CONSISTENT VARIABLE SELECTION IN ADDITIVE MODELS

Lan Xue

Oregon State University

Supplementary Material

S1. Assumptions

Let $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ denote independent pairs, each having the same distribution as $(\mathbf{X}, Y)$. The technical assumptions we need are as follows.

(A1) *The density function* $f(\mathbf{x})$ *of* $\mathbf{X}$ *is absolutely continuous and compactly supported. Without loss of generality, let its support* $\mathcal{X} = [0,1]^d$. *Also there exits constants* $0 < c_1 \leq c_2$, *such that* $c_1 \leq f(\mathbf{x}) \leq c_2$, *for all* $\mathbf{x} \in \mathcal{X}$.

(A2) *The* d *sets of knots denoted as* $k_l = \{0 = x_{l,0} \leq x_{l,1} \leq \cdots \leq x_{l,N_l} \leq x_{l,N_l+1} = 1\}, l = 1, \ldots, d$, *are quasi-uniform, that is, there exists* $c_3 > 0$, such that

$$\max_{l=1,\ldots,d} \frac{\max(x_{l,j+1} - x_{l,j}, j = 0, \ldots, N_l)}{\min(x_{l,j+1} - x_{l,j}, j = 0, \ldots, N_l)} \leq c_3.$$

Furthermore, *the number of interior knots* $N_l \asymp n^{1/(2p+3)}$, *where* p *is the degree of the spline and* '$\asymp$' *means both sides have the same order. In particular,* $h \asymp n^{-1/(2p+3)}$.

(A3) *For* $1 \leq l \leq d$, *the functions* $\alpha_l \in C^p([0,1])$, *where* $C^p([0,1])$ *denotes the space of* $p-$*times continuously differentiable functions on* $[0,1]$.

(A4) *The conditional variance function* $\sigma^2(\mathbf{x}) = \mathrm{Var}(Y|\mathbf{X} = \mathbf{x})$ *is bounded on* $\mathcal{X}$.

S2. Auxiliary lemmas

Denote $\mathcal{M}_{n,0} \subset \mathcal{M}_n$ *as*

$$\mathcal{M}_{n,0} = \left\{ m_n(\mathbf{x}) = \sum_{l=1}^{s} g_l(x_l); g_l \in \varphi_l^{0,n} \right\},$$

which is the approximation space knowing $\alpha_{l0} = 0$, for $l = s+1, \ldots, d$. To prove Theorem 1 , we will make use of two standard least square spline estimators, denoted as $\hat{m}^*_{n,0}$ and $\hat{m}^*_n$, which are the best least square approximation of m_0 in approximation spaces $\mathcal{M}_{n,0}$, and $\mathcal{M}_n$ respectively. That is, define

$$\hat{m}^*_{n,0} = \underset{m_{n,0} \in \mathcal{M}_{n,0}}{\operatorname{argmin}} \left\| Y - m_{n,0} \right\|^2_n, \quad \hat{m}^*_n = \underset{m_n \in \mathcal{M}_n}{\operatorname{argmin}} \left\| Y - m_n \right\|^2_n.$$

Here we cite some results regarding the standard polynomial spline estimation.

Lemma 1 *Under conditions (A1-A4), with* $\rho_n = 1/\sqrt{nh} + h^{p+1}$, *one has*

(i) $\left\| \hat{m}^*_{n,0} - m_0 \right\| = O_p(\rho_n)$, $\left\| \hat{m}^*_{n,0} - m_0 \right\|_n = O_p(\rho_n)$.

(ii) $\left\| \hat{m}^*_n - m_0 \right\| = O_p(\rho_n)$, $\left\| \hat{m}^*_n - m_0 \right\|_n = O_p(\rho_n)$.

Lemma 1 is the standard results regarding the mean square (or L_2)-convergence rate for standard polynomial spline estimators (e.g. Theorem 1 in Huang 1998).

Lemma 2 *Under conditions (A1-A2), as* $n \to \infty$, *one has*

$$\sup_{\phi_1 \in \mathcal{M}_n, \phi_2 \in \mathcal{M}_n} \left| \frac{\langle \phi_1, \phi_2 \rangle_n - \langle \phi_1, \phi_2 \rangle}{\|\phi_1\| \, \|\phi_2\|} \right| = O_p\left(\sqrt{\frac{\log^2(n)}{nh}} \right).$$

In particular, there exist constants $0 < c < 1 < C$ *such that, except on an event whose probability tends to zero as* $n \to \infty$, $c \|m_n\| \leq \|m_n\|_n \leq C \|m_n\|, \forall m_n \in \mathcal{M}_n$.

Lemma 2 is crucial to prove both Theorem1 and Theorem 2. It shows that the empirical and theoretical inner products are uniformly close over the approximation space $\mathcal{M}_n$. The general proof of Lemma 2 can be found in Xue and Yang (2006a) or Huang (1998).

Lemma 3 *Under condition (A1), let* $\delta = (1 - c_1/c_2)^{1/2}$, *and* $c_4 = \left(\frac{1-\delta}{2}\right)^{(d-1)/2} > 0$. *Then for any* $m = \sum_{l=1}^d \alpha_l \in \mathcal{M}$, *one has*

$$\|m\| \geq c_4 \sum_{l=1}^d \|\alpha_l\|.$$

Lemma 3 is the Lemma 1 in Stone (1985), which implies that the model space $\mathcal{M}$ is essentially identifiable (up to sets of Lebesgue measure zero). That is, for any $m \in \mathcal{M}$, there is essentially a unique additive representation $m = \sum_{l=1}^{d} \alpha_l$, with $\alpha_l \in H_l^0$. The next lemma follows immediately from Lemmas 2 and 3.

Lemma 4 *Under conditions (A1-A2), there exists a constant* $c_5 > 0$, *such that, except on an event whose probability tends to zero, as* $n \to \infty$, *for any* $m_n = \sum_{l=1}^{d} g_l \in \mathcal{M}_n$, *with* $g_l \in \varphi_l^{0,n}$,

$$\|m_n\|_n \geq c_5 \sum_{l=1}^{d} \|g_l\|_n \,.$$