Asymptotic property of additive B-spline regression

吉田 拓真 [1]　　内藤 貫太 [2]

[1] 島根大学大学院総合理工学研究科 [2] 島根大学総合理工学部

1 はじめに

スプライン平滑化はノンパラメトリック平滑化の代表的な手法であり, 説明変数が 1 次元の場合 (以下 Univariate model と言う) のみならず, 説明変数が多次元の場合の代表的なモデルである Additive model(Hastie and Tibshirani (1990)) においても有用である．本講演では，Additive model における p 次 B-スプラインを用いたときの罰則付きスプラインの漸近分布について議論した．Additive model における罰則付きスプラインの漸近挙動は各変数毎に見た場合に, Univariate model における罰則付きスプラインの漸近挙動と同等になること, 漸近正規性を有することなどを示した．また, 得られた結果をシミュレーションで検証した.

2 モデル設定と罰則付きスプライン

データ $\{(y_i, x_{i1}, x_{i2} : i = 1, \cdots, n)\}$ に対して，Additive model

$$y_i = f(x_{i1}, x_{i2}) + \varepsilon_i = f_1(x_{i1}) + f_2(x_{i2}) + \varepsilon_i \tag{1}$$

を考える．ただし, $(x_{i1}, x_{i2}) \in (0,1) \times (0,1)$ であり, $q_j(x_j)$ を説明変数 x_j の密度関数, $q(x_1, x_2)$ を同時密度とする．また, ε_i は互いに独立に $E[\varepsilon_i] = 0$, $V[\varepsilon_i] = \sigma^2(x_{i1}, x_{i2})$ に従う確率変数とする．いま，(1) を近似するモデルとして, B-スプライン回帰モデル

$$y_i = \sum_{k=-p+1}^{K(n)} B_k^{[p]}(x_{i1}) b_{1,k} + \sum_{k=-p+1}^{K(n)} B_k^{[p]}(x_{i2}) b_{2,k} + \varepsilon_i$$

を考える．ただし, $b_{j,k}(j = 1, 2, k = -p+1, \cdots, K(n))$ は未知パラメータである．また, $K(n) = O(n^\gamma), 0 < \gamma < 1$ であり，$B_j^{[p]}(x)$ は p 次の B-スプライン関数 (de Boor(2001)) である．(1) は $\boldsymbol{y} = (y_1 \ \cdots \ y_n)'$, $X_1 = (B_{i,-p+j}^{[p]}(x_{i1}))_{ij}$, $X_2 = (B_{i,-p+j}^{[p]}(x_{i2}))_{ij}$, $\boldsymbol{b}_1 = (b_{1,-p+1} \cdots \ b_{1,K(n)})'$, $\boldsymbol{b}_2 = (b_{2,-p+1} \cdots \ b_{2,K(n)})'$, $\boldsymbol{\varepsilon} = (\varepsilon_1 \ \cdots \ \varepsilon_n)'$ を用いると

$$\boldsymbol{y} = X_1 \boldsymbol{b}_1 + X_2 \boldsymbol{b}_2 + \boldsymbol{\varepsilon}$$

と書ける．このとき，$\boldsymbol{b}_1, \boldsymbol{b}_2$ の推定量 $\hat{\boldsymbol{b}}_1, \hat{\boldsymbol{b}}_2$ を

$$(\boldsymbol{y} - X_1\boldsymbol{b}_1 - X_2\boldsymbol{b}_2)'(\boldsymbol{y} - X_1\boldsymbol{b}_1 - X_2\boldsymbol{b}_2) + \sum_{j=1}^{2} \lambda_{jn} \boldsymbol{b}_j' Q_m \boldsymbol{b}_j \tag{2}$$

の $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ についての最小化によって構成する．ここで, λ_{jn} は平滑化パラメータ, Q_m は m 次差分行列である．さらに, $x_j \in (0,1)$ に対して, $f_j(x_j)$ の推定量を

$$\hat{f}_j(x_j) = \sum_{k=-p+1}^{K(n)} B_k^{[p]}(x_j) \hat{b}_{j,k}, j = 1, 2$$

と構成する．これを罰則付きスプラインという．$(x_1, x_2) \in (0,1)\times(0,1)$ における y の予測量は

$$\hat{y} = \hat{f}_1(x_1) + \hat{f}_2(x_2)$$

と構成される．ただし, (2) を最小にするような $\boldsymbol{b}_j (j=1,2)$ はそれぞれ

$$\boldsymbol{b}_1 = \Lambda_1^{-1} X_1'(\boldsymbol{y} - X_2\boldsymbol{b}_2), \quad \boldsymbol{b}_2 = \Lambda_2^{-1} X_2'(\boldsymbol{y} - X_1\boldsymbol{b}_1) \tag{3}$$

となることから, 実際, $\boldsymbol{b}_j$ の推定量は (3) を元にした Backfitting algorithm(Hastie and Tibshirani (1990)) によって構成される．

3　罰則付きスプラインの漸近理論

$\hat{f}_j(x_j)(j=1,2)$ の漸近挙動は

$$\hat{f}_{pen,j}(x_j) = \boldsymbol{B}(x_j)'(X_j'X_j + \lambda_{jn}Q_m)^{-1}X_j'\boldsymbol{y}$$

の漸近挙動と同等になることが示される．ただし, $\boldsymbol{B}(x_j) = (B^{[p]}_{-p+1}(x_j) \ \cdots \ B^{[p]}_{K(n)}(x_j))'$ である．一方, $\hat{f}_{pen,j}(x_j)$ は $\{(y_i, x_{ij} : i = 1, \cdots, n)\}$ に対する Univariate model

$$y_i = f_j(x_{ij}) + \varepsilon_i$$

を考えたときの罰則付きスプラインであり, $\hat{f}_{pen,j}(x_j)$ の漸近的性質は Claeskens et al. (2009) によって得られている．Claeskens et al. (2009) の結果を用いて, $\hat{f}_j(x_j)(j=1,2)$ の漸近バイアス, 漸近分散を導出できる．さらに, その結果を用いて, 次の Theorem が示される．

Theorem ある $\delta > 0$ に対して, $E[|\varepsilon_i|^{2+\delta}] < \infty$ とし, $f_j \in C^{p+1}$ とする． また, $\gamma > 1/3$, $\lambda_{1n} = o((n/K(n))^{1/2})$, $\lambda_{2n} = o((n/K(n))^{1/2})$ とする．このとき, $(x_1, x_2) \in (0,1)\times(0,1)$ に対して

$$\left\{ V \begin{bmatrix} \hat{f}_1(x_1) \\ \hat{f}_2(x_2) \end{bmatrix} \right\}^{-1/2} \begin{bmatrix} \hat{f}_1(x_1) - f_1(x_1) \\ \hat{f}_2(x_2) - f_2(x_2) \end{bmatrix} \xrightarrow{D} N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, I_2 \right)$$

が成立する．

4　シミュレーション

第 3 節の Additive model における罰則付きスプラインの漸近結果について, シミュレーションにより検証した．罰則付きスプライン $\hat{f}_j(x_j)(j=1,2)$ はそれぞれ n が大きいとき, $f_j(x_j)$ に漸近することが見て取れた．

参考文献

[1] Claeskens,G., Krivobokova,T. and Opsomer,J.D.(2009). Asymptotic properties of penalized spline estimators. *Biometrika*. **96**, 529-544.

[2] de Boor,C.(2001). *A Practical Guide to Splines*. Springer-Verlag.

[3] Hastie,T. and Tibshirani,R.(1990). *Generalized Additive Models*. London Chapman & Hall.