自然指数型分布族による空間重み付き経験ベイズ推定

東京大・経済・院　菅澤 翔之助
東京大・経済・院　川久保友超
東京工業大　社会理工　小笠原浩太

1　はじめに

十分な情報 (サンプルサイズ) が得られない地域の平均を推定する際, 単純な算術平均では分散が大きくなってしまう問題がある. そのような問題への統計的アプローチとして経験ベイズ法がある. カウントデータを含む幅広いデータに対して使用可能なモデルとして以下のようなものがある.

$$y_i|\theta_i \sim f(y_i|\theta_i) = \exp[n_i(\theta_i y_i - \psi(\theta_i)) + c(y_i, n_i)],$$
$$\theta_i \sim \pi(\theta_i|\nu, m_i) = \exp[\nu(m_i\theta_i - \psi(\theta_i))]C(\nu, m_i),$$

ここで μ_i を $\mu_i = E[y_i|\theta_i] = \psi'(\theta_i)$ と定義する. また $Var(y_i|\theta_i) = \psi''(\theta_i)/n_i = Q(\mu_i)/n_i$ という構造を考える. ただし $Q(x)$ は二次関数 $Q(x) = v_0 + v_1 x + v_2 x^2$ である. また $m_i = \psi'(\boldsymbol{x}'\boldsymbol{\beta})$ という構造を入れることで共変量があるケースも扱えるようになる. このモデルにおいて μ_i のベイズ推定量は

$$\widehat{\mu}_i^{\mathrm{B}} = \frac{n_i y_i + \nu m_i}{n_i + \nu}$$

となる. この推定量は未知パラメータ $\boldsymbol{\eta} = (\boldsymbol{\beta}^t, \nu)^t$ に依存しているので, これらを周辺尤度から以下のように推定する.

$$\widehat{\boldsymbol{\eta}} = \mathrm{argmax}_{\boldsymbol{\eta}} \sum_{i=1}^{m} \left[\log C(\nu, m_i) - \log C(n_i + \nu, \widehat{\mu}_i^{\mathrm{B}})\right].$$

この $\widehat{\boldsymbol{\eta}}$ をベイズ推定量 $\widehat{\mu}_i^{\mathrm{B}}$ に代入することで経験ベイズ推定量 $\widehat{\mu}_i^{\mathrm{EB}}$ を得る.

2　空間重み付き経験ベイズ推定

以上のパラメータ推定では空間異質性 (非定常性) を考慮していない. 地理的な情報が付随して得られているケースではその情報も反映させた方がより良い推定を行うことができると考えられる. そこで Sugasawa, Kawakubo and Ogasawara (2015) では $\boldsymbol{\eta}$ が空間的に滑らかに変化していると想定して, 各地点におけるそれらの値を推定する手法を提案した. それは以下のようにパラメータを推定するものである.

$$\widehat{\boldsymbol{\eta}}_i = \mathrm{argmax}_{\boldsymbol{\eta}} \sum_{k=1}^{m} w(u_i, u_k) \left[\log C(\nu, m_k) - \log C(n_k + \nu, \widehat{\mu}_k^{\mathrm{B}})\right].$$

ここで $w(u_i, u_k)$ は地点 u_k の尤度が地点 u_i の推定に与えるウエイト関数である. このように推定することでハイパーパラメータの空間非定常性を考慮した経験ベイズ推定を行うことができる. 本報告では, さらに, 提案した経験ベイズ推定量の MSE の評価およびその推定量を紹介し, シミュレーションおよび実データ解析を通して有用性を示す.

参考文献

[1] Sugasawa, S., Kawakubo, Y. and Ogasawara, K. (2015). Geographically weighted empirical Bayes estimation via natural exponential family.