

空間回帰モデルにおける 最尤推定量の一致性

北里大学 未来工学部 力丸佑紀
データサイエンスコンソーシアム, 慶應義塾大学 柴田里程



空間回帰モデル

誤差 $\mathbf{u}$ が定常性をもつような空間回帰モデル

$$\mathbf{y} = X(\lambda)\boldsymbol{\beta} + \mathbf{u}, \quad P_\lambda(T_1, T_2)u_v = \varepsilon_v, \quad \varepsilon_v \sim N(0, \sigma^2)$$

$$X(\lambda) = S(\lambda)^{-1}X$$

$$S(\lambda) = I - \lambda \sum_{k \in K} w_k U_{n_1}^{k_1} \otimes U_{n_2}^{k_2}, \quad (U_n)_{i,i+1} = 1, \text{ それ以外は } 0$$

$$P_\lambda(T_1, T_2) = 1 - \lambda \sum_{k \in K} w_k T_1^{k_1} T_2^{k_2}$$

周りの点からどんな影響を受けるか

(SLM (空間ラグモデル) の $\mathbf{u}$ に定常過程を導入したモデル)

$$\mathbf{y} = \lambda W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

尤度

- 対数尤度（近似尤度）

$$L_A(\beta, \lambda, \sigma) = \frac{1}{2} \log \det(A) - \frac{N}{2} \log 2\pi - \frac{1}{2} u^T A u$$

$$A = C(\lambda)^T C(\lambda) / \sigma^2$$

$$C(\lambda) = I_{n_1} \otimes I_{n_2} - \lambda \sum_{k \in \mathcal{K}} w_k C_{n_1}^{k_1} \otimes C_{n_2}^{k_2} \quad : \text{巡回行列}$$

:

プロフィール尤度

- プロフィール尤度

$$L_A(\hat{\beta}, \lambda, \sigma) = \frac{1}{2} \log \det(A) - \frac{N}{2} \log 2\pi - \frac{1}{2} \mathbf{y}^T (\mathbf{I} - P(\lambda)) A (\mathbf{I} - P(\lambda)) \mathbf{y}$$

$$P(\lambda) = X(\lambda) (X(\lambda)^T X(\lambda))^{-1} X(\lambda)$$

- $\frac{\partial L_A}{\partial \beta} = 0$ より $\hat{\beta} = (X(\lambda)^T X(\lambda))^{-1} X(\lambda) \mathbf{y}$
- $\hat{\beta} = (X(\lambda)^T X(\lambda))^{-1} X(\lambda) \mathbf{y}$ を $L_A(\beta, \lambda, \sigma)$ に代入すると, 上の式が得られる.

証明の流れ : $\hat{\lambda}$ と $\hat{\sigma}$ の一致性を証明してから, $\hat{\beta}$ の一致性を証明する

i.i.dの場合 (Ferguson, 1996)

$$U_N(\theta_0, \theta) = \frac{1}{N} (L(\theta) - L(\theta_0))$$

$$U_N(\theta_0, \theta) - \mu(\theta_0, \theta) \rightarrow 0 \text{ in prob. as } N \rightarrow \infty$$

大数法則

$\mu(\theta_0, \theta) = 0$ となるのは, $\theta = \theta_0$ のときのみ.

$$\forall \varepsilon > 0, \exists C < 0, \sup_{\theta \in S} \mu(\theta_0, \theta) < C \text{ for } S = \{\theta \in \Theta, \|\theta - \theta_0\| > \varepsilon\}$$

ならば, $\hat{\theta} \rightarrow \theta_0$ in prob. as $N \rightarrow \infty$

ただし, $U_N(\theta_0, \hat{\theta}) = \max_{\theta} U_N(\theta_0, \theta)$

$$\mu(\theta_0, \theta) = -K(\theta_0, \theta)$$

$K(\theta_0, \theta)$: カルバックライブラー情報量

$K(\theta_0, \theta)$ は常に非負.

i.i.d.でなくても一貫性が証明できる場合

$$U_N(\theta_0, \theta) = \frac{1}{N} (L(\theta) - L(\theta_0)) \text{ とする.}$$

1 $U_N(\theta_0, \theta) - \mu_N(\theta_0, \theta) \rightarrow 0 \text{ in prob. as } N \rightarrow \infty$

$\mu(\theta_0, \theta) : \mu_N(\theta_0, \theta)$ の極限

2-1 $\mu(\theta_0, \theta) = 0$ となるのは, $\theta = \theta_0$ のときのみ.

2-2 $\forall \varepsilon > 0, \exists C < 0, \sup_{\theta \in S} \mu(\theta_0, \theta) < C \text{ for } S = \{\theta \in \Theta, \|\theta - \theta_0\| > \varepsilon\}$

ならば, $\hat{\theta} \rightarrow \theta_0 \text{ in prob. as } N \rightarrow \infty$

ただし, $U_N(\theta_0, \hat{\theta}) = \max_{\theta} U_N(\theta_0, \theta)$

1 $U_N(\lambda_0, \sigma_0, \lambda, \sigma) - \mu_N(\lambda_0, \sigma_0, \lambda, \sigma) \rightarrow 0$ in prob. as $N \rightarrow \infty$

Lemma1

$$\frac{1}{N} u^T (I - P(\lambda)) C(\lambda)^T C(\lambda) (I - P(\lambda)) u - \frac{\sigma_0^2}{N} \text{tr}((C(\lambda_0)^T C(\lambda_0))^{-1} (C(\lambda)^T C(\lambda))) \\ \rightarrow 0 \text{ uniformly in prob. as } N \rightarrow \infty$$

$$U_N(\lambda_0, \sigma_0, \lambda, \sigma) = \frac{1}{N} (L(\lambda, \sigma) - L(\lambda_0, \sigma_0))$$

$$= \frac{1}{2N} \log \left(\frac{\det(C(\lambda)^T C(\lambda))}{\det(C(\lambda_0)^T C(\lambda_0))} \right) - \frac{1}{2} \log \frac{\sigma_0^2}{\sigma^2} - \frac{\sigma_0^2}{2N\sigma^2} \text{tr}((C(\lambda_0)^T C(\lambda_0))^{-1} C(\lambda)^T C(\lambda)) + \frac{1}{2}$$

$$= \frac{1}{2N} \sum_j \log(g_j(\lambda)) - \frac{1}{2} \log \frac{\sigma_0^2}{\sigma^2} - \frac{\sigma_0^2}{2N\sigma^2} \sum_j g_j(\lambda) + \frac{1}{2},$$

$$\mu(\lambda, \sigma) = \frac{1}{2} \int_0^1 \int_0^1 \left[\log \left\{ g(\lambda, \omega) \frac{\sigma_0^2}{\sigma^2} \right\} + 1 - g(\lambda, \omega) \frac{\sigma_0^2}{\sigma^2} \right] d\omega$$

$$g_j(\lambda) = \left| \frac{\lambda_j(C(\lambda))}{\lambda_j(C(\lambda_0))} \right|^2$$

$$g(\lambda, \omega) = \left| \frac{P_\lambda(\omega)}{P_{\lambda_0}(\omega)} \right|^2$$

フーリエ表現

2-1 $\mu(\theta_0, \theta) = 0$ となるのは, $\theta = \theta_0$ のときのみ.

Lemma2

仮定1,2のもとで, $g(\lambda, \omega)$ は $\lambda = \lambda_0$ の場合に限り, ω によらない関数である.

最大値の原理

D : 有界連結開集合

$f(z_1, z_2) : D$ 上で 0 にはならない正則関数, $\bar{D}$ で連続.

$|f(z_1, z_2)|$ が D の境界上で定数なら, $f(z_1, z_2)$ は D 上で定数関数である.

$Q_\lambda(z_1, z_2) = z_1^{p_1} z_2^{p_2} P_\lambda(z_1, z_2)$ とする. ただし, $-p_1, -p_2$ は $P_\lambda(z_1, z_2)$ の負の最大次数.

このとき, $Q_\lambda(z_1, z_2) : D = \{(z_1, z_2); |z_1| < 1, |z_2| < 1\}$ で 0 でない正則関数.

$\because$ 非負のべき乗だけの多項式なので, D で正則.

$\because Q_\lambda(0, 0) = w_{-p_1, -p_2} \neq 0$ で, D 上で 0 にならない.

$R_\lambda(z_1, z_2) = \frac{Q_\lambda(z_1, z_2)}{Q_{\lambda_0}(z_1, z_2)}$ も同様.

$R_\lambda(z_1, z_2)$ に対して, 最大値の原理を使う.

$|R_\lambda(z_1, z_2)|$ は $|z_1| = |z_2| = 1$ で定数なので、最大値の原理より、 $R_\lambda(z_1, z_2)$ は D で定数関数である。

$$R_\lambda(z_1, z_2) = \frac{Q_\lambda(z_1, z_2)}{Q_{\lambda_0}(z_1, z_2)} = \frac{P_\lambda(z_1, z_2)}{P_{\lambda_0}(z_1, z_2)} \quad \text{なので, } P_\lambda(z_1, z_2) = c P_{\lambda_0}(z_1, z_2).$$

よって、係数比較から、 $c = 1, \lambda = \lambda_0$.

次に、 $R_\lambda(z_1, z_2) = \frac{\bar{Q}_\lambda(z_1, z_2)}{Q_{\lambda_0}(z_1, z_2)}$ も同様に議論すると、 $c = 1, \bar{P}_\lambda(z_1, z_2) = P_{\lambda_0}(z_1, z_2)$

1. symmetric ($w_k = w_{-k}$) のとき、 $\bar{P}_\lambda(z_1, z_2) = P_\lambda(z_1, z_2)$. よって解は $\lambda = \lambda_0$

2. skew symmetric ($w_k = -w_{-k}$) のとき、 $\bar{P}_\lambda(z_1, z_2) = P_{-\lambda}(z_1, z_2)$. よって解は $\lambda = -\lambda_0$

3. それ以外するとき、 $\bar{P}_\lambda(z_1, z_2) \neq P_{\lambda_0}(z_1, z_2)$.

仮定2
(予稿参照)

「 $\mu(\lambda_0, \sigma_0, \lambda, \sigma) = 0$ は $\sigma = \sigma_0$ のときのみ」は $\mu(\lambda_0, \sigma_0, \lambda, \sigma)$ の形からわかる。

2-2

$\forall \varepsilon > 0, \exists C < 0, \sup_{\theta \in S} \mu(\theta_0, \theta) < C$ for $S = \{\theta \in \Theta, \|\theta - \theta_0\| > \varepsilon\}$

$$\textcircled{1} \quad \max_{\sigma} 2\mu(\lambda, \sigma) = \int_0^1 \int_0^1 \log \left(\frac{g(\lambda, \omega)}{g(\lambda)} \right) d\omega \leq \frac{-1}{2 \max_{\omega} g(\lambda, \omega)^2} \int_0^1 \int_0^1 (g(\lambda, \omega) - g(\lambda))^2 d\omega$$

仮定1より, $\max_{\omega} g(\lambda, \omega)^2 < C$.

$$v(\lambda) = \int_0^1 \int_0^1 (g(\lambda, \omega) - g(\lambda))^2 d\omega$$

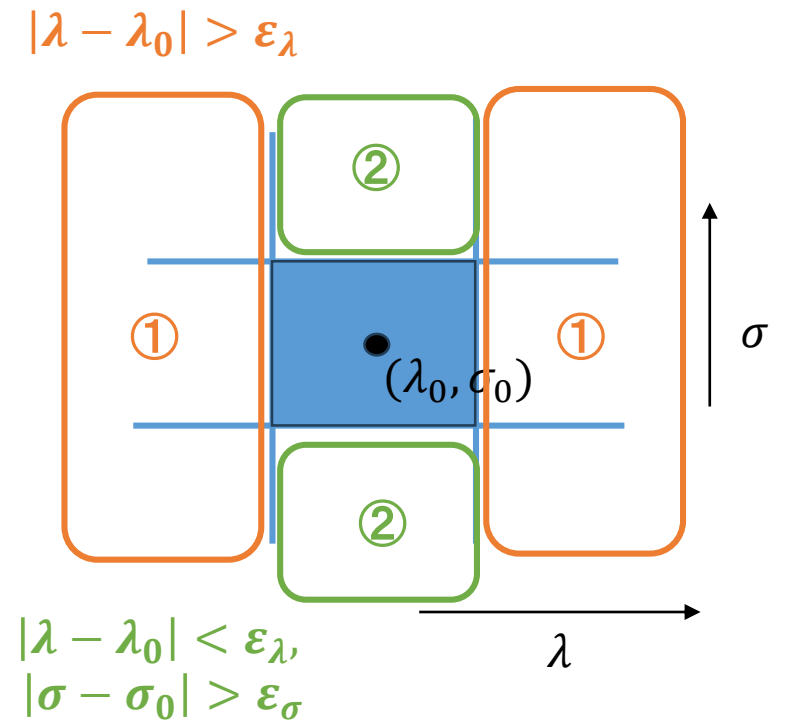
$$v(\lambda) = 0 \Leftrightarrow g(\lambda, \omega) = g(\lambda)$$

Lemma2 からこれは $\lambda = \lambda_0$ のときしか起きない.

$|\lambda - \lambda_0| > \varepsilon_{\lambda}$ では, $v(\lambda)$ が 0 ではない最小の値をもつ.

$$\textcircled{2} \quad 2\mu(\lambda, \sigma) \leq \log \left(\frac{\sigma_0}{\sigma} \right)^2 + g(\lambda) \left(1 - \left(\frac{\sigma_0}{\sigma} \right)^2 \right)$$

ε_{λ} と ε_{σ} の割合を調整するなどして, 証明.



$\hat{\beta}$ の一致性

$\hat{\beta} = (X(\lambda)^T X(\lambda))^{-1} X(\lambda) \mathbf{y}$ なので,

$\text{Var}(\hat{\beta}(\lambda_0)) = \sigma_0^2 (X(\lambda_0)^T X(\lambda_0))^{-1}$ が消えればよい.

$$\lambda_{\min}(X(\lambda_0)^T X(\lambda_0)) \geq \lambda_{\min}(X^T X) / \lambda_{\max}(S(\lambda_0) S(\lambda_0)^T)$$

であることに注意すれば, $\lambda_{\max}(S(\lambda_0) S(\lambda_0)^T)$ が有界であることと, 仮定3よりいえる.



一致性の証明

独立でも同分布でもない場合の一致性の証明

いろいろ苦労はありましたが、なんとか証明できました。

