

小地域推定におけるベイズ法

東京大・経済 久保川 達也

標本調査において、通常は調査区全体の特性を調べるために標本調査が行われるが、そのデータを利用して地域ごとの特性値を推定したい状況がしばしば生ずる。そのとき、狭い地域や人口が粗な地域に対しては十分なデータがとられていないため、その地域のデータだけでは特性値の十分な推測ができない。このような状況での推定問題は特に小地域推定と呼ばれる。

小地域推定において用いられる代表的なモデルは枝分かれ誤差分散モデル (NERM) である。これを有限母集団の枠組みで記述すると

$$Y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, N_i \quad (1)$$

と表される。 m が地域の個数、 N_i が地域 i の母集団の個体の個数、 \mathbf{x}_{ij} が共変量で N_i 個の個体についてすべて観測されているものとする。 v_i は地域効果として地域の差異を表すものとし、これを $v_i \sim \mathcal{N}(0, \sigma_v^2)$ に従う変量効果とし、 ε_{ij} は $\mathcal{N}(0, \sigma_e^2)$ に従う誤差項とする。各地域からは $\{Y_{ij}\}_{j=1, \dots, N_i}$ の一部が抽出され、一般性を失うことなく y_{i1}, \dots, y_{in_i} が観測されたものとする。推定したい特性値は母集団平均 $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$ であり、これを $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$ に基づいて推定することになる。 $\bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ とおくと、 \bar{Y}_i は

$$\bar{Y}_i = \frac{n_i}{N_i} \bar{y}_i + \frac{1}{N_i} \sum_{j=n_i+1}^{N_i} Y_{ij}$$

と書けるので、観測された項と観測されない項に分けられる。

v_i が地域で共通の変量として入っていることにより、 $j \neq j'$ に対して Y_{ij} と $Y_{ij'}$ の間には相関関係が入ることになるので、この相関を利用して \mathbf{y}_i に基づいて観測できない変量 Y_{ij} を条件付き期待値 $E[Y_{ij} | \mathbf{y}_i]$ により予測することができる。こうして得られる予測量は次のように書かれる。

$$\hat{\mu}_i^B(\boldsymbol{\beta}, \sigma_v^2, \sigma_e^2) = \frac{n_i}{N_i} \bar{y}_i + \frac{1}{N_i} \sum_{j=n_i+1}^{N_i} \left\{ \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \frac{n_i \sigma_v^2}{\sigma_e^2 + n_i \sigma_v^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}) \right\}$$

モデル (1) は固定効果 $\boldsymbol{\beta}$ と変量効果 v_i とから成る混合効果モデルで、線形混合モデルとして一般化される。全地域に共通の回帰係数 $\boldsymbol{\beta}$ を用いることにより全データのプーリングがなされ、その結果 $\boldsymbol{\beta}$ の安定した推定値が得られる。また地域効果 v_i を変量として扱うことにより縮小効果が生まれ、 \bar{y}_i は安定した推定値の方向へ縮小され、その結果、地域の差異を考慮しながら推定精度を改善する推定手法になる。

地域の差異 v_i を変量として扱うことはベイズの枠組みで捉えることができる。ベイズ流の考え方の良さは、現象を説明するためにモデルに知的で豊かな構造を組み入れ、しかも実行可能なアルゴリズムとして MCMC を用いることができる点である。小地域推定においても様々なベイズモデルが考案され、(1) σ_e^2 に逆ガンマ分布を想定して分散の異質性を組み入れた変量分散モデル、(2) 地域の誤差分散の異質性を分散関数により説明するモデル、(3) 変量効果 v_i を変量として扱うか、0 に退化したものとして扱うかを確率的に組み込んだ不確実的 NERM、(4) 所得・支出などの正のデータに線形混合モデルを適用させるための変換モデル、(5) 平均と分散のベンチマーク問題など、これまでに考えられてきたベイズモデルや、小地域問題から発生するいくつかの問題を解説する。