多変量小地域推定における信頼領域について

伊藤　翼　　久保川　達也
統計数理研究所　統計思考院　　東京大学大学院　経済学研究科

小地域推定問題において, 線形混合モデルから導出される経験最良線形不偏予測量 (EBLUP) は, 小地域の平均値に対する安定した予測量を与える. EBLUP の不確実性を測る指標として, 平均二乗予測誤差の 2 次不偏推定量の導出のみが多くの研究で行われてきたそこで, 我々は変量効果の共分散行列の構造が完全に未知であるといった設定の下で, 多変量 Fay-Herriot モデルを考え, 信頼水準を 2 次のオーダーで達成する信頼領域を EBLUP を中心としたマハラノビス距離にもとづいて構成した.

地域レベルの集計データ $(\boldsymbol{y}_1, \boldsymbol{X}_1), \ldots, (\boldsymbol{y}_m, \boldsymbol{X}_m)$ が観測されている, ただし, m は地域数であり, $\boldsymbol{y}_i$ は k-次元の集計値, $\boldsymbol{X}_i$ は $k \times s$ の共変量行列である. 小地域の平均の真値を $\boldsymbol{\theta}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{v}_i, i = 1, \ldots, m$ とすると, Fay (1987) で提案された多変量 Fay-Herriot モデルは,

$$\boldsymbol{y}_i \mid \boldsymbol{\theta}_i \sim \mathcal{N}_k(\boldsymbol{\theta}_i, \boldsymbol{D}_i), \quad \boldsymbol{\theta}_i \sim \mathcal{N}_k(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}).$$

と記述される. $\boldsymbol{\theta}_a$ の信頼領域の構成が主題である. $\widehat{\boldsymbol{\Psi}}$ を $\boldsymbol{\Psi}$ の一致推定量とすれば, $\boldsymbol{\theta}_a$ の EBLUP は $\widehat{\boldsymbol{\theta}}_a^{EB} = \boldsymbol{y}_a - \boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a - \boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})\}$ となる. ここで, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) = \{\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{y}$ である. このとき, $\widehat{\boldsymbol{\theta}}_a^{EB}$ の MSE 行列は,

$$\begin{aligned}
&\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\\
=&E[\{\widehat{\boldsymbol{\theta}}_a^* - \boldsymbol{\theta}_a\}\{\widehat{\boldsymbol{\theta}}_a^* - \boldsymbol{\theta}_a\}^\top] + E[\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*\}^\top] + E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top]\\
=&\boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + O(m^{-3/2}).
\end{aligned}$$

と 2 次近似できる. B_1, B_2, B_3 を以下で定義する.

$$\begin{aligned}
B_1 =&B_1(\boldsymbol{\Psi}) = -\frac{1}{2}\mathrm{tr}\left(E[(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))^2]\right),\\
B_2 =&B_2(\boldsymbol{\Psi}) = -\frac{1}{8}\left\{E[\mathrm{tr}^2(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))] + 2\mathrm{tr}\left(E[(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))^2]\right)\right\},\\
B_3 =&B_2(\boldsymbol{\Psi}) = \mathrm{tr}\,(\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{G}_{3a}(\boldsymbol{\Psi})).
\end{aligned}$$

ただし, $\boldsymbol{H}_a(\boldsymbol{\Psi}) = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi})$, $\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}) = \boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})$ である. このとき, 自由度 k のカイ二乗分布の分布関数と密度関数をそれぞれ $F_k(x)$, $f_k(x)$ とすれば, マハラノビス距離 $(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top\boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)$ の漸近分布は, $F_k(x) + 2(B_1 - B_2 - B_3)f_{k+2}(x) + 2B_2f_{k+4}(x)$ と 2 次近似される. この漸近分布を用いてバートレット補正を行えば, 補正項は $h^*(\boldsymbol{\Psi}) = -2\{(B_1 - B_3 - B_2)/k + B_2x/k(k+2)\}$ となる. よって, 信頼水準 α を 2 次のオーダーで達成する $\boldsymbol{\theta}_a$ の信頼領域は,

$$CR = \{\boldsymbol{\theta}_a \mid (\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a^{EB})^\top\boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a^{EB}) \leq \{1 + h^*(\widehat{\boldsymbol{\Psi}})\}\chi^2_{k,\alpha}\}$$

となる.