日本統計学会誌
第 37 巻，第 1 号，2007 年 9 月
151 頁～160 頁

縮小型推定量を用いた Universal Kriging Predictor

丸　山　祐　造*

A Universal Kriging Predictor Using Shrinkage Estimation

Yuzo Maruyama*

地球統計学において，観測された情報（観測値とその位置）に基づいて，未観測地点での値を予測する問題を考える．標準的には，最良線形不偏予測量が用いられるが，本論文では縮小型のベイズ予測量が，平均二乗誤差の意味で最良不偏予測量よりも良いことを示す．

For the problem of predicting values on unobserved points using observed information (values and locations), the best linear unbiased prediction is standard in geostatistics. In this paper, we show that a shrinkage Bayes prediction is superior to the best linear unbiased one in terms of mean squared error.
Key Words and Phrases: kriging, shrinkage method

1. イントロダクション

本論文では，地球統計学の中心的手法の一つであるクリギングについて通常仮定される不偏性の仮定を外し，縮小型推定量を用いた予測量を考える．縮小型推定量は James and Stein (1961) で提案され，当初人工的な推定量と考えられたが，Efron and Morris (1972) により経験ベイズ的解釈が与えられて以来，自然な推定量の一つと認識されている．予測問題においても，いくつかの典型的な問題設定において，縮小型の予測量が活躍することが分かっている (Stein, 1960; Baranchik, 1973; Copas, 1983; Gotway and Cressie, 1993)．その中でも，Gotway and Cressie (1993) はクリギングへの応用を意識しており，本論文では彼らの問題設定に沿って，縮小型推定量を用いた新たな予測量を考える．

まずクリギングの標準的な設定に従って，興味のある空間データを確率場 $Y(u), u \in D \subset R^d$ の実現値と考える．u は領域 D の中の位置を表し，次元 d は通常 2 または 3 である．確率場であるので，u 毎に $Y(u)$ は確率変数である．さらに空間データを解析するためには，いくつかの仮定を置いて $Y(u)$ をモデリングする必要がある．比較的緩い仮定として，全ての u について $E[Y(u)]$ と $\mathrm{Var}[Y(u)]$ が存在するとする．これより $Y(u)$ は，$m(u)$ を $E[Y(u)]=m(u)$ を満たす mean function，$e(u)$ を $E[e(u)]=0$ を満たす誤差項とし，

$$Y(u)=m(u)+e(u)$$

のように分解できる．さらに意味のある統計的推論を行うために，やや強い仮定を置く．まず $m(u)$ に線形構造を仮定する．具体的には $m(u)$ が p 個の既知の関数 $x_1(u),\dots, x_p(u)$ の線形

* 東京大学・空間情報科学研究センター：〒277-8568 千葉県柏市柏の葉 5-1-5

結合（ただし，$x_1(u)\equiv 1$ とする）

$$m(u)=\sum_{i=1}^{p}\beta_i x_i(u) \tag{1.1}$$

として書けるとする．また誤差項 $e(u)$ は平均が 0 の正規確率場であると仮定する．

本論文では，n 地点 $(u_1,\ldots,u_n)$ の観測値 $Y(u_1),\ldots,Y(u_n)$ が与えられているときに，未観測の k 地点 $(u_{01},\ldots,u_{0k})$ での値 $Y(u_{01}),\ldots,Y(u_{0k})$ を予測する問題を扱う．記述の簡単のため，$Z=(Y(u_1),\ldots,Y(u_n))'$, $Z_0=(Y(u_{01}),\ldots,Y(u_{0k}))'$, $\epsilon=(e(u_1),\ldots,e(u_n))'$, $\epsilon_0=(e(u_{01}),\ldots,e(u_{0k}))'$,

$$X=\begin{pmatrix} x_1(u_1) & \cdots & x_p(u_1) \\ \vdots & \ddots & \vdots \\ x_1(u_n) & \cdots & x_p(u_n) \end{pmatrix},\quad X_0=\begin{pmatrix} x_1(u_{01}) & \cdots & x_p(u_{01}) \\ \vdots & \ddots & \vdots \\ x_1(u_{0k}) & \cdots & x_p(u_{0k}) \end{pmatrix}$$

$\beta=(\beta_1,\ldots,\beta_p)'$ と書く．$e(v)$ が正規確率場であることから，ϵ と ϵ_0 の同時分布は，平均ベクトルが 0 の多変量正規分布となる．分散共分散行列は

$$\mathrm{Var}[(\epsilon',\epsilon_0')']=\begin{pmatrix} \Sigma_{ZZ} & \Sigma_{Z0} \\ \Sigma_{0Z} & \Sigma_{00} \end{pmatrix}\equiv\sigma^2\Sigma \tag{1.2}$$

（ここで Σ は既知の正定値行列で，σ^2 は未知の尺度母数）と仮定する．従って我々が扱う予測問題を整理すると，線形回帰モデル

$$Z=X\beta+\epsilon$$

から Z が観測値として得られているときに，線形回帰モデル

$$Z_0=X_0\beta+\epsilon_0$$

から発生する k 次元ベクトル Z_0 を観測値 Z の関数 $\delta(Z)=(\delta_1(Z),\ldots,\delta_k(Z))'$ で予測する問題を考えることに帰着する．予測の精度を評価する尺度として二乗誤差

$$L(\delta,Z_0)=\sigma^{-2}\sum_{i=1}^{k}(\delta_i(Z)-Z_{0,i})^2=\sigma^{-2}(\delta(Z)-Z_0)'(\delta(Z)-Z_0) \tag{1.3}$$

を採用し，その（Z と Z_0 に関する）期待値

$$r(\delta,Z_0;\beta,\sigma^2)=\frac{1}{\sigma^2}E[(\delta(Z)-Z_0)'(\delta(Z)-Z_0)] \tag{1.4}$$

で予測量を評価する．本論文では，(1.4) で定義される $r(\delta,Z_0;\beta,\sigma^2)$ を予測量 δ の期待予測誤差と呼ぶことにする．もし，任意の (β,σ^2) について $r(\delta_*,Z_0;\beta,\sigma^2)\le r(\delta,Z_0;\beta,\sigma^2)$ が成り立ち，ある (β,σ^2) について強い不等号が成り立つとき，δ_* は δ よりも良い予測量であると定義する．

本論文で扱う予測量のクラスは，β が既知という仮定の下で構成する最良予測量に推定量 β をプラグインするものである．2 節では，このタイプの予測量の期待予測誤差を評価する問題が，プラグインする推定量をある種の二乗損失関数で評価する推定問題に帰着することを示す．3 節では，その推定問題の正準形を与える．プラグインする推定量としては，一般化最小二乗推定量が最も自然であり，実際に不偏な推定量のクラスの中では最も良いことが分かる．しか

し$p \geq 3$のとき，不偏性の制約を外すと，一般化最小二乗推定量よりも良い推定量が存在することが知られている（スタイン現象）．4節では，ある種の縮小型推定量が一般化最小二乗推定量を改良するための十分条件を導く．5節では，その十分条件を満たす一般化ベイズ推定量のクラスを構築し，その中に非常に簡便な形を持つ縮小型ベイズ推定量が存在することを紹介する．6節では，今回の問題設定の欠点及び時系列で得られる空間データへの適用可能性について述べる．

2. 予測量のクラス

まずβが既知の場合に，線形予測量

$$\delta(Z)=BZ+c \tag{2.1}$$

（Bは$k \times n$の行列，cはk次元ベクトル）のクラスを考える．(1.4) に代入すると，その期待予測誤差は

$$\sigma^{-2}\{[(BX-X_0)\beta+c]'[(BX-X_0)\beta+c]\}+\mathrm{tr}(BZ_{ZZ}B'+\Sigma_{00}-2B\Sigma_{Z0}) \tag{2.2}$$

で与えられる．(2.2) をBとcに関して微分して0とおいた正規方程式を解くと，解は

$$B_{opt}=\Sigma_{0Z}\Sigma_{ZZ}^{-1}, \quad c_{opt}=(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\beta \tag{2.3}$$

となる．(2.3) で与えられる解が (2.2) の期待予測誤差を実際に最小化していることは容易に分かる (Stein, 1999)．従って，βが既知のもとでの最適な線形予測量は

$$\delta_1(Z)=B_{opt}Z+c_{opt}=\Sigma_{0Z}\Sigma_{ZZ}^{-1}Z+(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\beta \tag{2.4}$$

であり，最小化された期待予測誤差は，

$$r(\delta_1(Z), Z_0, \beta, \sigma^2)=\mathrm{tr}(\Sigma_{00}-\Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}) \tag{2.5}$$

となる．ところでβが既知という仮定は，mean function (1.1) が既知という仮定に対応している．mean function が既知の下での最良な予測量である (2.4) は，地球統計学の分野では simple kriging predictor と呼ばれる．

通常βは未知であるので，予測量の一般的なクラスとして (2.4) のβを推定量$\hat{\beta}$で置き換えた

$$\delta(Z,\hat{\beta})=\Sigma_{0Z}\Sigma_{ZZ}^{-1}Z+(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\beta} \tag{2.6}$$

が考えられる．ここで$\hat{\beta}$がβの不偏推定量ならば，

$$\begin{aligned}E[\delta(Z,\hat{\beta})]&=\Sigma_{0Z}\Sigma_{ZZ}^{-1}X\beta+(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)E[\hat{\beta}]\\&=X_0\beta=E[Z_0]\end{aligned}$$

であり，$\delta(Z,\hat{\beta})$ も不偏性を持つ予測量となることに注意する．

さて不偏性に重きを置く場合，一般化最小二乗推定量

$$\hat{\beta}_{GLS}=(X'\Sigma_{ZZ}^{-1}X)^{-1}X'\Sigma_{ZZ}^{-1}Z \tag{2.7}$$

で置き換えた

$$\delta(Z,\hat{\beta}_{GLS})=\Sigma_{0Z}\Sigma_{ZZ}^{-1}Z+(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\beta}_{GLS} \tag{2.8}$$

が最も自然だと考えられる．実は$\delta(Z,\hat{\beta}_{GLS})$は，不偏性を満たす線形予測量のクラスの中で，期待予測誤差を最小にすることが知られており，最良線形不偏予測量と呼ばれる．その期待予測誤差は

$$r(\delta(Z,\hat{\beta}_{GLS}),Z_0;\beta,\sigma^2)=\mathrm{tr}\{X'\Sigma_{ZZ}^{-1}XQ\}$$

（Qは下の（2.9）で与えられる）である．地球統計学の分野では，（2.8）はuniversal kriging predictorと呼ばれる．

本論文では不偏性にかかわらず，一般に任意の推定量$\hat{\beta}$をプラグインした予測量（2.6）の良さを議論したい．その目的のためにHarville（1985）によって得られた期待予測誤差に関する分解定理は非常に強力である．（Harville（1985）は，より一般的な設定での分解定理を与えている．）

定理 2.1 （Harville, 1985）．誤差項が正規分布で与えられるとする．このとき（2.6）で与えられる予測量の期待予測誤差は，

$$Q=(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)'(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X) \tag{2.9}$$

として

$$r(\delta,Z_0;\beta,\sigma^2)=\sigma^{-2}E[(\hat{\beta}-\beta)'Q(\hat{\beta}-\beta)]+\mathrm{tr}(\Sigma_{00}-\Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}) \tag{2.10}$$

で与えられる．

証明． 付録を参照． □

定理2.1の（2.10）に注目すると，第二項はβが既知の場合の期待予測誤差（2.5）に一致しており，予測量（2.6）に於ける$\hat{\beta}$の選び方に依存しない．つまり予測量（2.6）においてβを$\hat{\beta}$で置き換えることによる影響は，（2.10）の右辺の第一項のみに現れる．従って，正規性の下で（2.6）のクラスの予測量を扱う場合，Z_0の予測の良さを扱う問題は，βの推定量$\hat{\beta}$の良さを評価する問題に帰着する．具体的には，線形回帰モデル

$$Z=X\beta+\epsilon,\quad \epsilon\sim N_n(0,\sigma^2\Sigma_{ZZ}) \tag{2.11}$$

の回帰係数ベクトルβの推定問題を一般化二乗損失関数

$$L_Q(d,\beta)=\sigma^{-2}(d-\beta)'Q(d-\beta) \tag{2.12}$$

のもとで扱い，推定量$\hat{\beta}$の良さをリスク関数$\sigma^{-2}E[(\hat{\beta}-\beta)'Q(\hat{\beta}-\beta)]$で測る．予測問題における予測量$\delta(Z,\hat{\beta})$の期待予測誤差は（2.10）を通じて得られることに注意する．さて，よく知られているように，この種の推定問題において説明変数の個数pが3以上ならばスタイン現象が起こり，一般化最小二乗推定量β_{GLS}よりも（リスクを小さくするという意味で）良い推定量が存在する．本論文では，Maruyama and Strawderman（2005）に従って改良する十分条件及びベイズ推定量でその条件を満たすものを提案する．

3. 正準形

この節ではβの推定問題の正準形を与える．

線形回帰モデル（2.11）において，一般化最小二乗推定量$\hat{\beta}_{GLS}$と残差平方和$S=(Z-X\hat{\beta}_{GLS})'\Sigma_{ZZ}^{-1}(Z-X\hat{\beta}_{GLS})$が十分統計量である．また$\hat{\beta}_{GLS}$と$S$は互いに独立であり，その分布は

$$\hat{\beta}_{GLS} \sim N_p(\beta, \sigma^2(X'\Sigma_{ZZ}^{-1}X)^{-1}), \quad S \sim \sigma^2\chi^2_{n-p} \tag{3.1}$$

となる．X と $X_0-\Sigma_{0Z}\sigma_{ZZ}^{-1}X$ の階数がそれぞれ p であると仮定すると，$X'\Sigma_{ZZ}^{-1}X$ と Q は $p\times p$ の正定値行列である．このとき，$X'\Sigma_{ZZ}^{-1}X$ と Q を同時に対角化する M，つまり

$$M(X'\Sigma_{ZZ}^{-1}X)^{-1}M'=G=\mathrm{diag}(g_1,\ldots,g_p), \quad MM'=Q \tag{3.2}$$

（ここで $g_1\geq\cdots\geq g_p>0$）を満たす非特異行列 M が存在する．

$U=M'\hat{\beta}_{GLS}$ のように $\hat{\beta}_{GLS}$ を線形変換すると，U は多変量正規分布 $N_p(M'\beta, \sigma^2 G)$ に従う．また損失関数（2.12）は

$$(d-\beta)'Q(d-\beta)=(M'd-M'\beta)'(M'd-M'\beta) \tag{3.3}$$

と書ける．従って，線形変換 M により，前節の最後で述べた推定問題が見易くなり，統計量 U, S が互いに独立に

$$U\sim N_p(\theta, \sigma^2 G), \quad S\sim\sigma^2\chi^2_{n-p}$$

に従うとき，$\theta=M'\beta$ の推定問題を二乗損失関数

$$L_I(d,\theta)=\sigma^{-2}(d-\theta)'(d-\theta)$$

のもとで考えることに帰着する．θ の推定量 $\hat{\theta}$ から予測量に戻すには，

$$\Sigma_{0Z}\Sigma_{ZZ}^{-1}Z+(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)(M')^{-1}\hat{\theta} \tag{3.4}$$

とすればよい．

4. 改良する推定量のクラス

本節では，最良線形不偏推定量 $U=M'\hat{\beta}_{GLS}$ を改良するための十分条件を与える．改良する推定量のクラスとして，

$$\hat{\theta}=\left(I-\frac{S}{U'C^{-1}G^{-1}U}\phi\left(\frac{U'C^{-1}G^{-1}U}{S}\right)C^{-1}\right)U \tag{4.1}$$

（ここで C は対角行列 $\mathrm{diag}(c_1,\ldots,c_p)$ である）なる縮小推定量のクラスを考える．(4.1) に対して，以下の定理が成り立つ．

定理 4.1 (Maruyama and Strawderman, 2005)．(4.1) で与えられる推定量 $\hat{\theta}_\phi$ は $\phi'(w)\geq 0$ でかつ

$$0\leq\phi(w)\leq 2(n+2)^{-1}\left(\frac{\Sigma\{g_i/c_i\}}{\mathrm{Max}\{g_i/c_i\}}-2\right) \tag{4.2}$$

ならば，最良不偏推定量 U よりもリスクが小さい．

証明．付録を参照のこと． □

この種のリスクの改良に関する十分条件は，Baranchik (1970), Strawderman (1971), Efron and Morris (1976), Lehmann and Casella (1998) などでよく知られている．実際，定理 4.1 を実際に証明するための道具も Efron and Morris (1976) と同じである．ただし，既存論文と比べて新たな特徴が二つある．一つは (4.1) において $c_1,\ldots,c_p$ を自由に決められる点である．

全ての既存論文では，全て $c_1,\ldots,c_p$ が予め固有値 $g_1,\ldots,g_p$ の関数として縛られている縮小推定量のクラスを扱っている．もう一つは，二次形式 $U'C^{-1}G^{-1}U(=w)$ と各成分に対する縮小率 $1-\phi(w)/\{wc_i\}$ の組み合わせがベイズの視点から自然であることである．後者に関しては，次章においてベイズ推定量を導出する過程で明らかになる．前者について，数理統計学の優れた教科書である Lehmann and Casella (1998) では，$U\sim N_p(\theta,G)$ のときに損失関数 $(d-\theta)'(d-\theta)$ のもとで縮小推定量

$$(1-\phi(U'U)/U'U)U$$

が U を改良するための十分条件が「ϕ が単調非減少で，かつ $0\leq\phi\leq 2\Sigma_{i=1}^p\{g_i/g_1\}-4$ である」ことが紹介されている (Chapter 5 の Theorem 5.7)．しかし，$\Sigma_{i=1}^p g_i-2g_1\leq 0$ となる場合には，改良する推定量を提案出来ない．一方，我々の定理 4.1 では，例えば $c_i=g_i/g_p$ とすることにより，上限は $2(p-2)/(n-p+2)$ となる．つまり $p\geq 3$ であれば，どのような G に対しても，改良する推定量を提案することが出来るという利点がある．

5. ベイズ推定量

本節では，縮小型推定量 (4.1) の形を持つ一般化ベイズ推定量のクラスを構築する．事前分布として

$$\begin{aligned}&\theta|\lambda,\eta\sim N_p(0,\eta^{-1}G(\lambda^{-1}C-1)),\quad \text{for } \eta=\sigma^{-2},\\&\lambda\propto\lambda^a(1-\lambda)^b I_{(0,1)},\quad \eta\propto\eta^e\end{aligned}\tag{5.1}$$

を考える．これは，Strawderman (1971), Lin and Tsai (1973), Berger (1976, 1980), 及び Faith (1978) で提案された事前分布を一般化したものである．但し，任意の a,b,e に対して，一般化ベイズ推定量が定義されるわけではなく，以下の導出の過程で順次制約を加えていく．さて U, S, λ, η の周辺密度は，$w=u'C^{-1}G^{-1}u/s$ として

$$\begin{aligned}&\int\exp\left(-\frac{\eta}{2}\sum\left\{\frac{(u_i-\theta_i)^2}{g_i}+\frac{\lambda}{c_i-\lambda}\frac{\theta_i^2}{g_i}\right\}-\frac{\eta s}{2}\right)\\&\quad\times\eta^{p+n/2+e}\lambda^{p/2+a}\Pi(c_i-\lambda)^{-1/2}(1-\lambda)^b d\theta\\&=\int\exp\left(-\frac{\eta}{2}\sum\frac{c_i}{(c_i-\lambda)g_i}(\theta_i-(1-\lambda/c_i)u_i)^2-\frac{\eta}{2}\lambda\sum\frac{u_i^2}{c_ig_i}-\frac{\eta s}{2}\right)\\&\quad\times\eta^{p+n/2+e}\lambda^{p/2+a}\Pi(g_i^{-1/2}(c_i-\lambda)^{-1/2})(1-\lambda)^b d\theta\\&\propto\exp\left(-\frac{\eta s}{2}(1+\lambda w)\right)\eta^{p/2+n/2+e}\lambda^{p/2+a}(1-\lambda)^b\end{aligned}\tag{5.2}$$

に比例する．また二乗損失関数 L_I の下で，一般化ベイズ推定量は $E(\eta\theta|U,S)/E(\eta|U,S)$ で与えられる．$p/2+n/2+e+2>0$ のとき，η での積分に関して

$$\int_0^\infty\eta^{p/2+n/2+e+1}\exp\left(-\frac{\eta}{2}\lambda\sum\frac{u_i^2}{c_ig_i}-\frac{\eta s}{2}\right)d\eta\propto(1+\lambda w)^{-p/2-n/2-e-2}\tag{5.3}$$

となることに注意すると，一般化ベイズ推定量 $\hat{\theta}_{GB}$ が

$$\hat{\theta}_{GB}=\left(1-\frac{E(\lambda\eta|U,S)}{E(\eta|U,S)}C^{-1}\right)U=\left(I-\frac{\phi_{GB}(W)}{W}C^{-1}\right)U$$

(ここで

$$\phi_{GB}(w)=w\frac{\int_0^1\lambda^{p/2+a+1}(1-\lambda)^b(1+w\lambda)^{-p/2-n/2-e-2}d\lambda}{\int_0^1\lambda^{p/2+a}(1-\lambda)^b(1+w\lambda)^{-p/2-n/2-e-2}d\lambda} \tag{5.4}$$

である.）と表されることが分かる．ただし，(5.4) の分母分子の積分が定義されるために $a>-p/2-1$ 及び $b>-1$ を仮定する.

ところで (5.4) の分母と分子の積分の型において，$t=(1+w)\lambda/(1+w\lambda)$ なる変数変換を考えると，

$$\int_0^1\lambda^\alpha(1-\lambda)^\beta(1+w\lambda)^{-\gamma}d\lambda=\frac{1}{(w+1)^{\alpha+1}}\int_0^1 t^\alpha(1-t)^\beta\left\{1-\frac{tw}{w+1}\right\}^{-\alpha-\beta+\gamma-2}dt$$

なる等式を得る．これを (5.4) の分母分子に適用すると，$\phi_{GB}(w)$ の別表現

$$\phi_{GB}(w)=\frac{w}{1+w}\frac{\int_0^1 t^{p/2+a+1}(1-t)^b\{1-tw/(w+1)\}^{n/2+e-a-1-b}dt}{\int_0^1 t^{p/2+a}(1-t)^b\{1-tw/(w+1)\}^{n/2+e-a-b}dt} \tag{5.5}$$

が得られる．特に (5.5) において $b=n/2-a+e-1$ の場合を考えると，$\alpha=(p/2+a+1)/(b+1)=(p/2+a+1)/(n/2+e-a)$ として，

$$\begin{aligned}\phi_{GB}(w)&=\frac{w}{w+1}\frac{B(p/2+a+2,\ b+1)}{B(p/2+a+1,\ b+1)-\{w/(w+1)\}B(p/2+a+2,\ b+1)}\\&=\frac{w}{(w+1)(1+1/\alpha)-w}=\frac{\alpha w}{\alpha+1+w}\end{aligned} \tag{5.6}$$

となる．つまり $\phi_{GB}(w)$ は著しく簡潔な形に帰着し，簡潔な一般化ベイズ推定量

$$\hat{\theta}_{GB}=\left(I-\frac{\alpha}{\alpha+1+W}C^{-1}\right)U \tag{5.7}$$

が得られる．推定量 $\hat{\theta}_{GB}$ は，事前分布のモードであるゼロベクトル $0(\alpha\to\infty)$ と最良線形不偏推定量 $U(\alpha\to 0)$ の重みつき平均である．その U への重みは，固定した α のもとで 0 と U の二次形式 W で測った距離が遠い程重く，$W\to\infty$ のとき推定量は U に収束することが分かる．また固定した W のもとで，α が大きい程 0 への重みが重いことも分かる.

さて明らかに $\phi_{GB}(w)$ は w に関して単調増加関数であり，$w\to\infty$ のとき α に収束する．定理 4.1 と組み合わせると，次の結果が得られる.

系 5.1. α 及び $c_1,\ldots,c_p$ を

$$0<\alpha\leq\frac{2}{n+2}\left(\frac{\sum\{g_i/c_i\}}{\max\{g_i/c_i\}}-2\right) \tag{5.8}$$

を満たすように選べば，一般化ベイズ推定量 $\hat{\theta}_{GB}$ は U よりも良い推定量である.

先に書いたように $c_i=g_i/g_p$ とすれば，(5.8) の上限は $2(p-2)/(n+2)$ となる．従って，$\hat{\theta}_{GB}$ において $c_i=g_i/g_p,\ \alpha=(p-2)/(n+2)$ とした推定量

$$\left(I-\frac{g_p(p-2)}{n+p+(n+2)g_pU'G^{-2}U}G^{-1}\right)U \tag{5.9}$$

は，Uよりもリスクが小さい推定量になる．本論文では，(5.9) を (3.4) に代入した予測量を，期待予測誤差が最良線形不偏予測量 $\delta(Z,\hat{\beta}_{GLS})$ よりも小さい予測量として推奨する.

6. 考察

前節までの結果において，分散共分散行列 Σ を既知とする仮定が，実用上は都合が悪い．実際には分散共分散行列が既知であることはまれであり，データから推定する必要がある．残念ながら，推定された分散共分散行列をプラグインした場合に，既知の場合と対応する結果は一般に成立しない．Harville (1985) は，最良線形不偏予測量に推定した分散共分散行列をプラグインした予測量が不偏であるための十分条件を与えた．しかし，定理 4.1 で与えられた改良するための十分条件のような理論的な結果を導くのは，非常に困難であると思われる．従って，本論文で提案する縮小型予測量に推定された分散共分散行列をプラグインしたときの数値的評価は，理論的な結果のサポートとして必要であり，現在研究中である．Stein (1999) は，推定した分散共分散行列をプラグインした予測量の理論的性質（主に漸近的な性質）を丁寧に議論している．

また，(4.1) や (5.7) において原点方向に縮小することに関して疑問があるかもしれない．しかし，近付ける点は理論上 0 でなくても任意の点 θ_0 で構わないのである．具体的には，$W=\{(U-\theta_0)'C^{-1}G^{-1}(U-\theta_0)\}/S$ として

$$\hat{\theta}_\phi=(I-\{\phi(W)/W\}C^{-1})(U-\theta_0)+\{\phi(W)/W\}C^{-1}\theta_0 \tag{6.1}$$

や

$$\hat{\theta}_{GB}=\left(I-\frac{\alpha}{\alpha+1+W}C^{-1}\right)(U-\theta_0)+\frac{\alpha}{\alpha+1+W}C^{-1}\theta_0 \tag{6.2}$$

としても，定理 4.1 や系 5.1 は成立する．特にデータが時系列に得られている場合，θ_0 として前期の推定量を選べば，疑似カルマン=フィルタ推定量が得られる．この具体的な手続きを書き下して，本稿を締めることにする．

右下の添字 t は t 期を表し，Info_{t-1} は $t-1$ 期での全ての情報（β_{t-1} と η_{t-1} の事後分布）を表すことにすると，(2.11)，(5.1) はそれぞれ

$$Z_t=X_t\beta_t+\epsilon_t,\ \epsilon_t\sim N(0,\eta_t^{-1}I)$$
$$\beta_t|\eta_t,\lambda,\mathrm{Info}_{t-1}\sim N\left(\frac{E[\eta_{t-1}\beta_{t-1}|Z_{t-1},X_{t-1}]}{E[\eta_{t-1}|Z_{t-1},X_{t-1}]},\eta_t^{-1}(M_t')^{-1}G(\lambda^{-1}C-1)M_t^{-1}\right)$$
$$\lambda\sim\lambda^a(1-\lambda)^b,\ \eta_t\sim\eta_t^e$$

と書き直すことが出来る．また β の推定量は，

$$\cdots\rightarrow\frac{E[\eta_{t-1}\beta_{t-1}|Z_{t-1},X_{t-1}]}{E[\eta_{t-1}|Z_{t-1},X_{t-1}]}(=\hat{\beta}_{t-1}^{GB})\rightarrow\frac{E[\eta_t\beta_t|Z_t,X_t]}{E[\eta_t|Z_t,X_t]}(=\hat{\beta}_t^{GB})\rightarrow\cdots$$

のように更新されていく．通常のカルマン=フィルタと違う点は，前期の情報は $E[\eta_{t-1}\beta_{t-1}|Z_{t-1},X_{t-1}]/E[\eta_{t-1}\beta_{t-1}|Z_{t-1},X_{t-1}]=\hat{\beta}_{t-1}^{GB}$ だけに集約し，t 期の事前分布の平均として利用すること，結果として，$\hat{\beta}_{t-1}^{GB}$ と $\hat{\beta}_t^{GLS}$ の重みつき平均として表される t 期の推定量 $\hat{\beta}_t^{GB}$ は，その重みが前期（$t-1$ 期）の情報に依存せず，$\hat{\beta}_t^{GLS}$ のリスクを改良するように事前分布のパラメータを調整することにより配分されること，である．

付録

定理 2.1 の証明. まず誤差項の正規性の下では

$$E(Z_0|Z)=\Sigma_{0Z}\Sigma_{ZZ}^{-1}Z+(X_0-\Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\beta\equiv\delta(Z,\beta) \tag{A.1}$$

が成り立つことに注意する．期待予測誤差の σ^2 倍を展開すると

$$\begin{aligned}E[\|\delta(Z,\hat\beta)-Z_0\|^2]&=E[\|\delta(Z,\hat\beta)-\delta(Z,\beta)\|^2]+E[\|\delta(Z,\beta)-Z_0\|^2]\\&\quad+2E[\{\delta(Z,\hat\beta)-\delta(Z,\beta)\}'\{\delta(Z,\beta)-Z_0\}]\\&=E[(\hat\beta-\beta)'Q(\hat\beta-\beta)]+\mathrm{tr}\{\mathrm{Var}(Z_0|Z)\}\\&\quad+2E_Z[\{\delta(Z,\hat\beta)-\delta(Z,\beta)\}'\{E_{Z_0|Z}[E(Z_0|Z)-Z_0]\}]\\&=E[(\hat\beta-\beta)'Q(\hat\beta-\beta)]+\sigma^2\mathrm{tr}(\Sigma_{00}-\Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0})\end{aligned}$$

となる（期待値符号 E は，添字がない場合 Z と Z_0 の両方について期待値を取ることを意味する）．交差項は（A.1）より 0 となる．□

定理 4.1 の証明. $\hat\theta_\phi$ のリスクは

$$\begin{aligned}R(\theta,\sigma^2,\hat\theta_\phi)&=E[(\hat\theta_\phi-\theta)'(\hat\theta_\phi-\theta)/\sigma^2]\\&=R(\theta,\sigma^2,U)+E\left[\frac{S^2}{\sigma^2}\frac{\Sigma\{U_i^2/(c_i^2)\}}{(\Sigma\{U_i^2/(c_ig_i)\})^2}\phi^2\left(\frac{\Sigma\{U_i^2/(c_ig_i)\}}{S}\right)\right]\\&\quad-2E\left[\sum\frac{S}{\sigma^2}\frac{U_i}{c_i}(U_i-\theta_i)\sum\frac{U_i^2}{c_ig_i}\phi\left(\frac{\Sigma\{U_i^2/(c_ig_i)\}}{S}\right)\right]\end{aligned} \tag{A.2}$$

と表される．以下では $W=U'C^{-1}G^{-1}U/S$ と書く．（A.2）の第二項は，$S\sim\sigma^2\chi_n^2$ に対して成り立つカイ二乗等式

$$E[Sg(S)]=\sigma^2E[ng(S)+2Sg'(S)] \tag{A.3}$$

を用いると，

$$\begin{aligned}&E\left[\frac{U'C^{-2}U}{(U'C^{-1}G^{-1}U)^2}\frac{S}{\sigma^2}\left\{S\phi^2\left(\frac{U'C^{-1}G^{-1}U}{S}\right)\right\}\right]\\&=E\left[\frac{U'C^{-2}U}{(U'C^{-1}G^{-1}U)^2}\{nS\phi^2(W)+2S\phi^2(W)-4\phi(W)\phi'(W)U'C^{-1}G^{-1}U\}\right]\\&=E\left[\frac{U'C^{-2}U}{U'C^{-1}G^{-1}U}\left\{(n+2)\frac{\phi^2(W)}{W}-4\phi(W)\phi'(W)\right\}\right]\end{aligned}$$

のように（ ）の中に未知母数 θ，σ^2 を含まない形に書ける．（A.2）の第三項については，$U\sim N(\theta_i,g_i\sigma^2)$ のときに成り立つスタイン等式

$$E[(U_i-\theta_i)h(U)]=g_i\sigma^2E[(\partial/\partial U_i)h(U)] \tag{A.4}$$

を用いて

$$\begin{aligned}&\sum E\left[\frac{1}{c_i\sigma^2}(U_i-\theta_i)U_i\left(\frac{\Sigma\{U_i^2/(c_ig_i)\}}{S}\right)^{-1}\phi\left(\frac{\Sigma\{U_i^2/(c_ig_i)\}}{S}\right)\right]\\&=\sum E\left[\frac{g_i}{c_i}\left\{W^{-1}\phi(W)+2\frac{U_i^2}{c_ig_iS}(W^{-1}\phi'(W)-W^{-2}\phi(W))\right\}\right]\end{aligned}$$

$$=E\left[\sum\frac{g_i}{c_i}\frac{\phi(W)}{W}+2\frac{U'C^{-2}U}{S}\left\{\frac{\phi'(W)}{W}-\frac{\phi(W)}{W^2}\right\}\right]$$

のように，やはり未知母数を含まない形に書ける．$\phi'(w)\geq 0$ に注意すると，

$$\begin{aligned}
&R(\theta,\sigma^2,\hat{\theta}_\phi)\\
&\leq R(\theta,\sigma^2,\hat{\theta}_\phi)+E\left[\frac{\phi(W)}{W}\frac{U'C^{-2}U}{U'C^{-1}G^{-1}U}\left\{(n+2)\phi(W)-2\sum\frac{g_i}{c_i}\frac{U'C^{-1}G^{-1}U}{U'C^{-2}U}+4\right\}\right]\\
&\leq R(\theta,\sigma^2,U)+E\left[\frac{\phi(W)}{W}\frac{U'C^{-2}U}{U'C^{-1}G^{-1}U}\left\{(n+2)\phi(W)-2\frac{\Sigma\{g_i/c_i\}}{\max\{g_i/c_i\}}+4\right\}\right]\\
&\leq R(\theta,\sigma^2,U)
\end{aligned}$$

となる．　□

参考文献

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **41**, 642–645.

Baranchik, A. J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables, *Ann. Statist.*, **1**, 312–321.

Berger, J. O. (1980). A robust generalized Bayes estimator and con.dence region for a multivariate normal mean, *Ann. Statist.,* **8**(4), 716–761.

Berger, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.*, **4**(1), 223–226.

Copas, J. B. (1983). Regression, prediction and shrinkage, *J. Roy. Statist. Soc. Ser. B*, **45**(3), 311–354.

Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case, *J. Amer. Statist. Assoc.*, **67**, 130–139.

Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution, *Ann. Statist.*, **4**(1): 1121.

Faith, R. E. (1978). Minimax Bayes estimators of a multivariate normal mean, *J. Multivariate Anal.*, **8**(3), 372–379.

Gotway, C. A. and Cressie, N. (1993). Improved multivariate prediction under a general linear model, *J. Multivariate Anal.*, **45**(1), 56–72.

Harville, D. A. (1985). Decomposition of prediction error, *J. Amer. Statist. Assoc.*, **80**(389), 132–138.

James, W. and Stein, C, (1961). Estimation with quadratic loss, *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 361–379, Univ. California Press, Berkeley, Calif.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

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.

Maruyama, Y. and Strawderman, W. E. (2005). A new class of generalized Bayes minimax ridge regression estimators, *Ann. Statist.*, **33**(4): 1753–1770.

Stein, C. (1960). Multiple regression, *Contributions to probability and statistics*, 424–443, Stanford Univ. Press, Stanford, Calif..

Stein, M. L. (1999). *Interpolation of spatial data*, Springer Series in Statistics. Springer- Verlag, New York, .

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.*, **42**(1), 385–388.