CIRJE-J-18

母数制約と変量効果

——推定精度を高めるために——

東京大学大学院経済学研究科
久保川 達也

1999 年 10 月

Random Effects and Restriction of Parameters: A Review

Tatsuya Kubokawa
Faculty of Economics
University of Tokyo

Abstract

Random effects and restriction of parameters are important concepts in constructing statistical models based on small samples: Restricting the parameter spaces to subspaces constrained by inequalities or equalities yields pooling the data to get stablized estimators, and incorporating the random effects into the models produces shrinkage procedures more accurate than usual crude estimators. This paper gives a review on the random effects and the restriction of parameters and demonstrates the importance of these two concepts in an example of small area estimation. It is also explained that the shrinkage estimators have interesting relation with the factor analysis, the Kalman filter and the Spline smoothing method.

母数制約と変量効果

— 推定精度を高めるために —

久保川　達也*

1　はじめに

データから統計的モデルを構築する際，モデルのパラメタ数をむやみに多くしてはいけないことは，自由度調整済み決定係数，Mallows の C_p 統計量，赤池情報量規準などのモデル（変数）選択規準が教えるところである。それは，パラメタが増えるにつれてパラメタの推定誤差が大きくなり，その結果構築されたモデルが不安定になり，モデルの信頼性が損なわれることによる。このことは，特にデータ数が少ないときに問題となる。

データ数が少ないときの簡単な方策として

(1) 事前情報を取り入れて，母数に制約を入れること，

(2) 母数を変量として扱うこと

があげられる。(1) については，経験や知識などの事前に得られる情報から，例えば，いくつかの母平均が等しい（等号制約）と仮定できたり，またいくつかの母平均の間に順序関係（順序制約）が仮定できるときには，そのような制約を積極的に取り入れることによって，興味ある母平均の推定精度を高めることができる。このことは，母数に制約を入れることによってデータをプールすることができ，データ数の少ないことに原因する推定誤差の増大を，データのプーリングによって補うことができるからである。

(2) については，例えば繰り返し数が少なく水準数の大きい一元配置モデルでは，主効果をすべて母数にしてしまうとパラメタ数が多くなってしまうので，それらの主効果を変量として扱うことによって，推定量の不安定性が軽減される。こうした変量モデルでは，変量効果の通常の推定量は経験最良線形不偏予測量 (Empirical Best Linear Unbiased Predictor) であるが，実はこれは Stein 型縮小推定量となっている。すなわち，モデルの母数を変量効果として扱うことによって個々の推定量をより安定した推定量の方向へ縮小することができ，その結果，データの少ないことに原因する推定誤差の増大を軽減することができる。

この報告では，特にデータ数が少ないときに推定精度を高める方法としての,「母数制約とデータのプーリング」,「変量効果と縮小推定」，について概説する。

まず，母数制約に関する代表的な話題は平均の順序制約であり，多標本の平均の間に Simple Ordering (SO) や Simple Tree Ordering (STO) などの不等式制約が課されるときに，データのプーリングを通して平均についてのより精度の高い推定量を求めることである。通常は制限付き最尤 (REML) 推定量が使われが，理論的にも REML 推定量が制限無しの推定量を改良していることが順序制約 SO のときには示すことができる。しかしこ

*東京大学　大学院経済学研究科

のような主張は，順序制約 STO のときには必ずしも成り立たない。Hwang and Peddada (1985) は，順序制約 STO のときにも順序制約 SO での REML 推定量を使用することを主張し，また平均についての一般的な順序制約条件に対して有効な推定量を与えるための方法を提案した。2 節では，Hwang and Peddada (1985) の結果を中心に，平均の順序制約，等号制約に関する結果を概説する。

一方，分散の順序制約下での推定問題の研究は，平均のときほどは十分にはなされていないようである。しかし，分散の順序制約が，変量モデルにおいては自然に成立することは興味深い。すなわち，一元配置変量モデルでは 2 つの分散の不等式制約が生じ，また nested な二元配置変量モデルでは 3 つの分散の間に SO の制約条件が存在することになる。3 節では，分散の不等式制約下での推定及び分散の Stein 推定量との関係，変量モデルにおける分散成分の推定，その多変量モデルへの拡張について概説する。母数の等号制約など母数制約に関するその他の問題については 4 節で簡単に紹介する。

変量効果と縮小推定についての内容が 5 節以降で述べられる。5 節では，モデルに変量効果を組み入れることによって，Stein 型縮小推定量が経験最良線形不偏予測量 (EBLUP) として導出されることを説明する。Bayesian の側面からは，これは経験 Bayes 推定量として解釈されるが，完全な Bayesian の立場からの階層 Bayes 的縮小推定量や，縮小推定量の条件付き推測からの解釈についても言及する。

変量モデルでは，母数モデルで扱うところの母数平均を，共通な母数平均と変量効果の和として表現する。

$$(\text{母数モデルでの平均}) = (\text{共通な母数平均}) + (\text{変量効果})$$

共通平均は，平均の等号制約を意味しており，データのプーリングがなされる。一方，変量効果によって，推定量に縮小操作が取り込まれる。結局これらの 2 つの効果を組み合わせることによって，通常の推定値がプールされた安定した推定値へ縮小される。このことは，より安定した推定精度の高い推定量を与える可能性があることを意味しており，母数制約と変量効果は，データ数の少ないときのモデルを構築する際大切な点になると思われる。これに関する説明が，5.2 節に与えられる。

6 節では，縮小推定についてのその他の役割として，多重共線性のもとでの推定量の安定化，事前情報や知識を利用した縮小推定，仮説の信頼度に適応した推定量の分解をとりあげて説明する。7 節では，多変量混合モデルでの予測問題への拡張や，縮小推定と因子分析モデルとの関係，カルマンフィルターとの関係，データの平滑化との関係について概説する。

8 節では，縮小推定法の有効な応用例の 1 つとして知られる小地域推定の問題を扱い，Battese *et al.* (1988) による混合モデルを利用した穀物作付け面積の推定を例として，混合モデルにおける母数制約と変量効果が小地域推定にどのように役に立つのかについて説明する。また 2 値データに基づいた小地域推定の問題を取り上げ，一般線形混合モデルとポアソン-ガンマモデルを用いた死亡率の推定問題について簡単に紹介する。

第 1 部　母数制約とデータのプーリング

まず，母数に制約を課すときデータのプーリングが行われて推定精度を高めることができることを述べる。代表的な母数制約は平均に関する順序制約であるので，この推定問題を最初に紹介する。3 節で分散の順序制約と分散成分の推定について，また 4 節で母数の等号制約等について概説する。

2 平均の順序制約

2.1 制限付き最尤推定量

平均 μ_i, 分散 σ^2, 繰り返し数 r_i の k 標本問題において，$\mu_1, \ldots, \mu_k$ に関しての代表的な大小関係は，Simple Ordering (SO) と Simple Tree Ordering (STO) で，それぞれ

$$\begin{aligned} \text{SO} &: \mu_1 \le \mu_2 \le \ldots \le \mu_k, \\ \text{STO} &: \mu_1 \le \mu_i, \quad i = 2, \ldots, k \end{aligned}$$

で与えられる。このような順序制約は，例えば，順序制約 STO は，対照群の平均 μ_1 に対して何らかの処理を施すことによってその平均 μ_i が μ_1 より大きくなることが考えられる場合に仮定される。

いま X_i を i 番目の標本の標本平均とし，$\boldsymbol{X} = (X_1, \ldots, X_k)'$，$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ とするとき，$\boldsymbol{\Omega} = \mathrm{diag}\,(\omega_1, \ldots, \omega_k)$ に対して，順序制約のもとで $(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Omega}(\boldsymbol{X} - \boldsymbol{\mu})$ を最小にする解 $\boldsymbol{\mu}$ を，単調回帰 (Isotonic Regression, IR) 推定量という。$\mathcal{I} = \{1, 2, \ldots, k\}$ の部分集合 D に対して，

$$A_X(D) = \sum_{i \in D} \omega_i X_i / \sum_{i \in D} \omega_i$$

とおくとき，μ_i の IR 推定量は次の minimax 形式で与えられる (Robertson *et al.* (1988))。

$$\widehat{\mu}_i = \min_{\mathcal{L}_i} \max_{\mathcal{U}_i} A_X(\mathcal{L}_i \cap \mathcal{U}_i) = \max_{\mathcal{U}_i} \min_{\mathcal{L}_i} A_X(\mathcal{L}_i \cap \mathcal{U}_i)$$

但し，$\mathcal{L}_i$ は $\mu_j \le \mu_i$ なる $j \in \mathcal{I}$ で構成される任意の集合であり，$\mathcal{U}_i$ は，$\mu_i \le \mu_j$ なる $j \in \mathcal{I}$ で構成される任意の集合を表わす。ここで，$\boldsymbol{\Omega}^{-1} = \mathrm{diag}\,(\sigma_1^2/r_1, \ldots, \sigma_k^2/r_k)$ とるときには，$\widehat{\mu}_i$ は，μ_i の制限付き最尤 (REML) 推定量となっている。記号 $(s:t)$ を $\{s, s+1, \ldots, t\}$ なる整数の集合とすると，代表的な順序制約 SO, STO のもとでの IR 推定量は,

$$\begin{aligned} \widehat{\mu}_i^{SO} &= \min_{i \le t} \max_{s \le i} A_X(s:t) = \max_{s \le i} \min_{i \le t} A_X(s:t), \\ \widehat{\mu}_1^{STO} &= \min_{1 \in D} A_X(D), \\ \widehat{\mu}_i^{STO} &= \max(\widehat{\mu}_1^{STO}, X_i), \quad i = 2, \ldots, k \end{aligned}$$

で与えられる。例えば，$k = 3$ のときには,

$$\begin{aligned} \widehat{\mu}_1^{SO} &= \min\left\{X_1, \frac{\omega_1 X_1 + \omega_2 X_2}{\omega_1 + \omega_2}, \frac{\omega_1 X_1 + \omega_2 X_2 + \omega_3 X_3}{\omega_1 + \omega_2 + \omega_3}\right\}, \\ \widehat{\mu}_1^{STO} &= \min\left\{X_1, \frac{\omega_1 X_1 + \omega_2 X_2}{\omega_1 + \omega_2}, \frac{\omega_1 X_1 + \omega_3 X_3}{\omega_1 + \omega_3}, \frac{\omega_1 X_1 + \omega_2 X_2 + \omega_3 X_3}{\omega_1 + \omega_2 + \omega_3}\right\} \end{aligned}$$

と表され，$\widehat{\mu}_1^{STO} \le \widehat{\mu}_1^{SO}$ なる関係が成り立つ。順序制約によって推定量の結合（データのプーリング）がなされていることがわかる。

このような REML，IR 推定量は，制限無しの(不偏)推定量を改良してほしいわけであるが，一般には必ずしもそうなっていない。例えば，$\widehat{\mu}_1^{STO}$ は X_1 より優れているとはいえないことが，Lee (1988), Hwang and Peddada (1994) によって示された。しかし，このことは，順序制約という情報の不要さを示しているのではなく，むしろ順序制約の入り方によっては REML，IR 推定量が望ましくない場合があると考えた方がよい。実際順序制約 STO において，$\widehat{\mu}_1^{STO}$ の代わりに $\widehat{\mu}_1^{SO}$ を考えると，$\widehat{\mu}_1^{SO}$ は X_1 を改良することが示される。

Hwang and Peddada (1994) は，かなり一般的な条件のもとで $\widehat{\mu}_i^{SO}$ が X_i を改良することを示している。いま，すべての s に対して μ_i と μ_s の間に大小関係がつくとき，すなわち，$\mu_i \leq \mu_s$ もしくは $\mu_i \geq \mu_s$ のとき，μ_i は node (結び目) であるという。また $\boldsymbol{X}$ の密度関数が非増加関数 $g(\cdot)$ を用いて $g((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))$ で与えられるとし，$\boldsymbol{\Sigma}$ は対角行列，$\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ とする。このとき μ_i が node であれば，$\widehat{\mu}_i^{SO}$ は確率的に X_i を改良する。すなわち，すべての $c > 0$ に対して

$$P\left[|\widehat{\mu}_i^{SO} - \mu_i| < c\right] \geq P\left[|X_i - \mu_i| < c\right]$$

なる不等式が母数に関して一様に成り立つ。実はこの主張は，任意の非減少関数 $L(\cdot) \geq 0$ に対して

$$E\left[L(|\widehat{\mu}_i^{SO} - \mu_i|)\right] \leq E\left[L(|X_i - \mu_i|)\right]$$

が一様に成り立つことと同値であり (Hwang(1985))，これらの結果から直ちに，順序制約 SO のもとでは $\widehat{\mu}_i^{SO}$ が X_i を改良し，順序制約 STO のもとでは $\widehat{\mu}_1^{SO}$ が X_1 を改良することがいえる。

2.2 Hwang-Peddada による推定法

一般に順序関係が複雑に入っているときには REML 推定量の導出は大変になるだろうし，得られた REML 推定量が必ずしも X_i を改良しているとは限らない。このような点から REML 推定量に取って代わる推定量として，Hwang and Peddada (1994) は $\widehat{\mu}_i^{SO}$ に基づいた推定方法を提案している。それは，例えば Loop Ordering $\mu_1 \leq \mu_i \leq \mu_k,\ i = 2, \ldots, k-1$, においては，

(A) まず，μ_1, μ_k は node なので，μ_1, μ_k はそれぞれ $\widehat{\mu}_1^{SO}, \widehat{\mu}_k^{SO}$ で推定され，それらは X_1, X_k を改良している。ここで $\widehat{\mu}_1^{SO}, \widehat{\mu}_k^{SO}$ は $\mu_2, \ldots, \mu_{k-1}$ の間に $\mu_2 \leq \ldots \leq \mu_{k-1}$ なる順序関係が想定された上で得られていることに注意しよう。もしこの想定が正しければかなりよい推定がなされているだろうし，たとえ間違っていても X_1, X_k を改良しているわけで，実害はないものとみなされる。もし事前に順序について漠然としながらも何らかの情報が得られているのなら，それを用いて $\widehat{\mu}_1^{SO}, \widehat{\mu}_k^{SO}$ を構成することが望ましい。

(B) node になっていない母数，例えば μ_2 を推定するときには，残りの $\mu_3, \ldots, \mu_{k-1}$ を取り除いてできる順序関係 $\mu_1 \leq \mu_2 \leq \mu_k$ を考えると，μ_2 は node になり，推定量 $(\widehat{\mu}_1^{SO}, X_2, \widehat{\mu}_k^{SO})$ に重み $\boldsymbol{\Omega} = \mathrm{diag}\,(w, \omega_2, w)$ をとった IR 推定量 $\widehat{\mu}_2^{SO}(w)$ を作り，その極限として μ_2 の推定量

$$\widehat{\mu}_2^* = \lim_{w\to\infty} \widehat{\mu}_2^{SO}(w) = \min\{\max(\widehat{\mu}_1^{SO}, X_2), \widehat{\mu}_k^{SO}\}$$

を定義する。同様にして $\widehat{\mu}_3^*, \ldots, \widehat{\mu}_{k-1}^*$ が得られ，Loop Ordering については，すべての推定量が求まったことになる。

(C) もしはじめに与えられた順序関係の中に node がないときには，その部分集合の中に node が存在する最大の順序関係を見つけて上記の (A), (B) の方法で推定量を求め，後はこのことを繰り返していけば，すべての平均に推定量を与えることができる。

STO のとき，μ_i の推定量をこの方法で求めると $\widehat{\mu}_i^* = \max(\widehat{\mu}_1^{SO}, X_i),\ i = 2, \ldots, k$, となり，REML 推定量との違いは $\widehat{\mu}_1^{STO}$ を $\widehat{\mu}_1^{SO}$ で置き換えただけである。母数の間に適当な条件があれば $\widehat{\mu}_i^*$ が X_i を改良することが主張できる。Hwang and Peddada (1994) は，$\boldsymbol{X}$ の共分散行列 $\boldsymbol{\Sigma}$ が対角とは限らない場合の議論も展開している。

2.3 同時推定と線形関数の推定

順序制約下での平均の推定問題において，REML 推定量自体を改良する縮小推定量が得られている。共分散行列 $\boldsymbol{\Sigma}$ をもった多変量正規分布モデル $\boldsymbol{Z}=(Z_1,\ldots,Z_p)'\sim\mathcal{N}_p(\boldsymbol{\theta},\boldsymbol{\Sigma})$ において，$\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)'$ が正象限制限 (Positive Orthant Restriction)

$$\mathrm{POR}:\theta_i\geq 0,\quad i=1,\ldots,p$$

をみたしているとする。このとき $\boldsymbol{\theta}$ の REML 推定量は次のようにして与えられる。

$\mathcal{J}=\{1,\ldots,p\}$ の部分集合 a に対して，$|a|$ を a の要素の個数，$a'=\mathcal{J}\setminus a$ を a の補集合とし，$\boldsymbol{Z}_a$ を $i\in a$ なる Z_i から構成される $|a|$-次元ベクトルとする。また

$$\mathbf{Cov}\left[\begin{pmatrix}\boldsymbol{Z}_a\\ \boldsymbol{Z}_{a'}\end{pmatrix}\right]=\begin{pmatrix}\boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{aa'}\\ \boldsymbol{\Sigma}_{aa'} & \boldsymbol{\Sigma}_{a'a'}\end{pmatrix}$$

とし，

$$\boldsymbol{Z}_{a:a'}(\boldsymbol{\Sigma})=\boldsymbol{Z}_a-\boldsymbol{\Sigma}_{aa'}\boldsymbol{\Sigma}_{a'a'}^{-1}\boldsymbol{Z}_{a'},\quad \boldsymbol{\Sigma}_{a:a'}=\boldsymbol{\Sigma}_{aa}-\boldsymbol{\Sigma}_{aa'}\boldsymbol{\Sigma}_{a'a'}^{-1}\boldsymbol{\Sigma}_{a'a}$$

とおくと，$\boldsymbol{R}^p$ の 2^p 個の disjoint な分割

$$\mathcal{Z}_a(\boldsymbol{\Sigma})=\left\{\boldsymbol{Z}\in\boldsymbol{R}^p\mid\boldsymbol{\Sigma}_{a'a'}^{-1}\boldsymbol{Z}_{a'}\leq\boldsymbol{0},\boldsymbol{Z}_{a:a'}(\boldsymbol{\Sigma})>0\right\}$$

が定義される (Kudo(1963))。ここで $\cup_{a\in\mathcal{J}}\mathcal{Z}_a(\boldsymbol{\Sigma})=\boldsymbol{R}^p$ である。また $\boldsymbol{u}=(u_1,\ldots,u_{|a|})'$, $\boldsymbol{v}=(v_1,\ldots,v_{p-|a|})'$ に対して，k 次元ベクトル $\boldsymbol{\eta}_a(\boldsymbol{u},\boldsymbol{v})$ を

$$[\boldsymbol{\eta}_a(\boldsymbol{u},\boldsymbol{v})]_i=\begin{cases}u_i & \text{if}\ \ i\in a\\ v_i & \text{if}\ \ i\in a'\end{cases}$$

と定めると $\boldsymbol{\theta}$ の REML 推定量は，

$$\widehat{\boldsymbol{\theta}}^{REML}=\sum_{a\subset\mathcal{J}}\boldsymbol{\eta}_a(\boldsymbol{Z}_{a:a'}(\boldsymbol{\Sigma}),\boldsymbol{0})I(\boldsymbol{Z}\in\mathcal{Z}_a(\boldsymbol{\Sigma}))$$

で与えられる。但し，$I(A)$ は集合 A の定義関数である。

平均ベクトル $\boldsymbol{\theta}$ を 2 乗損失 $(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ のもとで同時推定する枠組みにおいて，Chang (1981, 82), Sengupta and Sen (1991) は $\widehat{\boldsymbol{\theta}}^{REML}$ を改良する縮小推定量を導出した。

$$\widehat{\boldsymbol{\theta}}^{RS}=\sum_{a\subset\mathcal{J}}I(\boldsymbol{Z}\in\mathcal{Z}_a(\boldsymbol{\Sigma}))\left\{1-\frac{c_a}{\{\boldsymbol{Z}_{a:a'}(\boldsymbol{\Sigma})\}'\boldsymbol{\Sigma}_{a:a'}^{-1}\boldsymbol{Z}_{a:a'}(\boldsymbol{\Sigma})}\right\}\boldsymbol{\eta}_a(\boldsymbol{Z}_{a:a'}(\boldsymbol{\Sigma}),\boldsymbol{0})$$

とおくと，$0\leq c_a\leq 2\max(|a|-2,0)$ なる c_a に対して，$\widehat{\boldsymbol{\theta}}^{RS}$ は $\widehat{\boldsymbol{\theta}}^{REML}$ を改良することが示される。$\boldsymbol{\Sigma}$ が未知のときには，上述の推定量において $\boldsymbol{\Sigma}$ をその推定量 $\widehat{\boldsymbol{\Sigma}}$ で置き換えることによって同様な結果が得られることを，Sengupta and Sen (1991) が示し，また Kuriki and Takemura (1997) は，$\boldsymbol{\Sigma}=\boldsymbol{I}_p$ で平均ベクトル $\boldsymbol{\theta}$ が滑らかな境界をもった閉凸集合に制約されているときに，その改良結果を示した。

この結果を用いると順序制約 SO, STO において REML 推定量を改良する推定量を導出することができる。例えば，2.1 節で取り上げられた順序制約 SO は，$\theta_1=\mu_1$, $\theta_i=\mu_i-\mu_{i-1}$, $i=2,\ldots,k$, とおくことにより，$-\infty<\theta_1<\infty$, $\theta_2\geq 0,\ldots,\theta_k\geq 0$ となり suborthant model として表現することができる。これに対応して $Z_1=X_1$, $Z_i=X_i-X_{i-1}$,

$i=2,\ldots,k$, とおくと，$\boldsymbol{Z}=(Z_1,\ldots,Z_k)'$ の共分散行列 $\boldsymbol{\Sigma}$ は三重対角行列となる。$\mathcal{J}^0=\{2,\ldots,k\}$ とおくと，$\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)'$ の REML 推定量は

$$\widehat{\boldsymbol{\theta}}^{REML}=\sum_{a\subset\mathcal{J}^0} I(\boldsymbol{Z}\in\mathcal{Z}_a(\boldsymbol{\Sigma}))\boldsymbol{\eta}_{(1,a)}(\boldsymbol{Z}_{(1,a):a'},\boldsymbol{0})$$

となるが，縮小推定量

$$\widehat{\boldsymbol{\theta}}^{RS}=\sum_{a\in\mathcal{J}^0} I(\boldsymbol{Z}\in\mathcal{Z}_a(\boldsymbol{\Sigma}))\left\{1-\frac{c_a}{\{\boldsymbol{Z}_{(1,a):a'}(\boldsymbol{\Sigma})\}'\boldsymbol{\Sigma}^{-1}_{(1,a):a'}\boldsymbol{Z}_{(1,a):a'}(\boldsymbol{\Sigma})}\right\}\boldsymbol{\eta}_{(1,a)}(\boldsymbol{Z}_{(1,a):a'},\boldsymbol{0})$$

によって，$0\le c_a\le 2\max(|a|-1,0)$ のとき改良される。

ここで取り上げられた suborthant model での縮小推定は，線形回帰モデル $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ において，特定の説明変数に対しては回帰係数 β_i が $\beta_i\ge 0$ や $\beta_i\le 1$ などの制約が前提とされる場面において有効であると思われるが，こうした応用はまだあまり研究されていない。

正象限制約 POR をもつ平均の推定に関連して，線形関数 $\boldsymbol{c}'\boldsymbol{\theta}=\sum_{i=1}^p c_i\theta_i$ を推定する問題を紹介する。これについては，いくつかの論文で研究されてきたが，最近 Shinozaki and Chang (1999) は明解な結論を与えることに成功した。そこでは，$\boldsymbol{\Sigma}=\boldsymbol{I}_p$ のとき，REML 推定量 $\sum_{i=1}^p c_i\max(Z_i,0)$ が $\sum_{i=1}^p c_iZ_i$ を改良するための必要十分条件は，

$$(\pi+1)\sum_{i\in T}c_i^2-\left(\sum_{i\in T}c_i\right)^2\ge 0$$

が任意の $T\subset\{i|c_i>0\}$ 及び任意の $T\subset\{i|c_i<0\}$ に対して成り立つということが示され，その結果，

$$E\left[\left\{\sum_{i=1}^p c_i\max(Z_i,0)-\boldsymbol{c}'\boldsymbol{\theta}\right\}^2\right]\le E\left[\{\boldsymbol{c}'\boldsymbol{Z}-\boldsymbol{c}'\boldsymbol{\theta}\}^2\right]\ \text{for all}\ \boldsymbol{c}\in\boldsymbol{R}^p\ \text{and all}\ \boldsymbol{\theta}\in\boldsymbol{R}^p \quad (2.1)$$

が成り立つための必要十分条件が $p\le 4$ であることが導かれる。(2.1) の主張は，Lehmann (1983) によって取り上げられた強い基準であり，分布の正規性をはずしたときの必要十分条件や一般的な orthant cone のもとでの条件が，Fernandez *et al.* (1999) によって議論されている。

3 分散の順序制約と分散成分の推定

3.1 制限付き最尤推定量

k 個の母分散の間に順序関係が入っているときの推定問題を考えよう。いま $T_1,\ldots,T_k$ を互いに独立な確率変数で，T_i/σ_i^2 が $\chi^2_{m_i}$-分布に従うとする。このとき，平均の順序関係の類推から，例えば

$$\begin{aligned}\text{SO}\ &:\ \sigma_1^2\le\sigma_2^2\le\ldots\le\sigma_k^2,\\ \text{STO}\ &:\ \sigma_1^2\le\sigma_i^2,\quad i=2,\ldots,k\end{aligned}$$

なる順序関係が思いつく。これらの順序制約のもとでの REML 推定量は次のように与えられる。$\mathcal{I}=\{1,2,\ldots,k\}$ の部分集合 D に対して

$$B_T(D)=\sum_{i\in D}T_i/\sum_{i\in D}m_i$$

とおくとき，順序制約 SO, STO のもとでの REML 推定量は，

$$\begin{aligned}\hat{\sigma}_i^{2SO} &= \min_{i\le t}\max_{s\le i} B_T(s:t) = \max_{s\le i}\min_{i\le t} B_T(s:t),\\ \hat{\sigma}_1^{2STO} &= \min_{1\in D} B_T(D),\\ \hat{\sigma}_i^{2STO} &= \max(\hat{\sigma}_1^{2STO}, m_i^{-1}T_i), \quad i=2,\ldots,k\end{aligned}$$

となる。これらの有効性に関する性質は，Hwang and Peddada (1994) が STO において集中確率の意味で $\hat{\sigma}_1^{2UB} = m_1^{-1}T_1$ は $\hat{\sigma}_1^{2STO}$ によってではなく $\hat{\sigma}_1^{2SO}$ によって改良されることを示したが，それ以外，平均の推定問題ほどは十分に研究されていない。

分散の順序制約は，何らかのノイズ等の影響で i 番目のバラツキ σ_i^2 が σ_1^2 (もしくは σ_{i-1}^2) より大きくなることが想定されるときに考えられるだろうが，変量，混合モデルなどを考えると分散のこうした順序関係がモデルから自動的に生ずることがわかる。例えば一元配置変量モデル

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i=1,\ldots,k,\ j=1,\ldots,r \tag{3.1}$$

を考えよう。ここで，α_i, ε_{ij} は互いに独立な確率変数で $\alpha_i \sim \mathcal{N}(0,\sigma_A^2)$, $\varepsilon_{ij} \sim \mathcal{N}(0,\sigma_1^2)$ とし，μ, σ_A^2, σ_1^2 を未知母数とする。$\overline{y}_{i\cdot} = r^{-1}\sum_{j=1}^r y_{ij}$, $\overline{y}_{\cdot\cdot} = (kr)^{-1}\sum_{i=1}^k\sum_{j=1}^r y_{ij}$,

$$S_1 = \sum_{i=1}^k\sum_{j=1}^r (y_{ij} - \overline{y}_{i\cdot})^2, \quad S_2 = r\sum_{i=1}^k (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2, \tag{3.2}$$

とおくと，最小十分統計量 $\overline{y}_{\cdot\cdot}$, S_1, S_2 は互いに独立で，$\overline{y}_{\cdot\cdot} \sim \mathcal{N}(\mu, (\sigma_1^2 + r\sigma_A^2)/rk)$,

$$\begin{aligned}S_1/\sigma_1^2 &\sim \chi_{n_1}^2, \quad n_1 = k(r-1)\\ S_2/\sigma_2^2 &\sim \chi_{n_2}^2, \quad n_2 = k-1, \quad \sigma_2^2 = \sigma_1^2 + r\sigma_A^2\end{aligned} \tag{3.3}$$

なる分布に従い，σ_1^2 と σ_2^2 の間には $\sigma_1^2 \le \sigma_2^2$ なる不等式制約が入っていることがわかる。

また，枝分かれ (nested) 二元配置変量モデル

$$y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ijk}, \quad i=1,\ldots,I,\ j=1,\ldots,J.\ k=1,\ldots,K$$

において，α_i, γ_{ij}, ε_{ijk} は互いに独立な確率変数で，$\alpha_i \sim \mathcal{N}(0,\sigma_A^2)$, $\gamma_{ij} \sim \mathcal{N}(0,\sigma_{AB}^2)$, $\varepsilon_{ijk} \sim \mathcal{N}(0,\sigma_1^2)$ とし，$\mu, \sigma_A^2, \sigma_{AB}^2, \sigma_1^2$ を未知母数とする。このとき，$S_1 = \sum_{i,j,k}(y_{ijk} - \overline{y}_{ij\cdot})^2$, $S_2 = K\sum_{i,j}(\overline{y}_{ij\cdot} - \overline{y}_{i\cdot\cdot})^2$, $S_3 = JK\sum_i(\overline{y}_{i\cdot\cdot} - \overline{y}_{\cdot\cdot\cdot})^2$ とおくと，

$$\begin{aligned}S_1/\sigma_1^2 &\sim \chi_{n_1}^2, \quad n_1 = IJ(K-1),\\ S_2/\sigma_2^2 &\sim \chi_{n_2}^2, \quad n_2 = I(J-1), \quad \sigma_2^2 = \sigma_1^2 + K\sigma_{AB}^2,\\ S_3/\sigma_3^2 &\sim \chi_{n_3}^2, \quad n_3 = I-1, \quad \sigma_3^2 = \sigma_1^2 + K\sigma_{AB}^2 + JK\sigma_A^2\end{aligned}$$

なる分布に従い，$\sigma_1^2, \sigma_2^2, \sigma_3^2$ の間には

$$\sigma_1^2 \le \sigma_2^2 \le \sigma_3^2$$

なる順序制約 SO 制約が成り立っている。一般に nested な変量モデルを考えると，分散の順序制約 SO がモデルから導かれるが，nested でない変量モデルでは順序制約 SO が部分的に成り立っている。

3.2 分散の Stein 推定量との関係

分散の順序制約下での推定と分散の Stein 推定との関係を論ずることは興味深い。まず $k=2$ のときには，不等式制約 $\sigma_1^2 \le \sigma_2^2$ のもとで σ_1^2 の REML 推定量は，

$$\hat{\sigma}_1^{2SO}(T_1, T_2, m_1, m_2) = \min\left\{\frac{T_1}{m_1}, \frac{T_1+T_2}{m_1+m_2}\right\}$$

と表されるが，実は，これは分散の Stein 推定量と同じ形をしている。

一元配置モデル (3.1) において $\theta_i = \mu + \alpha_i$ が母数の場合を考えよう。このとき (3.2) の S_1, S_2 は

$$S_1/\sigma_1^2 \sim \chi^2_{n_1}, \quad S_2/\sigma_1^2 \sim \chi^2_{n_2}(r\sum_i \alpha_i^2/\sigma_1^2)$$

に従う。ここで $\chi^2_{n_2}(\lambda)$ は，非心度 λ，自由度 n_2 の非心カイ二乗分布を表わしている。このとき Stein (1964) は，

$$\hat{\sigma}_1^{2ST} = \min\left\{\frac{S_1}{n_1}, \frac{S_1+S_2}{n_1+n_2}\right\}$$

を提案し，エントロピー損失 $L(\hat{\sigma}_1^2/\sigma_1^2) = \hat{\sigma}_1^2/\sigma_1^2 - \log\hat{\sigma}_1^2/\sigma_1^2 - 1$ に関して

$$E\left[L\left(\hat{\sigma}_1^{2ST}/\sigma_1^2\right)\right] \le E\left[L\left(n_1^{-1}S_1/\sigma_1^2\right)\right] \tag{3.4}$$

が一様に成り立つことを示した。仮説 $H_0: \ \theta_1 = \cdots = \theta_k$ を検定するための統計量は，$n_1S_2/(n_2S_1)$ であり

$$\hat{\sigma}_1^{2ST} = \begin{cases} (n_1+n_2)^{-1}(S_1+S_2) & \text{if } n_1S_2/(n_2S_1) \le 1, \\ n_1^{-1}S_1 & \text{otherwise} \end{cases}$$

とかけるので，$\hat{\sigma}_1^{2ST}$ は，予備検定推定量になっている。また $\hat{\sigma}_1^{2ST}$ が経験 Bayes 推定量であることもわかる (Kubokawa *et al.* (1992))。

(3.4) の結果は，母数モデルにおいて Stein 推定量の優越性を示している強い結果である。この結果から，順序制約下での REML 推定量 $\hat{\sigma}_1^{2SO}$ の優越性が $\alpha_1, \ldots, \alpha_k$ を変量にすることによって導かれる。実際 (3.4) の両辺を $\alpha_1, \ldots, \alpha_k$ に関して期待値をとると，S_1, S_2 の分布は (3.3) となるので，(3.4) は

$$E\left[L\left(\hat{\sigma}_1^{2SO}(S_1, S_2, n_1, n_2)/\sigma_1^2\right)\right] \le E\left[L\left(n_1^{-1}S_1/\sigma_1^2\right)\right]$$

となる。これは，不等式制約下での $\hat{\sigma}_1^{2SO}(S_1, S_2, n_1, n_2)$ の $n_1^{-1}S_1$ に対する優越性を示している。すなわち母数を変量にすることによって， Stein (1964) の結果から分散の不等式制約下での結果が導かれる。

3.3 分散成分の推定

一元配置の変量モデル (3.1) において，分散成分 σ_1^2, σ_A^2 はそれぞれ分散の群内成分，群間成分と呼ばれる。ここでは，(3.3) の S_1, S_2 に基づいてこれらの分散成分を推定する問題を扱う。

まず σ_1^2, σ_A^2 の不偏推定量は，

$$\hat{\sigma}_1^{2UB} = n_1^{-1}S_1, \quad \hat{\sigma}_A^{2UB} = r^{-1}\left(n_2^{-1}S_2 - n_1^{-1}S_1\right)$$

で与えられる。$\hat{\sigma}_A^{2UB}$ は, $\sigma_2^2 = \sigma_1^2 + r\sigma_A^2$ なる関係から $\hat{\sigma}_A^{2UB} = r^{-1}(\hat{\sigma}_2^{2UB} - \hat{\sigma}_1^{2UB})$ として求められる。しかし群間成分の不偏推定量 $\hat{\sigma}_A^{2UB}$ は正の確率で負の値をとってしまうという欠点がある。この問題について LaMotte (1973) は σ_A^2 の不偏でかつ非負な 2 次形式推定量が存在しないことを証明している。さらに Kleffe and Rao (1986) は σ_A^2 の不偏でない非負の 2 次形式推定量が群内の繰り替えし数 r を固定した上で群の数 k を大きくするという実際的な漸近論に関して一致性をもたないことも示している。したがって 2 次形式以外の非負な推定量を考える必要があろう。

そこで σ_1^2, σ_A^2 の REML 推定量を求めてみよう。σ_1^2, σ_2^2 の REML 推定量は, 3.1 節の議論から

$$\begin{aligned}\hat{\sigma}_1^{2REML} &= \min\left\{n_1^{-1}S_1, (n_1+n_2)^{-1}(S_1+S_2)\right\},\\ \hat{\sigma}_2^{2REML} &= \max\left\{n_2^{-1}S_2, (n_1+n_2)^{-1}(S_1+S_2)\right\}\end{aligned}$$

で与えられるから, $\hat{\sigma}_A^{2REML} = r^{-1}(\hat{\sigma}_2^{2REML} - \hat{\sigma}_1^{2REML})$ なる関係を用いると, σ_A^2 の REML 推定量は,

$$\hat{\sigma}_A^{2REML} = \max\left\{r^{-1}\left(n_2^{-1}S_2 - n_1^{-1}S_1\right), 0\right\}$$

となり, 非負な推定量が得られる。

REML 推定量の不偏推定量に対する優越性を示すために, (σ_1^2, σ_A^2) を $(\hat{\sigma}_1^2, \hat{\sigma}_A^2)$ で同時に推定するときの Kulback-Leibler 情報量損失

$$L(\hat{\sigma}_1^2, \hat{\sigma}_A^2; \sigma_1^2, \sigma_A^2) = n_1\left\{\frac{\hat{\sigma}_1^2}{\sigma_1^2} - \log\frac{\hat{\sigma}_1^2}{\sigma_1^2} - 1\right\} + n_2\left\{\frac{\hat{\sigma}_1^2 + r\hat{\sigma}_A^2}{\sigma_1^2 + r\sigma_A^2} - \log\frac{\hat{\sigma}_1^2 + r\hat{\sigma}_A^2}{\sigma_1^2 + r\sigma_A^2} - 1\right\}$$

を導入する。この損失関数は 2 つの変数 $\hat{\sigma}_1^2/\sigma_1^2$, $\hat{\sigma}_A^2/\sigma_A^2$ に関して凸関数になっている。この損失関数には, $\hat{\sigma}_2^2 = \hat{\sigma}_1^2 + r\hat{\sigma}_A^2$ なる関係が埋め込まれている。Kubokawa *et al.* (1999) では, 不偏推定量 $(\hat{\sigma}_1^{2UB}, \hat{\sigma}_A^{2UB})$ が Kullback-Leibler 損失関数のもとでミニマクスとなることを示している。

より一般的に, 母数 σ_1^2, σ_A^2 についての不偏推定量 $(\hat{\sigma}_1^{2UB}, \hat{\sigma}_A^{2UB})$ を改良する推定量のクラスを構成し, その中から非負な望ましい推定量を導くために,

$$\hat{\sigma}_1^2(\psi_1) = S_1\psi_1\left(\frac{S_2}{S_1}\right), \quad \hat{\sigma}_A^2(\psi_1, \psi_2) = \frac{1}{r}\left\{S_2\psi_2\left(\frac{S_1}{S_2}\right) - S_1\psi_1\left(\frac{S_2}{S_1}\right)\right\}$$

なる形の推定量を考えよう。ここで ψ_1, ψ_2 は, 不偏推定量をより望ましい形に修正するための関数である。この修正の程度をいかに決めるかが重要な問題であり, ここでは危険関数を導入することにより不偏推定量よりは悪くならない程度の修正の度合いを明確にすることが可能となる。Kullback-Leibler 損失関数に関する推定量 $(\hat{\sigma}_1^2(\psi_1), \hat{\sigma}_A^2(\psi_1, \psi_2))$ の危険関数は, $\omega = (\sigma_1^2, \sigma_A^2, \mu)$ に対して,

$$R(\omega; \hat{\sigma}_1^2(\psi_1), \hat{\sigma}_A^2(\psi_1, \psi_2)) = n_1R_1(\omega; S_1\psi_1\left(\frac{S_2}{S_1}\right)) + n_2R_2(\omega; S_2\psi_2\left(\frac{S_1}{S_2}\right)),$$

と分解することができる。ただし,

$$R_1(\omega; S_1\psi_1\left(\frac{S_2}{S_1}\right)) = E\left[\frac{S_1}{\sigma_1^2}\psi_1\left(\frac{S_2}{S_1}\right) - \log\frac{S_1}{\sigma_1^2}\psi_1\left(\frac{S_2}{S_1}\right) - 1\right],$$

$$R_2(\omega; S_2\psi_2\left(\frac{S_1}{S_2}\right)) = E\left[\frac{S_2}{\sigma_1^2 + r\sigma_A^2}\psi_2\left(\frac{S_1}{S_2}\right) - \log\frac{S_2}{\sigma_1^2 + r\sigma_A^2}\psi_2\left(\frac{S_1}{S_2}\right) - 1\right]$$

である。ここでの問題に対して Kubokawa *et al.* (1999) は, $i=1,2$ に対して, $R_i(\omega;S_i\psi_i) \le R_i(\omega;n_i^{-1}S_i)$ がすべての ω について一様に成り立つためには,

(a) $\psi_i(w)$ は非減少で, $\lim_{w\to\infty}\psi_i(w^{\delta_i}) = n_i^{-1}$,

(b) $i=1$ のときには $\{\psi_1(w)\} \ge \psi_1^*(w)$, $i=2$ のときには $\{\psi_2(w)\} \le \psi_2^*(w^{-1})$

なる十分条件をみたすことであることを示した。ここで, $\delta_1=1$, $\delta_2=-1$ であり,

$$\psi_i^*(w) = \frac{1}{n_1+n_2}\frac{\int_0^w x^{n_2/2-1}/(1+x)^{(n_1+n_2)/2}dx}{\int_0^w x^{(n_2-1-\delta_i)/2}/(1+x)^{(n_1+n_2+2)/2}dx}$$

とする。

この結果から, ψ_1, ψ_2 がそれらの条件をみたせば, $(\hat\sigma_1^2(\psi_1), \hat\sigma_A^2(\psi_1,\psi_2))$ が $(\hat\sigma_1^{2UB}, \hat\sigma_A^{2UB})$ を改良すること, すなわちミニマクスとなることがわかる。Kubokawa *et al.* (1999) では, REML 推定量 $(\hat\sigma_1^{2REML}, \hat\sigma_A^{2REML})$ がこのクラスに入ることが示されるとともに, 下で与えられる新しい形の推定量 $(\hat\sigma_1^{2EB}, \hat\sigma_A^{2EB})$, $(\hat\sigma_1^{2GB}, \hat\sigma_A^{2GB})$ もそのクラスに入ることを示した。ここで,

$$\begin{cases} \hat\sigma_1^{2EB} = \min\left\{\dfrac{S_1}{n_1}, \dfrac{S_1+n_2S_2/(n_2+2)}{n_1+n_2}\right\}, \\ \hat\sigma_A^{2EB} = \dfrac{1}{r}\left[\max\left\{\dfrac{S_2}{n_2}, \dfrac{(n_2+2)S_1/n_2+S_2}{n_1+n_2}\right\} - \hat\sigma_1^{2EB}\right], \end{cases}$$

$$\begin{cases} \hat\sigma_1^{2GB} = \dfrac{S_1}{n_1+n_2}\dfrac{\int_0^{S_2/S_1} x^{n_2/2-1}/(1+x)^{(n_1+n_2)/2}dx}{\int_0^{S_2/S_1} x^{n_2/2-1}/(1+x)^{(n_1+n_2+2)/2}dx}, \\ \hat\sigma_A^{2GB} = \dfrac{1}{r}\left[\dfrac{S_2}{n_1+n_2}\dfrac{\int_0^{S_2/S_1} x^{n_2/2-1}/(1+x)^{(n_1+n_2)/2}dx}{\int_0^{S_2/S_1} x^{n_2/2}/(1+x)^{(n_1+n_2)/2}dx} - \hat\sigma_1^{2GB}\right] \end{cases}$$

ここで明らかに $\hat\sigma_A^{2EB}$, $\hat\sigma_A^{2GB}$ は確率 1 で正の値をとる推定量である。また $(\hat\sigma_1^{2REML}, \hat\sigma_A^{2REML})$, $(\hat\sigma_1^{2EB}, \hat\sigma_A^{2EB})$ が経験 Bayes 推定量, $(\hat\sigma_1^{2GB}, \hat\sigma_A^{2GB})$ が一般化 Bayes 推定量となることも示される。シミュレーション実験による比較では, $(\hat\sigma_1^{2REML}, \hat\sigma_A^{2REML})$, $(\hat\sigma_1^{2EB}, \hat\sigma_A^{2EB})$ が $(\hat\sigma_1^{2UB}, \hat\sigma_A^{2UB})$ に比べてかなりよい推定量であることがわかる。また $(\hat\sigma_1^{2GB}, \hat\sigma_A^{2GB})$ は$\sigma_A^2 = 0.0$ のときには$(\hat\sigma_1^{2UB}, \hat\sigma_A^{2UB})$ と同じ危険関数の値を示すが, $\sigma_A^2/\sigma_1^2 \ge 0.5$ の範囲では 4 つの推定量の中で最良のものになっていることもわかる。こうしたシミュレーション実験とその結果の詳細や繰り返し数が等しくないときの結果については, Kubokawa *et al.* (1999) を参照されたい。

3.4 多変量混合モデルにおける分散成分の推定

多変量分散成分モデルもしくは多変量混合線形モデルにおける分散成分の推定は, Bock and Vandenberg (1968), Bock and Petersen (1975), Klotz and Putter (1969) らによって ML や REML 推定に関して古くから研究されてきた問題であり, 最近 Amemiya (1985), Amemiya and Fuller (1984), Anderson *et al.* (1986) らの論文によって再認識され, 理論研究の新たな展開をみせている。

多変量の一元配置モデル

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij}, \quad i=1,\dots,k,\ j=1,\dots,r,$$

において $\boldsymbol{\alpha}_i$, $\boldsymbol{e}_{ij}$ は独立な確率変数で, $\boldsymbol{\alpha}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_A)$, $\boldsymbol{e}_{ij} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_1)$ に従う。$\boldsymbol{\mu} \in \boldsymbol{R}^p$ は未知の共通平均, $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_1$ は未知の共分散行列とする。$\overline{\boldsymbol{y}}_{i.} = r^{-1}\sum_{j=1}^r \boldsymbol{y}_{ij}$, $\overline{\boldsymbol{y}}_{..} =$

$(rk)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{r}\boldsymbol{y}_{ij}$ として，$\boldsymbol{S}_1=\sum_{i=1}^{k}\sum_{j=1}^{r}(\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i.})(\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i.})'$, $\boldsymbol{S}_2=r\sum_{i=1}^{k}(\overline{\boldsymbol{y}}_{i.}-\overline{\boldsymbol{y}}_{..})(\overline{\boldsymbol{y}}_{i.}-\overline{\boldsymbol{y}}_{..})'$ とおくと，$\overline{\boldsymbol{y}}_{..}$, $\boldsymbol{S}_1$，$\boldsymbol{S}_2$ は最小十分統計量で，互いに独立に

$$\begin{aligned}\overline{\boldsymbol{y}}_{..} &\sim \mathcal{N}_p(\boldsymbol{\mu},(rk)^{-1}\boldsymbol{\Sigma}_2), \quad \boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}_1+r\boldsymbol{\Sigma}_A,\\ \boldsymbol{S}_1 &\sim \mathcal{W}_p(\boldsymbol{\Sigma}_1,n_1), \quad n_1=k(r-1),\\ \boldsymbol{S}_2 &\sim \mathcal{W}_p(\boldsymbol{\Sigma}_2,n_2), \quad n_2=k-1,\end{aligned}$$

に従う。共分散行列 $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ の間には

$$\boldsymbol{\Sigma}_1\le\boldsymbol{\Sigma}_2$$

なる不等式制約が入っている。ただし，この不等式は $\boldsymbol{\Sigma}_2-\boldsymbol{\Sigma}_1$ が非負定値であることを表わす。推定精度を高めるためにこの不等式制約の情報を使いたいのであるが，有限標本での推定量の優越性を主張することは技術的に困難であると思われてきた。しかし最近 Srivastava and Kubokawa (1999) は，この問題を解決するのに成功したので，彼らの主要部分を以下に紹介する。

まず群内分散 $\boldsymbol{\Sigma}_1$ の推定問題を扱う。$\boldsymbol{S}_2^{1/2}$ を $\boldsymbol{S}_2=(\boldsymbol{S}_2^{1/2})^2$ なる対称行列とし，$\boldsymbol{P}$ を

$$\boldsymbol{P}'\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}\boldsymbol{P}=\boldsymbol{\Lambda}=\mathrm{diag}\,(\lambda_1,\ldots,\lambda_p),$$

なる $p\times p$ 直交行列とする。ここで固有値は $\lambda_1\ge\lambda_2\ge\ldots\ge\lambda_p$ をみたすとする。非負関数 $\psi_i(\boldsymbol{\Lambda})$ に対して，$\boldsymbol{\Psi}(\boldsymbol{\Lambda})=\mathrm{diag}\,(\psi_1(\boldsymbol{\Lambda}),\ldots,\psi_p(\boldsymbol{\Lambda}))$ とおいて，$\boldsymbol{\Sigma}_1$ の推定量の一般形

$$\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})=\boldsymbol{S}_2^{1/2}\boldsymbol{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}'\boldsymbol{S}_2^{1/2}$$

を考えよう。この推定量を $\boldsymbol{S}_2$ の情報を用いて改善するために，打ち切りルール $[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}$ を

$$\begin{aligned}[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR} &= \mathrm{diag}\,(\psi_1^{TR}(\boldsymbol{\Lambda}),\ldots,\psi_p^{TR}(\boldsymbol{\Lambda})),\\ \psi_i^{TR}(\boldsymbol{\Lambda}) &= \min\left\{\psi_i(\boldsymbol{\Lambda}),\frac{\lambda_i+1}{n_1+n_2}\right\}, \quad i=1,\ldots,p,\end{aligned}$$

で定義する。そのとき打ち切り推定量は

$$\widehat{\boldsymbol{\Sigma}}_1([\boldsymbol{\Psi}]^{TR})=\boldsymbol{S}_2^{1/2}\boldsymbol{P}\mathrm{diag}\,(\psi_1^{TR}(\boldsymbol{\Lambda}),\ldots,\psi_p^{TR}(\boldsymbol{\Lambda}))\boldsymbol{P}'\boldsymbol{S}_2^{1/2},$$

と表わされる。Srivastava and Kubokawa (1999) は，エントロピー損失のもとで，推定量 $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ が $\widehat{\boldsymbol{\Sigma}}_1([\boldsymbol{\Psi}]^{TR})$ によって改良されることを証明した。

$\boldsymbol{S}_1=\boldsymbol{S}_2^{1/2}\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_2^{1/2}=\boldsymbol{S}_2^{1/2}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{S}_2^{1/2}$ であるから，不偏推定量 $\widehat{\boldsymbol{\Sigma}}_1^{UB}=n_1^{-1}\boldsymbol{S}_1$ は

$$\widehat{\boldsymbol{\Sigma}}_1^{UB}=\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi}^{UB}), \quad \boldsymbol{\Psi}^{UB}=\mathrm{diag}\,(n_1^{-1}\lambda_1,\ldots,n_1^{-1}\lambda_p)$$

と表わされる。上述の打ち切りルールを適用すると $\widehat{\boldsymbol{\Sigma}}_1^{REML}=\widehat{\boldsymbol{\Sigma}}_1([\boldsymbol{\Psi}^{UB}]^{TR})$,

$$[\boldsymbol{\Psi}^{UB}]^{TR}=\mathrm{diag}\,\left(\min\left\{\frac{\lambda_1}{n_1},\frac{\lambda_1+1}{n_1+n_2}\right\},\ldots,\min\left\{\frac{\lambda_p}{n_1},\frac{\lambda_p+1}{n_1+n_2}\right\}\right)$$

なる推定量が得られ不偏推定量を改良することがわかるが，これは REML 推定量になっている。

上で取り上げられた不偏推定量がミニマクスにならないことは，共分散行列の推定においてよく知られた結果である。我々の問題において1つのミニマクスな推定量はStein型推定量で $\widehat{\boldsymbol{\Sigma}}_1^{ST}=\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi}^{ST})$,

$$\boldsymbol{\Psi}^{ST}(\boldsymbol{\Lambda})=\mathrm{diag}\,(d_1\lambda_1,\ldots,d_p\lambda_p),\quad d_i=(n_1+p+1-2i)^{-1},\quad i=1,\ldots,p.$$

で与えられる。この推定量に打ち切りルールを適用すると $\widehat{\boldsymbol{\Sigma}}_1^{STTR}=\widehat{\boldsymbol{\Sigma}}_1([\boldsymbol{\Psi}^{ST}]^{TR})$,

$$[\boldsymbol{\Psi}^{ST}(\boldsymbol{\Lambda})]^{TR}=\mathrm{diag}\left(\min\left\{d_1\lambda_1,\frac{\lambda_1+1}{n_1+n_2}\right\},\ldots,\min\left\{d_p\lambda_p,\frac{\lambda_p+1}{n_1+n_2}\right\}\right),$$

となる。推定量 $\widehat{\boldsymbol{\Sigma}}_1^{STTR}$ がリスクの意味で優れていることが Srivastava and Kubokawa (1999) によって数値的に示されている。

一方，同様な打ち切りルールが群間分散 $\boldsymbol{\Sigma}_A$ の推定においても得られ，不偏推定量に対して

$$\widehat{\boldsymbol{\Sigma}}_A^{REML}=r^{-1}\boldsymbol{S}_2^{1/2}\boldsymbol{P}\mathrm{diag}\left(\max\left\{\frac{1}{n_2}-\frac{\lambda_i}{n_1},0\right\},i=1,\ldots,p\right)\boldsymbol{P}'\boldsymbol{S}_2^{1/2},$$

や，Stein 型ミニマクス推定量に対して

$$\widehat{\boldsymbol{\Sigma}}_A^{STR}=r^{-1}\boldsymbol{S}_2^{1/2}\boldsymbol{P}\mathrm{diag}\,(\max\{e_i-d_i\lambda_i,0\},i=1,\ldots,p)\,\boldsymbol{P}'\boldsymbol{S}_2^{1/2}$$

などの，非負定値な打ち切り推定量が導かれる。ここで $e_i=1/\{n_2-(p+1-2i)\}$ である。詳しくは，Srivastava and Kubokawa (1999) に書かれている。また，Calvin and Dykstra (1991a,b) は多変量の分散成分が3つ以上ある場合，例えばその場合

$$\boldsymbol{\Sigma}_1\le\boldsymbol{\Sigma}_2\le\boldsymbol{\Sigma}_3$$

などのような順序制約が入ることになるが，そのときの REML 推定量や最小2乗推定量についての議論を行っている。

4 母数の等号制約と他のトピック

まず，いくつかの母数の間に等号制約が課せられる問題について述べる。例えば，平均 μ_i, 分散 σ_i^2, 繰り返し数 r_i の k 標本問題において，母数の間に，

H_0: $\mu_1=\cdots=\mu_k=\mu,\ \sigma_1^2=\cdots=\sigma_k^2=\sigma^2$,
H_1: $\mu_1=\cdots=\mu_k=\mu$,
H_2: $\sigma_1^2=\cdots=\sigma_k^2=\sigma^2$,

などの等号制約が考えられる。H_0 は最も強い制約であるが，H_2 は比較的受け入れられる制約である。H_1 は共通平均 μ の推定問題になっており，$\overline{X}_i$ を μ_i の標本平均，$\hat{\sigma}_i^2$ を σ_i^2 の推定量とすると，μ の2段階一般化最小2乗 (FGLS) 推定量は

$$\hat{\mu}^{FGLS}=\left(\sum_{i=1}^k\frac{r_i}{\hat{\sigma}_i^2}\right)^{-1}\sum_{i=1}^k\frac{r_i}{\hat{\sigma}_i^2}\overline{X}_i$$

で与えられ，各標本平均をその精度の比で内分した形をしている。この形の $\overline{X}_1,\ldots,\overline{X}_k$ の結合推定量は，Graybill and Deal (1959) 推定量とも呼ばれ，例えば $\widehat{\mu}^{FGLS}$ がすべての $\overline{X}_i$ を改良するための十分条件は，$r_i \geq 11$ である (Shinozaki (1978), Bhattacharya (1984))。また特に $k=2$ のときの決定理論的議論が Brown and Cohen (1974), Cohen and Sackrowitz (1974), Kubokawa (1987) などによって展開された。この問題はまた，古くから研究されてきた，釣合型不完備ブロック計画におけるブロック間情報の回復問題とも関連している (Yates (1940), Seshadri (1963), Shah (1964), Stein (1966)，広津 (1976))。

共分散行列の級内相関構造は $r \times r$ 共分散行列 $\boldsymbol{\Sigma}$ の 1 つの制約と考えることができるだろう。これは，$\boldsymbol{e}_r = (1,\ldots,1)' \in \boldsymbol{R}^r$ に対して，

$$\boldsymbol{\Sigma} = \sigma^2\left\{(1-\rho)\boldsymbol{I}_r + \rho\boldsymbol{e}_r\boldsymbol{e}_r'\right\}$$

で与えられる。ρ は級内相関係数 (Intraclass Correlation Coefficient) と呼ばれる。この構造は，変量モデルの共分散行列に対応している。(3.1) において $\boldsymbol{y}_i = (y_{i1},\ldots,y_{ir})'$ とおくと，

$$\mathbf{Cov}\,(\boldsymbol{y}_i) = \sigma_1^2\boldsymbol{I}_r + \sigma_A^2\boldsymbol{e}_r\boldsymbol{e}_r' = (\sigma_1^2+\sigma_A^2)\left\{(1-\rho)\boldsymbol{I}_r + \rho\boldsymbol{e}_r\boldsymbol{e}_r'\right\}$$

と表わされる。ここで，$\rho = \sigma_A^2/(\sigma_1^2+\sigma_A^2)$ である。また，級内相関構造は，繰り返し測定 (Repeated Measurements) を扱うモデルの 1 つとして考えることもできる。繰り返し測定では，時点間の測定値に相関を考慮するので，上記以外にも様々な相関構造を共分散行列に入れるモデルが考えられる。例えば，測定時点間の距離が近いほど相関を高くするモデルとしては

$$\boldsymbol{\Sigma} = \frac{\xi}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{r-2} & \rho^{r-1} \\ \rho & 1 & \cdots & \rho^{r-3} & \rho^{r-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{r-2} & \rho^{r-3} & \cdots & 1 & \rho \\ \rho^{r-1} & \rho^{r-2} & \cdots & \rho & 1 \end{pmatrix}$$

なる $AR(1)$ モデルの構造を想定するのが基本的である。その他測定値が 2 値データのときの解析モデルなど詳しい内容については Lindsey (1993) などを参照されたい。

家族データを解析するモデルは，平均及び共分散行列の中に母数間の等号制約が課せられる例の 1 つであろう。いま N 組の家族の測定データがあり，i 番目の家族の子供の数が k_i，母 (父) 親の測定値を x_{i1}，子供たちの測定値を $x_{i2},\ldots,x_{i,k_i+1}$ として，$\boldsymbol{x}_i = (x_{i1},x_{i2},\ldots,x_{i,k_i+1})'$ とおくとき，$\boldsymbol{x}_i$ が k_i+1 次元正規分布 $\mathcal{N}_{k_i+1}(\boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*)$ に従うモデルを考えることが自然であるかもしれない。ここで，$\boldsymbol{\mu}_i^*$, $\boldsymbol{\Sigma}_i^*$ は，$\boldsymbol{e}_{k_i} = (1,\ldots,1)' \in \boldsymbol{R}^{k_i}$ とおいて，

$$\boldsymbol{\mu}_i^* = \begin{pmatrix} \mu_m \\ \mu_s \\ \vdots \\ \mu_s \end{pmatrix}, \quad \boldsymbol{\Sigma}_i^* = \begin{pmatrix} \sigma_m^2 & \sigma_{ms}\boldsymbol{e}_{k_i}' \\ \sigma_{ms}\boldsymbol{e}_{k_i} & \sigma_s^2\left\{(1-\rho_{ss})\boldsymbol{I}_{k_i} + \rho_{ss}\boldsymbol{e}_{k_i}\boldsymbol{e}_{k_i}'\right\} \end{pmatrix}$$

で与えられる。$\rho_{ms} = \sigma_{ms}/\sigma_m\sigma_s$ は級間相関係数 (Interclass Correlation Coefficient) と呼ばれ，ρ_{ss} は級内相関係数である。ρ_{ms} は親子の間の共通な相関係数，ρ_{ss} は子供どうしの共通な相関係数，μ_s は子供たちの共通平均として，ある意味で母数に等号制約が課されているモデルととらえることができる。群間相関，群内相関の推定・検定に関する研究が，Rosner (1979), Srivastava (1984), Srivastava and Katapa (1986), Konishi and Gupta

(1989), Konishi and Khatri (1990), Konishi and Shimizu (1994), Minami and Shumizu (1998) 等によってなされてきたので，詳しくはそれらの文献を参照してほしい。

一般に，母数に等号制約が入った場合曲がったモデルが生ずるが，その推測理論の微分幾何学的構造がAmari (1982) などにより研究された。変動係数一定の問題やナイル問題はその代表的なものであり，いずれも適当な群構造を当てはめることにより，補助統計量が最大不変量として解釈されて最良共変推定量が求まり，それが MLE を改良することがわかる (Kariya (1989), Kariya *et al.* (1988), Marchand (1994))。

その他，不良率や信頼度の推定は，ある確率変数 X について $P[X > c]$ なる確率の推定であり，$0 \leq P[X > c] \leq 1$ なる条件をみたす推定方式が望まれる (藤野 (1987))。その意味から判別分析における誤判別率の推定も同様な条件がをもった問題であり，種々の推定法が提案されている (小西，本田 (1992))。また非心母数 λ の推定は，$\lambda > 0$ なる条件が課せられる問題である。$\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}_p)$, $S \sim \sigma^2 \chi^2_n$ とする線形モデルの標準形を考えると，非心母数は $\lambda = ||\boldsymbol{\theta}||^2/\sigma^2$ であり，λ の不偏推定量は

$$\hat{\lambda}^{UB} = (n-2)||\boldsymbol{X}||^2/S - p$$

となり，$\hat{\lambda}^{UB} < 0$ なる値を取りえてしまうことがわかる。$\hat{\lambda} > 0$ なる条件をみたし，かつ $\hat{\lambda}^{UB}$ を平均 2 乗誤差の意味で改良する推定量として，Kubokawa *et al.* (1993) は，

$$\hat{\lambda}^{TR} = \max\left(\hat{\lambda}^{UB}, \frac{2(n-2)}{p+2}\frac{||\boldsymbol{X}||^2}{S}\right)$$

なる打ち切り推定量を提案した。

第 2 部　変量効果と縮小推定

4 節の平均の等号制約 H_0 は，極めて強い制約であり，平均 μ が全体平均で推定され個々の標本の特徴は反映されない。そこで，もう少し弱い制約として，μ_i を $\mathcal{N}(\mu, \sigma^2)$ に従う確率変数としてはどうであろうか。このような μ_i を変量効果と呼ぶと，変量効果を導入することによって各標本の特徴を反映した $\overline{X}_i$ を全体平均の方向へ縮小した推定量が得られる。こうした点から，母数的等号制約は「強い制約」，変量効果は「弱い制約」ということばで表現してもよいだろう。5 節では変量モデルと縮小推定の関係について，6 節では縮小推定の役割，7 節では縮小推定と因子分析やカルマンフィルターなどとの関係，8 節では縮小推定の有効な応用例である小地域推定の問題について概説する。

5　縮小推定

5.1　変量効果と Stein 現象

まず一元配置変量モデル

$$\begin{aligned} y_{ij} &= \theta_i + \varepsilon_{ij}, \quad i = 1, \ldots, k,\ j = 1, \ldots, r && (5.1) \\ \theta_i &= \mu + \alpha_i && (5.2) \end{aligned}$$

を考えよう。ここで，α_i, ε_{ij} は互いに独立な確率変数で $\alpha_i \sim \mathcal{N}(0, \sigma^2_A)$, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_1)$ とし，μ, σ^2_A, σ^2_1 を未知母数とする。$\overline{y}_{\cdot\cdot} = (kr)^{-1}\sum_{i=1}^k \sum_{j=1}^r y_{ij}$, また S_1, S_2 を (3.2) とおくと，$\overline{y}_{\cdot\cdot}$, S_1, S_2 は互いに独立で，$\overline{y}_{\cdot\cdot} \sim \mathcal{N}(\mu, (\sigma^2_1 + r\sigma^2_A)/rk)$, $S_1/\sigma^2_1 \sim \chi^2_{n_1}$, $n_1 = k(r-1)$,

$S_2/\sigma_2^2 \sim \chi^2_{n_2}$, $n_2 = k-1$, $\sigma_2^2 = \sigma_1^2 + r\sigma_A^2$, なる分布に従い，$\sigma_1^2$ と σ_2^2 の間には $\sigma_1^2 \le \sigma_2^2$ なる不等式制約が入っている。

上記のモデルにおいて θ_i, $i = 1, \ldots, k$ を予測 (推定) する問題を考えよう。Bayes の枠組みで考えると，θ_i の Bayes 推定量は，

$$\hat{\theta}_i^B = \overline{y}_{i\cdot} - \delta(\overline{y}_{i\cdot} - \mu), \quad \delta = \sigma_1^2/\sigma_2^2 = \sigma_1^2/(\sigma_1^2 + r\sigma_A^2)$$

となる。$\sigma_1^2 \le \sigma_2^2$ なる関係から，δ は $0 < \delta \le 1$ という制約を受ける。δ が 1 に近いということは $\theta_1, \ldots, \theta_k$ がほぼ等しいことを意味し，δ が 0 に近ければそれらがばらついていることを表わすので，$\hat{\theta}_i^B$ はそのバラツキを反映した形をしている。μ を $\overline{y}_{\cdot\cdot}$ で推定し，δ を (3.2), (3.3) で与えられる S_1, S_2 の関数 $\widehat{\delta}(S_1, S_2)$ で推定すると，経験 Bayes 推定量

$$\hat{\theta}_i^{EB}(\widehat{\delta}) = \overline{y}_{i\cdot} - \widehat{\delta}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})$$

が得られる。これは, 最良線形不偏予測量 (BLUP) において分散成分の推定量を代入した，いわゆる経験最良線形不偏予測量 (EBLUP) と呼ばれるものである。変量効果が入ることによって，$\overline{y}_{i\cdot}$ を，強い等号制約によりプールされた推定量 $\overline{y}_{\cdot\cdot}$ の方向へ縮小した形をしている。$\widehat{\delta}$ を 0, 1 に近づけると $\hat{\theta}_i^{EB}(\widehat{\delta})$ はそれぞれ $\overline{y}_{i\cdot}$, $\overline{y}_{\cdot\cdot}$ に近づき，$\hat{\theta}_i^{EB}(\widehat{\delta})$ は $\overline{y}_{i\cdot}$ と $\overline{y}_{\cdot\cdot}$ の間の値をとっていることがわかる。この $\widehat{\delta}$ は $\overline{y}_{i\cdot}$ を $\overline{y}_{\cdot\cdot}$ へ縮小する量をコントロールする (縮小) 関数であり，どのように推定するかが重要な問題である。平均 2 乗誤差による $\widehat{\boldsymbol{\theta}}^{EB} = (\hat{\theta}_1^{EB}, \ldots, \hat{\theta}_k^{EB})'$ のリスクを計算すると，

$$R(\widehat{\boldsymbol{\theta}}^{EB}; \delta, \mu) = E\left[\sum_{i=1}^k \left(\hat{\theta}_i^{EB} - \theta_i\right)^2 / \sigma_1^2\right] = \frac{1}{r} E\left[\frac{(\widehat{\delta} - \delta)^2}{\sigma_1^2} S_2\right] + \frac{1}{r}(k - n_2\delta)$$

となり，分散の比 $\delta = \sigma_2^2/\sigma_1^2$ を制約 $0 < \delta \le 1$ のもとで損失関数 $(\widehat{\delta} - \delta)^2 S_2/\sigma_1^2$ に関して推定する問題に帰着される。

δ を S_1/S_2 の定数倍で推定したときの最適な定数は $c_0 = (n_2 - 2)/(n_1 + 2)$ であり，$\widehat{\delta}_0 = c_0 S_1/S_2$ を経験 Bayes 推定量に代入すると

$$\hat{\theta}_i^{JS} = \hat{\theta}_i^{EB}(\widehat{\delta}_0) = \overline{y}_{i\cdot} - \frac{n_2 - 2}{n_1 + 2} \frac{S_1}{r\sum_{i=1}^k (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2} (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})$$

となる。これは Lindley の推定量と呼ばれ，James and Stein (1961) の推定量を変形したものである。もちろん最初のモデル (5.1) において $\mu = 0$ としたならば，通常の James-Stein 推定量が導かれる。縮小推定量 $\hat{\theta}_i^{JS}$ は，θ_i の推定精度を高めるために $\overline{y}_{i\cdot}$ 以外にその周りの推定値 $\overline{y}_{1\cdot}, \ldots, \overline{y}_{i-1\cdot}, \overline{y}_{i+1\cdot}, \ldots, \overline{y}_{k\cdot}$ を利用する形をしている。$n_2 \ge 3$ 即ち $k \ge 4$ のとき，この推定量 $\hat{\theta}_i^{JS}$ が $\overline{y}_{i\cdot}$ を改良することは James and Stein (1961), Efron and Morris (1972), Stein (1973, 81) などによって母数モデルにおいて示されたが，このことは変量モデルでも成り立っている。さらに $\delta \le 1$ なる制約を考慮すると $\widehat{\delta}_0^+ = \min(\widehat{\delta}_0, 1)$ なる推定量がとられるべきであり，そのとき

$$\hat{\theta}_i^{JS+} = \hat{\theta}_i^{EBP}(\widehat{\delta}_0^+) = \overline{y}_{\cdot\cdot} + \max\left(1 - \widehat{\delta}_0, 0\right)(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}) \tag{5.3}$$

となり，いわゆる positive-part Stein 推定量 が導かれ，$\hat{\theta}_i^{JS}$ を改良することもわかる。なお Stein 問題の詳しい説明については，Berger (1985), Brandwein and Strawderman (1990), Kubokawa (1998), 竹内 (1979), 篠崎 (1991), 久保川 (1995,96) などが参照される。

縮小推定量が変量モデル (5.1) において経験 Bayes 推定量もしくは EBLU 予測量として自然に得られるということは，逆に，このような変量モデルが想定される場合に縮小推定量が利用されることが望ましいことを意味している。このことは, Robinson (1991) などによって指摘されており，また Efron and Morris (1975) で取り上げられた応用例もその指摘に沿ったものになっている。

Stein 問題は，3つ以上の平均を同時に推定すると，最尤推定量が縮小推定量によって改良されてしまうというもので，多くの理論家の興味を引いた。しかし，何でも3つ以上の平均を，たとえ全く無関係なものでも，同時に推定する枠組みで扱えば改良結果が得られるという現象自体が，不自然で人工的な印象を与え，実務家を遠ざける理由の1つになっていると思われる。例えば，多変量回帰モデルは標準形では $p\times m$ 行列 $\boldsymbol{X}$ が $\boldsymbol{X}\sim\mathcal{N}_{p,m}(\boldsymbol{\Theta},\boldsymbol{\Sigma},\boldsymbol{I}_m)$ に従うと表わされるが，$\boldsymbol{X}$ の列ベクトルを結合してもまた行ベクトルを結合しても，$m\geq 3$ か $p\geq 3$ のどちらかがみたされれば改良できてしまい，どちらを縮小するべきかについて困惑してしまう。

これに対して，縮小推定に関する上述の指摘は，推定量を縮小することやデータを結合することが許される状況や根拠を説明している。後述する穀物の収穫量の推定問題においては，いくつかの隣り合った小地域は類似性があるから，地域効果 (ブロック効果，群効果) を正規分布に従う変量とみなすことがもっともらしい。また小地域のデータ数は少ないので，全ての小地域の差異を母数として取り込むよりも変量として扱った方が目的の推定の精度を高めることができ，結果として類似した周辺地域のデータを結合することができる合理的な方法といえよう。縮小推定法は，当にこうした問題設定において有効な方法である。

また2節3節の平均及び分散の不等式制約下での推定問題との関係について述べると，(5.3) において James-Stein 推定量 $\hat{\theta}_i^{JS}$ を $\hat{\theta}_i^{JS+}$ で修正すること，すなわち，δ の推定量 $\widehat{\delta}_0=c_0S_1/S_2$ を $\min(\widehat{\delta}_0,1)$ で修正することは，σ_1^2 の推定量 $m_1^{-1}T_1$ を $\hat{\sigma}_1^{2SO}(T_1,T_2,m_1,m_2)$ で修正すること，また分散 σ_1^2 の推定量 $n_1^{-1}S_1$ を Stein 推定量 $\hat{\sigma}_1^{2ST}$ で修正することに対応しており，これらの3つの問題は，推定問題として同じ構造を持っていることがわかる。Rukhin (1992) は，この3つの問題は，正規分布の正の平均 ($\mu>0$) の推定問題と漸近的に同値であることを示した。また IERD 法を用いて一般化 Bayes 推定量などを含んだ改良型推定量のクラスを求めると，いずれの問題でもそのクラスに入る推定量が同等な性質をもっていることがわかる (Kubokawa (1999))。

5.2 母数制約と変量効果

5.1 節で説明されているように，(5.2) の θ_i が母数なら θ_i は $\overline{y}_{i\cdot}$ で推定されるが，$\theta_i=\mu+\alpha_i$ において μ を共通母数とすることによってデータのプーリングが行われて μ は $\overline{y}_{\cdot\cdot}$ で推定され，α_i を変量とすることによって $\overline{y}_{i\cdot}$ は $\overline{y}_{\cdot\cdot}$ の方向へ縮小されることになる。もしさらに「$\mu=\mu_0$ なる既知の定数」なる強い制約を課すなら，得られる縮小推定量は

$$\hat{\theta}_i^{JS}(\mu_0)=\overline{y}_{i\cdot}-\frac{k-2}{n_1+2}\frac{S}{r\sum_{i=1}^k(\overline{y}_{i\cdot}-\mu_0)^2}(\overline{y}_{i\cdot}-\mu_0)$$

となる。こうした考え方から，母数の制約と変量効果を組み合わせることによって，安定した推定値や強い確信に基づいた値への縮小がなされ，その結果精度の高い推定量を求めることができることになる。

このことを，小地域推定の問題で取り上げられている混合モデルについてみてみる。そこでは，$\boldsymbol{x}_{ij}$ を q-次元の既知の横ベクトル，$\alpha_i,\varepsilon_{ij}$ を互いに独立な確率変数で，それぞれ $\mathcal{N}(0,\sigma_A^2)$, $\mathcal{N}(0,\sigma_1^2)$ に従うものとするとき，

$$y_{ij}=\boldsymbol{x}_{ij}\boldsymbol{\beta}+\alpha_i+\varepsilon_{ij},\quad i=1,\ldots,k,\ j=1,\ldots,r_i \tag{5.4}$$

なる混合モデルが扱われ，各郡(小地域)の平均 $\theta_i = r_i^{-1}\sum_{j=1}^{r_i}\boldsymbol{x}_{ij}\boldsymbol{\beta} + \alpha_i$ の推定が行われる。このモデルにおいて，q 次元の回帰係数ベクトル $\boldsymbol{\beta}$ は共通な母数として入っており，繰り返し数 $r_1, \ldots, r_k$ が小さくてもプーリングによってより安定した推定値を求めることが可能であり，変量効果によって $\overline{y}_{i\cdot}$ をその安定した値へ縮小することができる。また k が大きく，全体がいくつかの中地域に分割できるときには，その中地域毎に共通の回帰係数を設定して中地域の中でのプーリングを行うなど，回帰係数の制約を工夫することによって柔軟な縮小推定量を構成することができる。

いままで説明してきた縮小推定量においては，縮小する率は各成分に関して一定であった。しかし，経験 Bayes による縮小の良さは，繰り返し数 r_i が不揃いのときに顕著に現われる。例えば，r_i が等しいとは限らない一元配置変量モデルを考えてみよう。これは，上のモデルで $\boldsymbol{x}_{ij}\boldsymbol{\beta} = \mu$ とした場合に対応していて，$\theta_i = \mu + \alpha_i$ の Bayes 推定量は

$$\hat{\theta}_i^B = \overline{y}_{i\cdot} - \frac{\sigma_1^2}{\sigma_1^2 + r_i\sigma_A^2}(\overline{y}_{i\cdot} - \mu)$$

で与えられる。μ を $\overline{y}_{\cdot\cdot}$ で推定し，σ_1^2, σ_A^2 を周辺分布に基づいて推定すると，経験 Bayes 推定量

$$\hat{\theta}_i^{EB} = \overline{y}_{i\cdot} - \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + r_i\hat{\sigma}_A^2}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})$$

が導かれる。この推定量において，縮小関数 $\hat{\sigma}_1^2/(\hat{\sigma}_1^2 + r_i\hat{\sigma}_A^2)$ は，r_i が小さければ大きくなり，r_i が大きければ小さくなっている。すなわち，r_i が小さければ $\overline{y}_{i\cdot}$ の推定誤差が大きいため $\overline{y}_{\cdot\cdot}$ の方へより大きく縮小され，r_i が大きければ $\overline{y}_{i\cdot}$ の推定誤差が小さいのであまり縮小されないように働いている。従って，$\hat{\theta}_i^{EB}$ は，個々の推定量のバラツキの程度に応じて縮小率を調整することによって，より精度の高い推定値を与えることを可能にしている。実際，$\hat{\theta}_i^{EB}$ が $\overline{y}_{i\cdot}$ を平均 2 乗誤差の意味で改良しているという理論研究については，Shinozaki and Chang (1993, 96) に詳しく論じられている。

5.3 条件付き推測からの解釈

5.1 節で与えられた経験 Bayes 推定量は，相関構造が入っているときの条件付き推測によっても導出されることを説明しよう。いま，第 i ブロックの j 番目のデータ y_{ij} が，$i = 1, \ldots, k$, $j = 1, \ldots, r$ にわたって与えられていて，これらに基づいて第 i ブロックにおける予測量を構成することを考える。$\boldsymbol{j}_r = (1, \ldots, 1)' \in \boldsymbol{R}^r$ とし，$\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ir})'$ とおくとき，y_{i0} を第 i ブロックの(予測されるべき)確率変数とすると，y_{ij} と y_{i0} は同じ平均と共分散構造を持つと考えて，

$$\begin{pmatrix} \boldsymbol{y}_i \\ y_{i0} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu\boldsymbol{j}_r \\ \mu \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \sigma_{22} \end{pmatrix}\right),$$

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2\boldsymbol{I}_r + \sigma_A^2\boldsymbol{j}_r\boldsymbol{j}_r' & \sigma_A^2\boldsymbol{j}_r \\ \sigma_A^2\boldsymbol{j}_r' & \sigma_1^2 + \sigma_A^2 \end{pmatrix}$$

なるモデルが考えられる。このとき，相関構造を利用して $\boldsymbol{y}_i$ を与えたときの y_{i0} の条件付き期待値を計算すると

$$\begin{aligned} E[y_{i0}|\boldsymbol{y}_i] &= \mu + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_i - \mu\boldsymbol{j}_r) \\ &= \mu + \frac{\sigma_A^2}{\sigma_1^2}\boldsymbol{j}_r'\left\{\boldsymbol{I}_r - \frac{\sigma_A^2}{\sigma_1^2 + r\sigma_A^2}\boldsymbol{j}_r\boldsymbol{j}_r'\right\}(\boldsymbol{y}_i - \mu\boldsymbol{j}_r) \\ &= \overline{y}_{i\cdot} - \frac{\sigma_1^2}{\sigma_1^2 + r\sigma_A^2}(\overline{y}_{i\cdot} - \mu) \end{aligned}$$

となり，5.1 節の Bayes 推定量が得られる。μ，$\sigma_1^2, \sigma_1^2 + r\sigma_A^2$ を $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ から推定して代入すれば経験 Bayes 推定量が求まることになる。

5.4 経験 Bayes 法と階層 Bayes 法

モデル (5.1) のすべての y_{ij}, θ_i の組を便宜上 $\boldsymbol{y}$, $\boldsymbol{\theta}$ と表わすと，(5.1), (5.2) は，それぞれ条件付き分布として $p(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_1^2)$, $p(\boldsymbol{\theta}|\mu, \sigma_A^2)$ のように表わされる。すると周辺分布は $\int p(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_1^2)p(\boldsymbol{\theta}|\mu, \sigma_A^2)d\boldsymbol{\theta} = p(\boldsymbol{y}|\mu, \sigma_1^2, \sigma_2^2)$ となるので，これに基づいて μ, σ_1^2, σ_2^2 を推定し Bayes 推定量に代入したものが，経験 Bayes 推定量となる。これに対して，μ, σ_1^2, σ_2^2 にさらに事前分布 $\pi_1(\mu)$, $\pi_2(\sigma_1^2, \sigma_2^2)$ を仮定して Bayes 推定量を求める方法を階層 Bayes 法という。

例えば，$\mu \sim \mathcal{N}(0, \sigma_\mu^2)$ を仮定してみると，2 乗損失関数 $\sum_{i=1}^k (\hat{\theta}_i - \theta)^2/\sigma_1^2$ に関する階層 Bayes 推定量は，事後分布の期待値 $E^{\sigma_1^2, \sigma_2^2|S_1, S_2}[\cdot|S_1, S_2]$ を使って，

$$\hat{\theta}_i^{HB} = \overline{y}_{i\cdot} - \frac{E^{\sigma_1^2,\sigma_2^2|S_1,S_2}\left[\sigma_2^{-2}\left\{\overline{y}_{i\cdot} - kr\sigma_\mu^2(\sigma_2^2 + kr\sigma_\mu^2)^{-1}\overline{y}_{\cdot\cdot}\right\}\middle|S_1, S_2\right]}{E^{\sigma_1^2,\sigma_2^2|S_1,S_2}\left[\sigma_1^{-2}\middle|S_1, S_2\right]}$$

と表わされるが，σ_μ^2 をデータから推定するには自由度が足りないことがわかる。σ_μ^2 に関して不変測度で積分してしまうのも手ではあるが，ここでは $\sigma_\mu^2 \to \infty$ とした場合，すなわち μ の分布として Lebesgue 測度 $d\mu$ をとってみると

$$\hat{\theta}_i^{HB} = \overline{y}_{i\cdot} - \frac{E^{\sigma_1^2,\sigma_2^2|S_1,S_2}\left[\sigma_2^{-2}\middle|S_1, S_2\right]}{E^{\sigma_1^2,\sigma_2^2|S_1,S_2}\left[\sigma_1^{-2}\middle|S_1, S_2\right]}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})$$

となって自然な形が表われる。ここで，階層事前分布 $d\mu$ に関しても期待値をとった Bayes リスクを考えると，これは発散してしまうので数値計算を行うときには注意を要する。

次に σ_1^2, σ_2^2 の事前分布として，$\eta = \sigma_1^{-2}, \delta = \sigma_1^2/\sigma_2^2$ とおくとき，$\pi_2(\eta, \delta) = \eta^{-1}d\eta\delta^{-2}d\delta$ を考えよう。このとき，

$$\begin{aligned}\phi_0(S_2/S_1) &= \frac{E^{\sigma_1^2,\sigma_2^2|S_1,S_2}\left[\sigma_2^{-2}\middle|S_1, S_2\right]}{E^{\sigma_1^2,\sigma_2^2|S_1,S_2}\left[\sigma_1^{-2}\middle|S_1, S_2\right]} = \frac{E^{\eta,\delta}\left[\eta\delta|S_1, S_2\right]}{E^{\eta,\delta}\left[\eta|S_1, S_2\right]} \\ &= \frac{\int_0^w x^{n_2/2-1}/(1+x)^{(n_1+n_2)/2+1}dx}{\int_0^w x^{n_2/2-2}/(1+x)^{(n_1+n_2)/2+1}dx}\end{aligned}$$

となる。Kubokawa (1994) は，母数モデルにおいて，$\hat{\theta}_i(\phi) = \overline{y}_{i\cdot} - \phi(S_2/S_1)(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})$ が $\hat{\theta}_i^{JS}$ を改良するための十分条件は，$\phi(w)$ が単調非減少で $\lim_{w\to\infty}\phi(w) = (n_2-2)/(n_1+2)$，かつ $\phi(w) \geq \phi_0(w)$ であることを示した。この条件の導出には，Kubokawa (1994, 95), Kubokawa and Robert (1994) などで使われた IERD 法（*The Integral-Expression-of-Risk-Difference Method*）という，リスクの差を積分表現するテクニックが用いられた。ここで得られた階層 Bayes 推定量は，これらの条件をみたしており，$\hat{\theta}_i^{JS}$ を母数モデルにおいて改良していることがわかる。推定量の優越性が母数モデルにおいて成立するときには，すべての α_i に関して期待値をとると変量モデルになるので，その優越性は変量モデルにおいても保証されることになる。

また，$\mu = 0$ のときには，この階層事前分布は，母数モデルにおいて

$$\boldsymbol{\theta} \mid \eta,\ \delta \sim \mathcal{N}_p\left(\boldsymbol{0}, \frac{1-\delta}{\delta}\eta^{-1}\boldsymbol{I}_p\right), \quad (\eta, \delta) \sim \eta^{-1}d\eta\delta^{-2}d\delta$$

なる事前分布を想定したものに一致する。これは，Lin and Tsai (1973) によって提案され, また σ^2 が既知のときには Strawderman (1971) によって許容的ミニマクスな推定量を求めるに導入されたものであり，彼らの事前分布を直接眺めると人工的な感じがするが，変量モデルの観点からは自然な事前分布の入り方であることがわかる。

6　縮小推定の役割

5節では，変量効果と縮小推定の関係について述べてきたが，この節では，もう少し広い視野に立って，縮小推定の役割や有用性についてまとめてみたい。

6.1　データのプーリングと縮小による推定量の安定化

5節では，母数制約と変量効果の導入によって縮小推定量を生み出すことができることを説明したが，推定量を安定化させるために縮小推定量がもっている機能について整理すると，次の2つ点にまとめられる。

1つは，母数制約によるデータのプーリングであり，繰り返し数 r_i が小さいことから $\overline{y}_{i\cdot}$ の推定誤差が大きくなってしまう場合, 例えば8節で扱う小地域の母集団平均の推定ではこのことが問題となるが，この場合混合モデルを考えることによって周辺地域のデータをプールし，$\overline{y}_{i\cdot}$ をそのプールした推定値に縮小して推定値の安定化を図り，推定精度を高めることができる。

もう1つは，r_i が大きければ $\overline{y}_{i\cdot}$ は推定精度が高いとしてあまり縮小しない一方，r_i が小さければプールされた推定値の方向へより大きく縮小して安定化を図っている点である。

このように，小標本でのバラツキが大きいデータの解析に縮小推定の考え方が有効であることを最初に示したのは Efron and Morris (1975) であろう。彼らは, エルサルバドルの 36 都市のトキソプラズマ症の発生率の推定に縮小推定量の適用を試み, 標準誤差の大きい推定値に対しては大幅に縮小され標準誤差の小さいものに対してはあまり縮小されていないことを図示して，縮小推定の啓蒙を行った。

6.2　多重共線性のもとでの推定量の安定化

線形回帰モデル $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_N)$ において, 回帰係数 $\boldsymbol{\beta}\in\boldsymbol{R}^p$ を推定する際，説明変数間の多重共線性の存在が，計量経済学の分野などではしばしば問題となる。これは，$\boldsymbol{X}$ の中に高い相関をもつ説明変数が存在すると $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ が不安定になってしまうという問題で，具体的には，$\boldsymbol{X}'\boldsymbol{X}$ を $p\times p$ 直交行列 $\boldsymbol{H}$ で対角化したときの固有値を $\lambda_1\geq\cdots\geq\lambda_p$ とするとき，すなわち，$\boldsymbol{HX}'\boldsymbol{XH}'=\mathrm{diag}\,(\lambda_1,\ldots,\lambda_p)$ のとき，小さい固有値がほとんど 0 に近い値になってしまい，その結果, 通常の最小 2 乗推定量

$$\widehat{\boldsymbol{\beta}}^{OLS}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

が不安定になってしまう。そこで, Hoerl and Kennard (1970) はリッジ回帰推定量

$$\widehat{\boldsymbol{\beta}}^{R}(k)=(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y},\quad k>0$$

を提案した。しかし, k の決め方には恣意性があり，また $\widehat{\boldsymbol{\beta}}^{R}(k)$ は $\widehat{\boldsymbol{\beta}}^{OLS}$ を一様には改良しないので, k をデータから適当に推定することにより，$\widehat{\boldsymbol{\beta}}^{R}(\hat{k})$ が $\widehat{\boldsymbol{\beta}}^{OLS}$ を一様に改良するように $\hat{k}$ を定めることができる。しかし, Casella (1980) により指摘されたように, $\widehat{\boldsymbol{\beta}}^{R}(\hat{k})$ は再び多重共線性の影響を受けて不安定性が生じてしまう。こうして, 両方の性質

を合わせもつ推定量を得ることができなかった。そこで，両方の性質を適当に調整する基準とその基準に沿った新たな縮小推定量の提案が望まれ，Casella (1985) は, 安定性とミニマクス性を同時にみたすための, 計画行列 $\boldsymbol{X}$ の条件を導出し, また Hill and Judge (1990) は, 縮小推定量の危険関数の不偏推定量に基づいて取り除く変数を決定する方法を提案した。

最近，Shinozaki and Chang (1993, 96) の結果を用いると上記の問題が解決されることがわかった。$\boldsymbol{H}\widehat{\boldsymbol{\beta}}^{OLS} \sim \mathcal{N}_p(\boldsymbol{H\beta}, \mathrm{diag}\,(\lambda_1^{-1},\ldots,\lambda_p^{-1})\sigma^2)$ とするとき，未知の階層母数 A に対して $\boldsymbol{H\beta} \sim \mathcal{N}_p(\boldsymbol{0}, A\boldsymbol{I}_p)$ なる事前分布を仮定すると，$\boldsymbol{H\beta}$ の Bayes 推定量は

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{B}(A) &= \boldsymbol{H}'\mathrm{diag}\left(\frac{\lambda_1 A}{\lambda_1 A+1},\ldots,\frac{\lambda_p A}{\lambda_p A+1}\right)\boldsymbol{H}\widehat{\boldsymbol{\beta}}^{OLS} \\ &= A\left(\boldsymbol{X}'\boldsymbol{X}A+\boldsymbol{I}_p\right)^{-1}\boldsymbol{X}'\boldsymbol{y}\end{aligned}$$

となり，リッジ回帰推定量が出てくる。$S=(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$, $\quad n=N-p$, $\boldsymbol{H}\widehat{\boldsymbol{\beta}}^{OLS}=(X_1,\ldots,X_p)'$ とおくとき，Shinozaki and Chang (1993) では，A を

$$\sum_{i=1}^{p}\frac{\lambda_i X_i^2}{\lambda_i A+1}=\frac{p-2}{n+2}S$$

の解 $\hat{A}$ として求め，これを Bayes 推定量 $\widehat{\boldsymbol{\beta}}^{B}(A)$ に代入した経験 Bayes 推定量 $\widehat{\boldsymbol{\beta}}^{EB}=\widehat{\boldsymbol{\beta}}^{B}(\hat{A})$ が $\widehat{\boldsymbol{\beta}}^{OLS}$ を平均 2 乗誤差の意味で改良していることを示した。従って $\widehat{\boldsymbol{\beta}}^{OLS}$ を一様に改良していてしかも安定性をもったリッジ回帰推定量が得られたことになる。

6.3 事前情報や知識の利用

事前情報や経験，知識から，母数についてのある仮説が成り立っていると予想される場合，その情報を用いてその仮説の方向へ縮小する推定量が考えられる。ここでは，線形回帰モデル $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}\sim\mathcal{N}_N(\boldsymbol{0},\sigma^2\boldsymbol{I}_N)$ において, 線形仮説

$$H_0:\boldsymbol{C\beta}=\boldsymbol{d}$$

が成立していることが期待される場合を考えよう。ここで $\boldsymbol{X}(N\times p)$, $\boldsymbol{C}(q\times p)$ はそれぞれランク $p, q\ (q<p)$ の既知の行列，$\boldsymbol{d}$ は q 次元の既知のベクトル，$\boldsymbol{\beta}$, σ^2 が未知母数とする。

$\boldsymbol{\beta}$ の経験 Bayes 推定量を求めるために，$\boldsymbol{H}_1=(\boldsymbol{CC}')^{-1/2}\boldsymbol{C}$ とし，$\boldsymbol{H}=(\boldsymbol{H}_1',\boldsymbol{H}_2')'$ を $p\times p$ 直交行列，$\boldsymbol{\theta}_1=\boldsymbol{H}_1\boldsymbol{\beta}$, $\boldsymbol{\theta}_2=\boldsymbol{H}_2\boldsymbol{\beta}$, $\boldsymbol{\theta}=(\boldsymbol{\theta}_1',\boldsymbol{\theta}_2')'=\boldsymbol{H\beta}$ とし，

$$\boldsymbol{\nu}_{10}=(\boldsymbol{CC}')^{-1/2}\boldsymbol{d}$$

とする。$\boldsymbol{\beta}$ の OLS 推定量 $\widehat{\boldsymbol{\beta}}^{OLS}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ に対して，$\widehat{\boldsymbol{\theta}}_i^{OLS}=\boldsymbol{H}_i\widehat{\boldsymbol{\beta}}^{OLS}$, $i=1,2$, とおく。$\boldsymbol{\theta}$ の事前分布として

$$\boldsymbol{\theta}\sim\mathcal{N}_p\left(\boldsymbol{\nu},\sigma_A^2(\boldsymbol{HX}'\boldsymbol{XH}')^{-1}\right)$$

を仮定する。ここで $\boldsymbol{\nu}=(\boldsymbol{\nu}_{10}',\boldsymbol{\nu}_2')'$ とし $\boldsymbol{\nu}_2$ は $(p-q)$ 次元の未知ベクトル，σ_A^2 は未知母数とする。$\boldsymbol{y}$, $\boldsymbol{\theta}$ の同時密度関数の指数の部分については

$$(\boldsymbol{y}-\boldsymbol{XH}'\boldsymbol{\theta})'(\boldsymbol{y}-\boldsymbol{XH}'\boldsymbol{\theta})/\sigma^2+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{HX}'\boldsymbol{XH}'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)/\sigma_A^2$$

$$
\begin{aligned}
= \; & \frac{\sigma^2+\sigma_A^2}{\sigma^2\sigma_A^2}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}^B\right)'\boldsymbol{H}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{H}'\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}^B\right)+\frac{1}{\sigma^2}S \\
& +\frac{1}{\sigma^2+\sigma_A^2}\left(\widehat{\boldsymbol{\theta}}_2^{REML}-\boldsymbol{\nu}_2\right)'\boldsymbol{K}_{22}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}\right) \\
& +\frac{1}{\sigma^2+\sigma_A^2}\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right)'\boldsymbol{K}_{11.2}\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right),
\end{aligned}
$$

と分解できる。ここで，$\boldsymbol{K}_{ij}=\boldsymbol{H}_i\boldsymbol{X}'\boldsymbol{X}\boldsymbol{H}_j'$, $i,j=1,2$, $\boldsymbol{K}_{11.2}=\boldsymbol{K}_{11}-\boldsymbol{K}_{12}\boldsymbol{K}_{22}^{-1}\boldsymbol{K}_{21}$, $S=\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}\right)'\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}\right)$,

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}^B &= \widehat{\boldsymbol{\theta}}^{OLS}-\frac{\sigma^2}{\sigma^2+\sigma_A^2}\left(\widehat{\boldsymbol{\theta}}^{OLS}-\boldsymbol{\nu}\right), \\
\widehat{\boldsymbol{\theta}}_2^{REML} &= \widehat{\boldsymbol{\theta}}_2^{OLS}+\boldsymbol{K}_{22}^{-1}\boldsymbol{K}_{21}\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right),
\end{aligned}
$$

である。$\widehat{\boldsymbol{\theta}}^B$ は Bayes 推定量であり，$\widehat{\boldsymbol{\theta}}_2^{REML}$ は仮説 H_0 のもとでの $\boldsymbol{\theta}_2$ の REML 推定量となっている。従って $\boldsymbol{\theta}$ の事後分布は，

$$
\boldsymbol{\theta}|\boldsymbol{y}\sim\mathcal{N}_p\left(\widehat{\boldsymbol{\theta}}^B,\frac{\sigma^2+\sigma_A^2}{\sigma^2\sigma_A^2}\boldsymbol{H}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{H}'\right)
$$

となる。また周辺分布については，$\widehat{\boldsymbol{\theta}}_2^{REML}$, S, $(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10})'\boldsymbol{K}_{11.2}(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10})$ が互いに独立で，$S/\sigma^2\sim\chi_n^2$,

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_2^{REML} &\sim \mathcal{N}_{p-q}\left(\boldsymbol{\nu}_2,(\sigma^2+\sigma_A^2)^{-1}\boldsymbol{K}_{22}\right), \\
\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right)'&\boldsymbol{K}_{11.2}\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right)/(\sigma^2+\sigma_A^2)\sim\chi_q^2
\end{aligned}
$$

となる。そこで $\boldsymbol{\nu}_2$ を $\widehat{\boldsymbol{\theta}}_2^{REML}$ で推定し，$\sigma^2/(\sigma^2+\sigma_A^2)$ を $\widehat{\delta}=\widehat{\delta}(F)$,

$$
F=\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right)'\boldsymbol{K}_{11.2}\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right)/S,
$$

なる関数で推定すると，$\boldsymbol{\theta}$ の経験 Bayes 推定量は，

$$
\widehat{\boldsymbol{\theta}}^{EB}=\begin{pmatrix}\widehat{\boldsymbol{\theta}}_1^{EB}\\ \widehat{\boldsymbol{\theta}}_2^{EB}\end{pmatrix}=\begin{pmatrix}\widehat{\boldsymbol{\theta}}_1^{OLS}-\widehat{\delta}\left(\widehat{\boldsymbol{\theta}}_1^{OLS}-\boldsymbol{\nu}_{10}\right)\\ \widehat{\boldsymbol{\theta}}_2^{OLS}-\widehat{\delta}\left(\widehat{\boldsymbol{\theta}}_2^{OLS}-\widehat{\boldsymbol{\theta}}_2^{REML}\right)\end{pmatrix}
$$

として得られる。$\boldsymbol{\theta}_1$ の推定は，$\widehat{\boldsymbol{\theta}}_1^{OLS}$ を仮説 H_0 の方向へ縮小し，$\boldsymbol{\theta}_2$ の推定は，$\widehat{\boldsymbol{\theta}}_1^{OLS}$ を仮説 H_0 が成り立つときの REML 推定量 $\widehat{\boldsymbol{\theta}}_2^{REML}$ の方向へ縮小した形をしていることがわかる。

$\boldsymbol{\theta}=\boldsymbol{H}\boldsymbol{\beta}$ なる関係から，もとの回帰係数 $\boldsymbol{\beta}$ の推定の形で書き直すと，

$$
\widehat{\boldsymbol{\beta}}^{EB}=\widehat{\boldsymbol{\beta}}^{OLS}-\widehat{\delta}\left(\widehat{\boldsymbol{\beta}}^{OLS}-\boldsymbol{H}_1'\boldsymbol{\nu}_1-\boldsymbol{H}_2'\widehat{\boldsymbol{\theta}}_2^{REML}\right)
$$

となる。ここで

$$
\boldsymbol{K}_{11.2}=(\boldsymbol{C}\boldsymbol{C}')^{1/2}\left\{\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right\}^{-1}(\boldsymbol{C}\boldsymbol{C}')^{1/2},
$$

$\boldsymbol{H}_1'\boldsymbol{H}_1 + \boldsymbol{H}_2'\boldsymbol{H}_2 = \boldsymbol{I}$, $\boldsymbol{H}_1 = (\boldsymbol{C}\boldsymbol{C}')^{-1/2}\boldsymbol{C}$ に注意して計算すると

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{H}_1'\boldsymbol{\nu}_1 - \boldsymbol{H}_2'\widehat{\boldsymbol{\beta}}_2^{REML}\\
&= \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right) - \boldsymbol{H}_2'\boldsymbol{K}_{22}^{-1}\boldsymbol{K}_{21}(\boldsymbol{C}\boldsymbol{C}')^{-1/2}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)\\
&= \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)\\
&\quad -(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{H}_1', \boldsymbol{H}_2')\begin{pmatrix} \boldsymbol{K}_{11} - \boldsymbol{K}_{11.2} \\ \boldsymbol{K}_{21} \end{pmatrix}(\boldsymbol{C}\boldsymbol{C}')^{-1/2}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)\\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\left\{\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right\}^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)
\end{aligned}$$

となる。また,

$$F = \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)'\left\{\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right\}^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)/S$$

と書き直せるので，$\boldsymbol{\beta}$ の James-Stein 推定量は,

$$\widehat{\boldsymbol{\beta}}^{JS} = \widehat{\boldsymbol{\beta}}^{OLS} - \frac{q-2}{n+2}\frac{1}{F}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\left\{\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right\}^{-1}\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{d}\right)$$

と表わされる。F は線形仮説 H_0 を検定する統計量である。

特殊な線形仮説として，$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, $\boldsymbol{\beta}_2' \in \boldsymbol{R}^q$ に対して，$H_0^* : \boldsymbol{\beta}_2 = \boldsymbol{0}$ なる仮説を想定してみる。これは，Ghosh *et al.* (1989) によって調べられており，例えば，主効果と交互作用効果を含んだモデルにおいて交互作用効果が無視できると想定される状況に対応する。この場合 $\boldsymbol{C} = (\boldsymbol{0} : \boldsymbol{I}_q)$, $\boldsymbol{d} = \boldsymbol{0}$ とおくことになるので， $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, $\boldsymbol{X}_2(N \times q)$, と分割し,

$$\begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{V}^{11} & \boldsymbol{V}^{12} \\ \boldsymbol{V}^{21} & \boldsymbol{V}^{22} \end{pmatrix}$$

とおくと，$\widehat{\boldsymbol{\beta}}^{JS}$ は,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{JS*} &= \widehat{\boldsymbol{\beta}}^{OLS} - \frac{q-2}{n+2}\frac{1}{F^*}\begin{pmatrix} \boldsymbol{V}^{12} \\ \boldsymbol{V}^{22} \end{pmatrix}\left(\boldsymbol{V}^{22}\right)^{-1}\widehat{\boldsymbol{\beta}}_2^{OLS}\\
&= \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1^{OLS} - (q-2)(n+2)^{-1}F^{*-1}(\widehat{\boldsymbol{\beta}}_1^{OLS} - \widehat{\boldsymbol{\beta}}_1^{REML}) \\ \{1 - (q-2)(n+2)^{-1}F^{*-1}\}\,\widehat{\boldsymbol{\beta}}_2^{OLS} \end{pmatrix}
\end{aligned}$$

と表わされる。ここで，$F^* = (\widehat{\boldsymbol{\beta}}_2^{OLS})'(\boldsymbol{C}^{22})^{-1}\widehat{\boldsymbol{\beta}}_2^{OLS}/S$ であり，REML 推定量 $\widehat{\boldsymbol{\beta}}_1^{REML}$ は

$$\widehat{\boldsymbol{\beta}}_1^{REML} = \widehat{\boldsymbol{\beta}}_1^{OLS} - \boldsymbol{V}^{12}\left(\boldsymbol{V}^{22}\right)^{-1}\widehat{\boldsymbol{\beta}}_2^{OLS}$$

と書かれる。上述のモデルを変量モデルの形式で記述すると

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + (\boldsymbol{X}_1, \boldsymbol{X}_2)\boldsymbol{\alpha} + \boldsymbol{\epsilon},$$

$\boldsymbol{\alpha} \sim \mathcal{N}_p(\boldsymbol{0}, \sigma_A^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$, $\boldsymbol{\epsilon} \sim \boldsymbol{N}_N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_N)$ と表わされ，変量係数モデルの形をしていることがわかる。

6.4 推定量の適応型分解

二元配置変量モデルの各平均の推定については，縮小推定量は，通常の推定量を２つの主効果と１つの交互作用効果に分解し，それぞれの信頼度によって重み付けた形をしている。具体的には，

$$y_{ijk} = \theta_{ij} + \varepsilon_{ijk}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K$$

において，平均 θ_{ij} が４つの要素 μ, α_i, β_j, γ_{ij} から構成されていて，

$$\theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

とかけ，ε_{ijk}, α_i, β_j, γ_{ij} が互いに独立に $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$, $\alpha_i \sim \mathcal{N}(0, \sigma_A^2)$, $\beta_j \sim \mathcal{N}(0, \sigma_B^2)$, $\gamma_{ij} \sim \mathcal{N}(0, \sigma_{AB}^2)$ に従うと仮定する。ここで μ, σ^2, σ_A^2, σ_B^2, σ_{AB}^2 は未知母数である。このとき θ_{ij} の Bayes 推定量は，

$$\begin{aligned} \hat{\theta}_{ij}^{B} &= \bar{y}_{\cdots} + \left(1 - \frac{\sigma^2}{\tau_A^2}\right)(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + \left(1 - \frac{\sigma^2}{\tau_B^2}\right)(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\ &\quad + \left(1 - \frac{\sigma^2}{\tau_{AB}^2}\right)\left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right) + \frac{\sigma^2}{\tau_A^2 + \tau_B^2 - \tau_{AB}^2}(\mu - \bar{y}_{\cdots}) \end{aligned} \tag{6.1}$$

となる。ただし，$\tau_A^2 = \sigma^2 + K\sigma_{AB}^2 + JK\sigma_A^2$, $\tau_B^2 = \sigma^2 + K\sigma_{AB}^2 + IK\sigma_B^2$, $\tau_{AB}^2 = \sigma^2 + K\sigma_{AB}^2$ であり，それらの間には，

$$\sigma^2 \le \tau_{AB}^2 \le \tau_A^2, \quad \sigma^2 \le \tau_{AB}^2 \le \tau_B^2$$

なる順序関係がある。いま $S = \sum_{i,j,k}(y_{ijk} - \bar{y}_{ij\cdot})^2$, $T_A = JK\sum_i(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$, $T_B = IK\sum_j(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$, $T_{AB} = K\sum_{i,j}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2$ とおくと，S, T_A, T_B, T_{AB} は互いに独立に

$$\begin{aligned} S/\sigma^2 &\sim \chi_n^2, & n &= IJ(K-1), \\ T_A/\tau_A^2 &\sim \chi_{n_A}^2, & n_A &= I - 1, \\ T_B/\tau_B^2 &\sim \chi_{n_B}^2, & n_B &= J - 1, \\ T_{AB}/\tau_{AB}^2 &\sim \chi_{n_{AB}}^2, & n_{AB} &= (I-1)(J-1) \end{aligned}$$

に従う (広津 (1976))。μ を $\bar{y}_{\cdots}$ で推定し，σ^2, τ_A^2, τ_B^2, τ_{AB}^2 を S, T_A, T_B, T_{AB} に基づいて推定すると，

$$\begin{aligned} \hat{\theta}_{ij}^{EB} &= \bar{y}_{\cdots} + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_A^2}\right)(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_B^2}\right)(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\ &\quad + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_{AB}^2}\right)\left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right) \end{aligned} \tag{6.2}$$

なる経験 Bayes 推定量が得られる。従って，問題は順序制約のある分散の比 σ^2/τ_A^2, σ^2/τ_B^2, σ^2/τ_{AB}^2 の推定に帰着される。１つの自然な推定量は，James-Stein 型推定量

$$\begin{aligned} \hat{\theta}_{ij}^{S+} &= \bar{y}_{\cdots} + \left(1 - \frac{n_A - 2}{n+2}\frac{S}{T_A}\right)^+ (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + \left(1 - \frac{n_B - 2}{n+2}\frac{S}{T_B}\right)^+ (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\ &\quad + \left(1 - \frac{n_{AB} - 2}{n+2}\frac{S}{T_{AB}}\right)^+ \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right) \end{aligned} \tag{6.3}$$

であるが，これは竹内 (1979) によって提案されたものである。ここで，a^+ は $\max(a, 0)$ を意味している。θ_{ij} を母数と考えたときの不偏推定量は $\overline{y}_{ij\cdot}$ であり，Sun (1996) は $\hat{\theta}_{ij}^{S+}$ が $\overline{y}_{ij\cdot}$ を改良することを母数モデルにおいて証明した。もちろん変量モデルでもこの優越性は成り立つ。

(6.2) に対応して $\overline{y}_{ij\cdot}$ は

$$\overline{y}_{ij\cdot} = \overline{y}_{\cdots} + (\overline{y}_{i\cdot\cdot} - \overline{y}_{\cdots}) + (\overline{y}_{\cdot j\cdot} - \overline{y}_{\cdots}) + \left(\overline{y}_{ij\cdot} - \overline{y}_{i\cdot\cdot} - \overline{y}_{\cdot j\cdot} + \overline{y}_{\cdots}\right)$$

と分解され，この式の右辺の第 1 項，第 2 項，第 3 項がそれぞれ 3 つの効果 α_i, β_j, γ_{ij} に対応する項である。しかし，繰り返し数 K が小さいときには，このままでは γ_{ij} の推定に当たる第 3 項は不安定になってしまう。そこで，各効果の仮説についての信頼性に応じて重み付けすることによって推定量を安定化することが考えられるが，(6.2), (6.3) は，そのように重み付けした結合推定量の形をしている。

上記の議論においては，α_i, β_j, γ_{ij} をすべて変量にしてきたが，その 1 部が母数である混合モデルについても対応する Bayes 及び経験 Bayes 推定量が得られる。例えば，すべての γ_{ij} について $\gamma_{ij} = 0$ が成り立つときには, $\sigma_{AB}^2 \to 0$ として

$$\hat{\theta}_{ij}^{EB} = \overline{y}_{\cdots} + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_A^2}\right)(\overline{y}_{i\cdot\cdot} - \overline{y}_{\cdots}) + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_B^2}\right)(\overline{y}_{\cdot j\cdot} - \overline{y}_{\cdots})$$

なる経験 Bayes 推定量が得られ，またすべての α_i を変量ではなく母数として扱うときには，$\sigma_A^2 \to \infty$ とすればよく，そのとき経験 Bayes 推定量は,

$$\hat{\theta}_{ij}^{EB} = \overline{y}_{i\cdot\cdot} + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_B^2}\right)(\overline{y}_{\cdot j\cdot} - \overline{y}_{\cdots}) + \left(1 - \frac{\hat{\sigma}^2}{\hat{\tau}_{AB}^2}\right)\left(\overline{y}_{ij\cdot} - \overline{y}_{i\cdot\cdot} - \overline{y}_{\cdot j\cdot} + \overline{y}_{\cdots}\right)$$

と表わされる。

7 モデルの拡張と関連する問題

7.1 多変量混合モデルでの予測

5 節の結果は，多変量混合線形モデルにおける予測問題に拡張できる。

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta} \boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, r,$$

において，$\boldsymbol{y}_{ij}$ は p 変量観測ベクトル，$\boldsymbol{\beta}$ は共通な未知の $p \times q$ 回帰係数行列，$\boldsymbol{b}_{ij}$ は q 次元の共変量ベクトル，$\boldsymbol{\alpha}_i$ をブロック効果で $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$ に従う変量とし，観測誤差 $\boldsymbol{\epsilon}_{ij}$ は $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_1)$ に従うものとする。これは，(5.4) を多変量へ拡張したモデルである。

いま $\overline{\boldsymbol{y}}_{i\cdot} = r^{-1} \sum_{j=1}^r \boldsymbol{y}_{ij}$, $\overline{\boldsymbol{b}}_{i\cdot} = r^{-1} \sum_{j=1}^r \boldsymbol{b}_{ij}$ に対して, $\boldsymbol{u}_{ij} = \boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i\cdot}$, $\boldsymbol{c}_{ij} = \boldsymbol{b}_{ij} - \overline{\boldsymbol{b}}_{i\cdot}$ とし，$\boldsymbol{U} = (\boldsymbol{u}_{11}, \ldots, \boldsymbol{u}_{1r}; \cdots; \boldsymbol{u}_{k1}, \ldots, \boldsymbol{u}_{kr})$, $\boldsymbol{C} = (\boldsymbol{c}_{11}, \ldots, \boldsymbol{c}_{1r}; \cdots; \boldsymbol{c}_{k1}, \ldots, \boldsymbol{c}_{kr})$ とおく。このとき,

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_1 &= \boldsymbol{U}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^-, \\ \boldsymbol{S}_1 &= (\boldsymbol{U} - \widehat{\boldsymbol{\beta}}_1 \boldsymbol{C})(\boldsymbol{U} - \widehat{\boldsymbol{\beta}}_1 \boldsymbol{C})' \end{aligned}$$

とする。ここで，$\boldsymbol{A}^-$ は行列 $\boldsymbol{A}$ の一般化逆行列を表わす。また $\overline{\boldsymbol{Y}} = (\overline{\boldsymbol{y}}_{1\cdot}, \ldots, \overline{\boldsymbol{y}}_{k\cdot})$, $\overline{\boldsymbol{B}} = (\overline{\boldsymbol{b}}_{1\cdot}, \ldots, \overline{\boldsymbol{b}}_{k\cdot})$ とおいて

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_2 &= \overline{\boldsymbol{Y}}\,\overline{\boldsymbol{B}}'(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^-, \\ \boldsymbol{S}_2 &= r(\overline{\boldsymbol{Y}} - \widehat{\boldsymbol{\beta}}_2 \overline{\boldsymbol{B}})(\overline{\boldsymbol{Y}} - \widehat{\boldsymbol{\beta}}_2 \overline{\boldsymbol{B}})' \end{aligned}$$

とする。rank $(\boldsymbol{C}\boldsymbol{C}') = q_1 \le q$, rank $(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}') = q_2 \le q$ と仮定すると，$\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$, $\boldsymbol{S}$, $\boldsymbol{W}$ は互いに独立に

$$\begin{aligned}
\boldsymbol{S}_1 &\sim \mathcal{W}_p(\boldsymbol{\Sigma}_1, n_1), \quad n_1 = k(r-1) - q_1, \\
\boldsymbol{S}_2 &\sim \mathcal{W}_p(\boldsymbol{\Sigma}_2, n_2), \quad n_2 = k - q_2, \\
\widehat{\boldsymbol{\beta}}_1 &\sim \mathcal{N}_{p\times q}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_1, (\boldsymbol{C}\boldsymbol{C}')^-), \\
\widehat{\boldsymbol{\beta}}_2 &\sim \mathcal{N}_{p\times q}(\boldsymbol{\beta}, r^{-1}\boldsymbol{\Sigma}_2, (\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^-), \quad \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A
\end{aligned}$$

に従う。2つの共分散行列 $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ の間に

$$\boldsymbol{\Sigma}_1 \le \boldsymbol{\Sigma}_2$$

なる不等式制約が入っていることになる。この不等式は $\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1$ が非負定値であることを意味する。

各群の平均 $\boldsymbol{\theta}_i = \boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot} + \boldsymbol{\alpha}_i$ の予測問題を考えよう。これを Bayes の枠組みで考えると，$\boldsymbol{\theta}_i$ の Bayes 推定量は

$$\widehat{\boldsymbol{\theta}}_i^B = \overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\Delta}\left(\overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\beta}\overline{\boldsymbol{b}}_i\right), \quad \boldsymbol{\Delta} = \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1} \tag{7.1}$$

となるので，$\boldsymbol{\beta}$ を $\widehat{\boldsymbol{\beta}}_2$ で推定し，$\boldsymbol{\Delta}$ を $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ の関数 $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{S}_1, \boldsymbol{S}_2)$ で推定すると，

$$\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\Delta}}) = \overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\Delta}}\left(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_i\right) \tag{7.2}$$

なる経験 Bayes 推定量が得られる。この推定量の平均2乗誤差を計算すると

$$\begin{aligned}
&\sum_{i=1}^{k} E[(\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\Delta}}) - \boldsymbol{\theta}_i)'\boldsymbol{\Sigma}_1^{-1}(\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\Delta}}) - \boldsymbol{\theta}_i)] \\
&\quad = r^{-1}E_\omega\left[\mathrm{tr}\left\{(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})'\boldsymbol{\Sigma}_1^{-1}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\boldsymbol{S}_2\right\}\right] + r^{-1}\left(pk - n_2\mathrm{tr}\,\boldsymbol{\Delta}\right)
\end{aligned}$$

となるので，共分散行列の比 $\boldsymbol{\Delta}$ を $\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1^{1/2} \le \boldsymbol{I}_p$ なる不等式制約のもとで推定する問題に帰着される。Kubokawa and Srivastava (1999) は，通常の推定量 $\widehat{\boldsymbol{\Delta}}_0 = (n_1 + p + 1)^{-1}(n_2 - p - 1)\boldsymbol{S}_1\boldsymbol{S}_2^{-1}$ を改良する推定量として Bartlett 分解に基づいた James-Stein 型推定量，Stein 型直交不変推定量，Efron-Morris 型推定量を求めた。例えば，Efron-Morris 型推定量は，

$$\widehat{\boldsymbol{\Delta}}^{EM} = \frac{n_2 - p - 1}{n_1 + p + 1}\boldsymbol{S}_1\boldsymbol{S}_2^{-1} + \frac{(p-1)(p+2)(n_1+n_2)}{(n_1+p+1)(n_1-p+3)}\frac{1}{\mathrm{tr}\,\boldsymbol{S}_1\boldsymbol{S}_2^{-1}}\boldsymbol{I}_p$$

となる。この推定量を代入した $\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\Delta}}^{EM})$ の優越性は，変量モデルでもまた母数モデルでも示すことができ，母数モデルでの結果は Konno (1992) に一致する。(7.2) で与えられた $\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\Delta}})$ では，$\boldsymbol{\beta}$ の推定量として $\widehat{\boldsymbol{\beta}}_2$ のみが使われているが，$\widehat{\boldsymbol{\beta}}_1$ の平均も $\boldsymbol{\beta}$ であり，共通な $\boldsymbol{\beta}$ を2つの推定量を結合して推定し $\widehat{\boldsymbol{\theta}}_i^B$ に代入することが合理的なように思える。実際，8節で扱われる問題では，そのような結合推定量，もっと一般的に一般化最小2乗推定量が用いられる。しかし漸近的近似を行うか数値実験を行わなければ，こうした推定量の性質を議論することはできない。

7.2 因子分析モデル

多変量解析の伝統的な分析方法である因子分析と縮小推定の関係を議論することは興味深い。p 変量の観測ベクトル $\boldsymbol{y}_i, i=1,\ldots,N,$ に k 個の潜在変数をもった因子分析モデル

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{f}_i + \boldsymbol{\epsilon}_i, \quad i=1,\ldots,N$$

を当てはめるとする。ここで，$\boldsymbol{f}_i$, $\boldsymbol{\epsilon}_i$ は独立な確率変数で，$\boldsymbol{f}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{I}_k)$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Psi})$ とし，$\boldsymbol{\mu} \in \boldsymbol{R}^p$, $p\times k$ 行列 $\boldsymbol{A}$, 対角行列 $\boldsymbol{\Psi} = \mathrm{diag}\,(\psi_1,\ldots,\psi_p)$ を未知母数とする。$\boldsymbol{f}_i$ は因子得点，$\boldsymbol{A}$ は因子負荷量と呼ばれる。$\boldsymbol{y}_i$ の共分散行列を $\boldsymbol{\Sigma}$ とすると $\mathbf{Cov}\,(\boldsymbol{y}_i) = \boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'+\boldsymbol{\Psi}$ なる関係が成り立つ。

このモデルは，変量 $\boldsymbol{f}_i$ をもった変量モデルで，また $\boldsymbol{f}_i$ に事前分布が想定されている Bayes モデルとしても解釈される。いま i 番目の観測変量の条件付き期待値 $\boldsymbol{\theta}_i = E[\boldsymbol{y}_i|\boldsymbol{f}_i] = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{f}_i$ を推定することを考えると，$\boldsymbol{\theta}_i$ の Bayes 推定量は

$$\widehat{\boldsymbol{\theta}}_i^B = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}) = \boldsymbol{y}_i - \boldsymbol{\Delta}(\boldsymbol{y}_i - \boldsymbol{\mu}), \quad \boldsymbol{\Delta} = \boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}$$

となる。$\boldsymbol{\mu}$ は $\overline{\boldsymbol{y}} = N^{-1}\sum_{i=1}^N \boldsymbol{y}_i$ で推定され，$\boldsymbol{\Psi}$, $\boldsymbol{\Sigma}$ は $\boldsymbol{S} = \sum_{i=1}^N (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})'$ に基づいて推定される。$\boldsymbol{\Psi}$，$\boldsymbol{A}$ の MLE は

$$q(\boldsymbol{\Psi}, \boldsymbol{A}) = \mathrm{tr}\,(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) - \log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}|$$

を最小にする解として与えられるが，この解の性質等については，丘本 (1986), 竹内 (1986), 柳井他 (1990) 等で論じられている。Akaike (1987) は，事前分布を組み入れて

$$q^*(\boldsymbol{\Psi}, \boldsymbol{A}) = \mathrm{tr}\,(\boldsymbol{S}(\boldsymbol{A}\boldsymbol{A}'+\boldsymbol{\Psi})^{-1}) - \log|\boldsymbol{S}(\boldsymbol{A}\boldsymbol{A}'+\boldsymbol{\Psi})^{-1}| \;+\; \delta\,\mathrm{tr}\,\boldsymbol{A}'\boldsymbol{\Psi}^{-1}\boldsymbol{A}$$

を最小化する解を求めることを提案し，不適解の処理を検討した。右辺の第 3 項は, ψ_i のどれかが 0 に近づくにつれて大きくなり, $\widehat{\psi}_i = 0$ となる解が生じないようなペナルティーとしてはたらいていることがわかる。

いずれにしても，何らかの方法で $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma}$ が推定されると，経験 Bayes 推定量

$$\widehat{\boldsymbol{\theta}}_i^{EB} = \boldsymbol{y}_i - \widehat{\boldsymbol{\Delta}}\,(\boldsymbol{y}_i - \overline{\boldsymbol{y}}), \quad \widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Sigma}}^{-1} \tag{7.3}$$

が得られる。$\widehat{\boldsymbol{\theta}}_i^{EB}$ は，(7.2) と同様な形をしており，$\boldsymbol{y}_i$ の条件付き期待値は，$\boldsymbol{y}_i$ を標本平均 $\overline{\boldsymbol{y}}$ の方向へ縮小した値で推定されることがわかる。このことは，因子分析モデルが縮小推定を生む構造を有していることを意味していて興味深い。$\widehat{\boldsymbol{\theta}}_i^{EB}$ と $\boldsymbol{\theta}_i$ の間の平均 2 乗誤差を計算すると

$$E\left[\sum_{i=1}^N (\widehat{\boldsymbol{\theta}}_i^{EB} - \boldsymbol{\theta}_i)'\boldsymbol{\Psi}^{-1}(\widehat{\boldsymbol{\theta}}_i^{EB} - \boldsymbol{\theta}_i)\right] = E\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})'\boldsymbol{\Psi}^{-1}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\boldsymbol{S}\right] + \mathrm{tr}\,\boldsymbol{\Delta}$$

となり，この量を小さくするように，$\boldsymbol{\Delta} = \boldsymbol{\Psi}(\boldsymbol{\Psi} + \boldsymbol{A}\boldsymbol{A}')^{-1}$ を推定することになるが，このような議論はほとんどなされていない。

因子分析の本来の目的からは，因子得点 $\boldsymbol{f}_i$ の推定 (予測) の方が関心があるわけで，この Bayes 推定量は

$$\hat{\boldsymbol{f}}_i^B = \boldsymbol{A}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}),$$

経験 Bayes 推定量は

$$\hat{\boldsymbol{f}}_i^{EB} = \widehat{\boldsymbol{A}}'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})$$

となるが，この理論的性質の議論もあまりなされていないように思う。

7.3 カルマンフィルター

Bayes 統計の現実的な使用の1つにカルマンフィルターがある。これは，$t=1,2,\ldots$ に対して

$$
\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{H}\boldsymbol{\theta}_t + \delta_t, \quad \boldsymbol{\delta}_t \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{R}) \\
\boldsymbol{\theta}_t &= \boldsymbol{F}\boldsymbol{\theta}_{t-1} + \boldsymbol{G}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{Q})
\end{aligned}
$$

なる状態空間表現で与えられるシステムにおいて，各 t 時点で出力される p 次元データ $\boldsymbol{y}_t$ を順次組み入れて，$\boldsymbol{\theta}_t$ の予測を効率よく update させていく方法で幅広く応用されてきた。詳しくは，Kalman (1960), Zellner (1971, 83), West and Harrison (1997), Leonard and Hsu (1999), 尾崎他 (1998) などを参照されたい。

いま $t-1$ 時点で，$\boldsymbol{Y}_{t-1} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})$ を与えたときの $\boldsymbol{\theta}_{t-1}$ の事後分布が

$$\boldsymbol{\theta}_{t-1}|\boldsymbol{Y}_{t-1} \sim \mathcal{N}_p(\boldsymbol{\theta}^*_{t-1},\ \boldsymbol{D}_{t-1})$$

で与えられるものとする。このとき次の t 時点で $\boldsymbol{\theta}_t$ の推定値 $\boldsymbol{\theta}^*_t$ が $\boldsymbol{\theta}^*_{t-1}$ からどのように更新されるかをみてみよう。$\boldsymbol{\theta}_t = \boldsymbol{F}\boldsymbol{\theta}_{t-1} + \boldsymbol{G}\boldsymbol{\epsilon}_t$ より $\boldsymbol{\theta}_t$ の事後分布は,

$$\boldsymbol{\theta}_t|\boldsymbol{Y}_{t-1} \sim \mathcal{N}_p(\boldsymbol{F}\boldsymbol{\theta}^*_{t-1},\ \boldsymbol{F}\boldsymbol{D}_{t-1}\boldsymbol{F}' + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}')$$

となる。これは，1期先の状態空間の予測分布を表わしており，その平均と共分散は

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_{t|t-1} &= \boldsymbol{F}\boldsymbol{\theta}^*_{t-1} \\
\boldsymbol{D}_{t|t-1} &= \boldsymbol{F}\boldsymbol{D}_{t-1}\boldsymbol{F}' + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}'
\end{aligned}
$$

と計算される。一方 $\boldsymbol{y}_t|\boldsymbol{\theta}_t \sim \mathcal{N}_p(\boldsymbol{H}\boldsymbol{\theta}_t, \boldsymbol{R})$ より，$\boldsymbol{Y}_t = (\boldsymbol{Y}_{t-1}, \boldsymbol{y}_t)$ を与えたときの $\boldsymbol{\theta}_t$ の事後分布は,

$$
\begin{aligned}
\pi(\boldsymbol{\theta}_t|\boldsymbol{Y}_t) &\propto \exp\left\{(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_{t|t-1})'\boldsymbol{D}^{-1}_{t|t-1}(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}_{t|t-1}) + (\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{\theta}_t)'\boldsymbol{R}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{\theta}_t)\right\} \\
&\propto \exp\left\{(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}^*_t)'\boldsymbol{D}^{-1}_t(\boldsymbol{\theta}_t - \widehat{\boldsymbol{\theta}}^*_t)\right\}
\end{aligned}
$$

より $\boldsymbol{\theta}_t|\boldsymbol{Y}_t \sim \mathcal{N}_p(\boldsymbol{\theta}^*_t, \boldsymbol{D}_t)$ となる。ただし,

$$
\begin{aligned}
\boldsymbol{\theta}^*_t &= \widehat{\boldsymbol{\theta}}_{t|t-1} + \left(\boldsymbol{D}^{-1}_{t|t-1} + \boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\boldsymbol{R}^{-1}\left(\boldsymbol{y}_t - \boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}\right), \\
\boldsymbol{D}_t &= \left(\boldsymbol{D}^{-1}_{t|t-1} + \boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{H}\right)^{-1}
\end{aligned}
$$

である。ここでカルマンゲイン $\boldsymbol{K}_t$ を

$$\boldsymbol{K}_t = \left(\boldsymbol{D}^{-1}_{t|t-1} + \boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\boldsymbol{R}^{-1}$$

で定義すると，行列演算によって

$$\boldsymbol{K}_t = \boldsymbol{D}_{t|t-1}\boldsymbol{H}'\left(\boldsymbol{H}\boldsymbol{D}_{t|t-1}\boldsymbol{H}' + \boldsymbol{R}\right)^{-1}$$

と変形でき，$\boldsymbol{\theta}^*_t$, D_t は,

$$
\begin{aligned}
\boldsymbol{\theta}^*_t &= \widehat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{K}_t\left(\boldsymbol{y}_t - \boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}\right), \\
\boldsymbol{D}_t &= (\boldsymbol{I}_p - \boldsymbol{K}_t\boldsymbol{H})\,\boldsymbol{D}_{t|t-1}
\end{aligned}
$$

と表現できる。$\boldsymbol{\theta}_t^*$ は，$\boldsymbol{y}_t$ を $\boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}$ の方向へ縮小した形をしている。

縮小推定との関係をもう少し明確にするために，$E[\boldsymbol{y}_t|\boldsymbol{\theta}_t] = \boldsymbol{H}\boldsymbol{\theta}_t$ の安定した推定量を求めることを考えてみる。t 時点での観測値 $\boldsymbol{y}_t$ は，$\mathcal{N}(\boldsymbol{\theta}_t, \boldsymbol{R})$ に従って変動しているので，この値を，過去の観測値 $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ を使って修正することが合理的であろう。$\boldsymbol{\theta}_t^*$ を用いてより安定した推定量を求めると，

$$\begin{aligned}\hat{\boldsymbol{y}}_t &= \boldsymbol{H}\boldsymbol{\theta}_t^* \\ &= \boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1} + \boldsymbol{H}\boldsymbol{D}_{t|t-1}\boldsymbol{H}'\left(\boldsymbol{H}\boldsymbol{D}_{t|t-1}\boldsymbol{H}' + \boldsymbol{R}\right)^{-1}\left(\boldsymbol{y}_t - \boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}\right) \\ &= \boldsymbol{y}_t - \boldsymbol{R}\left(\boldsymbol{H}\boldsymbol{D}_{t|t-1}\boldsymbol{H}' + \boldsymbol{R}\right)^{-1}\left(\boldsymbol{y}_t - \boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}\right) \end{aligned} \tag{7.4}$$

となる。また，システムの共分散行列 $\boldsymbol{R}, \boldsymbol{Q}$ が未知のときには，適当な事前分布を入れて推定値を計算することが可能である (West and Harrison (1997), Leonard and Hsu (1999))。こうして得られた Bayes 及び経験 Bayes 推定量は，(7.1), (7.2) と同様な形をしており，t 時点でのデータ $\boldsymbol{y}_t$ を過去のデータに基づいた推定値 $\boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}$ の方向へ縮小していることがわかる。

時系列モデルを状態空間表現を用いて書くことができることは，多くの時系列の教科書で説明されている (Harvey (1989), Brockwell and Davis (1991), West and Harrison (1997), 尾崎他 (1998))。例えば，$AR(p)$ モデル

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{7.5}$$

については，

$$\begin{aligned}\boldsymbol{\theta}_t &= \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad \boldsymbol{H} = (1, 0, \ldots, 0) \\ \boldsymbol{F} &= \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & 1 & 0 \end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\end{aligned}$$

とおくと，(7.5) は，

$$\begin{aligned} y_t &= \boldsymbol{H}\boldsymbol{\theta}_t \\ \theta_t &= \boldsymbol{F}\boldsymbol{\theta}_{t-1} + \boldsymbol{G}\boldsymbol{\epsilon}_t \end{aligned}$$

のように状態空間表現することができ，通常の状態空間表現において $\boldsymbol{R} = \boldsymbol{0}$ とした特殊な場合になっている。このときカルマンゲインは，

$$\boldsymbol{K}_t = \frac{1}{\boldsymbol{H}\boldsymbol{D}_{t|t-1}\boldsymbol{H}'}\boldsymbol{D}_{t|t-1}\boldsymbol{H}'$$

となるので，予測量 $\boldsymbol{\theta}_t^*$ は，

$$\boldsymbol{\theta}_t^* = \widehat{\boldsymbol{\theta}}_{t|t-1} + \frac{1}{\boldsymbol{H}\boldsymbol{D}_{t|t-1}\boldsymbol{H}'}\boldsymbol{D}_{t|t-1}\boldsymbol{H}'\left(y_t - \boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}\right)$$

と書ける。予測量は，y_t を過去のデータによる安定した推定値 $\boldsymbol{H}\widehat{\boldsymbol{\theta}}_{t|t-1}$ の方向へ縮小した形をしており，この意味において時系列モデルも縮小推定の概念が埋め込まれたモデルであるということもできるかもしれない。

7.4 データ平滑化法としての縮小推定

点 $t_1,\ldots,t_n$ $(0 \le t_1 < \cdots < t_n \le 1)$ において, 観測値 $y_1,\ldots,y_n$ が得られているとき, これらのデータに対して回帰モデル

$$\begin{aligned} y_i &= f(t_i) + e_i, \quad i = 1,\ldots,n, \\ \boldsymbol{e} &= (e_1,\ldots,e_n)' \sim \mathcal{N}_n(0, \sigma^2 \boldsymbol{I}_n) \end{aligned}$$

を当てはめることを考える。但し, f は 1 次導関数 $f^{(1)}$ が絶対連続で, 2 次導関数 $f^{(2)} \in L^2[0,1]$ なる関数とする。通常は, 平滑化パラメータ $\lambda > 0$ に対して,

$$\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int_0^1 \{f^{(2)}(t)\}^2 dt \tag{7.6}$$

を最小にする関数 f を求める。この解は3次の自然スプライン関数であり, それは, 3次B-スプライン関数 $B_1(t),\ldots,B_n(t)$ の一次結合

$$f(t) = \sum_{i=1}^{n} \gamma_j B_j(t) \tag{7.7}$$

で一意的に表わすことができる。$\mu_i = f(t_i)$, $B_{ij} = B_j(t_i)$ に対して, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)'$, $\boldsymbol{B} = (B_{ij})$ とおき, また$\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_n)'$ とおくと,

$$\boldsymbol{\mu} = \boldsymbol{B}\boldsymbol{\gamma}$$

と表わされるので, $\boldsymbol{\mu}$ を推定すれば, $\boldsymbol{\gamma}$ が求められ (7.7) で与えられるスプライン関数が得られることになる。$\Omega_{ij} = \int_0^1 B_i^{(2)}(t) B_j^{(2)}(t)dt$, $\boldsymbol{\Omega} = (\Omega_{ij})$ に対して$\boldsymbol{D} = \boldsymbol{B}'^{-1}\boldsymbol{\Omega}\boldsymbol{B}^{-1}$ とおき, $\boldsymbol{y} = (y_1,\ldots,y_n)'$ とすると, (7.6) は

$$||\boldsymbol{y} - \boldsymbol{\mu}||^2 + \lambda \boldsymbol{\mu}'\boldsymbol{D}\boldsymbol{\mu}$$

と書き直すことができる。$\lambda > 0$ を与えたときの解は

$$\boldsymbol{\mu}(\lambda) = \boldsymbol{M}(\lambda)\boldsymbol{y}, \quad \boldsymbol{M}(\lambda) = (\boldsymbol{I}_n + \lambda \boldsymbol{D})^{-1}$$

で与えられることが知られており, 従って, 一つの平滑化スプライン関数 $f_\lambda(t)$ が得られる。

パラメータ λ は平滑化の程度を表わしており, これをどのように決定するかが重要な問題である。これには, 2つの方法が知られており, 1 つは, Wahba (1983) らの提唱する一般化交差検証法 (Generalized Cross Validation) で,

$$||(\boldsymbol{I}_n - \boldsymbol{M}(\lambda))\boldsymbol{y}||^2 / \{\mathrm{tr}\,(\boldsymbol{I}_n - \boldsymbol{M}(\lambda))\}^2$$

を最小にする $\lambda = \hat{\lambda}^{GCV}$ を選ぶというものである。もう 1 つは, Wahba (1985) らによって提案された経験 Bayes の方法である。Yanagimoto and Yanagimoto (1987) は, データが等間隔に得られているとき $(t_i = i)$, $\boldsymbol{\mu} \sim \mathcal{N}_{n-2}(\boldsymbol{0}, (\sigma^2/\lambda)\boldsymbol{D}^-)$ なる事前分布を想定した。但し, $\boldsymbol{D}$ は $\boldsymbol{\mu}'\boldsymbol{D}\boldsymbol{\mu} = \sum(\mu_i - 2\mu_{i+1} + \mu_{i+2})^2$ なる $n \times n$ 対称行列である。このとき, Bayes 推定量は $\widehat{\boldsymbol{\mu}}^B(\lambda) = \boldsymbol{M}(\lambda)\boldsymbol{y}$ で与えられ, λ の推定量 $\hat{\lambda}^{EB}$ は $\boldsymbol{y}$ の周辺尤度に σ^2 の推定量を代入したもの

$$U(\lambda) = (n-2)\log\{\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{M}(\lambda))\boldsymbol{y}\} - (n-2)\log\lambda + \log|\boldsymbol{I}_n + \lambda\boldsymbol{D}|$$

を最小にする λ を求めることで与えられる。

8 小地域推定

5節では分散成分モデルにおいて，平均の James-Stein 型縮小推定量が経験 Bayes 推定量もしくは経験最良線形不偏予測量として自然に得られることを述べた。この応用例として重要なのが小地域推定問題である。これは，全国的な標本調査から全国を構成する各地域の母集団値を推定する問題であり，小地域における標本数が少ないことからその推定誤差がきわめて大きくなることが知られている。このような状況において他(周辺)地域に関する調査データや別の目的のために得られる他調査データなどの補助情報を用いることでより効率的な推定を行うことができる可能性がある。ここでは分散成分モデルを利用する小地域推定(small area estimation)方法を中心に，時系列データや2値データを解析するモデルなども含めて紹介する。

小地域推定を扱った論文は，Bayes 統計学の隆盛や変量モデルの普及に呼応してここ 10 年の間に実に多く出版されてきた。代表的な論文としては Ghosh and Rao (1994), Datta and Ghosh (1991), Prasad and Rao (1990) があげられる。また 8.1, 8.2 節のより詳しい説明については，久保川，国友 (1996) を参照されたい。

8.1 混合モデルを利用した推定方法：穀物作付面積の推定を例として

混合モデルを利用して実際にデータ解析を行った Battese, Harter and Fuller (1988) が扱った小地域に対する穀物(とうもろこしと大豆)の作付面積の推定問題について紹介しておこう。そこで扱っている問題では標本が得られた地域はアイオワ州の北部中央地域の 12 の郡(county)であり，各郡はさらに約 250h の農作区画(segment)に細分されている。i 番目の郡における母集団区画の個数を N_i とし, そのうち n_i 個の区画が標本として抽出されている。n_i は 1~5 程度で, 12個の郡全体で 36 個の区画が標本としてランダムに抽出されている。標本として抽出された各々の区画では, 連邦農務省 (USDA) が直接農家にインタヴュー調査を行うことによって，その区画におけるとうもろこし(または大豆)の作付面積についてのデータが利用可能である。他方, 人工衛星 LANDSAT から観測することにより, 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, 12 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得られている。ここで連邦農務省の調査データと衛星データを用いて, 各々の区画におけるとうもろこし(または大豆)作付面積の郡母集団平均を効率的に推定することが分析の目的である。

各々の区画についてのデータを用いると連邦農務省の調査データと衛星データとの間にはほぼ線形関係が認められるため,

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_{ij}, \quad i = 1, \ldots, k, \ j = 1, \ldots, n_i \tag{8.1}$$

なるモデルが想定できる。ここで i は郡に対する指数, j は各郡の中の区画に対する指数とすれば y_{ij} は i 番目の郡における j 番目の区画のとうもろこし(または大豆)の作付面積についての調査データを表わしている。また x_{1ij}, x_{2ij} はそれぞれ, 各々の区画について衛星データによりとうもろこし及び大豆と識別されたピクセルの個数を表わしている。誤差項 u_{ij} は, 郡効果(地域別差異)を考慮して

$$u_{ij} = v_i + e_{ij} \tag{8.2}$$

なる構造をもつと考えられよう。各小地域の標本サイズが小さいため，v_i を母数効果として扱った場合には推定誤差が大きくなってしまうことがデータ分析上の大きな問題である。そこで v_i を変量効果として扱い，この変量がすべての小地域で同一の分布に従って

いると想定することによって，小地域の周囲の情報を組み込むことが可能となると云うのがBHFモデルのアプローチである。ここでは，v_i, e_{ij} はすべて互いに独立で正規分布

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

に従うとしよう。この仮定は，u_{ij} の共分散構造が

$$E[u_{ij}u_{st}] = \begin{cases} \sigma_v^2+\sigma_e^2, & i=s, j=t; \\ \sigma_v^2, & i=s, j\neq t; \\ 0, & i\neq s \end{cases}$$

で与えられることを意味している。この仮定により12の郡すべての情報が組み込まれるわけである。ここで例えばある小地域に対してそれに隣接する地域の情報のみを用いたいのであれば，郡効果の入れ方を適当に修正したモデルを考えればよいことを注意しておく。

ここで設定した分散成分モデルのもとで, i 番目の郡に対する各々の区画におけるとうもろこし作付面積の母集団郡平均を推定することが問題である。第 i 郡の母集団サイズを N_i，第 i 郡，第 j 区画の穀物の作付け面積を Y_{ij} として有限母集団(finite population)の枠組みで考えると,

$$Y_{ij} = \beta_0+\beta_1 x_{1ij}+\beta_2 x_{2ij}+v_i+e_{ij}, \quad i=1,\ldots,k,\ j=1,\ldots,N_i$$

より，推定すべき母数は

$$\mu_{i(p)} \equiv \bar{Y}_i = N_i^{-1}\sum_{j=1}^{N_i} Y_{ij} = \beta_0+\beta_1\bar{x}_{1i(p)}+\beta_2\bar{x}_{2i(p)}+v_i+\bar{e}_{i(p)} \tag{8.3}$$

と見なすことができる。ただし, $\bar{x}_{\ell i(p)} = N_i^{-1}\sum_{j=1}^{N_i} x_{\ell ij}$, $\ell=1,2$, $\bar{e}_{i(p)} = N_i^{-1}\sum_{j=1}^{N_i} e_{ij}$ である。有限母集団の枠組みで考えると郡の母集団サイズ N_i が大きい小地域については $\bar{e}_{i(p)}$ が 0 に近いと考えられるので，ここではまず

$$\mu_i \equiv \beta_0+\beta_1\bar{x}_{1i(p)}+\beta_2\bar{x}_{2i(p)}+v_i \tag{8.4}$$

を推定する問題を考えよう。

ここで問題の記述をより簡便にするために混合モデルの行列表現を与えよう。ベクトルと行列の大きさを適当に揃えて $\boldsymbol{y}_i=(y_{i1},\ldots,y_{in_i})'$, $\boldsymbol{y}=(\boldsymbol{y}_1',\ldots,\boldsymbol{y}_k')'$, $\boldsymbol{x}_{ij}=(1,x_{1ij},x_{2ij})$, $\boldsymbol{x}_i=(\boldsymbol{x}_{i1}',\ldots,\boldsymbol{x}_{in_i}')'$, $\boldsymbol{X}=(\boldsymbol{x}_1',\ldots,\boldsymbol{x}_k')'$, $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)'$ とし，$\boldsymbol{u}$ も $\boldsymbol{y}$ と同様に $\boldsymbol{u}_i=(u_{i1},\ldots,u_{in_i})'$, $\boldsymbol{u}=(\boldsymbol{u}_1',\ldots,\boldsymbol{u}_k')'$ と定義する。このとき式 (8.1) と (8.2) からなる混合モデルは

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}$$

と表わすことができる。ただし，誤差項 $\boldsymbol{u}$ の分散共分散行列はブロック対角行列 (block diag($\cdot$) で表わす。) により

$$\boldsymbol{V}(\omega)=E[\boldsymbol{uu}'] = \text{block diag}(\boldsymbol{V}_1,\ldots,\boldsymbol{V}_k),$$

$$\boldsymbol{V}_i=\boldsymbol{J}_i\sigma_v^2+\boldsymbol{I}_i\sigma_e^2$$

と表現される。ただし，母数 $\omega=(\sigma_v^2,\sigma_e^2)$, また $n_i\times n_i$ 行列 $\boldsymbol{J}_i=\boldsymbol{e}_i\boldsymbol{e}_i'$ ($\boldsymbol{e}_i$ はすべての要素が1の $n_i\times 1$ ベクトル), $\boldsymbol{I}_i$ (あるいは$\boldsymbol{I}_{n_i}$) は $n_i\times n_i$ の単位行列である。また (8.4) で与えられた, 推定したい郡母平均 μ_i は

$$\mu_i=\bar{\boldsymbol{x}}_{i(p)}\boldsymbol{\beta}+v_i$$

と表わされる。ただし $\bar{\boldsymbol{x}}_{i(p)} = N_i^{-1}\sum_{j=1}^{N_i} \boldsymbol{x}_{ij}$ とおいた。

2つの分散母数 $\omega = (\sigma_v^2, \sigma_e^2)$ が既知のときには誤差項の分散共分散行列が既知と見なせるので回帰分析の一般的議論から母係数 $\boldsymbol{\beta}$ の一般化最小2乗推定量 (GLSE) は

$$\widetilde{\boldsymbol{\beta}}(\omega) = (\boldsymbol{X}'\boldsymbol{V}(\omega)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\omega)^{-1}\boldsymbol{y} \tag{8.5}$$

で与えられる。さらに郡効果を表現している変量 v_i と $\bar{u}_i = n_i^{-1}\sum_{j=1}^{n_i} u_{ij}$ の共分散行列は簡単な計算から

$$\begin{pmatrix} \sigma_v^2 & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2 + n_i^{-1}\sigma_e^2 \end{pmatrix}$$

で与えられる。従って正規分布の性質を利用すると $\bar{u}_i$ を与えたときの v_i の条件付期待値は

$$E[v_i|\bar{u}_i] = \bar{u}_i \frac{\sigma_v^2}{\sigma_v^2 + n_i^{-1}\sigma_e^2}$$

となる。ここで, $\bar{\boldsymbol{x}}_i = n_i^{-1}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ とおくとき, $\bar{u}_i$ は $\bar{y}_i - \bar{\boldsymbol{x}}_i\widetilde{\boldsymbol{\beta}}(\omega)$ で推定されるので, 結局, μ_i の最良線形不偏推定量 (BLUE) は

$$\tilde{\mu}_i(\omega) = \bar{\boldsymbol{x}}_{i(p)}\widetilde{\boldsymbol{\beta}}(\omega) + \left(\bar{y}_i - \bar{\boldsymbol{x}}_i\widetilde{\boldsymbol{\beta}}(\omega)\right)\frac{\sigma_v^2}{\sigma_v^2 + n_i^{-1}\sigma_e^2} \tag{8.6}$$

で与えられることがわかる。

ここで求められた推定量は次のように直観的に解釈することが可能である。まず $\bar{\boldsymbol{x}}_{i(p)}\widetilde{\boldsymbol{\beta}}(\omega)$ を回帰総合推定量 (Regression Synthetic Estimator)

$$\bar{\boldsymbol{x}}_{i(p)}\widetilde{\boldsymbol{\beta}}(\omega) + \left(\bar{y}_i - \bar{\boldsymbol{x}}_i\widetilde{\boldsymbol{\beta}}(\omega)\right)$$

は $\bar{y}_i$ に依存するので調査回帰推定量 (Survey Regression Estimator) と呼ぶことにしよう。$\tilde{\mu}_i(\omega)$ はこれら2つの推定量を σ_v^2 と σ_e^2/n_i の分散比で内分した形をしている。即ち, v_i の e_{ij} に対する相対的バラツキが小さければ, $\bar{y}_i$ を $\bar{\boldsymbol{x}}_{i(p)}\widetilde{\boldsymbol{\beta}}(\omega)$ の方向へ縮小することによって単純平均のばらつきを修正し，点推定の安定化を図っていると解釈できるのである。

次に有限母集団の枠組みで各郡の母集団平均 $\mu_{i(p)} = \bar{Y}_i$ を推定する問題を考えよう。いま観測されたデータの組を $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$, 観測されないデータを $\boldsymbol{Y}_i^* = (Y_{i,n_i+1}, \ldots, Y_{i,N_i})'$ とおくと，$\boldsymbol{y}_i$ を与えたときの $\boldsymbol{Y}_i$ の条件付き期待値 $E[\boldsymbol{Y}_i|\boldsymbol{y}_i]$ を用いて $\boldsymbol{Y}_i$ を推定するのが最も自然と考えられよう。この条件付き期待値を具体的に求めるために $(\boldsymbol{y}_i', \boldsymbol{Y}_i^{*\prime})'$ の同時密度関数を書き表わすと,

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{Y}_i^* \end{pmatrix} \sim \mathcal{N}_{N_i}\left(\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{x}_i^* \end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right)$$

となる。ただし，$\boldsymbol{x}_{ij} = (1, x_{1ij}, x_{2ij})$ に対して $\boldsymbol{x}_i^* = (\boldsymbol{x}_{i,n_i+1}', \ldots, \boldsymbol{x}_{i,N_i}')'$ とし，また n_i 次元ベクトル $\boldsymbol{e}_i = (1, \ldots, 1)'$, $N_i - n_i$ 次元ベクトル $\boldsymbol{e}_i^* = (1, \ldots, 1)'$ を用いると分散共分散行列の各行列要素は

$$\boldsymbol{\Sigma}_{11} = \sigma_v^2\boldsymbol{e}_i\boldsymbol{e}_i' + \sigma_e^2\boldsymbol{I}_{n_i}, \quad \boldsymbol{\Sigma}_{12} = \sigma_v^2\boldsymbol{e}_i\boldsymbol{e}_i^{*\prime}, \quad \boldsymbol{\Sigma}_{22} = \sigma_v^2\boldsymbol{e}_i^*\boldsymbol{e}_i^{*\prime} + \sigma_e^2\boldsymbol{I}_{N_i-n_i}$$

と表される。従って $\boldsymbol{y}_i$ を与えたときの $\boldsymbol{Y}_i^*$ の条件付き分布は

$$\boldsymbol{Y}_i^*|\boldsymbol{y}_i \sim \mathcal{N}_{N_i-n_i}\left(\boldsymbol{x}_i^*\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)$$

となる。この条件付き分布の平均は，

$$E[\boldsymbol{Y}_i^*|\boldsymbol{y}_i] = \boldsymbol{x}_i^*\boldsymbol{\beta} + \frac{n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2}\boldsymbol{e}_i^*(\bar{y}_i - \bar{\boldsymbol{x}}_i\boldsymbol{\beta})$$

となるので，

$$\begin{aligned} E\left[\sum_{j=n_i+1}^{N_i} Y_{ij}\Big|\boldsymbol{y}_i\right] &= E\left[\boldsymbol{e}_i^{*\prime}\boldsymbol{Y}_i^*|\boldsymbol{y}_i\right] \\ &= (N_i - n_i)\left\{\bar{\boldsymbol{x}}_i^*\boldsymbol{\beta} + \frac{n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2}(\bar{y}_i - \bar{\boldsymbol{x}}_i\boldsymbol{\beta})\right\} \end{aligned}$$

と表わされる。ただし，$\bar{\boldsymbol{x}}_i^* = (N_i - n_i)^{-1}\sum_{j=n_i+1}^{N_i}\boldsymbol{x}_{ij}$ である。この $\boldsymbol{\beta}$ に一般化最小 2 乗推定量 $\tilde{\boldsymbol{\beta}}$ を代入すると，以上の計算から，ここで求める郡の母平均 $\mu_{i(p)} = \bar{Y}_i = N_i^{-1}\sum_{j=1}^{N_i}Y_{ij}$ の推定量は，

$$\begin{aligned} \tilde{\mu}_{i(p)}(\omega) &= E[\bar{Y}_i|\boldsymbol{y}_i] \\ &= \frac{n_i}{N_i}\bar{y}_i + \frac{1}{N_i}E\left[\sum_{j=n_i+1}^{N_i} Y_{ij}\Big|\boldsymbol{y}_i\right] \\ &= \frac{n_i}{N_i}\bar{y}_i + \frac{N_i - n_i}{N_i}\left\{\bar{\boldsymbol{x}}_i^*\tilde{\boldsymbol{\beta}} + \frac{n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2}(\bar{y}_i - \bar{\boldsymbol{x}}_i\tilde{\boldsymbol{\beta}})\right\} \end{aligned}$$

で与えられる。これは一般的な議論から実は最良線形不偏推定量にもなっていることがわかる。

これまでは分散成分 σ_v^2, σ_e^2 が既知のものと仮定して考察してきた。しかし実際の応用の場では分散成分が既知なことはきわめてまれであり，これらの母数をデータから推定する必要がある。与えられたデータからまず分散成分の推定量を求めて $\tilde{\mu}_i(\omega)$, $\tilde{\mu}_{i(p)}(\omega)$ に代入して得られた母数の推定量は 2 段階推定量と呼ばれている。

分散成分の 2 段階推定では，しばしば Henderson (1953) が提案した方法が使われている。この方法ではまず $\bar{x}_{\ell i} = \sum_j x_{\ell ij}/n_i$ とし，$y_{ij} - \bar{y}_i$ を説明変数 $\{x_{1ij} - \bar{x}_{1i}, x_{2ij} - \bar{x}_{2i}\}$ へ通常の最小 2 乗 (OLS) 回帰したときの残差を $\hat{e}_{ij}$ とする。また y_{ij} を $\{1, x_{1ij}, x_{2ij}\}$ へ最小 2 乗回帰したときの残差を r_{ij} としよう。さらに 2 つの残差 2 乗和を

$$S_1 = \sum_{i=1}^k\sum_{j=1}^{n_i}\hat{e}_{ij}^2, \quad S_2 = \sum_{i=1}^k\sum_{j=1}^{n_i}(r_{ij}^2 - \hat{e}_{ij}^2)$$

とおこう。ここで

$$\nu = \sum_{i=1}^k n_i - k - 2,$$

$$M = (k-1)^{-1}\left\{\sum_{i=1}^k n_i - \mathrm{tr}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{i=1}^k n_i^2\bar{\boldsymbol{x}}_i'\bar{\boldsymbol{x}}_i\right]\right\}$$

とおこう。このとき σ_e^2, σ_v^2 の推定量として，3.3 節で与えられた推定量 $(\hat{\sigma}_1^{2REML}, \hat{\sigma}_A^{2REML})$, $(\hat{\sigma}_1^{2EB}, \hat{\sigma}_A^{2EB})$ が $(\hat{\sigma}_1^{2UB}, \hat{\sigma}_A^{2UB})$ において，n_1, n_2, r をそれぞれ ν, $k-1$, M で置き換えた推定量を用いればよい。

8.2　小地域推定のための他のモデル

ここでは，小地域の推定のために提案されてきた他のモデルを紹介する。

[1] 一般的な混合線形モデル

前節で議論した小地域推定のための分散成分モデルは，より一般的な混合線形モデル

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e}$$

の1つの例になっていると見なすことができる。ここで，$\boldsymbol{y}$ は観測データのベクトル，$\boldsymbol{X}$, $\boldsymbol{Z}$ は既知の行列である。また $\boldsymbol{v}$, $\boldsymbol{e}$ は独立にそれぞれ多次元正規分布 $\mathcal{N}(\boldsymbol{0}, \boldsymbol{G})$, $\mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$ に従う確率変数で，共分散行列 $\boldsymbol{G}$, $\boldsymbol{R}$ はいくつかの分散成分 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$ に依存している。$\boldsymbol{y}$ の共分散行列は

$$\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{ZGZ}'$$

と表されるので，分散成分 $\boldsymbol{\theta}$ を既知とするとき，$\boldsymbol{\beta}$ の一般化最小2乗推定量は

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

で与えられる。このとき推定したい母数をベクトル $\boldsymbol{a}$, $\boldsymbol{b}$ を既知として例えば $\mu = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{v}$ とすれば，その最良線形不偏推定量は

$$\tilde{\mu}(\boldsymbol{\theta}) = \boldsymbol{a}'\widetilde{\boldsymbol{\beta}} + \boldsymbol{b}'\boldsymbol{GZ}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})$$

で与えられる。従って，前節で扱った混合モデルはここで説明した一般的な混合線形モデルの特殊な例となっていることはあきらかであろう。さらにこの混合線形モデルの枠組みの中でとらえることのできるものとして，ここでさらに3つのモデルを挙げておくことにしよう。

[2] 変量係数モデル

Dempster-Rubin-Tsutakawa (1981) は，回帰係数 β を変量的に扱う一般的なモデルを提案している。彼らのモデル (DRT モデルと略する) は直接には分散成分モデルとして開発しているが次のような意味で前節に議論した小地域推定と関連づけることが可能である。その統計モデルを分散成分モデルとして見ると

$$\begin{aligned} y_{ij} &= \beta_i x_{ij} + e_{ij} \\ &= \beta x_{ij} + v_i x_{ij} + e_{ij}, \quad i = 1, \ldots, k,\ j = 1, \ldots, n_i \end{aligned}$$

と表現することができる。ここで確率変数 v_i, e_{ij} は前節で定義されたものと同じである。各郡の係数は $\beta_i = \beta + v_i$ としているので，回帰係数が固定係数ではなく変量となっていることを意味している。このことからこの種のモデルは前節に議論したBHFの分散成分モデルを郡効果の分散が説明変数の水準に依存する形に拡張していることになる。

このDRTモデルでは第 i 地域の分散共分散行列は $\boldsymbol{V}_i = \sigma_e^2 \boldsymbol{I}_{n_i} + \sigma_v^2 \boldsymbol{x}_i \boldsymbol{x}_i'$ となっている。ただし $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{in_i})'$, $\boldsymbol{X} = (\boldsymbol{x}_1', \cdots, \boldsymbol{x}_k')'$ である。ここで行列の逆転公式を使えば

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} = \sum_{i=1}^{k} \frac{1}{\sigma_v^2 + \sigma_e^2 \left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}}$$

という関係を利用することができる。したがって係数の一般化最小２乗推定量は

$$\tilde{\beta} = \frac{\sum_{i=1}^{k}\left(\sigma_v^2 + \sigma_e^2\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}\right)^{-1}\sum_{i=1}^{k}\boldsymbol{x}_i'\boldsymbol{y}_i}{\sum_{i=1}^{k}\left[\sigma_v^2 + \sigma_e^2\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}\right]^{-1}}$$

である。

ここで関心のある小地域 i に対して推定したい平均は $\mu_i = \bar{X}_i\beta + \bar{X}_i v_i$ である。$\boldsymbol{u}_i = (v_i x_{i1} + e_{i1}, \cdots, v_i x_{in_i} + e_{in_i})'$ とおき，確率変数 v_i と $\boldsymbol{u}_i$ の分散共分散行列を計算すると条件付期待値が

$$E[v_i|\boldsymbol{u}_i] = \frac{\sigma_v^2\sum_{j=1}^{n_i}x_{ij}^2}{\sigma_v^2\sum_{j=1}^{n_i}x_{ij}^2 + \sigma_e^2}\left[\left(\sum_{j=1}^{n_i}x_{ij}^2\right)^{-1}\left(\sum_{j=1}^{n_i}x_{ij}y_{ij}\right) - \beta\right]$$

で与えられるので，μ_i の最良線形不偏推定量は

$$\hat{\mu}_i(\omega) = \bar{X}_i\hat{\beta} + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2/\sum_{j=1}^{n_i}x_{ij}^2}\bar{X}_i(\tilde{\beta}_i - \tilde{\beta})$$

となる。ここで $\omega = (\sigma_v^2, \sigma_e^2)$, $\tilde{\beta}_i = \sum_{j=1}^{n_i}x_{ij}y_{ij}/\sum_{j=1}^{n_i}x_{ij}^2$ である。

このモデルにおいて２段階推定量を得るための分散成分の不偏推定量は２つの最小２乗残差より構成することができる。いま各郡データに対して最小２乗法を適用して得られた最小２乗残差は

$$\hat{e}_{ij} = y_{ij} - x_{ij}\left(\sum_{j=1}^{n_i}x_{ij}^2\right)^{-1}\sum_{j=1}^{n_i}x_{ij}y_{ij}$$

で与えられる。また全体のデータを用いて最小２乗法を適用して得られる最小２乗残差を r_{ij} としよう。

$$n_* = \sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}^2 - \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}^2\right)^{-1}\sum_{i=1}^{n_i}\left(\sum_{j=1}^{n_i}x_{ij}^2\right)^2$$

とするとき，このDRTモデルにおける母分散成分の不偏推定量はそれぞれ

$$\hat{\sigma}_e^2 = \left(\sum_{i=1}^{k}n_i - k\right)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\hat{e}_{ij}^2,$$

$$\hat{\sigma}_v^2 = n_*^{-1}\left\{\sum_{i=1}^{k}\sum_{j=1}^{n_i}r_{ij}^2 - \sum_{i=1}^{k}(n_i - 1)\hat{\sigma}_e^2\right\}$$

で与えられる。さらに，こうして得られた郡母分散の不偏推定量は正の確率で負の値をとることがあることがわかる。そこで 3.3 節と同様にしてこれらを改良する非負な推定量を得ることができよう。

ここで説明したDRTモデルなどのような統計モデルは補助変数と母集団データとの関係が各郡，あるいは分類により異なると考えられる場合に有効であろう。

[3] Fay-Herriot モデル

Fay and Herriot (1979) は，小地域における個人所得を推定する問題において通常の標本平均を改善するモデル分析として Efron and Morris (1975) による経験ベイズ・モデルを取り上げている。彼らは実際の所得データへの応用をいわゆる多変量解析における母平均の推定に関して知られている Stein 効果の例としての視点からとりあげている。ここではより一般的にその統計的モデル (FH モデルと略する) を，既に説明した混合モデルの枠組みを使えば簡単にその本質的部分を理解できることを指摘しておこう。

このFHモデルはこれまでのように行列記号を使えば

$$\bar{y}_i = \mu_i + e_i,$$

$$\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i\ , \quad i = 1, \ldots, k,$$

と表現することができる。ここで，$\boldsymbol{x}_i$ は補助変数 (縦) ベクトル，e_i, v_i はすべて互いに独立に, それぞれ $\mathcal{N}(0, D_i), \mathcal{N}(0, A)$ にしたがう確率変数，A は未知母数である。さらに D_i は他の情報を用いる状況を想定しているので既知としておく。(例えば $\bar{y}_i$ に対応する標本サイズを n_i とするとき $D_i = \sigma^2/n_i$ で与えるのが自然であろう。この場合には σ^2 は他の情報を用いて推定されている状況を考えればよいであろう。) このとき A と D_i を既知とするときの $\boldsymbol{\beta}$ の一般化最小 2 乗推定量は $\boldsymbol{V} = \mathrm{diag}(A+D_1, \ldots, A+D_k)$, $\overline{\boldsymbol{y}} = (\bar{y}_1, \ldots, \bar{y}_k)'$ に対して

$$\begin{aligned}\widetilde{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\overline{\boldsymbol{y}} \\ &= \left[\sum_{i=1}^{k}(D_i + A)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1} \times \left[\sum_{i=1}^{k}(D_i + A)^{-1}\boldsymbol{x}_i\bar{y}_i\right]\end{aligned}$$

で与えられる。

また確率変数 v_i と u_i の分散共分散行列より条件付期待値は

$$E[v_i|u_i] = \frac{D_i}{D_i + A}u_i$$

となる。したがって，これまでの議論と同様にしてμ_i の最良線形不偏推定量は

$$\tilde{\mu}_i(A) = \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}} + \frac{D_i}{A + D_i}(\bar{y}_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})$$

で与えられることがわかる。この推定量は最小 2 乗推定を, 与えられた情報の方向に改善しようとする形と理解することができる。ここで母数 A の不偏推定量は最小 2 乗残差

$$\hat{u}_i = \bar{y}_i - \boldsymbol{x}_i'\left(\sum_{i=1}^{k}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\sum_{i=1}^{k}\boldsymbol{x}_i\bar{y}_i$$

を用いて,

$$\widehat{A} = (k-p)^{-1}\left\{\sum_{i=1}^{k}\hat{u}_i^2 - \sum_{i=1}^{k}D_i\left(1 - \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i\right)\right\}$$

で与えられる。そこでさらにこの推定量から適当に非負な値をとるように修正して分散を推定することが可能である。

[4] 時系列データを用いた小地域推定モデル

計量経済学などの研究分野においては時系列データに対する混合モデルが様々な形で研究されてきた。例えば計量経済学におけるこの種の研究については代表的文献としては Anderson and Hsiao (1981, 82) などを挙げることができよう。各小地域 i でのデータが時系列的に得られている場合については最近になり Rao and Yu (1994) が詳しい検討を始めている。彼らのモデルは

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_i + u_{it} + e_{it},$$

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad |\rho| < 1,$$

と表現される。ただし，$v_i \sim \mathcal{N}(0, \sigma_v^2)$, $e_{it} \sim \mathcal{N}(0, \sigma_e^2)$, $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$ であり，また確率変数列 $\{v_i\}, \{e_{it}\}, \{\epsilon_{it}\}$ は互いに独立である。この種のモデルも混合モデルとして表現できるので小地域の平均値の 2 段階推定量を導入することができる。彼らの研究では通常の 2 段階推定量の平均 2 乗誤差などの導出を行っているが，前節で報告した分散推定の改善を行うことも可能であろう。

8.3 2 値データに基づいた小地域推定

2 値データの解析において，変量効果を導入したモデルは，死亡率の推定や死亡指標の推定など観測されるデータ数が少ないためバラツキが著しい問題に応用され，近年活発に研究されてきた。

ここでは，地域が $i = 1, \ldots, I$ に，また年齢階級が $j = 1, \ldots, J$ に分割されているときに，i 地域 j 年齢階級の (特定の病気による) 死亡率 p_{ij} を推定することを考える。

y_{ij} : i 地域，j 年齢階級の観測死亡数
n_{ij} : i 地域，j 年齢階級の人口
N_j : 標準人口集団の j 年齢階級人口
$N = N_1 + N_2 + \cdots + N_J$

p_{ij} の通常の推定量は，$\hat{p}_{ij}^{ML} = y_{ij}/n_{ij}$ で，これは疾病地図を描く際の死亡指標として，年齢調整死亡率 (DAR)

$$DAR_i = \sum_{j=1}^{J} \hat{p}_{ij}^{ML} \frac{N_j}{N}$$

を構成するのに使われる。しかし，分割されたセルによっては n_{ij} が小さくなるため，y_{ij} の大小が $\hat{p}_{ij}^{ML}$ の推定に影響を与え，その影響が結果として DAR_i の値を大きく左右してしまうという問題があることが知られている。すなわち，DAR_i の大小が n_{ij} の小さいセルの観測度数 y_{ij} によって大きく影響されてしまうという不合理な性質をもっている。そこで，このような問題点を修正できる手法として，標準化死亡率 (SMR)

$$SMR_i = \frac{\sum_{j=1}^{J} y_{ij}}{\sum_{j=1}^{J} n_{ij} P_{0j}}, \quad P_{0j} = \frac{\sum_{i=1}^{I} y_{ij}}{N_j}$$

が用いられるのが一般的である。

この節では，こうした問題を解決するために変量効果を導入したモデルとして知られる一般線形混合モデル (GLMM) や，その 1 つの例であるポアソン-ガンマモデルについて簡単に紹介する。

[1] 一般線形混合モデル

y_{ij} が2項分布 $Bin(n_{ij}, p_{ij})$ に従うとき，$p_{ij} = \exp(\theta_{ij})/\{1+\exp(\theta_{ij})\}$ とおいて θ_{ij} が

$$\theta_{ij} = \log\frac{p_{ij}}{1-p_{ij}} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \quad i=1,\ldots,I,\ j=1,\ldots,J \tag{8.7}$$

なる加法モデルをもつとする。ここで α_i, β_j はそれぞれ地域，年齢階級に対応する変量効果で $\alpha_i \sim \mathcal{N}(0,\sigma_A^2)$, $\beta_j \sim \mathcal{N}(0,\sigma_B^2)$ とし，交互作用項 γ_{ij} は $\gamma_{ij} \sim \mathcal{N}(0,\sigma_{AB}^2)$ なる変量とすることが考えられる。また共変量 $\boldsymbol{x}_{ij}$ が存在するときには，

$$\theta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_{ij}, \quad i=1,\ldots,I,\ j=1,\ldots,J \tag{8.8}$$

$\alpha_i \sim \mathcal{N}(0,\sigma_A^2)$, $\beta_j \sim \mathcal{N}(0,\sigma_{AB}^2)$, なる線形混合モデルも役立つであろう (Waclawiw and Liang (1994))。ここで $\boldsymbol{\beta}$ は未知の回帰係数ベクトルである。

上述の θ_{ij} には Logit 変換 $\theta_{ij} = \log p_{ij}/(1-p_{ij})$ が使われているが，標準正規分布の分布関数 $\Psi(\cdot)$ を用いて Probit 変換 $\theta_{ij} = \Psi^{-1}(p_{ij})$ により与えられてもよい。MaCulloch (1994) は，より一般的な Probit-Normal モデル

$$\begin{aligned} y_i &= \boldsymbol{x}_i\boldsymbol{\beta} + \sum_{s=1}^{k} \boldsymbol{z}_{is}\boldsymbol{u}_s + \varepsilon_i, \quad i=1\ldots,n \\ w_i &= I(y_i > 0) \\ \varepsilon_i &\sim \mathcal{N}(0,1), \quad \boldsymbol{u}_s \sim \mathcal{N}_{q_s}\left(\mathbf{0}, \sigma_s^2\boldsymbol{I}_{q_i}\right) \end{aligned}$$

を考えて，分散成分の REML 推定量を求めるためのアルゴリズム等についての議論をしている。ここで，$I(\cdot)$ は，定義関数であり，$w_i = 1$ or $w_i = 0$ のみが観測され，y_i の値は観測されない。

y_{ij} の従う分布として，ポアソン分布を含めたより一般的な分布族への拡張が，McCullagh and Nelder (1989) の一般線形モデル (GLM)

$$f(y_{ij}|\theta_{ij},\phi_{ij}) = \exp\left[\phi_{ij}^{-1}\left(y_{ij}\theta_{ij} - \psi(\theta_{ij})\right) + \rho(y_{ij};\phi_{ij})\right],$$

であり，自然母数 θ_{ij} に (8.7), (8.8) なる構造を仮定することが考えられる。Canonical Link 関数 $h(\cdot)$ を用いて

$$h(\theta_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_{ij}, \quad i=1,\ldots,I,\ j=1,\ldots,J$$

なる構造を埋め込んだモデルは，一般線形混合モデル (GLMM) と呼ばれ，McCullagh and Nelder (1989, 14.5節), Breslow and Clayton (1993), Lee and Nelder (1996), Malec and Sedransk (1997), Jiang (1998), Sutradher and Qu (1998) などによって扱われ，種々の計算方法が提案検討されている。Ghosh, *et al.* (1998) では，ϕ_{ij} を既知の尺度母数とし，$\boldsymbol{\beta}$, σ_A^2, σ_{AB}^2 に適当な事前分布を想定したときに，θ_{ij} の同時事後分布が有限であることを示し，θ_{ij} を推定するための Gibbs sampler を明示して，Missouri 州の肺ガンの死亡率地図を作成した。

死亡率のような度数の小さいデータの解析において，変量効果モデル，経験 Bayes モデルが極めて有用であることを印象づけた論文が，Tsutakawa (1985), Tsutakawa *et al.* (1985) である。彼らは，Missouri 州の 84 都市の死亡率 p_i を推定する際，y_i/n_i が極めて小さい値をとることから，y_i に平均 n_ip_i のポアソン分布を仮定し，

$$\theta_i = \log p_i/(1-p_i) = \mu + \alpha_i, \quad \alpha_i \sim \mathcal{N}(0,\sigma_A^2)$$

なる変量モデルを想定することを考えた。この変量効果により $y_i/\boldsymbol{n}_i$ は共通平均 μ の推定値の方向へ縮小されて平滑化され，より安定した推定値が得られることを示した。

[2] ポアソン-ガンマモデル

一般線形混合モデルは，共変量が使える場合には特に有用であるが，Logit 変換，Probit 変換を用いた場合には Bayes 推定量等を陽に書き下すことができないため，どのような操作によって平滑化がなされるのか，そのイメージを描きにくい。ポアソン分布の共役事前分布はガンマ分布であることから，p_{ij} にガンマ分布を仮定すると p_{ij} の Bayes 推定量は陽に表現できる。このような点から，丹後 (1988), Clayton and Kaldor (1987), Tsutakawa (1988), Manton *et al.* (1989), Heisterkamp *et al.* (1993) などでは，死亡率推定のためにポアソン-ガンマモデルが取り上げられた。

観測度数 y_{ij} が平均 $n_{ij}p_{ij}$ のポアソン分布

$$y_{ij} \sim \mathcal{P}o(n_{ij}p_{ij}), \quad i = 1, \ldots, I, \; j = 1, \ldots, J$$

に従うとき，独立関係

$$M_0 : \; p_{ij} = \alpha_i \beta_j$$

を仮定する。ここで α_i が地域効果，β_j が年齢階級死亡率である。β_j を既知の定数 $\beta_j = P_{0j}$ として α_i の MLE を求めると

$$\widehat{\alpha}_i^{ML} = \frac{\sum_{j=1}^{J} y_{ij}}{\sum_{j=1}^{J} n_{ij} P_{0j}}$$

となり SMR_i が出てくる。すなわち，SMR は年齢階級死亡率が地域によらず一定という仮定のもとで導出された指標であるため，この仮定が成立しない場合での使用には問題があることを意味する。

丹後 (1988) は，

$$M_1 : \; p_{ij} = \beta_j \gamma_{ij}$$

なる構造を入れて，γ_{ij} にガンマ分布 $\mathcal{G}a(a_j, b_j^{-1})$

$$\{\Gamma(a_j)\}^{-1} b_j^{a_j} x^{a_j - 1} \exp\{-b_j x\}$$

を仮定して，p_{ij} の経験 Bayes 推定量

$$\hat{p}_{ij}^{EB} = \widehat{\beta}_j \frac{\hat{a}_j + y_{ij}}{\hat{b}_j + n_{ij}\widehat{\beta}_j} \tag{8.9}$$

を導出した。ただし，$\hat{a}_j$, $\hat{b}_j$, $\widehat{\beta}_j$ は周辺分布の MLE として与えられる。モデル M_1 を組み入れたポアソン-ガンマモデルは，

$$\theta_{ij} = \log p_{ij} = \log \beta_j + \log \gamma_{ij}$$

として，一般線形混合モデルの 1 つの例ではあるが，経験 Bayes 推定量が (8.9) の形で表わされており，推定量が平滑化される理由が理解しやすい。すなわち，$\hat{a}_j = \hat{b}_j = 0$ なら $\hat{p}_{ij}^{EB} = y_{ij}/n_{ij}$ であり，$\hat{a}_j > 0, \hat{b}_j > 0$ が組み込まれることによって，$\hat{p}_{ij}^{EB}$ はより安定した推定値を与えることができる。この $\hat{p}_{ij}^{EB}$ を用いると DAR の経験 Bayes 推定量は

$$EBDAR_i = \sum_{j=1}^{J} \widehat{\beta}_j \frac{\hat{a}_j + y_{ij}}{\hat{b}_j + n_{ij}\widehat{\beta}_j} \frac{N_j}{N}$$

となり，日本職業別年齢調整死亡率の推定, 東京都胃ガン都市別年齢調整死亡率の推定への適用と解析結果の検討が，丹後 (1988) に与えられている。

Tsutakawa (1988) は，

$$M_2 : \; p_{ij} = \beta_j \alpha_i \gamma_{ij}$$

なるモデルを考え，α_i を与えたときの γ_{ij} の条件付き分布にガンマ分布 $\mathcal{G}a(\sigma_j^{-2}, \sigma_j^2 \alpha_i \beta_j)$, α_i に母数 c をもった逆ガンマ分布を想定し，肺ガンによる男子都市別死亡率の推定を行った。モデル M_2 を対数線形モデルで書き直すと

$$\theta_{ij} = \log p_{ij} = \log \beta_j + \log \alpha_i + \log \gamma_{ij}$$

となり，地域効果 α_i と交互作用効果 γ_{ij} を変量として扱う混合モデルであることがわかる。

参考文献（英文）

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.

Amari, S. (1982). Differential geometry of curved exponential families - Curvature and information loss. *Ann. Statist.* , **10**, 357-385.

Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite ? *Amer. Statist.*, **39**, 112-117.

Amemiya, Y. and Fuller, W. (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. *Ann. Statist.*, **12**, 497-509.

Anderson, B.M., Anderson, T.W. and Olkin, I. (1986). Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance. *Ann. Statist.*, **14**, 405-417.

Anderson, T.W. and C. Hsiao (1981). Estimation of dynamic models with error components. *J. Amer. Statist. Assoc.*, **76**, 598-606.

Anderson, T.W. and C. Hsiao (1982). Formulation and estimation of dynamic models using Panel data. *J. Econometrics.*, **18**, 47-82.

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.

Berger, J.O.(1985). *Statistical Decision Theory and Bayesian Analysis.* 2nd Ed., Springer-Verlag, New York.

Bhattacharya, C.G.(1984). Two inequalities with an application. *Ann. Inst. Statist. Math.* **36**, 129-134.

Bock, R.D. and Petersen, A.C. (1975). A multivariate correction for attenuation. *Biometrika*, **62**, 673-678.

Bock, R.D. and Vandenberg, S.G. (1968). Components of heritable variation in mental test scores. In *Progress in Human Behavior Genetics* (S. G. Vandenberg, ed.) 233-260. Johns Hopkins Press, Baltimore.

Brandwein, A.C. and Strawderman, W.E. (1990). Stein estimation: The spherically symmetric case. *Statist. Science*, **5**, 356-369.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **88**, 9-25.

Brockwell, P.J., and Davis, R.A. (1991). *Time Series: Theory and Methods.* 2nd Ed. Springer-Verlag, New York.

Brown, L.D. and Cohen, A.(1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2**, 963-976.

Calvin, J.A., and Dykstra, R.L. (1991a). Maximum likelihood estimation of a set of covariance matrices under Loewner order restrictions with applications to balanced multivariate variance components models. *Ann. Statist.*, **19**, 850-869.

Calvin, J.A., and Dykstra, R.L. (1991b). Least squares estimation of covariance matrices in balanced multivariate variance components models. *J. Amer. Statist. Assoc.*, **86**, 388-395.
Casella, G. (1981). Minimax ridge regression estimation. *Ann. Statist.*, **8**, 1036-1056.
Casella, G. (1985). Condition numbers and minimax ridge regression estimators. *J. Amer. Statist. Assoc.*, **80**, 753-758.
Chang, Y.-T. (1981). Stein-type estimators for parameters restricted by linear inequalities. *Keio Sci. Tech. Rep.*, **34**, 83-95.
Chang, Y.-T. (1982). Stein-type estimators of parameters in truncated spaces. *Keio Sci. Tech. Rep.*, **35**, 185-193.
Cohen, A. and Sackrowitz, H.B.(1974). On estimating the common mean of two normal distributions. *Ann. Statist.* **2**, 1274-1282.
Datta, G.S., and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist.* **19**, 1748-1770.
Dempster, A.P., Rubin, D.B., and Tsutakawa, R.K. (1981). Estimation in covariance component models. *J. Amer. Statist. Assoc.*, **76**, 341-353.
Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika*, **59**, 335-347.
Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311-319.
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.
Fernandez, M.A., Rueda, C. and Salvador, B. (1999). Parameter estimation under orthant restrictions. *Canad. J. Statist.*, **27**, to appear.
Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998). Generalized linear models for small-area estimation. *J. Amer. Statist. Assoc.*, **93**, 273-282.
Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Science*, **9**, 55-93.
Ghosh, M., Saleh, A.K.Md.E. and Sen, P.K. (1989). Empirical Bayes subset estimation in regression models. *Statist. Decisions*, **7**, 15-35.
Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press,
Heisterkamp, S.H., Doornbos, G., and Gankema, M. (1993). Disease mapping using empirical Bayes and Bayes methods on mortality statistics in the Netherlands. *Statist. Medicine*, **12**, 1895-1913.
Henderson, C.R.(1953). Estimation of variance and covariance components. *Biometrics*, **9**, 226-252.
Hill, R.C. and Judge, G.G. (1990). Improved estimation under collinearity and squared error loss. *J. Multivariate Anal.*, **32**, 296-312.
Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55-68.
Hwang, J.T. (1985). Universal domination and stochastic domination. *Ann. Statist.*, **13**, 295-314.
Hwang, J.T. and Peddada, S.D. (1994). Confidence interval estimation subject to order restrictions. *Ann. Statist.*, **22**, 67-93.
James, W., and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* , **1**, 361-379. University of California Press, Berkeley.
Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **93**, 720-729.
Kalman, R.E.K. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82D, 35-45.

Kariya, T. (1989). Equivariant estimation in a model with an ancillary statistic. *Ann. Statist.*, **17**, 920-928.

Kariya, T., Giri, N.C. and Perron, F. (1988). Equivariant estimation of a mean vector $\boldsymbol{\mu}$ of $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = 1$ or $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = \boldsymbol{c}$ or $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\mu}'\boldsymbol{\mu}\boldsymbol{I}$. *J. Multivariate Anal.*, **27**, 270-283.

Kleffe, J. and Rao, J.N.K.(1986). The existence of asymptotically unbiased nonnegative quadratic estimates of variance components in ANOVA models. *J. Amer. Statist. Assoc.* **81**, 692-698.

Klotz, J. and Putter, J. (1969). Maximum likelihood estimation of multivariate covariance components for the balanced one-way layout. *Ann. Math. Statist.*, **40**, 1100-1105.

Konishi, S., and Gupta, A.K. (1989). Testing the equality of several intraclass correlation coefficients. *J. statist. Plann. Inf.*, **21**, 93-105.

Konishi, S., and Khatri, C.G. (1990). Inferences on interclass and intraclass correlations in nultivariate familial data. *Ann. Inst. Statist. Math.*, **42**, 561-580.

Konishi, S., and Shimizu, K. (1994). Maximum likelihood estimation of an intraclass correlation in a bivariate normal distribution with missing observations. *Commun. Statist. - Theory Methods*, **23**, 1593-1604.

Konno, Y. (1992). *Improved estimation of matrix of normal mean and eigenvalues in the multivariate F - distribution.* Doctoral Dissertation, Institute of Mathematics, University of Tsukuba.

Kubokawa, T. (1987). Admissible minimax estimation of a common mean of two normal populations. *Ann. Statist.*, **15**, 1245-1256.

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Ann. Statist.*, **22**, 290-299.

Kubokawa, T. (1995). Estimation of variance components in mixed linear models. *J. Multivariate Anal.*, **53**, 210-236.

Kubokawa, T. (1998). The Stein phenomenon in simultaneous estimation: A review. *Applied Statistical Science III* (eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha), NOVA Science Publishers, Inc., New York, to appear.

Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Commun. Statist. - Theory Methods*, **28**, 613-650.

Kubokawa, T. and Robert, C.P. (1994). New perspectives on linear calibration. *J. Multivariate Anal.*, **51**, 178-200.

Kubokawa, T., Robert, C. and Saleh, A.K.Md.E. (1992). Empirical Bayes estimation of the variance parameter of a normal distribution with unknown mean under an entropy loss. *Sankhya* (Ser. A), **54**, 402-410.

Kubokawa, T., Robert, C.P. and Saleh, A.K.Md.E. (1993). Estimation of noncentrality parameters. *Canad. J. Statist.*, **21**, 45-57.

Kubokawa, T., Saleh, A.K.Md.E., and Konno, Y. (1999). Bayes, minimax and nonnegative estimators of variance components under Kullback-Leibler loss. *Journal of Statistical Planning and Inference*, to appear.

Kubokawa, T. and Srivastava, M.S. (1999). Prediction in multivariate mixed linear models with equal replications. Discussion paper series, Faculty of Economics, University of Tokyo.

Kudo, A. (1963). A multivariate analogue of the sided test. *Biometrika*, **50**, 403-418.

Kuriki, S. and Takemura, A. (1997). James-Stein type estimator by shrinkage to closed convex set with smooth boundary. Discussion Paper 97-F-22, Faculty of Economics, University of Tokyo.

LaMotte, L.R.(1973). Invariant quadratic estimators in the random, one-way ANOVA model. *Biometrics* **32**, 793-804.

Lee, C.I.C. (1988). The quadratic loss of order restricted estimators for several treatment means and a control mean. *Ann. Statist.*, **16**, 751-758.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models. *J. Royal Statist. Soc. B*, **58**, 619-678.

Lehmann, E.L. (1983). *Theory of Point Estimation.* Wiley, New York.

Leonard, T., and Hsu, J.S.J. (1999). *Bayesian Methods.* Cambridge University Press.

Lin, P.-E. and Tsai, H.-L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.*, **1**, 142-145.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc., Ser. B*, **34**, 1-41.

Lindsey, J.K. (1993). *Models for Repeated Measurements.* Oxford Science Publications.

Manton, K.G., Woodbury, M.A., Stallard, E., Riggan, W.B., Creason, J.P. and Pellom, A.C. (1989). Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates. *J. Amer. Statist. Assoc.*, **84**, 637-650.

Marchand, E. (1994). On the estimation of the mean of a $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population with $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ known. *Statist. Probab. Lett.*, **21**, 69-75.

McCulloch, C.E. (1994). Maximum likelihood variance components estimation for binary data. *J. Amer. Statist. Assoc.*, **89**, 330-335.

Minami, M., and Shimizu, K. (1998). Estimation for a common correlation coefficient in bivariate normal distributions with missing observations. *Biometrics*, **54**, 1136-1146.

Morris, C.N. (1983). Parametric empirical Bayes inference: Theory and application. *J. Amer. Statist. Assoc.*, **78**, 47-65.

Patterson, H.D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545-554.

Patterson, H.D., and Thompson, R. (1975). Maximum likelihood estimation of components of variance. In *Proceedings Eight International Biometric Conference* (L.C.A. Corsten and T. Postelnica, eds), 197-207. Biometric Soc., Washington, D.C.

Peixoto, J.L. and Harville, D.A. (1986). Comparisons of alternative predictors under the balanced one-way random model. *J. Amer. Statist. Assoc.*, **81**, 431-436.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.

Rao, J.N.K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canad. J. Statist.*, **22**, 511-528.

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statist. Sci.*, **6**, 15-51.

Rosner, B. (1979). Maximum likelihood estimation of interclass correlations. *Biometrika*, **66**, 533-538.

Rukhin, A.L. (1992). Asymptotic risk behavior of mean vector and variance estimators and the problem of positive normal mean. *Ann. Inst. Statist. Math.*, **44**, 299-311.

Seshadri, V. (1963). Combining unbiased estimators. *Biometrics*, **19**, 163-170.

Shah, K.R. (1964). Use of inter-block information to obtain uniformly better estimators. *Ann. Math. Statist.*, **35**, 1064-1078.

Sengupta, D. and Sen, P.K. (1991). Shrinkage estimation in a restricted parameter space. *Sankhya, A*, **53**, 389-411.

Shinozaki, N.(1978). A note on estimating the common mean of k normal distributions and Stein problem. *Comm. Statist. A - Theory Methods* **7**, 1421-1432.

Shinozaki, N. and Chang, Y.-T. (1993). Minimaxity of empirical Bayes estimators of the means of independent normal variables with unequal variances. *Commun. Statist. - Theory Method*, **22**, 2147-2169.

Shinozaki, N. and Chang, Y.-T. (1996). Minimaxity of empirical Bayes estimators shrinking toward the grand mean when variances are unequal. *Commun. Statist. - Theory Method*, **25**, 183-199.

Shinozaki, N. and Chang, Y.-T. (1999). A comparison of maximum likelihood and best unbiased estimators in the estimation of kinear combinations of positive normal means. *Statist. and Decisions*, **17**, 125-136.

Srivastava, M.S. (1984). Estimation of interclass correlations in familial data. *Biometrika*, **71**, 177-186.

Srivastava, M.S., and Katapa, R.S. (1986). Comparison of estimators of interclass and intraclass correlations from familial data. *Canad. J. Statist.*, **14**, 29-42.

Srivastava, M.S. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. *Ann. Statist.*, to appear.

Stein, C.(1956). Some problems in multivariate analysis, Part I. Tech. Report. No. 6, Stanford University.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155-160.

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics* (Neyman Festschrift, F. N. David, ed.), 351-366. Wiley, New York.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symp. Asymptotic Statist.*, 345-381.

Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(In Russian)

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135-1151.

Strawderman. W.E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.*, **42**, 385-388.

Sun, L. (1996). Shrinkage estimation in two-way multivariate normal model. *Ann. Statist.*, **24**, 825-840.

Sutradhar, B.C. and Qu, Z. (1998). On approximate likelihood inference in a Poisson mixed model. *Canad. J. Statist.*, **26**, 169-186.

Tsutakawa, R.K. (1985). Estimation of cancer mortality rates: A Bayesian analysis of small frequencies. *Biometrics*, **41**, 69-70.

Tsutakawa, R.K. (1988). Mixed model for analyzing geographic variability in mortality rates. *J. Amer. Statist. Assoc.*, **83**, 37-42.

Tsutakawa, R.K., Shoop, G.L. and Marienfeld, C.J. (1985). Empirical Bayes estimation of cancer mortality rates. *Statist. Medicine*, **4**, 201-212.

Waclawiw, M.A. and Liang, K.-Y. (1994). Empirical Bayes estimation and inference for the random effects model with binary response. *Statist. Medicine*, **13**, 541-551.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross validated smoothing spline. *J. Royal Statist. Soc.* (Ser. B), **45**, 133-150.

Wahba, G. (1985). Comparison of GCV and GML for choosing the smoothing parameter in the generalized smoothing problem. *Ann. Statist.* , **13**, 1378-1402.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* 2nd Ed., Springer, New York.

Yanagimoto, T. and Yanagimoto, M. (1987). The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model. *Technometrics*, **29**, 95-101.

Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugenics*, **10**, 317-325.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.
Zellner A. (1983). *Applied Time Series Analysis of Economic Data.* U.S. Commerce, Washington, D.C.

参考文献（和文）

藤野和建 (1987). 正規分布の片側確率の推定. 応用統計学, **16**, 119-130.
広津千尋 (1976). 分散分析. 教育出版.
広津千尋 (1982). 離散データ解析. 教育出版.
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.
小西貞則, 本田正幸 (1992). 判別分析における誤判別率推定とブートストラップ法. 応用統計学, **21**, 67-100.
久保川達也 (1995). 縮小推定の理論と応用 (1). 経済学論集, **61**, 4 号, 2-31.
久保川達也 (1996). 縮小推定の理論と応用 (2). 経済学論集, **62**, 1 号, 41-61.
久保川達也，国友直人 (1996). 小地域推定の理論：分散成分モデルの利用.
丘本正 (1986). 因子分析の基礎. 日科技連出版社.
尾崎統，北川源四郎 (1998). 時系列解析の方法. 朝倉書店.
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.
竹内啓 (1979a). 現象と行動のなかの統計数理. 新曜社.
竹内啓 (1979b). Stein 推定量の意味とその応用. 応用統計学, **8**, 81-95.
竹内 啓 (1986). 因子分析モデルにおける最尤推定量の構造について. 応用統計学 **15**, 29-45.
丹後俊郎 (1988). 死亡指標の経験的ベイズ推定量について-疾病地図への適用-. 応用統計学 **17**, 81-96.
柳井晴夫，繁桝算男，前川眞一，市川雅教 (1990). 因子分析-その理論と方法-. 朝倉書店.