

# Structure discovery and semiparametric regression

鹿児島大学大学院理工学研究科 吉田拓真

はじめに 回帰分析において、未知の回帰関数がパラメトリックモデルとして表されるかどうかの検討は重要である。もし、それができないような複雑なデータにはノンパラメトリック回帰を行う必要がある。本講演では、想定したパラメトリックモデルが真の回帰関数と一致するとみなしてよいのかを判断し、同時に推定量を構築するセミパラメトリック法について述べる。

**セミパラメトリック推定量** 得られたデータ  $\{(y_i, x_i) : i = 1, \dots, n\}$  に対して回帰モデル

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

を考える。ただし、 $f$  は未知の回帰関数、 $\varepsilon_i$  は誤差である。このとき、 $f$  をセミパラメトリック法によって推定する。まず、パラメータベクトル  $\boldsymbol{\beta} \in \mathbb{R}^p$  によって規定されるパラメトリックモデル  $f(x|\boldsymbol{\beta})$  を用意し、 $\boldsymbol{\beta}$  の一致推定量  $\hat{\boldsymbol{\beta}}$  を構成する。このとき、回帰関数は  $f(x) = f(x|\hat{\boldsymbol{\beta}}) + r(x)$  と分解される ( $r(x) = f(x) - f(x|\hat{\boldsymbol{\beta}})$ )。次に、データ  $\{(y_i - f(x_i|\hat{\boldsymbol{\beta}}), x_i) : i = 1, \dots, n\}$  にノンパラメトリック法を適用し、 $r(x)$  を推定する。ここでは、スプライン法によって  $r(x)$  の推定量  $\hat{r}(x)$  を構成する。すなわち、スプラインモデル  $\mathbf{B}(x)^T \mathbf{b}$  ( $\mathbf{B}(x)$  は  $B$ -スプライン関数を要素を持つベクトル、 $\mathbf{b}$  はパラメータベクトル) に対して、

$$\sum_{i=1}^n (y_i - f(x_i|\hat{\boldsymbol{\beta}}) - \mathbf{B}(x_i)^T \mathbf{b})^2 + \lambda \sqrt{\mathbf{b}^T \mathbf{b}}$$

を最小にする  $\mathbf{b} = \hat{\mathbf{b}}$  を用いて、 $r(x)$  の推定量  $\hat{r}(x) = \mathbf{B}(x)^T \hat{\mathbf{b}}$  を構築する。ここで、 $\lambda$  は平滑化パラメータ、罰則項は Yuan and Lin (2006) が提案した group lasso である。最終的に、 $f(x)$  のセミパラメトリック推定量を  $\hat{f}(x) = f(x|\hat{\boldsymbol{\beta}}) + \hat{r}(x)$  と構成する。

Group lasso は 真の関数が 0、すなわち、 $r(x) \equiv 0$  であれば、 $\hat{\mathbf{b}} = \mathbf{0}$  と推定する。つまりこの場合、 $\hat{f}(x) = f(x|\hat{\boldsymbol{\beta}}) + \hat{r}(x) = f(x|\hat{\boldsymbol{\beta}})$  となる。よって、想定したモデルが正しければ  $f(x|\hat{\boldsymbol{\theta}})$  が推定量となり、正しくなければ  $\hat{r}(x) \neq 0$  となり、そのまま  $\hat{f}(x)$  が推定量となる。このように、提案手法はパラメトリック構造の有無の判定と推定量の構築を行うことができる。Fan et al. (2009) などから、 $f(x|\boldsymbol{\beta})$  が  $f(x)$  に一致せずとも近い場合は  $\hat{f}(x)$  は通常のノンパラメトリック推定量よりはよい挙動をすることがわかっている。提案した手法は加法モデルに拡張することができる。また、Zhang et al. (2011) では、線形モデルを想定するケースを扱っており、本手法は Zhang et al. (2011) の拡張と捉えることができる。

## Reference

- Fan, J., Wu, Y. and Feng, Y. (2009). Local quasi-likelihood with a parametric guide. *Ann. Statist.* **37** 4153–4183.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B* **68** 49–67.
- Zhang, H.H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *J. Amer. Statist. Assoc.* **106** 1099–1112.