円周上の局所多項式回帰モデルの漸近的性質

金沢大学人間社会環境研究科　　鶴田靖人
金沢大学経済学経営学系　　寒河江雅彦

1　はじめに

本稿では、円周上の角度として定義される説明変数 Θ_i (例：風向) と実数直線上で定義される目的変数 Y_i(例:風速) の間に存在する関係を、回帰モデルで推定することを考える．目的変数 Y_i は以下の式を考える：

$$Y_i = m(\Theta_i) + v^{1/2}(\Theta_i)\varepsilon_i,$$

ただし，誤差項 ε_i は，平均が 0，分散が 1 の互いに独立な確率変数であり，回帰関数 $m(\theta)$ は，周期性: $m(\theta) = m(\theta + 2\pi)$ を満たすものとする．

Di Marzio et al. (2009) は、$m(\theta)$ に関するノンパラメトリックな推定量として sine-local polynomial regression (S-LPR) を提案した．S-LPR $\hat{m}(\theta; p, \kappa)$ は、p 次の sin 関数のべき級数多項式を用いたカーネル型重み付き二乗誤差：

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1 \sin(\Theta_i - \theta) - \cdots - \beta_p \sin^p(\Theta_i - \theta)\}^2 \times K_\kappa(\Theta_i - \theta)$$

を最小にする切片 $\hat{\beta}_0$ で与えられる．ただし，κ は集中度パラメータと呼ばれる平滑化パラメータとし，$K_\kappa(\theta)$ は円周上に台をもつ対称カーネル関数とする．彼らはあるカーネル関数族を与え、S-LPR の条件付平均二乗誤差 (MSE) を導出した．しかし、彼らが与えたカーネル関数族は、集中度パラメータ κ とカーネル関数 $K_\kappa(\theta)$ を分離することができず、MSE の収束レートに関する一般的な性質を見つけることが難しい．

我々は，Hall et al. (1987) が与えたカーネル関数は平滑化パラメータとカーネル関数自体を分離できることを示す．この特性を用いて、S-LPR の MSE の収束レートに関する一般的な性質を導く．

2　局所多項式回帰の漸近的性質について

Hall et al. (1987) が導入したカーネル関数 $K_{(h)}(\theta)$ は，

$$K_{(h)}(\theta) := C_h^{-1}(L) L_h(\theta),$$

で定義される．ただし，バンド幅 $h = h(n) > 0$ は，$\lim_{n\to\infty} h(n) = 0$ かつ $\kappa = h^{-2}$ を満たすものとする．また，$L_h(\theta) := L(h^{-2}\{1 - \cos(\theta)\})$ とし，$C_h(L) := \int_{-\pi}^{\pi} L_h(\theta) d\theta$ は基準化定数とする．

n が十分大きいとき，カーネル関数 $K_{(h)}(\theta)$ は，

$$K_{(h)}(hz) = h^{-1}[L(z^2/2) + O_p(h^2)], \qquad (1)$$

$z \in [-\pi/h, \pi/h)$ となる．(1) を用いて S-LPR の条件付バイアスと分散に関する収束レートを求めることができる．バイアスの収束レートは次数 p に依存しており，

$$\mathrm{Bias}_Y[\hat{m}(\theta; p, h)|\mathbf{\Theta}_n] = \begin{cases} O_p(h^{p+1}) & p \text{ が奇数の場合}, \\ O_p(h^{p+2}) & p \text{ が偶数の場合}, \end{cases}$$

となる．分散の収束レートは

$$\mathrm{Var}_Y[\hat{m}(\theta; p, h)|\mathbf{\Theta}_n] = O_p(\{nh\}^{-1})$$

で与えられる．

本稿では、定義域全体のリスクを評価するために、条件付重み付平均積分二乗誤差 (MISE)：

$$\mathrm{MISE}[\hat{m}(\theta; p, h)|\mathbf{\Theta}_n] := \mathrm{E}_Y\left[\int_{-\pi}^{\pi} \{\hat{m}(\theta; p, h) - m(\theta; p, h)\}^2 f(\theta) d\theta |\mathbf{\Theta}_n\right]$$

を採用した．S-LPR の最適な MISE の収束レートは，p が奇数のとき，

$$\mathrm{MISE}[\hat{m}(\theta; p, h_*)|\mathbf{\Theta}_n] = O_p(n^{-(2p+2)/(2p+3)})$$

となり，p が偶数のとき，

$$\mathrm{MISE}[\hat{m}(\theta; p, h_*)|\mathbf{\Theta}_n] = O_p(n^{-(2p+4)/(2p+5)})$$

となる．

3　数値実験

数値実験の結果は当日発表する．

参考文献

[1] Di Marzio, M., Panzera, A. and Taylor, C. C. (2009). *Statistics & Probability Letters* **79**, 2066-2075.

[2] Hall, P., Watson, G. S. and Cabrera, J. (1987). *Biometrika*, **74**, 751-762.