Estimation of B-spline Nonparametric Regression Models using Information Criteria

Seiya Imoto and Sadanori Konishi
Graduate School of Mathematics, Kyushu University
6-10-1 Hakozaki, Higashi-ku
Fukuoka 812-8581, Japan
simoto@math.kyushu-u.ac.jp; konishi@math.kyushu-u.ac.jp

1. Introduction

Nonparametric regression modelling has received considerable attention and many methods have been proposed to draw information from data with complex structure. We consider the use of B-spline nonparametric regression models estimated by penalized likelihood methods. A crucial point in constructing the models is in the choice of a smoothing parameter and the number of knots, for which several attempts have been made by using cross-validation and Akaike's (1973) information criterion.

We investigate this problem from information-theoretic point of view, and introduce a criterion for evaluating B-spline nonparametric regression models in the context of generalized linear models. The proposed criterion is applied to choose the optimal value of a smoothing parameter and the number of knots. We compare various types of criteria to examine their properties, using Monte Carlo simulations.

2. *B*-spline Nonparametric Regression Models

Suppose that we have n observations $\{(x_\alpha, y_\alpha); \alpha = 1, \cdots, n\}$ and that the independent responses y_α are from the exponential family of distributions with densities

$$f(y_\alpha|x_\alpha;\xi_\alpha,\phi) = \exp\left[\{y_\alpha\xi_\alpha - u(\xi_\alpha)\}/\phi + v(y_\alpha,\phi)\right], \quad \alpha = 1,\cdots,n \tag{1}$$

where $u(\cdot)$ and $v(\cdot,\cdot)$ are specific functions and ϕ is a scale parameter. The conditional expectation $\mathrm{E}[Y_\alpha|x_\alpha] = \mu_\alpha \ (= u'(\xi_\alpha))$ is related to the predictor η_α by $h(\mu_\alpha) = \eta_\alpha$, where $h(\cdot)$ is a link function. It is assumed that the linear predictor is given by B-splines

$$\eta_\alpha = \sum_{j=1}^{m} a_j B_j(x_\alpha) = \boldsymbol{b}_\alpha' \boldsymbol{a}, \tag{2}$$

where $B_j(\cdot)$ are constructed from polynomial pieces, jointed at knots and $\boldsymbol{a} = (a_1,\cdots,a_m)'$, $\boldsymbol{b}_\alpha = (B_1(x_\alpha),\cdots,B_m(x_\alpha))'$ (de Boor, 1978; Eilers and Marx, 1996).

It follows from (1) and (2) that in the generalized linear model situation B-spline nonparametric regression model is given by

$$f(y_\alpha|x_\alpha;\boldsymbol{a},\phi) = \exp\left[\{y_\alpha r(\boldsymbol{b}_\alpha'\boldsymbol{a}) - s(\boldsymbol{b}_\alpha'\boldsymbol{a})\}/\phi + v(y_\alpha,\phi)\right], \quad \alpha = 1,\cdots,n \tag{3}$$

where $r(\cdot) = u'^{-1} \circ h^{-1}(\cdot)$, $s(\cdot) = u \circ u'^{-1} \circ h^{-1}(\cdot)$. The parameters $\boldsymbol{a}$ and ϕ in the model are estimated by maximizing the penalized log-likelihood function

$$l_\lambda(\boldsymbol{a},\phi) = \sum_{\alpha=1}^{n} \left[\{y_\alpha r(\boldsymbol{b}_\alpha'\boldsymbol{a}) - s(\boldsymbol{b}_\alpha'\boldsymbol{a})\}/\phi + v(y_\alpha,\phi)\right] - \frac{n\lambda}{2}\boldsymbol{a}'D_k'D_k\boldsymbol{a}, \tag{4}$$

where λ (> 0) is a smoothing parameter and D_k is an $(m-k) \times m$ matrix of kth differences operater. The maximum penalized likelihood estimator $\hat{\boldsymbol{a}}$ and $\hat{\phi}$ depend on the smoothing parameter λ and the number of knots.

3. Information Criterion

We consider the problem of choosing the smoothing parameter λ and the number of knots from information-theoretic point of view. Konishi and Kitagawa (1996) gave a general theory for constructing information criteria in model evaluation and selection problems. We derive an information criterion for evaluating B-spline nonparametric regression models (3) estimated by the penalized likelihood methods in the following:

$$\mathrm{IC} = -2 \sum_{\alpha=1}^{n} \left[\{ y_\alpha r(\boldsymbol{b}_\alpha' \hat{\boldsymbol{a}}) - s(\boldsymbol{b}_\alpha' \hat{\boldsymbol{a}}) \} / \hat{\phi} + v(y_\alpha, \hat{\phi}) \right] + 2\mathrm{tr}\left(I J^{-1} \right), \tag{5}$$

where $(m+1) \times (m+1)$ matrices I and J are

$$I = \frac{1}{n} \sum_{\alpha=1}^{n} \frac{\partial \psi(y_\alpha | x_\alpha; \boldsymbol{a}, \phi)}{\partial \boldsymbol{\theta}} \frac{\partial \log f(y_\alpha | x_\alpha; \boldsymbol{a}, \phi)}{\partial \boldsymbol{\theta}'} \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}, \quad J = -\frac{1}{n} \sum_{\alpha=1}^{n} \frac{\partial^2 \psi(y_\alpha | x_\alpha; \boldsymbol{a}, \phi)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}},$$

with $\boldsymbol{\theta} = (\boldsymbol{a}', \phi)'$ and $\psi(y_\alpha | x_\alpha; \boldsymbol{a}, \phi) = \log f(y_\alpha | x_\alpha; \boldsymbol{a}, \phi) - (\lambda/2) \boldsymbol{a}' D_k' D_k \boldsymbol{a}$. We choose a smoothing parameter and the number of knots for which the value of the information criterion is minimized.

4. Numerical Comparison

We use Monte Carlo simulations to examine the properties of various types of criteria including generalized cross-validation, Akaike's (1980) Baysian information criterion, modified AIC (Eilers and Marx, 1996), and an improved AIC (Hurvich, Simonoff and Tsai, 1998). The proposed method is also applied to the analysis of a real data.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd Inter. Symp. on Information Theory* (Petrov,B.N. and Csaki,F. eds.), Akademiai Kiado, Budapest, pp.267-281.

Akaike, H. (1980): Likelihood and the Bayes procedure, *Bayesian Statistic* (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), Univ. Press, Valencia, Spain.

de Boor, C. (1978). *A Practical Guide to Splines.* Springer, Berlin.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, **11**, pp. 89-121.

Hurvich, C. M., Simonoff, J. S. and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. Roy. Statist. Soc. Ser.*, **B 60**: pp. 271-293.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* **83**, pp. 875-890.