# Bayesian generalized lasso 正則化を用いた空間パラメータモデルに対するモデル選択

書川　侑子　　総合研究大学院大学 統計科学専攻 博士課程(3年次編入) 4年

## 1. SVCモデル (空間パラメータモデル)

$$y_i = \sum_{m=1}^{l} I(\psi_i = m)\boldsymbol{x}_i^{(m)^T}\boldsymbol{\theta} + \epsilon_i$$

- 目的変数ベクトル　$\boldsymbol{y} = (y_1, \ldots, y_n)^T$
- $l$個の地域のうち地域$m$に対する説明変数ベクトル　$\boldsymbol{x}_1^{(m)}, \ldots, \boldsymbol{x}_n^{(m)}$
- 変数の種類の個数 $p_1$
- 回帰係数ベクトル $\boldsymbol{\theta} = (\theta_{1,1}, \ldots, \theta_{1,p_1}, \theta_{2,1}, \ldots, \theta_{2,p_1}, \ldots, \theta_{l,1}, \ldots, \theta_{l,p_1})^T$
- 誤差　$\epsilon_i \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \sigma^2)$
- 指示関数 $I(\cdot)$
- $(y_i, \boldsymbol{x}_i)$が地域$m$で観測される時 $\psi_i = m$をとる確率変数 $\psi_i \overset{\text{i.i.d}}{\sim} \mathrm{Multi}(1, l)$

## 2. Bayesian generalized lasso正則化 (Zhao and Bondell, 2020, Kyung et al., 2010)

隣り合う地域に属する回帰係数の差に対してラプラス分布を仮定することで、地域の近さによる変数の類似性を予測に反映させる

- $E$: 全ての回帰係数を頂点とし，隣接する回帰係数間に辺を持つグラフ$G=(V,E)$の辺集合

$(j, k) \in E$ に対し

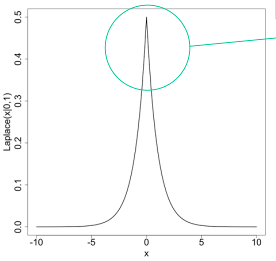
①回帰係数の種類に関わらず同じパラメータを持つラプラス分布を仮定

$$\theta_{r,j} - \theta_{r,k} \sim \mathrm{Laplace}(\lambda),\ r = 1, \ldots, p_1$$

②回帰係数の種類ごとに異なるパラメータを持つラプラス分布を仮定

$$\theta_{r,j} - \theta_{r,k} \sim \mathrm{Laplace}(\lambda_r),\ r = 1, \ldots, p_1$$

モデルの複雑さを変えると
事前分布のハイパーパラメータ数が変化



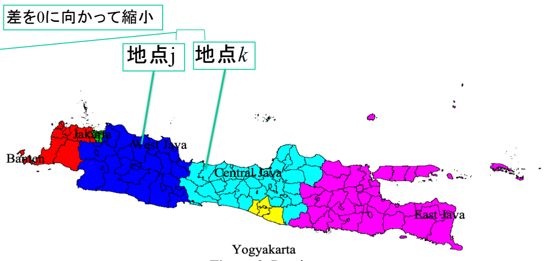




Figure 3. Provinces

(Rahardiantoro and Sakamoto, 2021)

ラプラス分布

$$\mathrm{Laplace}(\lambda) : \frac{1}{2\lambda}\exp\left(-\frac{|x|}{\lambda}\right)$$

## 3. 事前分布強調型情報量規準PIICの適用

$$z_1, z_2, \ldots, z_n, \tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n \overset{\text{i.i.d.}}{\sim} f(\cdot \mid \boldsymbol{\theta}), \qquad \boldsymbol{\theta} \sim \pi_n(\cdot; \boldsymbol{\xi})$$

$z_i$: 現在得られている$i$番目のサンプル, $z_i = (y_i, \boldsymbol{x}_i, \boldsymbol{\psi_i})$ , $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$
$\tilde{z}_i$: 将来のデータを表す$z_i$のコピー
$\boldsymbol{\theta} \in \mathbb{R}^p$: パラメータ, $p = lp_1$
$\boldsymbol{\xi} \in \mathbb{R}^q$: ハイパーパラメータ
$\boldsymbol{\theta}$の事後分布に対する期待値を$E_{\boldsymbol{\theta}|\boldsymbol{z};\boldsymbol{\xi}}(\cdot)$とし，ベイズ予測分布を$f(z \mid \boldsymbol{z}, \boldsymbol{\xi}) := E_{\boldsymbol{\theta}|\boldsymbol{z};\boldsymbol{\xi}}\{f(z \mid \boldsymbol{\theta})\}$ とする

WAIC (Watanabe, 2010):
ベイズ予測分布と真の分布間のKLダイバージェンスから定数を差し引いたものの漸近評価

$$-\sum_{i=1}^{n}\log \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{z};\boldsymbol{\xi}}\{f(z_i \mid \boldsymbol{\theta})\} + \sum_{i=1}^{n}\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{z};\boldsymbol{\xi}}\left[\{\log f(z_i \mid \boldsymbol{\theta})\}^2\right] - \sum_{i=1}^{n}\left[\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{z};\boldsymbol{\xi}}\{\log f(z_i \mid \boldsymbol{\theta})\}\right]^2$$

事前分布に対する罰則項がない
→事前分布が複雑なものを必ず選ぶ
→WAICでは最も複雑なモデルが必ず選ばれ，オーバーフィットする

解決策: Prior Intensified Information Criterion (Ninomiya, Y., 2021)

①事前分布のオーダーを$O(n)$とする　$\pi_n(\boldsymbol{\theta}, \boldsymbol{\xi}) \propto \pi(\boldsymbol{\theta}, \boldsymbol{\xi})^{n/n_0}$

$$\mathrm{PIIC1}(\boldsymbol{\xi}) \equiv -\sum_{i=1}^{n}\log f(z_i \mid \boldsymbol{z}; \boldsymbol{\xi}) + \mathrm{tr}\left\{\hat{\boldsymbol{I}}_{1,\boldsymbol{z},\boldsymbol{\xi}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{z},\boldsymbol{\xi}}\right)^{-1}\hat{\boldsymbol{I}}_{2,\boldsymbol{z},\boldsymbol{\xi}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{z},\boldsymbol{\xi}}\right)\right\}$$

$$\log g(z_i, \boldsymbol{\theta}; \boldsymbol{\xi}) \equiv \log f(z_i \mid \boldsymbol{\theta}) + \frac{1}{n_0}\log \pi(\boldsymbol{\theta}; \boldsymbol{\xi}), \log g(\boldsymbol{z}, \boldsymbol{\theta}; \boldsymbol{\xi}) \equiv \sum_{i=1}^{n}\log g(z_i, \boldsymbol{\theta}; \boldsymbol{\xi})$$

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{z},\boldsymbol{\xi}} \equiv \underset{\boldsymbol{\theta}}{\mathrm{argmax}}\{\log g(\boldsymbol{z}, \boldsymbol{\theta}; \boldsymbol{\xi})\}$$

$$\hat{\boldsymbol{I}}_{1,\boldsymbol{z},\boldsymbol{\xi}}(\boldsymbol{\theta}) \equiv -\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log g(\boldsymbol{z}, \boldsymbol{\theta}; \boldsymbol{\xi})$$

②PIIC1を最小にする$\hat{\boldsymbol{\xi}}_{\boldsymbol{z}}$によるバイアスの補正項を追加

$$\mathrm{PIIC2} \equiv \mathrm{PIIC1}(\hat{\boldsymbol{\xi}}_{\boldsymbol{z}}) + \mathrm{tr}\left\{\hat{\boldsymbol{J}}_{1,\boldsymbol{z}}\left(\hat{\boldsymbol{\xi}}_{\boldsymbol{z}}\right)^{-1}\hat{\boldsymbol{J}}_{2,\boldsymbol{z}}\left(\hat{\boldsymbol{\xi}}_{\boldsymbol{z}}\right)\right\}$$

$$\hat{\boldsymbol{J}}_{1,\boldsymbol{z}}(\boldsymbol{\xi}) \equiv -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'}\log f(z_i \mid \boldsymbol{z}, \boldsymbol{\xi})$$

$$\hat{\boldsymbol{J}}_{2,\boldsymbol{z}}(\boldsymbol{\xi}) \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\xi}}\log f(z_i \mid \boldsymbol{z}; \boldsymbol{\xi})\frac{\partial}{\partial\boldsymbol{\xi}'}\log f(z_i \mid \boldsymbol{z}; \boldsymbol{\xi})$$

参考文献

Zhao, Y. and Bondell, H. (2020). Solution paths for the generalized lasso with applications to spatially varying coefficients regression. Computational Statistics & Data Analysis, 142, 106821.
Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. Bayesian Analysis, 5(2), 369–411.
Rahardiantoro, S., & Sakamoto, W. (2021, March). Clustering regions based on socio-economic factors which affected the number of COVID-19 cases in Java Island. In Journal of Physics: Conference Series (Vol. 1863, No. 1, p. 012014). IOP Publishing.
Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research, 11 (12).
Ninomiya, Y. (2021). Prior intensified information criterion. arXiv preprint arXiv:2110.12145.