [1] Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall, New York.
- [2] Basu, R., Ghosh, J.K., and Mukerjee, R. (2003). Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics. *Statist. Prob. Letters*, **63**, 197-203.
- [3] Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.
- [4] Das, K., Jiang, J. and Rao, J.N.K. (2004). Mean squared error of empirical predictor. *Ann. Statist.*, **32**, 818-840.
- [5] Datta, G.S., Kubokawa, T. and Rao, J.N.K. (2002). Estimation of MSE in small area estimation. Unpublished manuscript.
- [6] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.
- [7] Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Wiley.
- [8] Diggle, P., Liang, K.-Y., and Zeger, S.L. (1994). *Longitudinal Data Analysis*. Oxford Univ. Press.

- [9] Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311-319.
- [10] Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd ed. Springer, New York.
- [11] Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.
- [12] Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley.
- [13] Henderson, C.R. (1950). Estimation of genetic parameters. *Ann. Math. Statist.*, **21**, 309-310.
- [14] Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge University Press. (「ミクロ計量経済学の方法：パネル・データ分析」(2007) 国友直人訳，東洋経済新報社)
- [15] Kubokawa, T. (2000). Estimation of variance and covariance components in elliptically contoured distributions. *J. Japan Statist. Soc.*, **30**, 143-176.
- [16] Kubokawa, T. and Strawderman, W.E. (2007). On minimaxity and admissibility of hierarchical Bayes estimators. *J. Multivariate Analysis*, **98**, 829-851.
- [17] Kubokawa, T. and Tsai, M.-T. (2006). Estimation of covariance matrices in fixed and mixed effects linear models. *J. Multivariate Analysis*, **97**, 2242-2261.
- [18] Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.
- [19] Lawson, A.B. (2006). *Statistical Methods in Spacial Epidemiology*. 2nd ed. Wiley, England.
- [20] Lawson, A.B., Browne, W.J. and Vidal Rodeiro, C.L. (2003). *Disease Mapping with WinBUGS and MLwiN*. Wiley, England.
- [21] McCulloch, C.E. (2003). *Generalized Linear Mixed Models*. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 7. IMS, USA.
- [22] McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear and Mixed Models*. Wiley, New York.
- [23] Molenberghs, G. and Verbeke, G. (2006). *Models for Discrete Longitudinal Data*. Springer.
- [24] Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New Jersey.
- [25] Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*, Wiley, New York.
- [26] Srivastava, M.S. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. *Ann. Statist.*, **27**, 2008-2032.

- [27] Tsimikas, J.V. and Ledolter, J. (1997). Mixed model representation of state space models: New smoothing results and their application to REML estimation. *Statistica Sinica*, **7**, 973-991.
- [28] Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- [29] 広津千尋 (1992). 実験データの解析—分散分析を超えて—. 共立出版.
- [30] 久保川達也 (2007). 線形混合モデルと小地域の推定. 応用統計学, **35**, 139-161.
- [31] 佐々木義之 (2007). 変量効果の推定と BLUP 法. 京都大学学術出版会.
- [32] 下平英寿, 伊藤秀一, 久保川達也, 竹内啓 (2004). モデル選択 : 予測・検定・推定の交差点. 岩波書店
- [33] 笹瀬吉隆, 久保川達也 (2005). 経験ベイズ信頼区間の漸近補正と小地域推定への応用. 日本統計学会誌 (和文誌), **35**, 27-54.
- [34] 篠崎信雄 (1991). Stein タイプの縮小推定量とその応用. 応用統計学, **20**, 59-76.
- [35] 丹後俊郎 (1988). 死亡指標の経験的ベイズ推定量について. 応用統計学, **17**, 81-96.