Predictive information criterion for nonlinear regression model based on basis expansion methods

Kim, Daeju
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

出版情報：MI Preprint Series. 2011-6, 2011-02-26. 九州大学大学院数理学研究院
バージョン：
権利関係：

MI Preprint Series

Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

Predictive information criterion for nonlinear regression model based on basis expansion methods

Daeju Kim & Sadanori Konishi

MI 2011-6

(Received February 26, 2011)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Predictive information criterion for nonlinear regression model based on basis expansion methods

Daeju Kim[1] and Sadanori Konishi[2]

[1] *Graduate School of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.*

[2] *Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.*

E-mail: t-kin@math.kyushu-u.ac.jp, konishi@math.chuo-u.ac.jp

Abstract

Nonlinear regression modeling based on basis expansions has been widely used to explore data with complex structure. A crucial issue in nonlinear regression model is the choice of adjusted parameters including hyper-parameters for prior distribution and the number of basis functions. The selection of these parameters can be viewed as a model selection and evaluation problem. We derive an information criterion for the Bayesian predictive distribution in the case of both of regression coefficient and variance are unknown. Our proposed method make a selection of the appropriate value of hyper-parameters and the number of basis functions. Real data and simulation data analysis show that our proposed modeling strategy performs well in various situations.

Key Words: Basis expansion, the Bayesian predictive distribution, Predictive information criterion, Model selection, Nonlinear regression model

1 Introduction

Nonlinear regression models based on basis expansions provide a useful tool to analyze data with complex structure. The essential idea for basis expansions is to express a regression function as a linear combination of prescribed function, called basis functions (Hastie *et al.* (2001), Konishi and Kitagawa (2008)).

Since maximum likelihood methods yield unstable parameter estimates, the adopted model is usually estimated by those such as the method of maximum penalised likelihood (Good and Gaskins (1971), Green and Silverman (1994)), the method of regularisation or the Bayes approach and so on.

As one of the Bayesian approach, there are several studies on the Bayesian predictive distribution. Therefore choosing an appropriate prior distribution, the Bayesian predictive distribution can be configured better than plug-in predictive distribution (Komaki (1996)). Bishop (2006) introduced the Bayesian predictive distribution in the framework of regression models and showed a relationship between Maximum A Posteriori (MAP) estimator and Ridge estimator (Hoerl and Kennard (1970)).

One of the critical issue with Bayesian modeling is how to evaluate the goodness of the Bayesian predictive distributions. The goodness of Bayesian predictive distribution can be evaluated by the Kullback-Leibler information (Kullback and Leibler (1951)) between the true distribution and the Bayesian predictive distribution.

Kitagawa (1997) discussed the Bayesian predictive distribution and its evaluation method. Kitagawa (1997) consider the multivariate linear Gaussian model with an unknown mean and a known covariance matrix. Moreover, in this framework, he showed that the Bayesian predictive distribution is Gaussian distribution, and he derived PIC (Predictive Information Criterion) for evaluation of the Bayesian predictive distribution. Here, a significant difference from the case of maximum likelihood estimation is that the covariance matrix is assumed to be known.

In this paper, we derived the new model selection criteria based on Kitagawa (1997) to evaluate the Bayesian predictive distribution in framework of nonlinear regression model. The aim of this paper is to introduce procedures for choosing the values of hyper-parameters in the prior distribution and the number of basis functions simultaneously. Moreover, we use prior distribution introduced by Denison *et al.* (2002) and the Bayesian predictive distribution which can be obtained when both regression coefficient and variance are assumed to be unknown (Bishop (2006)).

The paper is organized as follow. In section 2, we describe a framework of nonlinear regression model based on basis expansions. In section 3, we derive the Bayesian predictive distribution for the nonlinear regression model when both of regression coefficients and regression variance are unknown. In section 4, we propose new model selection criterion for evaluating goodness of the Bayesian predictive distribution based on Kitagawa (1997). In section 5, we investigate the performance of the proposed model selection criterion by real data analysis and Monte Carlo simulations. Some concluding remarks are described in Section 6.

2 Nonlinear regression model based on basis expansion method

Suppose that $\{(y_i, \boldsymbol{x}_i); i = 1, ..., n\}$ be n sets of data obtained in terms of the response variable y and p-dimensional explanatory variables $\boldsymbol{x} = (x_1, ..., x_p)^T$. In order to draw information from the data, we consider the Gaussian nonlinear regression model

$$y_i = u(\boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1}$$

where $u(\cdot)$ is true smooth function and errors ε_i are sets of independent zero-mean normal samples with variance σ^2. Here, we consider basis expansions approximation. That is, the problem is to estimate the function $u(\cdot)$ from the observed data, and we use the basis functions. In the basis expansion approach (Eilers and Marx, 1996), the unknown $u(\boldsymbol{x}_i)$ is approximated by a linear combination of basis functions

$$u(\boldsymbol{x}_i) = w_0 + \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{x}_i) = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i), \tag{2}$$

where $\boldsymbol{\phi}(\boldsymbol{x}) = (1, \phi_1(\boldsymbol{x}), ..., \phi_m(\boldsymbol{x}))^T$ is a vector of basis functions and $\boldsymbol{w} = (w_0, w_1, ..., w_m)^T$ is an unknown coefficient parameter vector.

For several times, splines (Green and Silverman (1994)), B-splines (de Boor (2001), Imoto and Konishi (2003)) and radial basis functions (Bishop (1995), Ripley (1996)) are used here for basis functions.

From equation (1) and (2), for n independent observations, the nonlinear regression model based on basis functions $\phi_j(\boldsymbol{x})$ $(j = 1, ..., m)$ is expressed as

$$y_i = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, ..., n. \tag{3}$$

Since the nonlinear regression model (3) has a probability density function

$$f(y_i|\boldsymbol{x}_i; \boldsymbol{w}, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left[-\frac{\left\{y_i - \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i)\right\}^2}{2\sigma^2}\right], \quad i = 1, ..., n. \tag{4}$$

The unknown parameters in the model are the set of coefficients, $\boldsymbol{w} = (w_0, w_1, \cdots, w_m)^T$ and the regression variance σ^2.

3 Estimation

For n independent observations $\{(y_i, \boldsymbol{x}_i); i = 1, ..., n\}$, the nonlinear regression model based on basis expansions given in section 2 is expressed as

$$y_i = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, ..., n. \tag{5}$$

where $\boldsymbol{\phi}(\boldsymbol{x}_i) = (1, \phi_1(\boldsymbol{x}_i), ..., \phi_m(\boldsymbol{x}_i))^T$, $\boldsymbol{w} = (w_0, w_1, ..., w_m)^T$ and ε_i are error terms.

Then the maximum likelihood estimates of the coefficient vector $\boldsymbol{w}$ and σ^2 are respectively given by

$$\hat{\boldsymbol{w}}_{\text{MLE}} = (\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}, \quad \hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_{\text{MLE}})^T(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_{\text{MLE}}), \tag{6}$$

where $\mathbf{1} = (1, 1, ..., 1)^T$, $\Phi = (\mathbf{1}, \boldsymbol{\phi}(\boldsymbol{x}_1), \boldsymbol{\phi}(\boldsymbol{x}_2), \cdots, \boldsymbol{\phi}(\boldsymbol{x}_n))^T$ and $\boldsymbol{y} = (y_1, ..., y_n)^T$. However, when fitting a nonlinear regression model to data with a complex structure, the maximum likelihood method often yields unstable estimation and leads to overfitting. We therefore estimate $\boldsymbol{w}$, σ^2 by maximizing the posterior distribution,

$$\pi(\boldsymbol{w}, \sigma^2 \ |\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{w}, \sigma^2)\pi(\boldsymbol{w}, \sigma^2)}{\int\int f(\boldsymbol{y}|\boldsymbol{w}, \sigma^2)\pi(\boldsymbol{w}, \sigma^2)d\boldsymbol{w}d\sigma^2} \tag{7}$$

where, $\pi(\boldsymbol{w}, \sigma^2)$ is a prior distribuion. Moreover, we consider the parametric model given by equation (5), and make assume for the known variance. One of the typical forms for the conjugate prior distribution is given by a Gaussian distribution of the form,

$$\boldsymbol{w} \sim N(\mathbf{0}, (n\lambda)^{-1}I_m). \tag{8}$$

Then, the posterior distribution $\pi(\boldsymbol{w}|\boldsymbol{y})$ is still Gaussian distribution of mean vector $\boldsymbol{\xi}$ and variance-covariance matrix V,

$$\boldsymbol{\xi} = (\Phi^T\Phi + n\lambda\sigma^2 I_m)^{-1}\Phi^T\boldsymbol{y}, \quad V = \sigma^2(\Phi^T\Phi + n\lambda\sigma^2 I_m)^{-1}, \tag{9}$$

where λ is a hyperparameter. Then, we use the MAP estimator which is well-known as ridge estimator (Hoerl and Kennard (1970)),

$$\hat{\boldsymbol{w}} = (\Phi^T\Phi + n\lambda\sigma^2 I_m)^{-1}\Phi^T\boldsymbol{y}. \tag{10}$$

In many cases, we use the plug-in type model $f(\boldsymbol{y}|\hat{\boldsymbol{w}}, \hat{\sigma}^2)$ as a statistical model which is a nonlinear regression model based on basis expansion method. However, in this paper, we consider the Bayesian predictive distribution (Bishop (2006)) defined as

$$f(\boldsymbol{z}|\boldsymbol{y}) = \int\int f(\boldsymbol{z}|\boldsymbol{w}, \sigma^2)\pi(\boldsymbol{w}, \sigma^2 \ |\boldsymbol{y})d\boldsymbol{w}d\sigma^2. \tag{11}$$

where $\boldsymbol{z} = (z_1, z_2, ..., z_n)^T$ is a n-dimensional future data vector generated independently by the observed $\boldsymbol{y} = (y_1, y_2, ..., y_n)^T$.

Moreover, if regression variance σ^2 is known and we use the assumption of equation (3) and (8), as a result, the Bayesian predictive distribution is given by Gaussian distribution of mean vector $\boldsymbol{\mu}$ and variance-covariance matrix Σ analythically,

$$\boldsymbol{\mu} = \Phi\hat{\boldsymbol{w}}_n, \quad \Sigma = \sigma^2(2\Phi\Phi^T + n\lambda\sigma^2 I_n)(\Phi\Phi^T + n\lambda\sigma^2 I_n)^{-1}. \tag{12}$$

We consider another type of conjugate prior distribution, when both $\boldsymbol{w}$ and σ^2 are unknown, and one of correspondingly the conjugate prior distribution $\pi(\boldsymbol{w}, \sigma^2)$ which gives a Gaussian-inverse gamma distribution (Denison *et. al.* (2002)).

Now, we assume that the prior on the regression coefficients given the regression variance is a Gaussian distribution, and the prior on the regression variance is inverse-gamma distribution, by its specified following parameters.

$$\boldsymbol{w}|\sigma^2 \sim N(\mathbf{0}, \sigma^2(n\lambda I_m)^{-1}), \quad \sigma^2 \sim IG\left(\nu_0/2, \eta_0/2\right). \tag{13}$$

As a consequence of our assumptions, posterior distribution on the regression coefficients giving the regression variance is given by a Gaussian distribution, and the posterior distribution on the regression variance is an inverse-gamma distribution respectively as

$$\boldsymbol{w}|\sigma^2, \boldsymbol{y} \sim N(\hat{\boldsymbol{w}}_n, \sigma^2 \hat{A}_n), \quad \sigma^2|\boldsymbol{y} \sim IG\left(\hat{\nu}_n/2, \hat{\eta}_n/2\right), \tag{14}$$

where

$$\begin{aligned}
\hat{\eta}_n &= \eta_0 + (\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_{\text{MLE}})^T(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_{\text{MLE}}) + \hat{\boldsymbol{w}}_{\text{MLE}}^T \left\{n\lambda I_m + (\Phi^T\Phi)^{-1}\right\}^{-1} \hat{\boldsymbol{w}}_{\text{MLE}}, \\
\hat{\boldsymbol{w}}_n &= (\Phi^T\Phi + (n\lambda)^{-1}I_m)^{-1}\Phi^T\Phi\hat{\boldsymbol{w}}_{\text{MLE}} \\
\hat{A}_n &= (\Phi^T\Phi + n\lambda I_m)^{-1}, \quad \hat{\nu}_n = n + \nu_0, \ .
\end{aligned} \tag{15}$$

In this case, according to Bishop (2006), the predictive distribution was given by student t-distribution. The Bayesian predictive distribution is Student t-distribution with $\hat{\nu}_n$ degrees of freedom ,

$$\boldsymbol{z}|\boldsymbol{y} \sim St(\hat{\boldsymbol{\mu}}, \hat{\Sigma}^*, \hat{\nu}_n) \tag{16}$$

where, $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}^*$ are given by

$$\hat{\boldsymbol{\mu}} = \Phi\hat{\boldsymbol{w}}_n, \quad \hat{\Sigma}^* = \frac{\hat{\lambda}_n}{\hat{\nu}_n}(\Phi\hat{A}_n\Phi^T + I_n), \tag{17}$$

respectively. The Student t-distribution is well known to have more probability in its tails than a normal distribution with the same location and scale parameters.

Futhermore, by the assumption of equation (3) and (13), the Bayesian predictive distribution for nonlinear regression model based on basis expansion, $h(\boldsymbol{z}|\boldsymbol{y}; \lambda, \eta, \nu)$ is given by

$$\begin{aligned}
h(\boldsymbol{z}|\boldsymbol{y}; \lambda, \eta, \nu) = &\left\{\Gamma\left((n + \hat{\nu}_n)/2\right)/\Gamma\left(\hat{\nu}_n/2\right)\left(\pi\hat{\nu}_n\right)^{\frac{n}{2}}\right\} \left|(\hat{\lambda}_n/\hat{\nu}_n)(\Phi\hat{A}_n\Phi^T + I_n)\right|^{-\frac{1}{2}} \\
&\times \left[1 + (1/\hat{\nu}_n)(\boldsymbol{z} - \Phi\hat{\boldsymbol{w}}_n)^T \left\{(\hat{\lambda}_n/\hat{\nu}_n)(\Phi\hat{A}_n\Phi^T + I_n)\right\}^{-1} (\boldsymbol{z} - \Phi\hat{\boldsymbol{w}}_n)\right]^{-\left(\frac{n+\hat{\nu}_n}{2}\right)}
\end{aligned} \tag{18}$$

4 Model selection criterion

In the Bayesian nonlinear regression model, we have crucial problems to solve, such as the choices of a hyper-parameter λ for the prior distribution, and the number of basis functions m. Kitagawa (1997) proposed the predictive information criterion (PIC) for evaluating goodness of the Bayesian predictive distribution in framework of multivariate Gaussian linear model, under the assumption of unknown regression coefficient and known variance. According to Kitagawa (1997), PIC is given by

$$\mathrm{PIC} = -2 \log h(\boldsymbol{y}|\boldsymbol{y}, \lambda) + 2B_p \tag{19}$$

where B_p is bias term of the expected log-likelihood estimated by the log-likelihood.

In the case of the Bayesian predictive distribution (12) in section 3, according to Kitagawa (1997), PIC is derived as

$$\mathrm{PIC} = n \log(2\pi) + \log|\Sigma| + (\boldsymbol{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) + 2\mathrm{tr}\left\{(2\Phi\Phi^T + n\lambda\sigma^2 I)^{-1}\Phi\Phi^T\right\}$$

where,

$$\boldsymbol{\mu} = \Phi(\Phi^T\Phi + n\lambda\sigma^2 I)^{-1}\Phi^T\boldsymbol{y}, \quad \Sigma = \sigma^2(2\Phi\Phi^T + n\lambda\sigma^2 I)(\Phi\Phi^T + n\lambda\sigma^2 I)^{-1}. \tag{20}$$

In this paper, we derive the information criterion for the Bayesian predictive distribution in framework of nonlinear regression model where both of the regression coefficient and variance are unknown.

Since the Bayesian predictive distribution (18) which is derived in section 3, we consider applying the Laplace approximation method.

We use the Laplace's approximation method (Tierney and Kadane (1986), Davison (1986)) which has been extensively investigated as a useful tool for approximation the Bayesian predictive distributions. It aims to find a Gaussian approximation to a probability density defined over a set of continuous variables.

According to Konishi and Kitagawa (2008), if we regard $\hat{\boldsymbol{w}}_\star$ and $\hat{\sigma}^2_\star$ as the model of $n^{-1}\{\log f(\boldsymbol{y}|\boldsymbol{w}, \sigma^2) + \log \pi(\boldsymbol{w}, \sigma^2)\}$, the Laplace approximation method yields the Bayesian predictive distribution in framework of nonlinear regression model, in the form $h(\boldsymbol{z}|\boldsymbol{y}) = f(\boldsymbol{z}|\hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star)(1 + O_p(n^{-1}))$. Moreover, in this framework, we can analytically calculate mode of (14),

$$\hat{\boldsymbol{w}}_\star = (\Phi^T\Phi + n\lambda I_m)^{-1}\Phi^T\boldsymbol{y}, \ \hat{\sigma}^2_\star = C\left\{(\boldsymbol{y} - \Phi\boldsymbol{w}_\star)^T(\boldsymbol{y} - \Phi\boldsymbol{w}_\star) + n\lambda\boldsymbol{w}_\star^T\boldsymbol{w}_\star + \nu_0\right\}.$$

where $C = 1/(n + m + \nu_0 + 2)$.

By using the nonlinear regression model based on basis expansion methods, our proposed model selection criterion is given as

$$\mathrm{PIC} = n\log(2\pi) + n\log(\hat{\sigma}_\star^2) + \frac{1}{\hat{\sigma}_\star^2}(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_\star)^T(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_\star)$$
$$+2\left(\frac{\sigma^{2*}}{\hat{\sigma}_\star^2}\right)\mathrm{tr}\left[\Phi(\Phi^T\Phi + n\lambda I_m)^{-1}\Phi^T\right]. \qquad (21)$$

However, our proposed model selection criterion contains the unknown true regression variance σ^{2*}. Then, we use two types of (21), actually. We first consider the true variance, substituting the maximum likelihood estimator of the variance estimated from the data. As a consequence, PIC results can be expressed as following:

$$\mathrm{PIC}_{\mathrm{MLE}} = n\log(2\pi) + n\log(\hat{\sigma}_\star^2) + \frac{1}{\hat{\sigma}_\star^2}(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_\star)^T(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_\star)$$
$$+2\left(\frac{\hat{\sigma}_{\mathrm{MLE}}^2}{\hat{\sigma}_\star^2}\right)\mathrm{tr}\left[\Phi(\Phi^T\Phi + n\lambda I_m)^{-1}\Phi^T\right]. (22)$$

We next consider that the true variance substituting the mode of the posterior distribution.

$$\mathrm{PIC}_{\mathrm{Mode}} = n\log(2\pi) + n\log(\hat{\sigma}_\star^2) + \frac{1}{\hat{\sigma}_\star^2}(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_\star)^T(\boldsymbol{y} - \Phi\hat{\boldsymbol{w}}_\star)$$
$$+2\mathrm{tr}\left[\Phi(\Phi^T\Phi + n\lambda I_m)^{-1}\Phi^T\right]. (23)$$

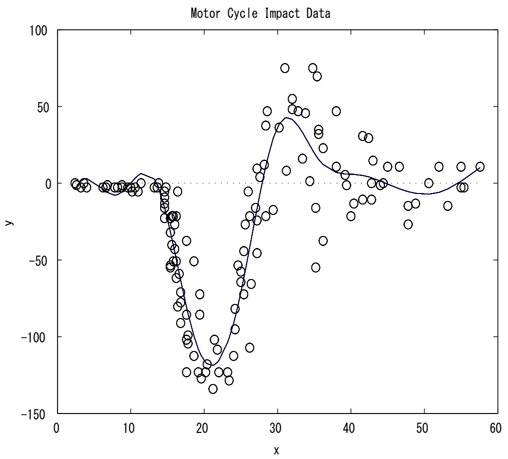
We select the optimal values of the number of basis functions and a hyper-parameter that minimize PIC.

5 numerical example

5.1 analysis of real data

We illustrate the proposed procedure to choose the smoothing parameter and the number of basis functions through the analysis of the motorcycle impact data (Silverman (1985), Härdle (1990), Eilers and Marx (1996)). The motorcycle impact data were simulated to investigate the efficacy of crash helmets and it comprised a series of measurements of head acceleration in units of gravity and times in milliseconds after impact.

We fit the Bayesian predictive distribution based on B-spline regression model with Gaussian noise (1) to the motorcycle impact data. Then we choose the number of basis functions m and the smoothing parameter λ that minimize the information criterion for the Bayesian predictive distribution $\mathrm{PIC}_{\mathrm{MLE}}$ and $\mathrm{PIC}_{\mathrm{Mode}}$ given by equation (22), (23). For the analysis of the motorcycle impact data, we set the candidate values of m and λ to $\{4, ..., 15\}$ and $\left\{10^{10(i-100)/99}; i = 1, ..., 100\right\}$, respectively. The criterion $\mathrm{PIC}_{\mathrm{MLE}}$ selected $m = 13$ and $\lambda = 1.62 \times 10^{-9}$, while $\mathrm{PIC}_{\mathrm{Mode}}$ selected $m = 13$ and $\lambda = 2.59 \times 10^{-9}$. The corresponding fitted curve is shown in Fig. 2 (solid curve).



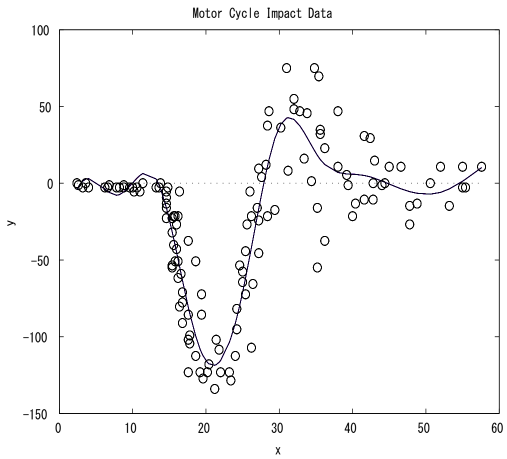


Figure 1: Real data examples: The motorcycle impact data: The left solid curved is the estimated curve based on $\mathrm{PIC}_{\mathrm{MLE}}$, The right solid curve is the estimated curve based on $\mathrm{PIC}_{\mathrm{Mode}}$.

5.2 curve fitting

For the second study, repeated random samples $\{(x_i, y_i); i = 1, ..., n\}$ with $n = 50,\ 100,\ 300$, were generated from a true regression model $y_i = u(x_i) + \varepsilon_i$. The design point x_i are uniformly distributed in $[0, 1]$ and the errors ε_i are independently, normally distributed with mean 0 and variance τ^2, where the standard deviation is taken as $\tau = 0.15R_y$ or $0.3R_y$ with R_y being the range of $u(x)$ over $x \in [0, 1]$. For the analysis of the Simulated data, we divide $[10^{-8}, 10^0]$ to 100 equal subintervals by points $\lambda_i (i = 0, ..., 100)$. We considered the following two cases for the true regression model:

$$u(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4, \tag{24}$$

$$u(x) = \sin(2\pi x^3). \tag{25}$$

We performed 300 repetitions, and then calculated average squared errors (ASE) defined by ASE $= \sum_{i=1}^{n} \{u(x_i) - \hat{y}_i\}^2 / n$ and predictive average squared errors (PASE) defined by PASE$=\sum_{i=1}^{n} \{z_i - \hat{y}_i\}^2 / n$ and deviations to assess the goodness of fit, respectively.

Where, z is the future observation data that generated from true regression model. Table 1 displays simulation results with ASE, PASE, the number of basis function m and hyper parameter λ.

Simulation results may be summarized as follows: Our proposed information criterion, $\mathrm{PIC}_{\mathrm{MLE}}$ and $\mathrm{PIC}_{\mathrm{Mode}}$, generally give good estimates in the sense of ASE and PASE, and yields stable hyper parameter estimates.

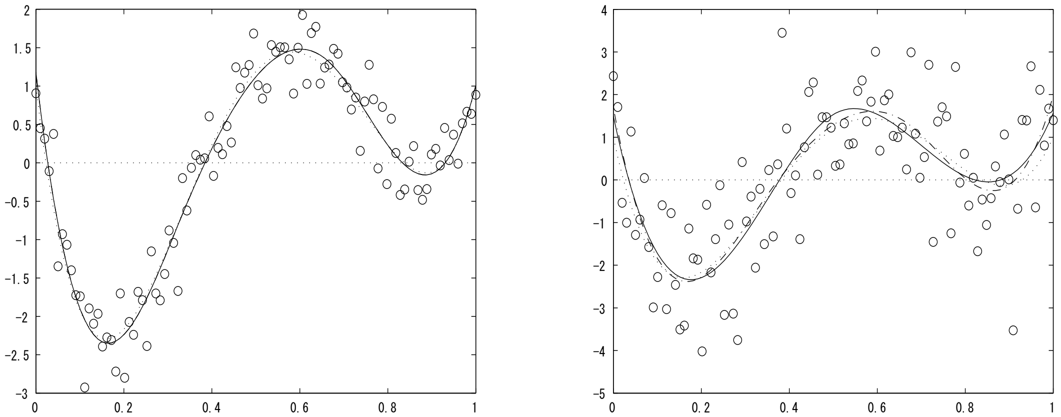


Figure 2: Examples of simulated data ($n = 100$; left: $\tau = 0.15R_y$; right: $\tau = 0.3R_y$): The dashed curve is the true regression curve $u(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$, while the solid curved and dot-dashed curvethe are estimated curve based on $\mathrm{PIC}_{\mathrm{MLE}}$ and $\mathrm{PIC}_{\mathrm{Mode}}$, respectively.

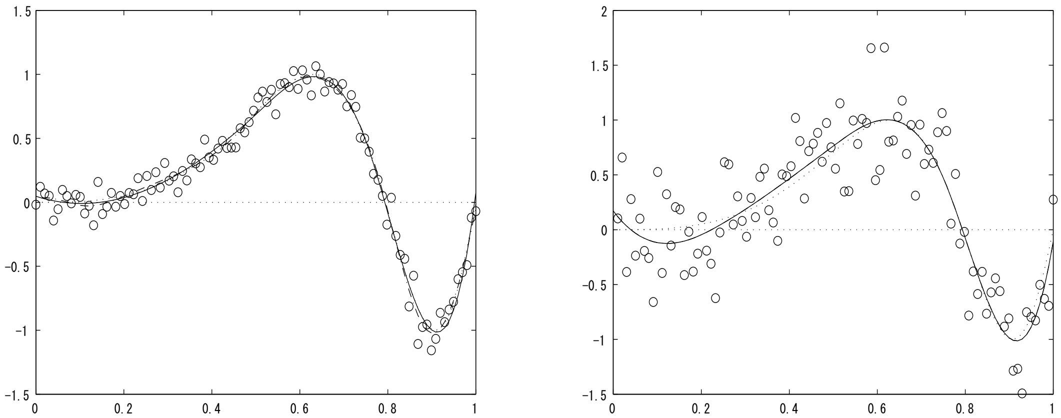


Figure 3: Examples of simulated data ($n = 100$; left: $\tau = 0.15R_y$; right: $\tau = 0.3R_y$): The dashed curve is the true regression curve $u(x) = \sin(2\pi x^3)$, while the solid curved and dot-dashed curvethe are estimated curve based on $\mathrm{PIC}_{\mathrm{MLE}}$ and $\mathrm{PIC}_{\mathrm{Mode}}$, respectively.

Table 1: $\tau^2 = 0.15$

Criterion	**Basis mean (SD)**	**hyper paramete mean (SD)**	**AMSE mean (SD)**	**APSE mean (SD)**
($n = 50$)				
GIC	8.37	3.03×10^{-4}	2.20×10^{-2}	1.19×10^{-1}
	(3.06)	(1.70×10^{-2})	(1.28×10^{-2})	(2.55×10^{-2})
mAIC	7.61	6.73×10^{-4}	2.04×10^{-2}	1.18×10^{-1}
	(2.79)	(2.70×10^{-3})	(1.22×10^{-1})	(2.48×10^{-2})
$\mathrm{PIC}_{\mathrm{MLE}}$	6.57	4.04×10^{-4}	1.79×10^{-2}	1.16×10^{-1}
	(2.12)	(2.00×10^{-3})	(1.11×10^{-2})	(2.45×10^{-2})
$\mathrm{PIC}_{\mathrm{Mode}}$	7.14	1.00×10^{-8}	1.93×10^{-2}	1.17×10^{-1}
	(2.38)	(6.29×10^{-23})	(1.19×10^{-2})	(2.45×10^{-2})
($n = 100$)				
GIC	7.88	3.03×10^{-4}	1.03×10^{-2}	1.076×10^{-1}
	(2.60)	(1.70×10^{-3})	(0.62-$\times 10^{-2}$)	(1.64-$\times 10^{-2}$)
mAIC	7.45	4.71×10^{-4}	1.00×10^{-2}	1.073×10^{-1}
	(2.47)	(2.10×10^{-1})	(6.30×10^{-3})	(1.67×10^{-2})
$\mathrm{PIC}_{\mathrm{MLE}}$	6.60	1.00×10^{-8}	9.20×10^{-3}	1.064×10^{-1}
	(1.68)	(6.29×10^{-23})	(5.50×10^{-3})	(1.66×10^{-2})
$\mathrm{PIC}_{\mathrm{Mode}}$	7.16	1.00×10^{-8}	9.60×10^{-3}	1.070×10^{-1}
	(2.14)	(6.29×10^{-23})	(5.90×10^{-3})	(1.69×10^{-2})
($n = 300$)				
GIC	7.53	3.36×10^{-5}	3.20×10^{-3}	1.018×10^{-1}
	(1.86)	(5.83×10^{-4})	(1.90×10^{-3})	(9.00×10^{-3})
mAIC	7.47	3.36×10^{-5}	3.20×10^{-3}	1.018×10^{-1}
	(1.76)	(5.83×10^{-4})	(1.90×10^{-3})	(9.00×10^{-3})
$\mathrm{PIC}_{\mathrm{MLE}}$	7.24	1.00×10^{-8}	3.10×10^{-3}	1.017×10^{-1}
	(1.52)	(6.29×10^{-23})	(1.80×10^{-3})	(8.90×10^{-3})
$\mathrm{PIC}_{\mathrm{Mode}}$	7.43	1.00×10^{-8}	3.20×10^{-3}	1.018×10^{-1}
	(1.71)	(6.29×10^{-23})	(1.90×10^{-3})	(9.00×10^{-3})

Table 2: $\tau^2 = 0.3$

Criterion	**number mean(SD)**	**hyper parameter mean(SD)**	**AMSE mean(SD)**	**APSE mean(SD)**
$(n = 50)$				
GIC	8.46	1.60×10^{-3}	3.28×10^{-1}	1.8932
	(3.57)	(3.70×10^{-3})	(2.15×10^{-1})	(4.14×10^{-1})
mAIC	8.14	2.90×10^{-3}	2.85×10^{-1}	1.8539
	(3.53)	(4.60×10^{-3})	(1.87×10^{-1})	(3.88×10^{-1})
$\mathrm{PIC}_{\mathrm{MLE}}$	9.75	7.80×10^{-3}	2.47×10^{-1}	1.8197
	(4.08)	(6.80×10^{-3})	(1.28×10^{-1})	(3.73×10^{-1})
$\mathrm{PIC}_{\mathrm{Mode}}$	7.94	3.10×10^{-3}	2.69×10^{-1}	1.8397
	(3.54)	(4.70×10^{-3})	(1.80×10^{-1})	(3.86×10^{-1})
$(n = 100)$				
GIC	7.13	7.40×10^{-4}	1.42×10^{-1}	1.6976
	(3.04)	(2.60×10^{-3})	$(1.00\text{-}\times10^{-1})$	$(2.64\text{-}\times10^{-1})$
mAIC	6.77	9.09×10^{-4}	1.32×10^{-1}	1.6876
	(2.86)	(2.90×10^{-23})	(9.59×10^{-2})	(2.61×10^{-1})
$\mathrm{PIC}_{\mathrm{MLE}}$	6.93	2.00×10^{-3}	1.21×10^{-1}	1.6794
	(3.15)	(4.00×10^{-3})	(7.97×10^{-2})	(2.60×10^{-1})
$\mathrm{PIC}_{\mathrm{Mode}}$	6.55	7.40×10^{-4}	1.27×10^{-1}	1.6850
	(2.69)	(2.60×10^{-3})	(9.22×10^{-2})	(2.63×10^{-1})
$(n = 300)$				
GIC	6.98	1.50×10^{-4}	4.49×10^{-2}	1.6200
	(2.56)	(4.11×10^{-4})	(3.03×10^{-2})	(0.14)
mAIC	6.78	1.40×10^{-4}	4.37×10^{-2}	1.6199
	(2.46)	(3.90×10^{-4})	(2.88×10^{-2})	(0.14)
$\mathrm{PIC}_{\mathrm{MLE}}$	6.50	1.51×10^{-4}	4.14×10^{-2}	1.6178
	(2.20)	(4.61×10^{-4})	(2.77×10^{-2})	(0.14)
$\mathrm{PIC}_{\mathrm{Mode}}$	6.59	1.13×10^{-4}	4.26×10^{-2}	1.6188
	(2.2904)	(3.50×10^{-4})	(2.88×10^{-2})	(0.14)

Table 3: $\tau^2 = 0.15$

Criterion	**number mean(SD)**	**hyper parameter mean(SD)**	**AMSE mean(SD)**	**APSE mean(SD)**
$(n = 50)$				
GIC	10.45	7.90×10^{-3}	2.10×10^{-3}	9.90×10^{-3}
	(3.08)	(2.38×10^{-2})	(9.78×10^{-4})	(2.10×10^{-3})
mAIC	9.68	1.57×10^{-2}	2.00×10^{-3}	9.80×10^{-3}
	(3.06)	(3.47×10^{-2})	(9.86×10^{-4})	(2.10×10^{-3})
PIC_{MLE}	7.89	1.00×10^{-8}	1.70×10^{-3}	9.50×10^{-3}
	(1.76)	(6.29×10^{-23})	(8.99×10^{-4})	(2.00×10^{-3})
PIC_{Mode}	9.04	1.00×10^{-8}	1.90×10^{-3}	9.80×10^{-3}
	(2.70)	(6.29×10^{-23})	(8.59×10^{-4})	(2.10×10^{-3})
$(n = 100)$				
GIC	9.70	4.10×10^{-3}	1.00×10^{-3}	9.00×10^{-3}
	(2.96)	(1.27×10^{-2})	(4.87×10^{-4})	(1.30×10^{-3})
mAIC	9.05	5.70×10^{-3}	1.00×10^{-3}	9.00×10^{-3}
	(2.82)	(1.64×10^{-2})	(4.90×10^{-4})	(1.30×10^{-3})
PIC_{MLE}	8.04	1.00×10^{-8}	9.11×10^{-4}	8.90×10^{-3}
	(1.99)	(6.29×10^{-23})	(4.6×10^{-4})	(1.30×10^{-3})
PIC_{Mode}	8.60	1.00×10^{-8}	9.56×10^{-4}	8.90×10^{-3}
	(2.49)	(6.29×10^{-23})	(4.73×10^{-4})	(1.40×10^{-3})
$(n = 300)$				
GIC	10.31	8.75×10^{-4}	3.87×10^{-4}	8.40×10^{-3}
	(2.96)	(3.60×10^{-3})	(1.54×10^{-4})	(7.49×10^{-4})
mAIC	9.86	9.76×10^{-4}	3.85×10^{-4}	8.40×10^{-3}
	(2.92)	(4.00×10^{-3})	(1.50×10^{-4})	(7.49×10^{-4})
PIC_{MLE}	9.41	1.00×10^{-8}	3.74×10^{-4}	8.40×10^{-3}
	(2.79)	(6.29×10^{-23})	(1.43×10^{-4})	(7.50×10^{-4})
PIC_{Mode}	9.70	1.00×10^{-8}	3.77×10^{-4}	8.40×10^{-3}
	(2.84)	(6.29×10^{-23})	(1.48×10^{-4})	(7.48×10^{-4})

Table 4: $\tau^2 = 0.3$

Criterion	number mean(SD)	hyper parameter mean(SD)	AMSE mean(SD)	APSE mean(SD)
($n = 50$)				
GIC	9.85	8.10×10^{-3}	2.89×10^{-2}	1.548×10^{-1}
	(2.95)	(1.48×10^{-2})	(1.55×10^{-2})	(3.28×10^{-2})
mAIC	9.80	1.41×10^{-2}	2.72×10^{-2}	1.534×10^{-1}
	(2.97)	(1.89×10^{-2})	(1.40×10^{-2})	(3.21×10^{-2})
PIC_{MLE}	13.95	1.32×10^{-2}	2.89×10^{-2}	1.548×10^{-1}
	(1.35)	(5.00×10^{-3})	(1.14×10^{-2})	(3.18×10^{-2})
PIC_{Mode}	11.84	8.10×10^{-3}	2.94×10^{-2}	1.553×10^{-1}
	(2.86)	(5.40×10^{-3})	(1.32×10^{-4})	(3.21×10^{-2})
($n = 100$)				
GIC	8.96	4.70×10^{-3}	1.38×10^{-2}	1.41×10^{-1}
	(2.68)	(9.70×10^{-3})	(7.90×10^{-3})	(2.18×10^{-2})
mAIC	8.88	6.20×10^{-3}	1.35×10^{-2}	1.40×10^{-1}
	(2.64)	(1.07×10^{-2})	(7.70×10^{-3})	(2.19×10^{-2})
PIC_{MLE}	13.21	9.60×10^{-3}	1.48×10^{-2}	1.41×10^{-1}
	(2.09)	(2.40×10^{-3})	(6.20×10^{-3})	(2.13×10^{-2})
PIC_{Mode}	10.22	4.90×10^{-3}	1.42×10^{-2}	1.41×10^{-1}
	(3.13)	(5.10×10^{-3})	(7.60×10^{-3})	(2.18×10^{-2})
($n = 300$)				
GIC	8.58	1.70×10^{-3}	2.20×10^{-3}	6.45×10^{-2}
	(2.45)	(4.40×10^{-3})	(1.20×10^{-3})	(5.80×10^{-3})
mAIC	8.50	2.00×10^{-3}	2.20×10^{-3}	6.44×10^{-2}
	(2.45)	(4.70×10^{-3})	(1.30×10^{-3})	(5.80×10^{-3})
PIC_{MLE}	7.74	1.00×10^{-8}	2.00×10^{-3}	6.43×10^{-2}
	(1.59)	(6.29×10^{-23})	(1.30×10^{-3})	(5.80×10^{-3})
PIC_{Mode}	7.84	1.00×10^{-8}	2.00×10^{-3}	6.43×10^{-2}
	(1.70)	(6.29×10^{-23})	(1.30×10^{-3})	(5.80×10^{-3})

6 Concluding Remarks

This article has proposed the nonlinear regression modeling based on basis expansion technique. We have introduced the information criterion for the Bayesian predictive distribution in nonlinear regression model based on basis expansion method. In particular, we derived PIC, when both of regression coefficient and regression variance are unknown.

7 Appendix

7.1 Derivation of PIC

As introduced by Kitagawa (1997), PIC is the derivation of our case is as follows:

$$\mathrm{PIC} = -2\log h(\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star) + B_p(q(\cdot), \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star) \tag{26}$$

The log likelihood for the Bayesian predictive distribution, $\log h(\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star)$ is given by

$$\log h(\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\hat{\sigma}^2_\star I| - \frac{1}{2\hat{\sigma}^2_\star}(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)^T(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star). \tag{27}$$

Then, the diffence between the log likelihood for the Bayesian predictive distribution and expected loglikelihood is calcurated by

$$\begin{aligned}
&B_p(q(\cdot), \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star) \\
&= E_{q(\boldsymbol{y})}\left[\log h(\boldsymbol{y}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star) - E_{q(\boldsymbol{z})}\left[\log h(\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}^2_\star)\right]\right] \\
&= -\frac{1}{2\hat{\sigma}^2_\star}\mathrm{tr}\left\{E_{q(\boldsymbol{y})}\left[(\boldsymbol{y} - B\hat{\boldsymbol{w}}_\star)(\boldsymbol{y} - B\hat{\boldsymbol{w}}_\star)^T - E_{q(\boldsymbol{z})}\left[(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)^T\right]\right]\right\}.
\end{aligned} \tag{28}$$

Moreover we assume that true distribution $q(\boldsymbol{z})$ is given by

$$q(\boldsymbol{z}) = f(\boldsymbol{z}|\boldsymbol{w}^*, \sigma^{2*}) \tag{29}$$

where, $\boldsymbol{w}^* \in R^m$, $\sigma^{2*} \in R^1$. Then we denoted this fact as

$$\boldsymbol{z}|\boldsymbol{w}^*, \sigma^{2*} \sim N(B\boldsymbol{w}^*, \sigma^{2*}I). \tag{30}$$

Futhermore, The expected that the value of the difference between the log-likelihood function and expected log-likelihood function is evaluated as

$$\begin{aligned}
E_{q(\boldsymbol{z})}\left\{(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)^T\right\} &= E_{f(\boldsymbol{z}|\boldsymbol{w}^*, \sigma^{2*})}\left\{(\boldsymbol{z} - B\boldsymbol{w}^*)(\boldsymbol{z} - B\boldsymbol{w}^*)^T\right\} \\
&\quad + (B\boldsymbol{w}^* - B\hat{\boldsymbol{w}}_\star)(B\boldsymbol{w}^* - B\hat{\boldsymbol{w}}_\star)^T.
\end{aligned} \tag{31}$$

Moreover, we have,

$$
\begin{aligned}
B\boldsymbol{w}^* - B\hat{\boldsymbol{w}}_\star &= (BB^T + n\lambda I_n)^{-1} BB^T (B\boldsymbol{w}^* - \boldsymbol{y}) + n\lambda (BB^T + n\lambda I_n)^{-1} B\boldsymbol{w}^* \\
\boldsymbol{y} - B\hat{\boldsymbol{w}}_\star &= n\lambda (B^T B + n\lambda I_m)^{-1} (\boldsymbol{y} - B\boldsymbol{w}^*) + n\lambda (B^T B + n\lambda I_m)^{-1} B\boldsymbol{w}^*. \qquad (32)
\end{aligned}
$$

Consequently, we correct bias term.

$$
\begin{aligned}
& E_{q(\boldsymbol{y})} \left[\log h(\boldsymbol{y}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}_\star^2) - E_{q(\boldsymbol{z})} \left[\log h(\boldsymbol{z}|\boldsymbol{y}, \hat{\boldsymbol{w}}_\star, \hat{\sigma}_\star^2) \right] \right] \\
&= -\frac{1}{2\hat{\sigma}^2} \mathrm{tr} \left[E_{q(\boldsymbol{y})} \left[(\boldsymbol{y} - B\hat{\boldsymbol{w}}_\star)(\boldsymbol{y} - B\hat{\boldsymbol{w}}_\star)^T - E_{q(\boldsymbol{z})} \left[(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)(\boldsymbol{z} - B\hat{\boldsymbol{w}}_\star)^T \right] \right] \right] \\
&= -\left(\frac{\sigma^{2*}}{\hat{\sigma}_\star^2} \right) \mathrm{tr} \left[B(B^T B + n\lambda I_m)^{-1} B^T \right]. \qquad (33)
\end{aligned}
$$

References

[1] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford Univerity Press.

[2] Bishop. C. M. (2006). Pattern Recognition and Machine Learning. *Springer.*

[3] de Boor, C. (2001). *A Practical Guide to Splines.* Revised Edition. Springer.

[4] Davison, A. C. (1986). Approximate predictive likelihood. *Biometrika.*, **73**, 323–32.

[5] Denison, D. G. T., Holmes, C.C., Mallick, B. K. and Smith, A. F. M (2002). Bayesian Methods for Nonlinear Classification and Regression. *Wiley.*

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

[7] Good, I. J. and Gaskins, R. A. (1971). Non parametric roughness penalites for probability densities. *Biometrika*, **58**, 255–77.

[8] Green , P.J. and Silverman, B.W. (1994) *Non parametric Regression and Generalized Linear Models.* Chapman & Hall.

[9] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press.

[10] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased Estimation for Nonorthogonal problems, *Technometrics*, **12**, 55–67.

[11] Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning.* Springer.

[12] Imoto, S. and Konishi, S. (2003). Selection of smoothing parameters in B-spline nonparametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics*, **55**. 671–687.

[13] Kitagawa, G. (1997). Information criteria for the predictive evaluation of Bayesian models. *Communications in Statistics-Theory and Methods*, **26**, 2223–2246.

[14] Komaki, F. (1996). On asymptotic properties of predictive distributions. *Biometrika*, **83**, 299–313.

[15] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer.

[16] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.

[17] Ripley, B. D. (1996). *Pattern Recognition and Nwural Networks.* Cambridge University Press.

[18] Silverman, B. W. (1985). Some aspects of the spline smoothing approzch to nonparametric regression curve fitting (with discussion). *Journal of the Royal Statistic Society*, B **36**, 1–52.

[19] Tierney, L and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82–86.

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The intial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in nfinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric τ-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications

MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2-projection

MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)

MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance

MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models

MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons

MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine

MI2010-6 Shuichi KAWANO, Toshihiro MISUMI & Sadanori KONISHI
Semi-supervised logistic discrimination via graph-based regularization

MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation

MI2010-8 Takahiro ITO
Abstract collision systems on groups

MI2010-9 Hiroshi YOSHIDA, Kinji KIMURA, Naoki YOSHIDA, Junko TANAKA & Yoshihiro MIWA
An algebraic approach to underdetermined experiments

MI2010-10 Kei HIROSE & Sadanori KONISHI
Variable selection via the grouped weighted lasso for factor analysis models

MI2010-11 Katsusuke NABESHIMA & Hiroshi YOSHIDA
Derivation of specific conditions with Comprehensive Groebner Systems

MI2010-12 Yoshiyuki KAGEI, Yu NAGAFUCHI & Takeshi SUDOU
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Poiseuille type flow

MI2010-13 Reiichiro KAWAI & Hiroki MASUDA
On simulation of tempered stable random variates

MI2010-14 Yoshiyasu OZEKI
Non-existence of certain Galois representations with a uniform tame inertia weight

MI2010-15 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of a Rotating Flow Driven by Precession of Arbitrary Frequency

MI2010-16 Yu KAWAKAMI & Daisuke NAKAJO
The value distribution of the Gauss map of improper affine spheres

MI2010-17 Kazunori YASUTAKE
On the classification of rank 2 almost Fano bundles on projective space

MI2010-18 Toshimitsu TAKAESU
Scaling limits for the system of semi-relativistic particles coupled to a scalar bose field

MI2010-19 Reiichiro KAWAI & Hiroki MASUDA
Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling

MI2010-20 Yasuhide FUKUMOTO, Makoto HIROTA & Youichi MIE
Lagrangian approach to weakly nonlinear stability of an elliptical flow

MI2010-21 Hiroki MASUDA
Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test

MI2010-22 Toshimitsu TAKAESU
A Generalized Scaling Limit and its Application to the Semi-Relativistic Particles System Coupled to a Bose Field with Removing Ultraviolet Cutoffs

MI2010-23 Takahiro ITO, Mitsuhiko FUJIO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Composition, union and division of cellular automata on groups

MI2010-24 Toshimitsu TAKAESU
A Hardy's Uncertainty Principle Lemma in Weak Commutation Relations of Heisenberg-Lie Algebra

MI2010-25 Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators

MI2010-26 Reiichiro KAWAI & Hiroki MASUDA
On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling

MI2010-27 Chikashi ARITA & Daichi YANAGISAWA
Exclusive Queueing Process with Discrete Time

MI2010-28 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Motion and Bäcklund transformations of discrete plane curves

MI2010-29 Takanori YASUDA, Masaya YASUDA, Takeshi SHIMOYAMA & Jun KOGURE
On the Number of the Pairing-friendly Curves

MI2010-30 Chikashi ARITA & Kohei MOTEGI
Spin-spin correlation functions of the q-VBS state of an integer spin model

MI2010-31 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and spike detection via Gaussian basis expansions

MI2010-32 Nobutaka NAKAZONO
Hypergeometric τ functions of the q-Painlevé systems of type $(A_2 + A_1)^{(1)}$

MI2010-33 Yoshiyuki KAGEI
Global existence of solutions to the compressible Navier-Stokes equation around parallel flows

MI2010-34 Nobushige KUROKAWA, Masato WAKAYAMA & Yoshinori YAMASAKI
Milnor-Selberg zeta functions and zeta regularizations

MI2010-35 Kissani PERERA & Yoshihiro MIZOGUCHI
Laplacian energy of directed graphs and minimizing maximum outdegree algorithms

MI2010-36 Takanori YASUDA
CAP representations of inner forms of $Sp(4)$ with respect to Klingen parabolic subgroup

MI2010-37 Chikashi ARITA & Andreas SCHADSCHNEIDER
Dynamical analysis of the exclusive queueing process

MI2011-1 Yasuhide Fukumoto & Alexander B. Samokhin
Singular electromagnetic modes in an anisotropic medium

MI2011-2 Hiroki Kondo, Shingo Saito & Setsuo Taniguchi
Asymptotic tail dependence of the normal copula

MI2011-3 Takehiro Hirotsu, Hiroki Kondo, Shingo Saito, Takuya Sato, Tatsushi Tanaka & Setsuo Taniguchi
Anderson-Darling test and the Malliavin calculus

MI2011-4 Hiroshi Inoue, Shohei Tateishi & Sadanori Konishi
Nonlinear regression modeling via Compressed Sensing

MI2011-5 Hiroshi Inoue
Implications in Compressed Sensing and the Restricted Isometry Property

MI2011-6 Daeju Kim & Sadanori Konishi
Predictive information criterion for nonlinear regression model based on basis expansion methods