行政院國家科學委員會專題研究計畫 期末報告

無母數迴歸使用 B-spline 時選取 knot 的一種方式

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 101-2118-M-004-001-
執 行 期 間 ：101 年 08 月 01 日至 102 年 07 月 31 日
執 行 單 位 ：國立政治大學統計學系

計畫主持人：黃子銘

計畫參與人員：碩士班研究生-兼任助理人員：何昕燁
碩士班研究生-兼任助理人員：江元淳
碩士班研究生-兼任助理人員：王彥智
博士班研究生-兼任助理人員：鄭宇翔

公 開 資 訊 ：本計畫可公開查詢

中 華 民 國 102 年 07 月 18 日

中 文 摘 要 ： 本研究中提出一種選取 spline 節點的方式. 使用此方法選取 spline 節點可得到迴歸函數的適應性估計, 並且此方法容許使用非等距排放的節點.

中文關鍵詞： 適應性估計, 迴歸樣條, 節點選取

英 文 摘 要 ： A knot selection method for regression splines is proposed. This method yields a least square spline estimator that adapts to the smoothness of the regression function, and non-equally-spaced knots are allowed.

英文關鍵詞： adaptive estimation, regression splines, knot selection

An adaptive knot selection method for regression splines via penalized minimum contrast estimation*

Tzee-Ming Huang

July 18, 2013

Abstract

In this report, a knot selection method for regression splines is proposed. This method yields a least square spline estimator that adapts to the smoothness of the regression function, and the knots are allowed to be not equally spaced. If the true regression function s belongs to a Sobolev space $W_m^2[0,1]$, then the proposed estimator can converge to s at the rate $O(a_n n^{-m/(1+2m)})$ in terms of L^2 norm in probability, where a_n can be any sequence such that $\lim_{n\to\infty} a_n = \infty$.

1 Introduction

One of the most popular methods in non-parametric regression is B-spline estimation. B-splines are piecewise polynomials joined smoothly at points called knots. For implementation, one has to choose the number of knots and the degrees of polynomials. The choice of knots is especially crucial. For functions that are m times continuously differentiable with the m-th derivatives bounded by a constant, Zhou, Shen and Wolfe [8] showed that the number of knots should grow at the rate $n^{1/(1+2m)}$ for the spline estimator for the regression function to achieve the optimal convergence rate $n^{-2m/(1+2m)}$ in integrated mean squared error.

It is possible to construct estimators for regression functions in the Sobolev space $W_m^2[0,1]$ that can achieve the rate $n^{-m/(1+2m)}$ with respect to the L^2 norm without knowing m. These estimators are known as adaptive estimators. Barron, Birgé and Massart [1] derived risk bounds for penalized minimum contrast estimators, which can be used to construct adaptive estimators for regression functions. Huang [4] applied an inequality in Yang and Barron [7], which is obtained using a similar approach in [1], to construct

*This research is supported by National Science Council in Taiwan, R.O.C. Grant number: NSC 101-2118-M-004-001-.

an adaptive estimator for regression function using B-splines with equally spaced knots assuming the errors are normally distributed.

The objective of this study is to construct an adaptive estimator for regression function using B-splines without requiring that the knots are equally spaced or that the errors are normally distributed. This objective is achieved by first establishing an exponential inequality to control the error of minimum contrast estimators and then applying the result to derive convergence rate for a spline estimator obtained via model selection. The exponential inequality is given in Section 2. The application to adaptive B-spline estimation is given in Section 3.

2 An Exponential Inequality for Minimum Contrast Estimators

In this section, an exponential inequality that gives error bounds for minimum contrast estimators is presented. The problem set-up for minimum contrast estimators is this: Consider the problem of estimating an unknown function s based on observed variables Z_1, …, Z_n using

$$\tilde{s} = \text{argmin}_{t \in S} n^{-1} \sum_{i=1}^{n} \gamma(Z_i, t).$$

Then the function γ is called the contrast function and the estimator $\tilde{s}$ is called the minimum contrast estimator over S with respect to γ. In the least square regression framework, suppose that (X_1, Y_1), …, (X_n, Y_n) are observed and

$$Y_i = s(X_i) + W_i, \qquad i = 1, \ldots, n, \tag{1}$$

where s is defined on an interval I_0, W_i are errors of mean zero that are independent of the X_i's. Consider estimating s using functions in a parametric family S, then the least square estimator $\tilde{s}$ is a minimum contrast estimator with $Z_i = (X_i, Y_i)$ with respect to the contrast function $\gamma(z, t) = (y - t(x))^2$ for $z = (x, y)$. Let

$$\nu_n(t) = n^{-1} \sum_{i=1}^{n} [\gamma(Z_i, t) - E\gamma(Z_i, t)],$$

then the exponential inequality given in (4) in Theorem 1 in this section gives control of $\nu_n(s) - \nu_n(u)$ for all $u \in S$.

In the least regression framework, while the inequality in Lemma 0 in [7] can also be applied to give control of $\nu_n(s) - \nu_n(u)$ (see [4] for example), the exponential inequality given in (4) is more direct. Lemma 0 in [7] gives control of the likelihood function, and in [4], the control of $\nu_n(s) - \nu_n(u)$ is achieved by assuming the W_i's are normally distributed. In contrast, the inequality in (4) is derived following the proof of Theorem 5 in Birgé and Massart [2], which only requires moment conditions of the W_i's.

Let $\|\cdot\|$ and $\|\cdot\|_\infty$ be the L^2 norm and sup norm on I_0. Then the assumptions of Theorem 1 are as follows.

Assumption M1. Suppose that the observed variables Z_1, ..., Z_n can be expressed as $Z_i = f(s, X_i, W_i)$, where (X_1, W_1), ..., (X_n, W_n) are independent. Suppose that there exist non-negative constants A, B, A_2, $\tilde{B}$, k_0, non-negative functions M, Δ, M_2, Δ_2 such that for u, $v \in S$,

$$|\gamma(z,u) - \gamma(z,v)| \le M(w)\Delta(x,u,v),$$

$$|\gamma(z,u) - \gamma(z,s)| \le M_2(w)\Delta_2(x,u,s),$$

and, for all $m \ge 2$,

$$E_s[M^m(W_i)] \le a_m A^m, \text{ for all } i = 1, \dots, n,$$

$$\frac{1}{n}\sum_{i=1}^n E_s[\Delta^m(X_i,u,v)] \le b_m k_0 \|u-v\|^2 B^{m-2}$$

$$E_s[M_2^m(W_i)] \le a_m A_2^m, \text{ for all } i = 1, \dots, n,$$

$$\frac{1}{n}\sum_{i=1}^n E_s[\Delta_2^m(X_i,u,s)] \le b_m k_0 \|u-s\|^2 \tilde{B}^{m-2}$$

with either $a_m = 1$, $b_m = m!/2$, for all $m \ge 2$, or $b_m = 1$, $a_m = m!/2$, for all $m \ge 2$.

Assumption M2. There exist two constants $B' \ge 1$ and $r > 0$ such that, for $\sigma > 0$ and $0 < \delta < \sigma/5$, for any ball $\mathcal{B} \subset S$ of radius σ with respect to $\|\cdot\|$, one can find T: a subset of $\mathcal{B}$ such that $|T| \le (B'\sigma/\delta)^D$ for some $D \ge 1$ and for every $u \in \mathcal{B}$, there exists $v \in T$ such that

$$\|u - v\| \le \delta$$

and

$$\|\Delta(\cdot, u, v)\|_\infty \le r\delta.$$

Under the above two assumptions, we have the following result.

Theorem 1 *Suppose that Assumptions M1 and M2 hold. Suppose that* $B \geq 1$, $rB^2 \leq n/4$, $r \geq 1$, $A = A_0 + 2B$ *for some* $A_0 > 0$, *and* $A_2 = A_0 + B = 2\tilde{B}$. *Suppose that* $\|u\|_\infty \leq B$ *for all* $u \in S$. *Let*

$$c_1(t) = \frac{t}{(1+\sqrt{1+t})^2} \text{ and } c(t) = c_1\left(\frac{t}{k_0(A_0+2)}\right)$$

for $t > 0$. *Then, for* $n \geq A_0^2$, $\sigma > 0$, $\tau > 0$, *if*

$$\theta \geq \left(\frac{A_0+2}{c(8\tau)}\right) \vee 5 \tag{2}$$

and

$$n\eta > 24B^2 r(1 + 2D\log 2), \tag{3}$$

then we have

$$P_s\left[\nu_n[\gamma(\cdot,s) - \gamma(\cdot,u)] > 9\tau\left(\frac{\sigma^2}{4} \vee \|s-u\|^2\right) + 2\eta \text{ for some } u \in S\right]$$

$$\begin{aligned} \leq\ & \exp\left(-\frac{2c_1(\tau/k_0)n\eta}{(A_0+B)^2}\right) \\ & +1.6\left(1 - \exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau) \wedge \frac{1}{3}\right]\right)\right)^{-1} ((B'\theta) \vee 2)^{3D} \times \\ & \left[\exp\left(-\frac{nc(\tau)\eta}{2AB}\right) + \exp\left(-\frac{n\eta}{24B^2 r}\right)\right]. \end{aligned} \tag{4}$$

In addition, for $u \in S$ *and* $\eta \geq 0$, *we have*

$$P_s\left[\nu_n[\gamma(\cdot,u) - \gamma(\cdot,s)] > \tau\|u-s\|^2 + \eta\right] \leq \exp\left(-\frac{2c_1(\tau/k_0)n\eta}{(A_0+B)^2}\right). \tag{5}$$

The proof of Theorem 1 is similar to that of Theorem 5 in [2] and is omitted.

Note that Theorem 1 can be used to provide error bounds for penalized minimum contrast estimators. To see this, suppose that s can be approximated well by some function s^* in S_{j^*} for some j^* in an index set Λ. Suppose that there exist positive constants A_0 and k_0 such that for each $j \in \Lambda$, when $S = S_j$, the conditions in Theorem 1 holds with $A_0 = A_0$, $k_0 = k_0$, and $B = B_j$, $B' = B'_j$, $r = r_j$, $D = D_j$ for some B_j, B'_j, r_j and D_j. Let

$$\hat{s} = \text{argmin}_{j\in\Lambda, u\in S_j}\left(\frac{1}{n}\sum_{i=1}^n \gamma(Z_i,u) + \eta_{1,j}\right),$$

where $\eta_{1,j} \geq 0$ is chosen so that (3) holds with $\eta = \eta_{1,j}$, $B = B_j$, $r = r_j$ and $D = D_j$, and $\sum_{j\in\Lambda} p_j(\eta_{1,j}) < \infty$, where the function $p_j(\eta)$ is the upper bound for the probability in (4) with $S = S_j$, $B = B_j$, $B' = B'_j$, $r = r_j$ and $D = D_j$. The estimator $\hat{s}$ is called a penalized minimum contrast estimator for s.

To give error bound for $\|\hat{s} - s\|$ using Theorem 1, it is assumed that there exist positive constants k_1 and k_2 such that

$$k_1\|u - s\|^2 \leq E[\gamma(Z_i, u) - \gamma(Z_i, s)] \leq k_2\|u - s\|^2. \tag{6}$$

for all $u \in S_j$ for all $j \in \Lambda$. Also, it is assumed that there exists $\{\delta_n\}$: a sequence of positive numbers such that

$$\lim_{n\to\infty} \delta_n = \infty \text{ and } \lim_{n\to\infty} \frac{\delta_n \log(n)}{n^\alpha} = 0 \text{ for all } \alpha > 0 \tag{7}$$

and

$$B_j^2 r_j \leq \delta_n \text{ for all } j \in \Lambda. \tag{8}$$

Let $q_j(\eta)$ be the upper bound for the probability in (5) with $S = S_j$, $B = B_j$, $B' = B'_j$, $r = r_j$ and $D = D_j$. For $\xi > 0$, take

$$\eta_j = \frac{1}{2}\left(\eta_{1,j} + \frac{B_j^2 r_j \xi}{n}\right),$$

then, with probability at least $1 - \sum_{j\in\Lambda} p_j(\eta_j)$,

$$\begin{aligned}
&k_1\|u - s\|^2 - 9\tau\left(\frac{\sigma^2}{4} \vee \|s - u\|^2\right) \\
&\leq E(\gamma(Z_i, u)) - E(\gamma(Z_i, s)) - 9\tau\left(\frac{\sigma^2}{4} \vee \|s - u\|^2\right) \\
&\leq \frac{1}{n}\sum_{i=1}^n \gamma(Z_i, u) + \eta_{1,j} + \frac{B_j^2 r_j \xi}{n} - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i, s) \\
&\leq \frac{1}{n}\sum_{i=1}^n \gamma(Z_i, u) + \eta_{1,j} + \frac{\delta_n \xi}{n} - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i, s)
\end{aligned}$$

for all $u \in S_j$ for all $j \in \Lambda$, and

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \gamma(Z_i, s^*) - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i, s) \\
&\leq E[\gamma(Z_i, s^*) - \gamma(Z_i, s)] + \tau\|s^* - s\|^2 + \eta_{j^*} \\
&\leq E[\gamma(Z_i, s^*) - \gamma(Z_i, s)] + \tau\|s^* - s\|^2 + \frac{\eta_{1,j^*}}{2} + \frac{\delta_n \xi}{2n}
\end{aligned}$$

with probability at least $1 - q_{j^*}(\eta_{j^*})$. Thus with probability at least $1 - q_{j^*}(\eta_{j^*}) - \sum_{j\in\Lambda} p_j(\eta_j)$, we have the error bound

$$(k_1 - 9\tau)\|\hat{s} - s\|^2 - 9\tau\left(\frac{\sigma^2}{4}\right) \le \frac{1}{n}\sum_{i=1}^{n}\gamma(Z_i, s^*) + \eta_{1,j^*} + \frac{\delta_n\xi}{n} - \frac{1}{n}\sum_{i=1}^{n}\gamma(Z_i, s)$$
$$\le 1.5\left(\eta_{1,j^*} + \frac{\delta_n\xi}{n}\right) + E[\gamma(Z_i, s^*) - \gamma(Z_i, s)] + \tau\|s^* - s\|^2$$
$$\le 1.5\eta_{1,j^*} + (k_2 + \tau)\|s^* - s\|^2 + \frac{1.5\delta_n\xi}{n}. \tag{9}$$

3 Application to Adaptive B-spline Estimation

In this section, we will apply Theorem 1 to obtain adaptive B-spline estimators for s in (1). Here the regression function s is assumed to be in the Sobolev space $W_m^2[0,1]$ for some $m \ge 1$ and each S_j is taken to be a collection of B-splines that are on $[0,1]$ with order q for some integer $q \ge 1$ and boundary knots at 0 and 1 and distinct internal knots $\xi_1, \ldots, \xi_k$ in $\{1/2^J, \ldots, (2^J-1)/2^J\}$ for some positive integer J. Also, the B-splines in S_j are bounded by b in absolute value for some positive integer b. Thus the index $j = (b, q, \xi_1, \ldots, \xi_k)$. Let

$$\tilde{\Delta}_{2,j} = \max_{1\le i\le k+1}(\xi_i - \xi_{i-1}), \tag{10}$$

$$\tilde{\Delta}_{1,j} = \min_{1\le i\le k+1}(\xi_i - \xi_{i-1}), \tag{11}$$

$$B'_j = \sqrt{2\pi e}\left(0.5 + \frac{q\sqrt{q}(2q+1)9^{q-1}\sqrt{\tilde{\Delta}_{2,j}}}{\sqrt{\tilde{\Delta}_{1,j}}}\right), \tag{12}$$

$$r_j = 1 \vee \frac{1}{q\sqrt{\tilde{\Delta}_{2,j}(k+q)}}, \tag{13}$$

and

$$J_j = \min\{J \ge 1 :\ \xi_1, \ldots, \xi_k \text{ are in } \{1/2^J, \ldots, (2^J-1)/2^J\}\ \},$$

where $\xi_0 = 0$ and $\xi_{k+1} = 1$. Let Λ be the set of all $j = (b, q, \xi_1, \ldots, \xi_k)$'s such that b and q are positive integers, $2^{J_j} + q + b \le n$ and $r_j(2b)^2 \le \delta_n$, where $\{\delta_n\}$ is chosen such that (7) holds. Then the estimator for s considered here is the penalized least square estimator

$$\hat{s} = \text{argmin}_{j\in\Lambda, u\in S_j}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - u(X_i))^2 + \eta_{1,j}\right), \tag{14}$$

where

$$\eta_{1,j} = \frac{a_n r_j (2b)^2}{n} \Big((k+q)\log(B'_j) + (\log 2) 2^{J_j} + q + b \Big) \tag{15}$$

and $\{a_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$. The L^2 convergence rate for $\hat{s}$ is given in Theorem 2.

Theorem 2 *Suppose that the regression model in (1) holds, where the regression function s is in $W_m^2[0,1]$, the observed pairs (X_1, Y_1), ..., (X_n, Y_n) are IID, and the W_i's are of mean zero and are independent of the X_i's. Suppose that $Ee^{\alpha|W_i|} < \Gamma$ for some $\alpha > 0$ and $\Gamma \geq 1$ and X_i has a Lebesgue density that is bounded above and bounded below from zero. Suppose that $\{a_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$. Suppose that $\{\delta_n\}$ is chosen so that (7) holds. Then, for the estimator $\hat{s}$ given in (14) with S_j defined above in this section and $\eta_{1,j}$ defined in (15), we have $E\|\hat{s} - s\|^2 = O(a_n n^{-2m/(1+2m)})$.*

Proof of Theorem 2. Theorem 2 is an application of Theorem 1 with $Z_i = (X_i, Y_i)$ and $\gamma(z,t) = \gamma((x,y),t) = (y - t(x))^2$. To apply Theorem 1, Assumptions M1 and M2 will be verified first with $S = S_j$.

Verification of Assumption M1. Since

$$\begin{aligned} |\gamma(z,u) - \gamma(z,v)| &= |u(x) - v(x)| \cdot |2y - u(x) - v(x)| \\ &= |u(x) - v(x)| \cdot |2(s(x) + w) - u(x) - v(x)|, \end{aligned} \tag{16}$$

for u, $v \in S_j$, we have

$$|\gamma(z,u) - \gamma(z,v)| \leq 2(|w| + \|s\|_\infty + b)|u(x) - v(x)|.$$

Take $M(w) = 2(|w| + \|s\|_\infty + b)$ and $\Delta(x,u,v) = |u(x) - v(x)|$, then

$$\begin{aligned} E[M^m(W_i)] &= E[2^m(|W_i| + \|s\|_\infty + b)^m] \\ &\leq 4^m \left(\frac{1}{2} E(|W_i|^m) + \frac{1}{2}(\|s\|_\infty + b)^m \right) \\ &\leq \frac{m!}{2} 4^m \left(\frac{\Gamma}{\alpha^m} + \frac{(\|s\|_\infty + b)^m}{m!} \right) \\ &\leq \frac{m!}{2} 4^m \left(\frac{\Gamma}{\alpha} + \|s\|_\infty + b \right)^m . \end{aligned}$$

Here the inequality $E|W_i|^m \leq \dfrac{m!\Gamma}{\alpha^m}$ follows from the assumption that $Ee^{\alpha|W_i|} < \Gamma$ for some $\alpha > 0$ and $\Gamma \geq 1$.

For $u \in S_j$ and $v = s$, (16) gives that

$$|\gamma(z,u) - \gamma(z,s)| \leq (2|w| + \|s\|_\infty + b)|u(x) - s(x)|.$$

Take $M_2(w) = 2|w| + \|s\|_\infty + b$ and $\Delta_2(x,u,s) = |u(x) - s(x)|$, then

$$\begin{aligned} E[M_2^m(W_i)] &= E[(|2W_i| + \|s\|_\infty + b)^m] \\ &\leq 2^m \left(\frac{1}{2} E(2^m |W_i|^m) + \frac{1}{2}(\|s\|_\infty + b)^m \right) \\ &\leq 2^{m-1} m! \left(\frac{\Gamma 2^m}{\alpha^m} + \frac{(\|s\|_\infty + b)^m}{m!} \right) \leq \frac{m!}{2} 2^m \left(\frac{2\Gamma}{\alpha} + \|s\|_\infty + b \right)^m . \end{aligned}$$

To control $\sum_{i=1}^n E_s[\Delta^m(X_i,u,v)]$ and $\sum_{i=1}^n E_s[\Delta_2^m(X_i,u,v)]$, note that for $u \in S_j$ and $v \in S_j \cup \{s\}$,

$$\frac{1}{n} \sum_{i=1}^n E_s[u(X_i) - v(X_i)]^m \leq E_s[u(X_1) - v(X_1)]^2 \|u - v\|_\infty^{m-2},$$

where

$$E_s[u(X_1) - v(X_1)]^2 \leq k_0 \|u - v\|^2$$

for some constant k_0 that does not depend on j since the density of X_i is bounded above.

From the above inequalities for $E[M^m(W_i)]$, $E[M_2^m(W_i)]$ and $\sum_{i=1}^n E_s[u(X_i) - v(X_i)]^m/n$ and the fact that $\|u-v\|_\infty \leq 2b$, $\|u - s\|_\infty \leq \|s\|_\infty + b$, Assumption M1 holds for $S = S_j$ with $b_m = 1$, $a_m = m!/2$, $B = 2b$, $\tilde{B} \geq \|s\|_\infty + b$,

$$A = 4\left(\frac{\Gamma}{\alpha} + \|s\|_\infty + b \right) \text{ and } A_2 \geq 2\left(\frac{2\Gamma}{\alpha} + \|s\|_\infty + b \right).$$

Take $A_0 = 4(\Gamma/\alpha + \|s\|_\infty)$, $\tilde{B} = (A_0 + B)/2$, $A_2 = A_0 + B$, then $B \geq 1$, $A = A_0 + 2B$, $A_2 = A_0 + B = 2\tilde{B}$, and $\|u\|_\infty \leq B$ for all $u \in S_j$. Since the constants B and $\tilde{B}$ depend on j, they will be denoted by B_j and $\tilde{B}_j$ respectively. Note that the constant A_0 does not depend on j.

Verification of Assumption M2. To identify the constants in Assumption M2, Facts 1 and 2 will be applied. These two facts are first stated and proved below.

Fact 1 *Let $\bar{S}$ be a D-dimensional subspace of $L_2 \cap L_\infty(\mu)$ spanned by some basis $\{\phi_i : i \in \{1, \ldots, D\}\}$. Let $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denotes the L_2-norm and the L_∞-norm with respect to μ. Let $|\cdot|_2$ and $|\cdot|_\infty$ denote the l_2-norm and the l_∞-norm in R^D.*

Suppose that there exist constants T_1 and T_2 such that for $(\theta_1, \ldots, \theta_D) \in R^D$,

$$\left\| \sum_{i=1}^D \theta_i \phi_i \right\|_\infty \leq T_1 |\theta|_\infty \tag{17}$$

and

$$\frac{T_2}{\sqrt{D}} |\theta|_2 \leq \left\| \sum_{i=1}^D \theta_i \phi_i \right\|_2 \leq \frac{T_3}{\sqrt{D}} |\theta|_2. \tag{18}$$

Take $r' \geq T_1/T_3$ and

$$B' = \sqrt{2\pi e}\left(0.5 + \max\left(\frac{T_3}{T_2}, 1\right)\right) \tag{19}$$

Then for $\mathcal{B}$: an L_2 ball of radius σ in $\bar{S}$ with $0 < \delta < \sigma/5$, there exists a finite set $T \subset \mathcal{B}$ such that T is a δ-net for $\mathcal{B}$ with respect to the L_2-norm and a $r'\delta$-net with respect to the L_∞ norm, and the number of elements in T is at most $(B'\sigma/\delta)^D$.

Proof. Suppose that the center of $\mathcal{B}$ is $\sum_{i=1}^D \theta_i^* \phi_i$. Let $\theta^* = (\theta_1^*, \ldots, \theta_D^*)$. Then it follows from the first inequality in (18) that $\mathcal{B}$ is contained $\{\sum_{i=1}^D \theta_i \phi_i : (\theta_1, \ldots, \theta_D) \in B_0\}$, where B_0 is the l_2 ball of center θ^* and radius $\sqrt{D}\sigma/T_2$. Since the volume for an l_2 ball in R^D with radius σ is bounded by $c(D)\sigma^D$ (cf. Proof of Lemma 2 in [2]), where

$$c(D) = (2\pi e/D)^{D/2}(\pi D)^{-1/2},$$

we can cover B_0 with cubes of edge length δ/T_3 such that the number of cubes is at most

$$\frac{c(D)(\sqrt{D}\sigma/T_2 + \sqrt{D}\delta/T_3)^D}{(\delta/T_3)^D} \leq (1 + (T_3\sigma)/(T_2\delta))^D (2\pi e)^{D/2} \leq (B'\sigma/\delta)^D$$

for the B' in (19) if $\sigma/5 > \delta > 0$. Choosing one point from each cube to form a set T_0, and take $T = \mathcal{B} \cap \{\sum_{i=1}^D \theta_i \phi_i : (\theta_1, \ldots, \theta_D) \in T_0\}$, then from the second inequality in (18), T is a δ-net for $\mathcal{B}$ with respect to the L_2-norm. From (17), T is a $r'\delta$-net for $\mathcal{B}$ with respect to the L_∞-norm for $r' \geq T_1/T_3$. The proof of Fact 1 is complete.

Fact 2 *Suppose that μ is the Lebesgue measure on $[0, 1]$. Let $\bar{S}$ be the space of B-splines on $[0, 1]$ with order q and k knots $\xi_1, \ldots, \xi_k$ with multiplicities $m_1, \ldots, m_k$, where $0 < \xi_1 < \cdots < \xi_k < 1$. Then $\bar{S}$ is a sub-space of $L_2 \cap L_\infty(\mu)$. Suppose that $0 < \tilde{\Delta}_1 \leq \xi_i - \xi_{i-1} \leq \tilde{\Delta}_2$ for $i = 1, \ldots, k+1$, where $\xi_0 = 0$ and $\xi_{k+1} = 1$. Let $K = m_1 + \cdots + m_k$ and $D = K + q$. Then (17) holds with $T_1 = 1$ and (18) holds with $T_2 = \sqrt{\tilde{\Delta}_1 D}/(\sqrt{q}(2q+1)9^{q-1})$ and $T_3 = q\sqrt{\tilde{\Delta}_2 D}$.*

Proof. Let

$$(y_1, \ldots, y_{K+2q}) = (\underbrace{0, \ldots, 0}_{q \text{ times}} \underbrace{\xi_1, \ldots, \xi_1}_{m_1 \text{ times}}, \ldots, \underbrace{\xi_k, \ldots, \xi_k}_{m_k \text{ times}} \underbrace{1, \ldots, 1}_{q \text{ times}})$$

and let ϕ_i be the (normalized) B-spline basis of order q associated with knots $y_i, \ldots, y_{i+q}$ for $i = 1, \ldots, D$. Then $\phi_1, \ldots, \phi_D$ spans $\bar{S}$. It follows from

Equation (4.80) in Schumaker [5] that (17) holds with $T_1 = 1$, so it remains to check (18).

To check that the first inequality in (18) holds with the T_2 specified above, note that from (4.79) and (4.86) in [5], we have that for $f = \sum_{i=1}^{K+q} \theta_i \phi_i$,

$$|\theta_i| \leq (2q+1)^2 9^{2(q-1)} \tilde{\Delta}_1^{-1/2} \|f\|_{L_2[y_i, y_{i+q}]},$$

where ϕ_i is supported on $[y_i, y_{i+q}]$ which implies that

$$\begin{aligned} \sum_{i=1}^{K+q} \theta_i^2 &\leq (2q+1)^2 9^{2(q-1)} \tilde{\Delta}_1^{-1} \sum_{i=1}^{K+q} \|f\|^2_{L_2[y_i, y_{i+q}]} \\ &\leq (2q+1)^2 9^{2(q-1)} \tilde{\Delta}_1^{-1} q \|f\|_2^2, \end{aligned}$$

which implies that the first inequality in (18) holds with $T_2 = \sqrt{\tilde{\Delta}_1 D}/(\sqrt{q}(2q+1)9^{q-1})$.

To check that the first inequality in (18) holds with the T_3 specified above, we follow the approach in the proof of Lemma 4.2 in Ghosal, Ghosh and Van der Vaart [3], which is originally given in Stone [6]. For $f = \sum_{j=1}^{K+q} \theta_j \phi_j$, we have that for $x \in [y_i, y_{i+1})$ and $q+1 \leq i \leq q+K$, $f(x) = \sum_{j=i+1-q}^{i} \theta_j \phi_j(x)$ (cf. [5], Equations (4.25) and (4.29)), so it follows from Schwartz inequality that for $x \in [y_i, y_{i+1})$,

$$f^2(x) \leq q \sum_{j=i+1-q}^{i} \theta_j^2 \phi_j^2(x) \leq q \sum_{j=i+1-q}^{i} \theta_j^2,$$

which gives

$$\int_0^1 f^2(x)dx = \sum_{i=q+1}^{q+K} \int_{y_i}^{y_{i+1}} f^2(x)dx \leq q \sum_{i=q+1}^{q+K} \sum_{j=i+1-q}^{i} \theta_j^2 (y_{i+1} - y_i) \leq q^2 \tilde{\Delta}_2 \sum_{j=1}^{D} \theta_j^2$$

and the second inequality in (18) holds with $T_3 = q\sqrt{\tilde{\Delta}_2 D}$. The proof of Fact 2 is complete.

From Facts 1 and 2, for $S = S_j$, Assumption M2 holds with $D = k + q$,

$$B' = \sqrt{2\pi e} \left(0.5 + \max\left(\frac{q\sqrt{q}(2q+1)9^{q-1}\sqrt{\tilde{\Delta}_{2,j}}}{\sqrt{\tilde{\Delta}_{1,j}}}, 1 \right) \right) \stackrel{(12)}{=} B'_j$$

and

$$r = 1 \vee \frac{1}{q\sqrt{\tilde{\Delta}_{2,j}(k+q)}} \stackrel{(13)}{=} r_j,$$

where $\tilde{\Delta}_{2,j}$ and $\tilde{\Delta}_{2,j}$ are defined in (10) and (11) respectively. It is clear that $r_j \geq 1$ for $j \in \Lambda$ and $r_j B_j^2 \leq n/4$ for $j \in \Lambda$ if n is large enough so that $\delta_n \leq n/4$, as required in Theorem 1.

Now the error bound in (9) holds if (6) holds and (3) holds with $\eta = \eta_j$, $B = 2b$, $D = k + q$ for all $j \in \Lambda$. Here (6) holds with constants k_1 and k_2 that do not depend on j since the density for the distribution of X_i is supported on $[0, 1]$ and is bounded below from zero and bounded above on $[0, 1]$ and

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - t(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n}(Y_i - s(X_i))^2\right) = E(s(X_i) - t(X_i))^2.$$

To verify (3), note that with $B = 2b$ and $D = k + q$,

$$\frac{n\eta_j}{r_j B^2} > a_n(k+q)\log(B_j') \geq a_n D \log(3.5\sqrt{2\pi e})$$

and $24(1 + 2D\log 2) \leq 24(1 + 2\log 2)D$, so if n is large enough so that

$$a_n \log(3.5\sqrt{2\pi e}) \geq 24(1 + 2\log 2), \tag{20}$$

then (3) holds with $\eta = \eta_j$, $B = 2b$, $D = k + q$ for all $j \in \Lambda$.

As a result, for $\xi > 0$, if n is large enough so that (20) holds, then (9) holds with $\theta = \left(\frac{A_0 + 2}{c(8\tau)}\right) \vee 5$ with probability that is at least

$$1 - q_{j^*}(\eta_{j^*}) - \sum_{j \in \Lambda} p_j(\eta_j),$$

where

$$\begin{aligned}
\eta_j &= \eta_{1,j} + \frac{r_j(2b)^2\xi}{n} \\
&= \frac{a_n r_j (2b)^2}{n}\Big((k+q)\log(B_j') + (\log 2)2^{J_j} + q + b\Big) + \frac{r_j(2b)^2\xi}{n},
\end{aligned}$$

$$\begin{aligned}
q_j(\eta_j) &= \exp\left(-\frac{2c_1(\tau/k_0)n\eta_j}{(A_0+2b)^2}\right) \\
&\leq \exp\Big(-a_n c_9[(\log 2)2^{J_j} + q + b] - c_9\xi\Big),
\end{aligned} \tag{21}$$

and for $0 < \sigma < 1$,

$$p_j(\eta_j) = \exp\left(-\frac{2c_1(\tau/k_0)n\eta_j}{(A_0+2b)^2}\right)$$

$$
\begin{aligned}
&+1.6\left(1-\exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau)\wedge\frac{1}{3}\right]\right)\right)^{-1}\left((B_j'\theta)\vee 2\right)^{3(k+q)}\times\\
&\left[\exp\left(-\frac{nc(\tau)\eta_j}{2(A_0+2b)(2b)}\right)+\exp\left(-\frac{n\eta_j}{24r_j(2b)^2}\right)\right]\\
\leq\ &\exp\left(-\frac{c_3 n\eta_j}{(2b)^2}\right)+\frac{c_4}{\sigma^2}\exp\left(-\frac{c_5 n\eta_j}{(2b)^2}\right)\left(B_j'\theta\right)^{3(k+q)}
\end{aligned}
$$

for some constants c_9, c_3, c_4, c_5 that depend only on τ and A_0. For $0<\sigma<1$, if n is large enough so that

$$
c_5 a_n \geq 3+\frac{3\log\theta}{\log(3.5\sqrt{2\pi e})}, \tag{22}
$$

then for $\xi>0$ and $c_6=c_3\wedge c_5$,

$$
p_j(\eta_j)\leq 2\left(1\vee\frac{c_4}{\sigma^2}\right)\exp\left(-c_6\left[a_n[(\log 2)2^{J_j}+q+b]+\xi\right]\right).
$$

For n large enough so that $a_n c_6>2$,

$$
\begin{aligned}
&\sum_{j\in\Lambda}\exp\left(-c_6 a_n[(\log 2)2^{J_j}+q+b]\right)\\
&\leq\sum_b\sum_q\sum_J\sum_{k=1}^{2^J-1}\frac{2^J-1}{k!(2^J-1-k)!}\exp\left(-2[(\log 2)2^J+q+b]\right)\\
&\leq\sum_b\sum_q\sum_J 2^{-2^J}e^{-2b}e^{-2q}\stackrel{\text{def}}{=}c_7<\infty,
\end{aligned}
$$

so for $c_8=2c_7$,

$$
\sum_{j\in\Lambda}p_j(\eta_j)\leq c_8\left(1\vee\frac{c_4}{\sigma^2}\right)\exp\left(-c_6\xi\right).
$$

Therefore, for $0<\sigma<1$ and for n large enough so that $a_n c_6>2$ and (22) and (20) hold, we have

$$
(k_1-9\tau)\|\hat{s}-s\|^2>9\tau\left(\frac{\sigma^2}{4}\right)+1.5\eta_{1,j^*}+(k_2+\tau)\|s^*-s\|^2+\frac{1.5\delta_n\xi}{n}
$$

with probability at most

$$
\exp(-c_9\xi)+c_8\left(1\vee\frac{c_4}{\sigma^2}\right)\exp\left(-c_6\xi\right)
$$

for $\xi > 0$. Let

$$U = (k_1 - 9\tau)\|\hat{s} - s\|^2 - \left(9\tau\left(\frac{\sigma^2}{4}\right) + 1.5\eta_{1,j^*} + (k_2+\tau)\|s^* - s\|^2\right),$$

then for $\xi_0 > 0$, $0 < \sigma < 1 \wedge \sqrt{c_4}$, $c_{10} = c_6 \wedge c_9$ and $c_{11} = c_4(1+c_8)$, we have

$$\begin{aligned} E\left(\frac{nU}{1.5\delta_n}\right) &\leq \xi_0 + \int_{\xi_0}^{\infty}\left[\exp(-c_9\xi) + c_8\left(1 \vee \frac{c_4}{\sigma^2}\right)\exp\left(-c_6\xi\right)\right]d\xi \\ &\leq \xi_0 + \frac{c_{11}}{\sigma^2}\int_{\xi_0}^{\infty}\exp\left(-c_{10}\xi\right)d\xi \\ &\leq \xi_0 + \frac{c_{11}}{c_{10}\sigma^2}\exp\left(-c_{10}\xi_0\right). \end{aligned}$$

Take $\xi_0 = \delta_n^{-1} n^{1/(1+2m)}$ and $\sigma = n^{-m/(1+2m)}$, then

$$n^{2m/(1+2m)}E(U) \leq 1.5\delta_n n^{-1/(1+2m)}\left[\xi_0 + \frac{c_{11}}{c_{10}\sigma^2}\exp\left(-c_{10}\xi_0\right)\right] \to 0 \text{ as } n \to \infty,$$

so

$$(k_1 - 9\tau)E\|\hat{s} - s\|^2 \leq 1.5\eta_{1,j^*} + (k_2+\tau)\|s^* - s\|^2 + Cn^{-2m/(1+2m)} \quad (23)$$

for some constant $C > 0$. Let $b^* = \lfloor \|s\|_\infty \rfloor + 1$, $q^* = m+1$ and $J^* = 1 + \lfloor (1+2m)^{-1}\log_2(n)\rfloor$, where $\lfloor\cdot\rfloor$ is the floor function. Let

$$j^* = \left(b^*, q^*, \frac{1}{2^{J^*}}, \ldots, \frac{2^{J^*}-1}{2^{J^*}}\right),$$

then $2^{J^*} = O(n^{1/(1+2m)})$, $\|s^* - s\|^2 = O(n^{-2m/(1+2m)})$ (cf. Theorem 6.25 in [5]) and

$$\begin{aligned} \eta_{1,j^*} &= \frac{a_n r_{j^*}(2b^*)^2}{n}\left((2^{J^*} - 1 + q^*)\log(B'_{j^*}) + (\log 2)2^{J^*} + q^* + b^*\right) \\ &= O(a_n n^{-2m/(1+2m)}) \end{aligned}$$

since

$$r_{j^*} = 1 \vee \frac{1}{q^*\sqrt{2^{-J^*}(2^{J^*}-1+q^*)}} = 1$$

and

$$B'_{j^*} = \sqrt{2\pi e}\left(0.5 + q^*\sqrt{q^*}(2q^*+1)9^{q^*-1}\right).$$

Therefore, it follows from (23) that $E\|\hat{s}-s\|^2 = O(a_n n^{-2m/(1+2m)})$ by choosing $\tau < k_1/9$. The proof of Theorem 2 is complete.

4 Conclusion

An adaptive estimator for regression function using B-splines that allows non-equally spaced knots is successfully constructed. The estimator is a penalized least square estimator and it achieves the L^2 convergence rate $O(a_n n^{-2m/(1+2m)})$ for any sequence $\{a_n\}$ such that $\lim_{n\to\infty} a_n = \infty$ when the regression function is in $W_m^2[0,1]$ for some $m \geq 1$.

References

[1] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization (Disc: P415-419). *Probability Theory and Related Fields*, 113:301–413, 1999.

[2] Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli*, 4:329–375, 1998.

[3] Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, 2000.

[4] Tzee-Ming Huang. Convergence rates for posterior distributions and adaptive estimation. *The Annals of Statistics*, 32(4):1556–1593, 2004.

[5] Larry L. Schumaker. *Spline Functions: Basic Theory.* Wiley-Interscience, 1981.

[6] Charles J. Stone. The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, 14:590–606, 1986.

[7] Yuhong Yang and Andrew R. Barron. An asymptotic property of model selection criteria. *IEEE Transactions on Information Theory*, 44(1):95–116, 1998.

[8] S. Zhou, X. Shen, and D. A. Wolfe. Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26(5):1760–1782, 1998.

國科會補助計畫衍生研發成果推廣資料表

日期:2013/07/18

<table>
<tr><td rowspan="3">國科會補助計畫</td><td>計畫名稱: 無母數迴歸使用B-spline時選取knot 的一種方式</td></tr>
<tr><td>計畫主持人: 黃子銘</td></tr>
<tr><td>計畫編號: 101-2118-M-004-001- 學門領域: 半母(參)數模型與平滑方法</td></tr>
<tr><td colspan="2">無研發成果推廣資料</td></tr>
</table>

101 年度專題研究計畫研究成果彙整表

1-2118-M-004-001-

使用 B-spline 時選取 knot 的一種方式

備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
政 大 統 計 系 技 術 報 告 , http://stat.nccu.edu.tw/download.php?filename=1065_e16da070.pdf&dir=writing&title=--%E9%99%84%E4%BB

3 人於不同時期擔任助理

名稱或內容性質簡述

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

　　□實驗失敗

　　□因故實驗中斷

　　□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 ■未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

目前已經證明本研究中提出的 knot 選擇方法有良好的理論性質，下一步需要進行模擬實驗，和其他 knots 估計方法比較，以進一步了解此方法的運算時間和小樣本性質，以及相對於現有方法的優勢.