Spline Methods

Feng Li
feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

Revision: November 7, 2013

Today we are going to talk about...

1. Piecewise Polynomials
2. Avoiding knots selection with smoothing splines
3. Multi-dimensional splines
4. Discussions

Piecewise Polynomials

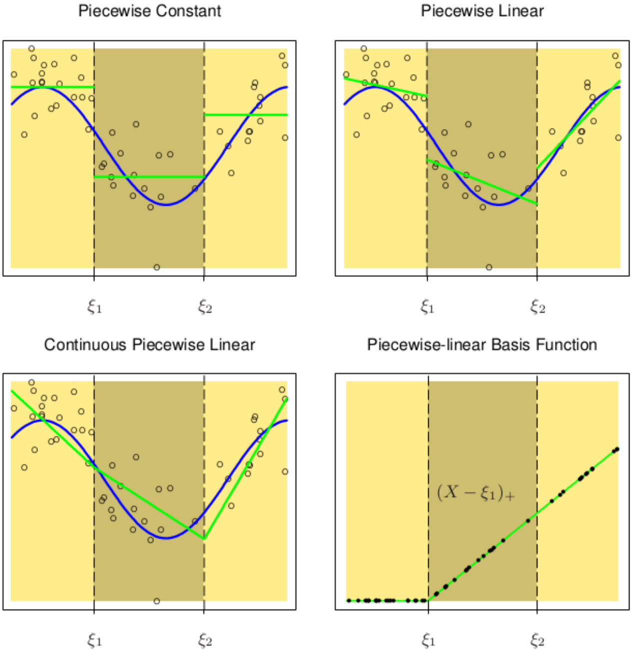
- Consider the regression model with only y and X
- Let

$$h_i(X) = (X - \xi_i)_+ = \begin{cases} X - \xi_i, & X > \xi_i \\ 0, & elsewhere \end{cases}$$

for $i = 1, ..., p$

- Then set up a regression model

$$y = \beta_0 + \beta_1 X + \alpha_1 h_1(X) + ... + + \alpha_p h_p(X) + \epsilon$$

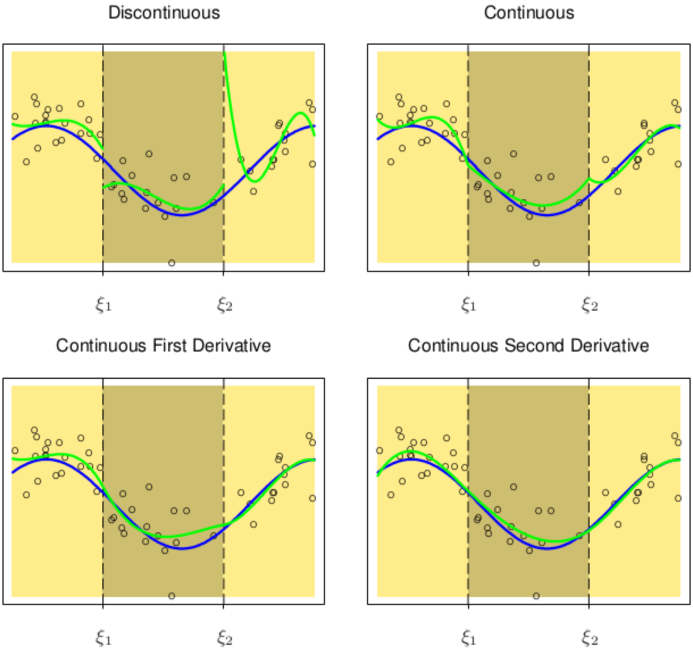

Piecewise Constant
Piecewise Linear
Continuous Piecewise Linear
Piecewise-linear Basis Function
$(X - \xi_1)_+$
ξ_1
ξ_2

Higher order piecewise polynomials

- Let

$$h_j = (X) = X^{j-1}, j = 1, ..., M$$
$$h_{M+l}(X) = (X - \xi_l)_+^{M-1}, l = 1, ..., K$$

- Then use all $h()$ with Y to setup a regression model.
- Terminologies: **basis functions**, **knots**, **knots locations**.
- **Regression splines**: when the knots are fixed


Discontinuous
Continuous
Continuous First Derivative
Continuous Second Derivative
ξ_1
ξ_2

Natural cubic splines

- Define

$$d_k(X) = \frac{(X - \xi_k)_+^3 - (X - \xi_K)_+^3}{\xi_K - \xi - k}$$

- And the spline is defined as

$$h_k(X) = d_k(X) - d_{K-1}(X)$$

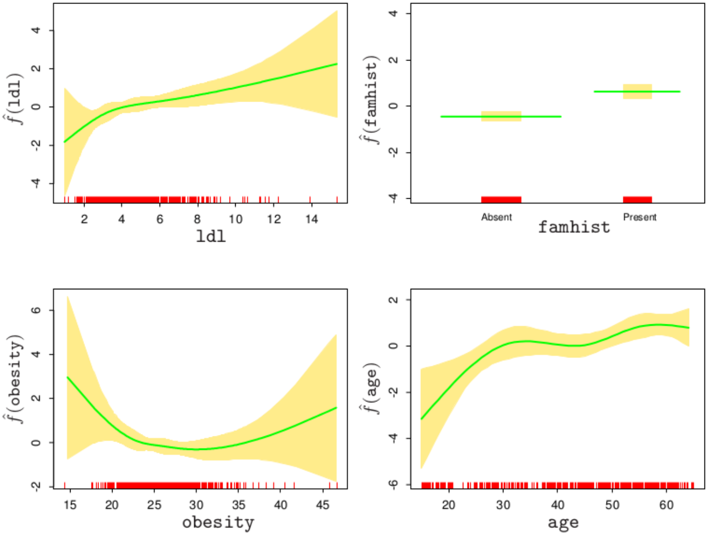


FIGURE 5.4. *Fitted natural-spline functions for each of the terms in the final model selected by the stepwise procedure. Included are pointwise standard-error bands. The* rug plot *at the base of each figure indicates the location of each of the sample values for that variable (jittered to break ties).*

B-splines

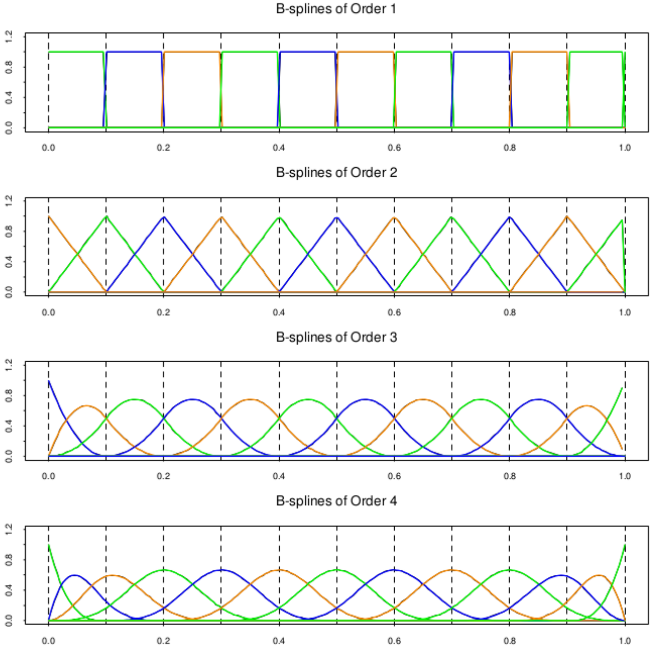
- Assume we have the two boundary knots $\xi_0 < \xi_1$ and $\xi_K < \xi_{K+1}$.
- We define a knot sequence τ such that

$$\tau_1 \leqslant \tau_2 \leqslant ... \leqslant \tau_M \leqslant \xi_o$$
$$\tau_{j+M} = \xi_j, j = 1, ..., K$$
$$\xi_{K+1} \leqslant \tau_{K+M+1} \leqslant \tau_{K+M+2} \leqslant ...\tau_{K+2M}$$

- Let $B_{i,m}(x)$ the ith B-spline function of order $m < M$ for the knot-sequence τ.

$$B_{i,1}(x) = \begin{cases} 1, & \tau_i \leqslant x \leqslant t_{i+1} \\ 0, otherwise \end{cases}$$
$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+i}} B_{i+1,m-1}(x)$$

- Properties of B-spline
  - A B-spline is a continuous function at the knots.
  - Any spline function of degree k on a given set of knots can be expressed as a linear combination of B-splines.


B-splines of Order 1
B-splines of Order 2
B-splines of Order 3
B-splines of Order 4
0.0
0.4
0.8
1.2
0.0
0.2
0.4
0.6
0.8
1.0

Avoiding knots selection with smoothing splines

- The **smoothing spline** is to minimize

$$\mathrm{RSS}(f,\lambda) = \sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt$$

  where λ is the **smoothing parameter**

  - $\lambda = 0$ the usual spline fitting with not penalty.
  - $\lambda \to \infty$ the curve is moving from rough to very smooth till a regression line without knots (very heavy penalty)

Smoothing example with natural cubic splines

- For natural cubic splines

$$f(x) = \sum_{j=1}^{K} \theta_j h_j(x)$$

- The RSS is now as

$$\text{RSS}(\theta, \lambda) = \sum_{i=1}^{N} \{y_i - \sum_{j=1}^{K} h_j(x_i)\}^2 + \lambda \int \left\{ \frac{\partial^2 \sum_{j=1}^{K} h_j(t_i)}{\partial t^2} \right\}^2 dt$$

- And the solution is

$$\hat{\theta} = \left(N'N + \lambda \int \{f''(t)\}^2 dt \right)^{-1} N'y$$

where N is the design matrix with all the data and basis functions.

- And the degree of freedom (no. of free parameters) is obtained through the trace of the hat matrix.

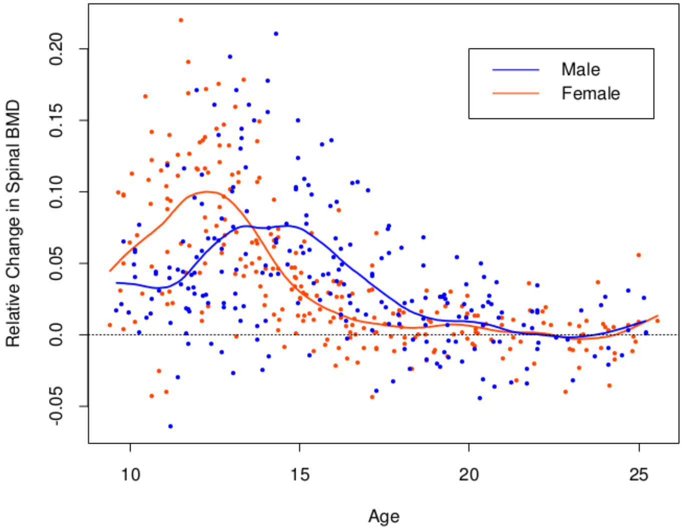


FIGURE 5.6. *The response is the relative change in bone mineral density measured at the spine in adolescents, as a function of age. A separate smoothing spline was fit to the males and females, with* $\lambda \approx 0.00022$*. This choice corresponds to about* 12 *degrees of freedom.*

Spline methods in logistic regression

- Recall the logistic regression

$$\log \frac{\Pr(G = k|X = x)}{\Pr(G = K|X = x)} = f(x)$$

  where $k = 1, 2, ..., K - 1$.
- Splines can also be used in $f(x)$.
- Need maximum likelihood method to obtain $\hat{\beta}$.
- Newton-Raphson algorithm is exactly of the same.

Multi-dimensional splines

- All the cases we considered are univariate splines.
- Multivariate splines are not so rich.
- People usually use thinplate splines

$$g\left(x_1, ...x_q, \boldsymbol{\xi}_j\right) = \left\|\boldsymbol{x} - \boldsymbol{\xi}_j\right\|^2 \ln \left\|\boldsymbol{x} - \boldsymbol{\xi}_j\right\|$$

- Can handle the interactions but the model complexity increase dramatically with the interactive knots.

Discussions

- How do you choose from different splines?
- How do we avoid overfitting in spline method?
- How to apply shrinkage methods like LASSO in splines?
- How to choose λ with smoothing splines
- Do we obtain unbiased estimators in spline methods?
- What is the bias-variance trade off?