# 方向統計学における多変量ノンパラメトリック回帰モデルとその漸近的性質

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

周期性を持つ確率変数は，角度変数 $\Theta_i$ と呼ばれ，単位円周上 $\mathbb{S}_1$ で定義される．周期性を持つ現象の構造を明らかにするためによく用いられる方法は，実数直線上で定義された目的変数 $Y_i$ と角度変数 $\Theta_i$ を含む説明変数ベクトル $\boldsymbol{U}_i$ の関係を回帰モデルで推定することである．

Di Marzio et al. (2009) は、説明変数ベクトル $\boldsymbol{\Theta}_i$ が $q$ 次元トーラス $\mathbb{S}_1^q$ 上のベクトル $\boldsymbol{\Theta}_i = (\Theta_{i1}, \ldots, \Theta_{iq})^T$ で与えられるノンパラメトリック回帰を提案した．また、Qin et al. (2011) は、説明変数ベクトルが $d$ 次元実数空間 $\mathbb{R}^d$ 上のベクトルと 1 次元角度変数からなるベクトル $\boldsymbol{U}_i = (\boldsymbol{X}_i, \Theta_i)^T$ で与えられるノンパラメトリック回帰を提案した．これら 2 つのノンパラメトリック回帰は、角度変数上のカーネル関数と平滑化パラメータを分離することができず、一般的な性質を求めることが難しい．

本稿では、Di Marzio et al. (2009) と Qin et al. (2011) のノンパラメトリック回帰を一般化し、説明変数ベクトルが $\mathbb{R}^d$ 上のベクトル $\boldsymbol{X}_i$ と $\mathbb{S}_1^q$ 上のベクトル $\boldsymbol{\Theta}_i$ からなる $d+q$ 次元ベクトル $\boldsymbol{U}_i = (\boldsymbol{X}_i^T, \boldsymbol{\Theta}_i^T)^T$ で与えられるノンパラメトリック回帰を提案し、その理論的性質を与える．

## 2　提案するノンパラメトリック回帰

目的変数 $Y_i$ は

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

とする．ただし，誤差項 $\varepsilon_i$ は，平均が 0，分散が 1 の互いに独立な確率変数，$m$ は回帰関数，$v(\boldsymbol{U}_i)$ は $Y_i$ の条件付分散とする．

回帰係数を $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{d+q})^T$ としたとき，提案する実数トーラス上のノンパラメトリック回帰 $\hat{m}(\boldsymbol{u};\boldsymbol{H})$ は，次の重み付き二乗誤差：

$$\sum_{i=1}^{n}\{Y_i - \boldsymbol{\beta}^T\boldsymbol{U_u}\}^2 K_{\boldsymbol{H}}(\boldsymbol{U_i} - \boldsymbol{u})$$

を最小にする切片 $\hat{\beta}_0$ で与えられる．ただし，

$$\boldsymbol{U_u} := \begin{bmatrix} 1 & (\boldsymbol{X}_1 - \boldsymbol{x})^T & \sin(\boldsymbol{\Theta}_1 - \boldsymbol{\theta}) \\ \vdots & \vdots & \vdots \\ 1 & (\boldsymbol{X}_n - \boldsymbol{x})^T & \sin(\boldsymbol{\Theta}_n - \boldsymbol{\theta}) \end{bmatrix}$$

は $n \times (d+q+1)$ デザイン行列を表す．$\boldsymbol{H} := \mathrm{diag}\{h_1^2, \ldots, h_d^2, h_{d+1}^2, \ldots, h_{d+q}^2\}$ はバンド幅行列，カーネル関数 $K_{\boldsymbol{H}}(\boldsymbol{u})$ は、2 つのプロダクトカーネル関数の積として，

$$K_{\boldsymbol{H}}(\boldsymbol{u}) := K_{(\boldsymbol{H}_d)}(\boldsymbol{x}) \times K_{(\boldsymbol{H}_q)}(\boldsymbol{u}),$$

と定義される．ただし，

$$K_{(\boldsymbol{H}_d)}(\boldsymbol{x}) := \prod_{j=1}^{d} K(x_j/h_j)/h_j$$

であり、

$$K_{(\boldsymbol{H}_q)}(\boldsymbol{x}) := \prod_{j=1}^{q} K_{(h_j)}(\theta_j)$$

である．ただし，$K_{(h_j)}(\theta_j)$ は，Hall et al. (1987) が定義した $K_{(h_j)}(\theta_j) = C_h^{-1}(L)L(h\{1-\cos(\theta_j)\})$ である．ただし，$C_h(L)$ は関数 $L()$ の基準化定数である．

我々は，Hall et al. (1987) が与えたカーネル関数は $h$ と関数 $L()$ を分離可能であることを示す．この性質を用いて、$\hat{m}(\boldsymbol{u};\boldsymbol{H})$ の条件付バイアスと分散の収束レートを導出できる．$h_j = h$ とおいたとき，$\hat{m}(\boldsymbol{u};\boldsymbol{H})$ のバイアスと分散はそれぞれ，

$$\mathrm{Bias}_Y[\hat{m}(\boldsymbol{u};\boldsymbol{H})|\boldsymbol{\Theta}_n] = O_p(h^2),$$
$$\mathrm{Var}_Y[\hat{m}(\boldsymbol{u};\boldsymbol{H})|\boldsymbol{\Theta}_n] = O_p(n^{-1}h^{-(d+q)}),$$

となる．バイアスと分散の導出によって，MISE の収束レート、バンド幅選択や漸近正規性などの議論が可能になる。

## 3　数値実験

数値実験の結果を当日に発表する．

## 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.

[3] Qin, X., Zhang, J. S., and Yan, X. D. (2011). A *Computers & Mathematics with Applications*, **62**, 3048-3055.