Arijit Chakrabarti · Jayanta K. Ghosh

Optimality of AIC in inference about Brownian motion

Abstract In the usual Gaussian White-Noise model, we consider the problem of estimating the unknown square-integrable drift function of the standard Brownian motion using the partial sums of its Fourier series expansion generated by an orthonormal basis. Using the squared L_2 distance loss, this problem is known to be the same as estimating the mean of an infinite dimensional random vector with l_2 loss, where the coordinates are independently normally distributed with the unknown Fourier coefficients as the means and the same variance. In this modified version of the problem, we show that Akaike Information Criterion for model selection, followed by least squares estimation, attains the minimax rate of convergence.

Keywords Nonparametric regression · Minimax · AIC · Oracle · Brownian motion · White-noise

1 Introduction

The Akaike Information Criterion (AIC), the now well-known penalized likelihood model selection criterion, was introduced and studied by Akaike (1973,1978). Different asymptotic optimality properties of AIC have been proved in the literature by several authors in the last three decades. In the first line of work, Shibata (1981, 1983) proved the optimality of AIC as a model selection rule in the infinite dimensional problem of nonparametric regression, where the goal is to find out

A. Chakrabarti (✉)
Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road,
Calcutta 700108, India
E-mail: arijit_v@isical.ac.in

J. K. Ghosh
Department of Statistics, Purdue University, 150 North University Street,
West Lafayette, IN 47907-2067, USA
Statistics and Mathematics Unit, Indian Statistical Institute,
203 B.T.Road, Calcutta 700108, India

the optimum number of terms to retain, for the purpose of prediction, in the Fourier series expansion of the unknown function generated by a given orthonormal sequence. Shibata (1983) has shown that AIC does as well as an oracle introduced by him in this problem. In a second line of work, Li (1987) and Shao (1997) proved the asymptotic optimality of AIC as a model selection rule in the context of selection of variables from a given set of variables in a Linear model setup. But the optimal rate of convergence of AIC has not been studied in the literature. The novelty of our paper is to show that model selection by AIC followed by least squares estimation achieves the minimax rate of convergence in one form of nonparametric function estimation problem.

We study AIC in the following problem of inference about an unknown signal or drift $f \in L_2[0,1]$ of a Brownian motion and prove it attains the optimal rate of convergence in two different senses. Given n, one observes $\{Z(t)\}$ given by

$$\mathrm{d}Z(t) = f(t)\mathrm{d}t + \frac{dB(t)}{\sqrt{n}}, 0 \le t \le 1, \tag{1}$$

where $B(t)$ is the standard Brownian motion. This is essentially the problem (Eq. 31) of Ibragimov and Has'minskii (1981, p. 345). In problem (Eq. 1), we consider a complete orthonormal basis $\{\phi_i, i = 1, 2, \ldots\}$ of $L_2[0,1]$. Then one can write

$$f(t) = \sum_{i=1}^{\infty} \theta_i \phi_i(t), \tag{2}$$

with equality in the sense of L_2 convergence, where θ_i's are the Fourier coefficients given by

$$\theta_i = \int_0^1 \phi_i(t) f(t)\,\mathrm{d}t, \quad \text{and} \quad \sum_{i=1}^{\infty} \theta_i^2 < \infty. \tag{3}$$

Then we need to study the somewhat simpler problem as follows:

$$y_i = \theta_i + \frac{\epsilon_i}{\sqrt{n}}, \epsilon_i \overset{i.i.d.}{\sim} N(0,1), \tag{4}$$

where $y_i = \int_0^1 \phi_i(t)\,\mathrm{d}Z(t)$ for $i = 1, 2, \ldots$.

Let $\hat{\theta} = \{\hat{\theta}_i\}$ be an estimate of θ and let $\hat{f}(t) = \sum_{i=1}^{\infty} \hat{\theta}_i \phi_i(t)$ the corresponding estimate of f. Then by Parseval's theorem,

$$\|f - \hat{f}\|^2 = \sum_{i=1}^{\infty} (\theta_i - \hat{\theta}_i)^2, \tag{5}$$

where $\|.\|$ is the usual L_2 norm. So estimating f in model (Eq. 1) is the same as estimating θ in model Eq. (4) in terms of the above losses. We use the setup of Eq. (4) in this paper and use the squared error l_2 loss. We show that model selection by AIC followed by least squares estimates attains the minimax rate of convergence for convergence in probability over the usual Sobolev balls $E_q(B)$ (defined in Sect. 2), for any $B > 0$. This result is based on a strong property of AIC with lower truncation. Under lower truncation it is shown that AIC is asymptotically equivalent to an oracle uniformly in $E_q(B)$, where the oracle provides a lower bound to the loss in a certain class of decision rules. We also show that model selection by AIC with upper truncation followed by least squares estimation, attains the minimax rate of convergence, i.e, $n^{-2q/2q+1}$, over the Sobolev balls mentioned before.

It is worthwhile to mention here that the definition of AIC [see Eq. (7)] does not require the knowledge of the order of smoothness q or the constant B appearing in the definition $E_q(B)$ (in Sect. 2) of the class of functions being considered. Yet model selection by AIC followed by least squares estimation yields the minimax rate over $E_q(B)$; showing that AIC is adaptive.

It is not hard to show that the Bayes Information Criterion (BIC) cannot have this kind of optimality. A counter-example is presented in Sect. 4.

Problem (1) has been shown in Brown and Low (1996) to be an equivalent version in a decision theoretic sense, upto the minimax rate of convergence, of the following nonparametric regression problem

$$Y_i = f\left(\frac{i}{n+1}\right) + \epsilon_i, \epsilon_i \overset{i.i.d}{\sim} N(0,1), i = 1, 2, \dots, n. \tag{6}$$

Using Eq. (1) through Eq. (6), Zhao (2000) has pointed out that nonparametric regression can in principle be studied through the y_i's. Her main result is to introduce a hierarchical prior on the parameter space and show that the corresponding Bayes estimator achieves the minimax rate of convergence. The relation between Eqs. (1) and (6) suggests that our AIC for Eq. (1) can be lifted in principle to provide an asymptotically minimax method of estimation for nonparametric regression. This is discussed in the last section.

Section 4 also includes a discussion on how to use the theoretical results derived for continuous path data, when one observes the process $\{Z(t)\}$ only at a finite number of equally spaced points.

2 Preliminaries, notations and theorems

Suppose, as in Eq. (4), one has random variable y_i's which are independent $N(\theta_i, 1/n)$, $i = 1, 2, 3, \dots$, where $\sum_{i=1}^{\infty} \theta_i^2 < \infty$. Using y_i's one has to come up with estimates $\hat{\theta}_i$ and the loss is $L = \sum_{i=1}^{\infty} (\hat{\theta}_i - \theta_i)^2$. We consider a restricted parameter space in our study, as in Zhao (2000), which is a Sobolev-type subspace of l_2 given by $E_q = \{\theta = \{\theta_i\} : \sum_{i=1}^{\infty} i^{2q}\theta_i^2 < \infty\}$, $q > 1/2$. We then study the asymptotic rate of convergence of model selection by AIC followed by least squares estimation in the Sobolev ball

$$E_q(B) = \left\{\theta : \sum_{i=1}^{\infty} i^{2q}\theta_i^2 \leq B\right\}.$$

With respect to the usual trigonometric basis, for q an integer, E_q corresponds to all periodic $L_2[0, 1]$ functions with absolutely continuous $(q-1)$th derivatives and q-th derivatives with bounded L_2 norm.

The AIC is not well defined in this case since we have an infinite sequence of observations. However, if we take $\hat{\theta}_i = 0$ for all $i > n$, the contribution to error for θ in $E_q(B)$ is

$$\sum_{i>n} \theta_i^2 = \sum_{i>n} \frac{i^{2q}\theta_i^2}{i^{2q}} \leq B(n+1)^{-2q} = o\left(n^{-\frac{2q}{2q+1}}\right) \text{ as } n \to \infty,$$

since $\theta \in E_q(B)$. So at least for the problem of finding decision rules that attain the minimax rate, we can ignore observations beyond the nth. With this modification one can define AIC as follows. Let

$$m^{\text{AIC}} = \underset{1\leq m\leq n}{\operatorname{argmin}}\, S(m) \quad \text{where } S(m) = \sum_{m+1}^{n} y_i^2 + \frac{2m}{n}. \tag{7}$$

The estimate of θ_i is y_i for $i \leq m^{\text{AIC}}$ and zero thereafter. The loss is $\sum_1^{m^{\text{AIC}}} (y_i - \theta_i)^2 + \sum_{m^{\text{AIC}}+1}^{n} \theta_i^2 + \sum_{n+1}^{\infty} \theta_i^2$.

One may interpret this as first choosing a model M_m for which $\theta_i = 0$ for $i > m$ and then estimating θ_i by least squares, i.e., by y_i for $i \leq m$.

We will now introduce some notations before we state our theorems. Define $L_n(m)$ by

$$L_n(m) = \sum_1^m (y_i - \theta_i)^2 + \sum_{m+1}^{\infty} \theta_i^2, \quad 1 \leq m \leq n, \tag{8}$$

the loss in choosing model M_m and then using least squares estimates. Let

$$r_m(\theta) = \frac{m}{n} + \sum_{m+1}^{\infty} \theta_i^2, \tag{9}$$

the risk of the estimate described above.

We next define two oracles based on $L_n(m)$ and $r_m(\theta)$ as follows.

Define m_1 as

$$m_1 = \underset{1\leq m\leq n}{\operatorname{argmin}}\, L_n(m). \tag{10}$$

Note that $L_n(m_1)$ is a lower bound to the loss of any decision rule that first picks a model M_m and then estimates θ_i by zero if $i > m$ and by y_i for $i \leq m$.

Define the second oracle m_0 as

$$m_0 = \underset{1\leq m\leq n}{\operatorname{argmin}}\, r_m(\theta). \tag{11}$$

Intuitively one expects m_1 and m_0 to be close but m_0 is easier to deal with. Note that both m_0 and m_1 depend on θ.

Let AIC^u be the model selection procedure which is AIC with upper truncation. It chooses the model M_{m^u} where

$$m^u = \operatorname*{argmin}_{1 \le m \le [n^{\frac{1}{2q+1}}]} S(m). \tag{12}$$

where $[n^{1/2q+1}]$ denotes the largest integer less than or equal to $n^{1/2q+1}$.

Notation. Henceforth "$a_n \sim b_n$ asymptotically", will mean that there exist positive constants $0 < k_1 < k_2$ such that for all sufficiently large n, $k_1 b_n \le a_n \le k_2 b_n$.

Consider now any sequence $\{m_n\}$ of integers such that $m_n \to \infty$ as $n \to \infty$ as slowly as we wish but $m_n \le m^*$ where $m^* \sim n^{1/2q+1}$ asymptotically, and is defined in the proof of Theorem 2.1.

Now define m_0^l and m_1^l as

$$m_0^l = \operatorname*{argmin}_{m_n \le m \le n} r_m(\theta), \quad m_1^l = \operatorname*{argmin}_{m_n \le m \le n} L_n(m) \tag{13}$$

and define m^l as

$$m^l = \operatorname*{argmin}_{m_n \le m \le n} S(m). \tag{14}$$

So m^l is the model chosen by AIC^l, the model selection procedure which is AIC with lower truncation as described in Eq. (14).

We now state the main results proved in this paper. (Note that we are suppressing the dependence of $\hat{\theta}$ on n for notational convenience.)

Theorem 2.1 *For the case $m \ge m_n$, we have,* (a)

$$L_n(m^l)(1 + o_p(1)) \le L_n(m_0^l)(1 + o_p(1)), \tag{15}$$

where the $o_p(1)$ terms on both sides of Eq. (15) tend to 0 in probability as $n \to \infty$ uniformly in $\theta \in E_q(B)$, and

$$L_n(m^l) = L_n(m_1^l)(1 + o_p(1)), \tag{16}$$

where the equality Eq. (16) holds on a set whose probability tends to 1 as $n \to \infty$ uniformly in $\theta \in E_q(B)$ and the $o_p(1)$ term on the r.h.s. of Eq. (16) tends to 0 in probability as $n \to \infty$ uniformly in $\theta \in E_q(B)$.

Also, for this case,

(b) $$n^{\frac{2q}{2q+1}} L_n(m^l) = O_p(1) \text{ uniformly in } \theta \in E_q(B).$$

Theorem 2.2 *Uniformly in $\theta \in E_q(B)$, we have*

$$n^{\frac{2q}{2q+1}} L_n(m^{\mathrm{AIC}}) = O_p(1).$$

Theorem 2.3 *Let $\hat{\theta}$ be the estimate of θ after a model is chosen by AIC^u, i.e., $\hat{\theta}_i = 0$ for $i > m^u$ and $\hat{\theta}_i = y_i$ for $i \le m^u$. Then $\hat{\theta}$ achieves the minimax rate of convergence, i.e., for any $B < \infty$,*

$$\lim_{n\to\infty} \sup_{\theta\in E_q(B)} n^{\frac{2q}{2q+1}} \; E(||\hat{\theta} - \theta||^2) < \infty.$$

Remark The lower truncation in Theorem 2.1 cannot be removed. This is easy to see by considering what happens for $\theta_n = (1/\sqrt{n}, 0, 0, \dots)$.

Proofs are given in the next section.

3 Proofs

Proof of Theorem 2.1. In the following, $e_i = y_i - \theta_i, i = 1, 2, \dots$ and $B = 1$ without any loss of generality. The proof has been divided into three steps for the purpose of clarity.

Step 1. In this step we will look at a simple minimax rule as follows. Consider rule r_m for a fixed m: For $m \le n$, estimate θ_i by y_i for $1 \le i \le m$ and for $i > m$, estimate θ_i by 0.

Risk of r_m at θ is $r_m(\theta)$ as defined in Eq. (3). Note that we can write

$$r_m(\theta) = \frac{m}{n} + \sum_{m+1}^{\infty} \frac{1}{i^{2q}} i^{2q} \theta_i^2.$$

Then,

$$\sup_{\theta\in E_q(1)} r_m(\theta) = \frac{m}{n} + \sup_{\theta\in E_q(1)} \left(\sum_{m+1}^{\infty} \frac{1}{i^{2q}} i^{2q} \theta_i^2 \right) = \frac{m}{n} + \frac{1}{(m+1)^{2q}}.$$

Now choose m^* as

$$m^* = \operatorname*{argmin}_{1\le m\le n} \left\{ \frac{m}{n} + \frac{1}{(m+1)^{2q}} \right\}. \tag{17}$$

It is easy to show that $m^* \sim n^{1/2q+1}$ asymptotically, whence the maximum risk of the rule r_{m^*} is

$$\sup_{\theta\in E_q(1)} r_{m^*}(\theta) = \frac{m^*}{n} + \frac{1}{(m^*+1)^{2q}} \sim n^{\frac{-2q}{(2q+1)}} \text{ asymptotically,}$$

$$\text{i.e., } \lim_{n\to\infty} n^{\frac{-2q}{(2q+1)}} \sup_{\theta\in E_q(1)} r_{m^*}(\theta) < \infty. \tag{18}$$

Thus r_{m^*} is a rule which attains the asymptotic minimax rate of convergence. Now note that $r_{m_0}(\theta) \le r_{m^*}(\theta), \forall\theta \in E_q(1)$, i.e.,

$$\sup_{\theta\in E_q(1)} r_{m_0}(\theta) \le \sup_{\theta\in E_q(1)} r_{m^*}(\theta) \sim n^{\frac{-2q}{(2q+1)}} \text{ asymptotically,} \tag{19}$$

whence the rule r_{m_0} based on the oracle m_0 does at least as well as the rule r_{m^*} asymptotically.

Step 2. In this step we will consider the lower truncated AIC and derive several properties associated with it.

First recall the definitions of $L_n(m)$ and $S(m)$ from Eqs. (8) and (7) respectively. Note,

$$L_n(m) = \sum_1^m e_i^2 + \sum_{m+1}^{\infty} \theta_i^2 = r_m(\theta) + \left(\sum_1^m e_i^2 - \frac{m}{n}\right).$$

Again,

$$\begin{aligned}
S(m) &= \sum_{m+1}^{n} e_i^2 + \sum_{m+1}^{n} \theta_i^2 + 2\sum_{m+1}^{n} \theta_i e_i + \frac{2m}{n} \\
&= \sum_1^n e_i^2 - \sum_1^m e_i^2 + \sum_{m+1}^{n} \theta_i^2 + 2\sum_{m+1}^{n} \theta_i e_i + \frac{2m}{n} \\
&= L_n(m) + R_n(m) + \sum_1^n e_i^2 - \sum_{n+1}^{\infty} \theta_i^2 \\
&= L_n(m) + R_n(m) + (\text{"constants" independent of } m),
\end{aligned}$$

where

$$R_n(m) = 2\sum_{m+1}^{n} \theta_i e_i - 2\left(\sum_1^m e_i^2 - \frac{m}{n}\right).$$

Hence, minimizing $S(m)$ with respect to m is equivalent to minimizing $L_n(m) + R_n(m)$ over m. So we have, with m^l, m_0^l and m_1^l as in Eqs. (13) and (14),

$$L_n(m^l) + R_n(m^l) \leq L_n(m_0^l) + R_n(m_0^l) \tag{20}$$

and

$$L_n(m^l) + R_n(m^l) \leq L_n(m_1^l) + R_n(m_1^l). \tag{21}$$

Let us now prove three lemmas which essentially show that the remainder terms $R_n(m^l)$, $R_n(m_0^l)$ and $R_n(m_1^l)$ in Eqs. (20) and (21) are negligible. These lemmas are crucial for proving the theorem.

Lemma 3.1

$$\sum_1^m e_i^2 - \frac{m}{n} = o_p(r_m(\theta)) \tag{22}$$

and

$$\sum_{m+1}^{n} \theta_i e_i = o_p(r_m(\theta)), \tag{23}$$

uniformly $\theta \in E_q(1)$ *and for* $m_n \le m \le n$ *as* $n \to \infty$.

So, for such a sequence $\{m_n\}$, *we also have,*

$$R_n(m) = o_p(r_m(\theta)) \tag{24}$$

uniformly in $\theta \in E_q(1)$ *and for* $m_n \le m \le n$ *as* $n \to \infty$.

Proof $E\left(\sum_1^m e_i^2 - m/n\right) = 0$. Fix $\epsilon > 0$. Then,

$$P\left(\left|\frac{\sum_1^m e_i^2 - m/n}{r_m(\theta)}\right| > \varepsilon\right) \le \frac{E\left(\sum_1^m e_i^2 - m/n\right)^2}{\varepsilon^2 r_m^2(\theta)}.$$

But,

$$E\left(\sum_1^m e_i^2 - \frac{m}{n}\right)^2 = \mathrm{Var}\left(\sum_1^m e_i^2 - \frac{m}{n}\right) = m\,\mathrm{Var}\left(e_i^2\right) = \frac{2m}{n^2}.$$

Also,

$$r_m^2(\theta) = \left(\sum_{m+1}^\infty \theta_i^2\right)^2 + \frac{m^2}{n^2} + \frac{2m}{n}\left(\sum_{m+1}^\infty \theta_i^2\right).$$

So,

$$\begin{aligned}\frac{E\left(\sum_1^m e_i^2 - m/n\right)^2}{\epsilon^2 r_m^2(\theta)} &= \frac{1}{\epsilon^2}\cdot\frac{2m/n^2}{\left(\sum_{m+1}^\infty \theta_i^2\right)^2 + m^2/n^2 + 2m/n\left(\sum_{m+1}^\infty \theta_i^2\right)}\\ &< \frac{2}{\epsilon^2 m} \le \frac{2}{\epsilon^2 m_n} \to 0 \quad \text{as } n \to \infty,\end{aligned}$$

for each $\theta \in E_q(1)$, proving Eq. (22). Similarly,

$$E\left(\sum_{m+1}^n \theta_i e_i\right) = 0 \quad \text{and Var}\left(\sum_{m+1}^n \theta_i e_i\right) = \left(\sum_{m+1}^n \theta_i^2\right)\frac{1}{n},$$

whereby

$$\begin{aligned}\frac{\mathrm{Var}\left(\sum_{m+1}^n \theta_i e_i\right)}{r_m^2(\theta)} &= \frac{\left(\sum_{m+1}^n \theta_i^2\right)\frac{1}{n}}{m^2/n^2 + \left(\sum_{m+1}^\infty \theta_i^2\right)^2 + 2m/n\left(\sum_{m+1}^\infty \theta_i^2\right)}\\ &< \frac{1}{2m} \le \frac{1}{2m_n} \to 0 \text{ as } n \to \infty,\end{aligned}$$

for each $\theta \in E_q(1)$, proving Eq. (23). Equation (24) now follows trivially from Eqs. (22) and (23) as

$$R_n(m) = 2\sum_{m+1}^n \theta_i e_i - 2\left(\sum_1^m e_i^2 - \frac{m}{n}\right).$$

So, Lemma 3.1 is proved. □

Corollary 3.1

$$L_n(m_1^l)(1 + o_p(1)) = r_{m_0^l}(\theta)(1 + o_p(1)) \tag{25}$$

almost surely, where the $o_p(1)$ terms on both sides of Eq. (25) tend to zero in probability as $n \to \infty$ uniformly in $\theta \in E_q(B)$.

Proof Using Lemma 3.1, we get

$$L_n(m) = r_m(\theta) + \left(\sum_1^m e_i^2 - \frac{m}{n}\right) = r_m(\theta)(1 + o_p(1)),$$

uniformly in $\theta \in E_q(1)$ and for $m_n \le m \le n$ as $n \to \infty$.

As m_0^l is a nonrandom integer in $[m_n, n]$, it also follows from the above observation that

$$\begin{aligned} L_n(m_0^l) &= r_{m_0^l}(\theta) + \left(\sum_1^{m_0^l} e_i^2 - \frac{m_0^l}{n}\right) \\ &= r_{m_0^l}(\theta)(1 + o_p(1)) \text{ uniformly in } \theta \in E_q(1) \text{ as } n \to \infty. \end{aligned}$$

Now observe that, from definition, $L_n(m_1^l) \le L_n(m_0^l)$, $r_{m_0^l}(\theta) \le r_{m_1^l}(\theta)$. Also note that

$$\sum_1^{m_1^l} e_i^2 - \frac{m_1^l}{n} = o_p(r_{m_1^l}(\theta)), \text{ uniformly in } \theta \in E_q(1) \text{ as } n \to \infty.$$

The last statement follows using the same argument employed in proving $R_{n1}(m^l) = o_p(r_{m^l}(\theta))$ uniformly in $\theta \in E_q(1)$ as $n \to \infty$ in Lemma 3.3. Combining all the above facts, one gets after some algebra,

$$\begin{aligned} L_n(m_1^l) &\le r_{m_0^l}(\theta)(1 + o_p(1)) \le r_{m_1^l}(\theta)(1 + o_p(1)) \\ &= L_n(m_1^l)(1 + o_p(1)) \text{ almost surely,} \end{aligned}$$

where all the $o_p(1)$ terms tend to 0 in probability as $n \to \infty$ uniformly in $\theta \in E_q(B)$. The proof of Eq. (25) now follows immediately from the above sequence of inequalities. □

Lemma 3.2 *$R_n(m) = o_p(L_n(m))$ uniformly in $\theta \in E_q(1)$ and for $m_n \le m \le n$ and as $n \to \infty$.*

Proof Fix $0 < \epsilon < 1$.

$$\begin{aligned} P\left\{\left|\frac{R_n(m)}{L_n(m)}\right| > \epsilon\right\} &= P\left\{\left|\frac{R_n(m)}{L_n(m)}\right| > \epsilon, \left|\frac{L_n(m)}{r_m(\theta)}\right| < 1 - \epsilon\right\} \\ &\quad + P\left\{\left|\frac{R_n(m)}{L_n(m)}\right| > \epsilon, \left|\frac{L_n(m)}{r_m(\theta)}\right| > 1 + \epsilon\right\} \\ &\quad + P\left\{\left|\frac{R_n(m)}{L_n(m)}\right| > \epsilon, 1 - \epsilon \le \left|\frac{L_n(m)}{r_m(\theta)}\right| \le 1 + \epsilon\right\}. \end{aligned} \tag{26}$$

The first two terms on the r.h.s. of Eq. (26) converge to zero as $L_n(m)/r_m(\theta) = 1 + o_p(1)$, uniformly in $\theta \in E_q(1)$ and for $m_n \le m \le n$ as $n \to \infty$. The third term is less than

$$P\left\{\left|\frac{R_n(m)}{r_m(\theta)}\right| > \epsilon(1-\epsilon)\right\} \to 0 \text{ as } n \to \infty,$$

by Lemma 3.1 uniformly in $\theta \in E_q(1)$ and for $m_n \le m \le n$. This proves Lemma 3.2.

Lemma 3.3 $R_n(m^l) = o_p(L_n(m^l))$ *uniformly in* $\theta \in E_q(1)$ *as* $n \to \infty$.

Proof We first prove that

$$R_n(m^l) = o_p(r_{m^l}(\theta)), \text{ uniformly in } \theta \in E_q(1) \text{ as } n \to \infty. \tag{27}$$

Now write,

$$R_n(m) = R_{n1}(m) + R_{n2}(m),$$

where

$$R_{n1}(m) = -2\left(\sum_1^m e_i^2 - \frac{m}{n}\right) \text{ and } R_{n2}(m) = 2\sum_{m+1}^n \theta_i e_i.$$

Fix $\epsilon > 0$. Then,

$$P\left\{\left|\frac{R_{n1}(m^l)}{r_{m^l}(\theta)}\right| > \epsilon\right\} \le P\left\{\max_{m_n \le m \le n}\left|\frac{R_{n1}(m)}{r_m(\theta)}\right| > \epsilon\right\}$$
$$\le \sum_{m_n \le m \le n} P\left\{\left|\frac{R_{n1}(m)}{r_m(\theta)}\right| > \epsilon\right\} \le \sum_{m_n \le m \le n} \frac{1}{\epsilon^4}\frac{E(R_{n1}^4(m))}{r_m^4(\theta)}.$$

Noting that $\left\{\sum_1^m e_i^2 - m/n,\ \ m \ge 1\right\}$ is a Martingale, we have, by a result proved in Dharmadhikari et al. (1968),

$$E\left(\sum_1^m e_i^2 - \frac{m}{n}\right)^4 \le D_1 m^2 E\left(e_1^2 - \frac{1}{n}\right)^4 \text{ for a positive constant } D_1$$
$$= \frac{D_1 m^2}{n^4} E\left(ne_1^2 - 1\right)^4$$
$$= \frac{D_2 m^2}{n^4}, \text{ for some new constant } D_2 \text{ as } ne_1^2 \sim \chi^2_{(1)}.$$

In the above $\chi^2_{(1)}$ refers to a central chi-square distribution with one degree of freedom.

So,
$$\sum_{m_n \le m \le n} \frac{E(R_{n1}^4(m))}{r_m^4(\theta)} \le \sum_{m_n \le m \le n} \frac{16 D_2 m^2/n^4}{\left(m/n + \sum_{m+1}^\infty \theta_i^2\right)^4}$$
$$\le \sum\nolimits_{m_n \le m \le n} \frac{16 D_2}{m^2} \to 0$$

as $n \to \infty$, whence $R_{n1}(m^l) = o_p(r_{m^l}(\theta))$, uniformly in $\theta \in E_q(1)$ as $n \to \infty$.

Now consider $R_{n2}(m)$ and note that

$$\frac{1}{2} R_{n2}(m) = \sum_{m+1}^{n} \theta_i e_i \sim N\left(0, \left(\sum_{m+1}^{n} \theta_i^2\right) \frac{1}{n}\right).$$

So we get, $E\left(\sum_{m+1}^{n} \theta_i e_i\right)^4 = 3\left(\sum_{m+1}^{n} \theta_i^2\right)^2 \cdot 1/n^2$. This implies, by a simple algebra, that

$$E \frac{\left(\sum_{m+1}^{n} \theta_i e_i\right)^4}{r_m^4(\theta)} < \frac{1}{2m^2}.$$

Using the last inequality in the same way as we did for $R_{n1}(m^l)$, we have $R_{n2}(m^l) = o_p(r_{m^l}(\theta))$ uniformly in $\theta \in E_q(1)$ as $n \to \infty$, proving Eq. (27).

We are done if we can show that $L_n(m^l) = r_{m^l}(\theta)(1 + o_p(1))$ uniformly in $\theta \in E_q(1)$ as $n \to \infty$, because then $R_n(m^l) = o_p(L_n(m^l))$ will follow by using exactly the same logic as in the proof of Lemma 3.2.

Fix $\epsilon > 0$. Then, by a simple argument,

$$P\left\{\left|\frac{L_n(m^l)}{r_{m^l}(\theta)} - 1\right| > \epsilon\right\} \le \sum_{m_n \le m \le n} \frac{1}{\epsilon^4} \cdot \frac{E\left(\sum_1^m e_i^2 - m/n\right)^4}{r_m^4(\theta)} \to 0$$

as $n \to \infty$ for all $\theta \in E_q(1)$, as already shown before. So, Lemma 3.3 is proved. □

Step 3. In this step we combine the results in Step 1 and Step 2 to finally prove Theorem 2.1.

Equation (15) of Part (a) of Theorem 2.1 follows by applying Lemma 3.3 to the left-hand side of Eq. (20) and Lemma 3.2 to the right-hand side of Eq. (20), by just noting that m_0^l is a nonrandom integer in $[m_n, n]$.

Equation (16) is then proved as follows. By Eqs. (15), (25) and the facts that $L_n(m_0^l) = r_{m_0^l}(1 + o_p(1))$ uniformly in $\theta \in E_q(1)$ as $n \to \infty$ and $L_n(m_1^l) \le L_n(m^l)$, one has

$$L_n(m_1^l)(1+o_p(1)) \le L_n(m^l)(1+o_p(1)) \le r_{m_0^l}(1+o_p(1)) = L_n(m_1^l)(1+o_p(1)),$$

where the above holds on a set whose probability tends to 1 as $n \to \infty$ uniformly in $\theta \in E_q(B)$. (Note that all the $o_p(1)$ terms in the above statement tend to 0 in probability as $n \to \infty$ uniformly in $\theta \in E_q(B)$.) Equation (16) then follows immediately from the above.

We shall now prove Part (b) of Theorem 2.1. To prove this, we first recall the definitions of m_0^l, m^* and the asymptotic minimaxity property of r_{m^*} as in Eqs. (13), (17) and (18) respectively. Then it follows that

$$r_{m_0^l}(\theta) \le r_{m^*}(\theta) \le D n^{-2q/(2q+1)}$$

for some $D > 0$ for each $\theta \in E_q(1)$ for all sufficiently large n in the same way as one shows in Eq. (19). Part (b) then follows easily by applying this fact together with Corollary 3.1 in part (a) of Theorem 2.1.

Remark If we choose $m_n = m^*$ in Theorem 2.1 and then combine with it the result of Theorem 2.3, Theorem 2.2 follows immediately. To see this one has to note that

$$\text{(i) } P\left\{n^{\frac{2q}{2q+1}} L_n(m^u) > K\right\} \leq n^{\frac{2q}{2q+1}} \frac{E(L_n(m^u))}{K}, \qquad \text{for each } K > 0,$$

$$\text{(ii) } n^{\frac{2q}{2q+1}} E(L_n(m^u)) \text{ is bounded for each } \theta \in E_q(1) \qquad \text{for each } n, \text{ and}$$

$$\text{(iii) } L_n(m^{\text{AIC}}) \leq \max\left\{L_n(m^u), L_n(m^l), L_n\left(m^{\text{AIC}}_{[n^{\frac{1}{2q+1}}]<m<m^*}\right)\right\} \quad \text{if } m^* > n^{\frac{1}{2q+1}},$$

$$L_n(m^{\text{AIC}}) \leq \max\left\{L_n(m^u), L_n(m^l)\right\} \qquad \text{if } m^* \leq n^{\frac{1}{2q+1}},$$

and $E(L_n(m)) \leq F n^{-\frac{2q}{2q+1}}$ for all large enough n, if $[n^{\frac{1}{2q+1}}] < m < m^*$,

where F is a positive number. In the above,

$$m^{\text{AIC}}_{[n^{\frac{1}{2q+1}}]<m<m^*} = \underset{[n^{\frac{1}{2q+1}}]<m<m^*}{\text{argmin}} \quad S(m),$$

where $S(m)$ is as in Eq. (7). □

But we present a more direct proof of Theorem 2.2 which does not require Theorem 2.3 and which explains the interesting behaviour of AIC for relatively small m.

Proof of Theorem 2.2 Fix $\epsilon > 0$ and $\eta > 0$ arbitrary. We can choose $\tilde{m}$ large enough such that

$$P\left\{2\left|\sum_{\tilde{m}+1}^{n} \theta_i e_i\right| < \epsilon L_n(\tilde{m})\right\} > 1 - \frac{\eta}{6}, \quad \forall n \geq \tilde{m}, \quad \forall \theta \in E_q(B)$$

and

$$P\left\{n^{\frac{2q}{2q+1}} L_n\left(m^{\text{AIC}}_{m\geq\tilde{m}}\right) \leq K_\epsilon\right\} > 1 - \eta, \text{ for all large enough } n > \tilde{m}, \forall\theta \in E_q(B)$$

for some suitably chosen K_ϵ, where $m^{\text{AIC}}_{m\geq\tilde{m}}$ is defined as

$$m^{\text{AIC}}_{m\geq\tilde{m}} = \underset{\tilde{m}\leq m\leq n}{\text{argmin}}\, S(m).$$

The above two probability statements follow directly from the arguments used in the proof of Theorem 2.1. We shall henceforth write $m^{\text{AIC}}_{\text{old}} = m^{\text{AIC}}_{m\geq\tilde{m}}$. For each $\theta \in E_q(B)$, define

$$K(\theta) = \max\left\{I : \sum_{i=I}^{\tilde{m}} \theta_i^2 > \epsilon n^{-2q/(2q+1)}\right\},$$

where the maximum is taken over the range $[1, \tilde{m}]$.

Now note, for each $\theta \in E_q(B)$, the following two cases occur.

Case 1. $\sum_{m^1+1}^{\tilde{m}} \theta_i^2 > \epsilon n^{-2q/(2q+1)}$ if $1 \le m^1 \le K(\theta) - 1$.

Case 2. $\sum_{m^1+1}^{\tilde{m}} \theta_i^2 \le \epsilon n^{-2q/(2q+1)}$ if $K(\theta) \le m^1 < \tilde{m}$.

Consider Case 2 first. Let $\tilde{m} > m^1 \ge K(\theta)$. Then

$$\begin{aligned} L_n(m^1) &= L_n(\tilde{m}) - R_n^1(m^1), \qquad \text{where} \\ R_n^1(m^1) &= \sum_{m^1+1}^{\tilde{m}} (y_i - \theta_i)^2 - \sum_{m^1+1}^{\tilde{m}} \theta_i^2. \end{aligned}$$

Now fix $\eta^1 > 0$, arbitrarily small. It is easy to show, by noting that $n \sum_{m^1+1}^{\tilde{m}} (y_i - \theta_i)^2 \sim \chi^2_{(\tilde{m}-m^1)}$ and $\tilde{m}$ is fixed, that for all large enough n,

$$P\left\{ \left| n^{2q/(2q+1)} R_n^1(m^1) \right| \le 2\epsilon \right\} > 1 - \eta^1, \quad \forall \theta \in E_q(B) \text{ and } K(\theta) \le m^1 < \tilde{m}.$$

Letting $\tilde{S}(m) = S(m) - \left(\sum_1^n e_i^2 - \sum_{n+1}^{\infty} \theta_i^2\right)$, henceforth, we have

$$\tilde{S}(m^1) = L_n(m^1) + 2\sum_{m^1+1}^{\tilde{m}} \theta_i e_i + 2\sum_{\tilde{m}+1}^{n} \theta_i e_i - 2\left(\sum_1^{m^1} e_i^2 - \frac{m^1}{n}\right).$$

Now note that $\tilde{m}$ is a fixed number, $\theta_i e_i \sim N(0, \theta_i^2/n)$ independently and $e_i^2 - 1/n$ are independently distributed with mean 0 and variance $2/n^2$. Using these facts and the last equality, it is easy to show that

$$n^{2q/(2q+1)} L_n(m^1) \le n^{2q/(2q+1)} \tilde{S}(m^1) + 2\epsilon + 2n^{2q/(2q+1)} \left| \sum_{\tilde{m}+1}^{n} \theta_i e_i \right|,$$

for all large enough n with probability at least $1 - 2\eta^1$, for each $\theta \in E_q(B)$ and $K(\theta) \le m^1 < \tilde{m}$.

We now consider Case 1. Note that

$$S(\tilde{m}) = S(m^1) + \left\{ 2\left(\frac{\tilde{m} - m^1}{n}\right) - \sum_{m^1+1}^{\tilde{m}} y_i^2 \right\}$$

and

$$\sum_{m^1+1}^{\tilde{m}} y_i^2 - 2\left(\frac{\tilde{m} - m^1}{n}\right) \stackrel{D}{=} \frac{1}{n}\left\{ \chi^2_{(\tilde{m}-m^1-1)} + W^2 - 2\left(\tilde{m} - m^1\right) \right\}$$

where

$$W \sim N\left(\sqrt{n \sum_{m^1+1}^{\tilde{m}} \theta_i^2}\ , 1 \right).$$

Since, $m^1 \le K(\theta) - 1$, $n \sum_{m^1+1}^{\tilde{m}} \theta_i^2 > \epsilon n^{1/(2q+1)} \to \infty$ as $n \to \infty$. So, again, we can choose n large enough so that

$$P\left\{\sum_{m^1+1}^{\tilde{m}} y_i^2 - \frac{2\left(\tilde{m} - m^1\right)}{n} > 0\right\} > 1 - \eta^1,$$

for all $\theta \epsilon E_q(B)$ and $1 \le m^1 \le K(\theta) - 1$, implying

$$S(\tilde{m}) < S(m^1), \text{ i.e., } \quad m^{\text{AIC}}_{\text{new}, m^1} = m^{\text{AIC}}_{\text{old}},$$
$$\text{where } m^{\text{AIC}}_{\text{new},m^1} = \underset{\substack{m \ge \tilde{m}, \\ m = m^1}}{\operatorname{argmin}} S(m), \text{ with}$$

probability bigger than $1 - \eta^1$ each $1 \le m^1 \le K(\theta) - 1$.

Finally note that the fact $\tilde{S}(m^{\text{AIC}}_{\text{old}}) = L_n(m^{\text{AIC}}_{\text{old}}) + R_n(m^{\text{AIC}}_{\text{old}})$, where $R_n(m)$ is as defined in the proof of Theorem 2.1, implies, by an easy argument, that each of the following events

$$n^{\frac{2q}{2q+1}} L_n\left(m^{\text{AIC}}_{\text{old}}\right) \le n^{\frac{2q}{2q+1}} \frac{1}{1-\epsilon} \tilde{S}\left(m^{\text{AIC}}_{\text{old}}\right) + \frac{4\epsilon}{1-\epsilon}$$

and

$$\tilde{S}\left(m^{\text{AIC}}_{\text{old}}\right) \le (1+\epsilon) L_n\left(m^{\text{AIC}}_{\text{old}}\right)$$

holds with probability bigger than $1 - \eta/3$, for all $\theta \in E_q(B)$ and $\forall n \ge \tilde{m}$.

Now, consider, for each $\theta \in E_q(B)$, the probability of the following occurring simultaneously

$$\begin{aligned}
&\Big\{ n^{\frac{2q}{2q+1}} L_n\left(m^1\right) \le n^{\frac{2q}{2q+1}} \tilde{S}(m^1) + 2\epsilon + 2n^{\frac{2q}{2q+1}} \left|\sum_{\tilde{m}+1}^{n} \theta_i e_i\right| \quad \text{for all } K(\theta) \le m^1 < \tilde{m}, \\
&\left| n^{\frac{2q}{2q+1}} R_n^1\left(m^1\right)\right| \le 2\epsilon \quad \text{for all } K(\theta) \le m^1 < \tilde{m}, \\
&L_n\left(m^{\text{AIC}}_{\text{new},m^1}\right) = L_n\left(m^{\text{AIC}}_{\text{old}}\right) \quad \text{for } 1 \le m^1 \le K(\theta) - 1, \\
&n^{\frac{2q}{2q+1}} L_n\left(m^{\text{AIC}}_{\text{old}}\right) \le n^{\frac{2q}{2q+1}} \frac{1}{1-\epsilon} \tilde{S}\left(m^{\text{AIC}}_{\text{old}}\right) + \frac{4\epsilon}{1-\epsilon}, \\
&2\left|\sum_{\tilde{m}+1}^{n} \theta_i e_i\right| < \epsilon L_n(\tilde{m}),\ n^{\frac{2q}{2q+1}} L_n\left(m^{\text{AIC}}_{\text{old}}\right) \le K_\epsilon, \\
&\tilde{S}(m^{\text{AIC}}_{\text{old}}) \le (1+\epsilon) L_n(m^{\text{AIC}}_{\text{old}}) \Big\}.
\end{aligned}$$

The probability of this event can be shown to be larger than $1 - 11/6\eta - 4\tilde{m}\eta^1$ for all large enough n for each $\theta \in E_q(B)$. So the above event, in turn implies, the following event,

$$n^{\frac{2q}{2q+1}} L_n\left(m^{\text{AIC}}\right) \le \frac{1+\epsilon}{1-\epsilon} K_\epsilon + \frac{4\epsilon}{1-\epsilon}.$$

The above statement follows by observing that

$$m^{\text{AIC}} = m^{\text{AIC}}_{\text{old}} \text{ or } m^{\text{AIC}}_{\text{new},m^1} \quad \text{for some } 1 \le m^1 < \tilde{m}.$$

Now we are done, as $\tilde{m}$ is a fixed number for a given η and ϵ and so η^1 can be chosen to be $\eta/24\tilde{m}$, to start with, making the quantity $1 - 11/6\eta - 4\tilde{m}n^1$ equal to $1 - 2\eta$. This completes the proof of Theorem 2.2. □

Proof of Theorem 2.3 Fix any $C > 0$. Define $\lambda(\theta)$ as in Zhao (2000), i.e.,

$$\lambda(\theta) = \max\left\{ I : \sum_{i=I}^{\infty} \theta_i^2 \ge (B + C)n^{-\frac{2q}{2q+1}} \right\}.$$

It is easy to see, vide Zhao (2000), that $\lambda(\theta) \le n^{1/(2q+1)}$.

Now recall the definition of m^u from Eq. (12). Note that

$$\begin{aligned} E\left(||\hat{\theta} - \theta||^2\right) &= E(L_n(m^u)) \\ &= E\left\{ \sum_{1}^{m^a} (y_i - \theta_i)^2 + \sum_{m^a+1}^{\infty} \theta_i^2 \right\}. \end{aligned}$$

Now $E(L_n(m^u)) = E(L_n(m^u)1_{(m^a \le \lambda(\theta))}) + E(L_n(m^u)1_{(m^a > \lambda(\theta))})$. But,

$$\begin{aligned} E(L_n(m^u)1_{(m^a > \lambda(\theta))}) = \; & E\left\{ \sum_{1}^{m^a} (y_i - \theta_i)^2 1_{(m^a > \lambda(\theta))} \right\} \\ & + E\left\{ \left(\sum_{m^a+1}^{\infty} \theta_i^2 \right) 1_{(m^a > \lambda(\theta))} \right\}. \end{aligned} \tag{28}$$

The second term in the r.h.s. of Eq. (28) is trivially less than $(B + C)n^{-2q/(2q+1)}$. The first term is less than or equal to $E\left\{\sum_1^{[n^{1/2q+1}]}(y_i - \theta_i)^2\right\} = [n^{1/(2q+1)}]/n \le n^{-2q/(2q+1)}$, whence $E(L_n(m^u)1_{(m^a > \lambda(\theta))}) < (B + C + 1)n^{-2q/(2q+1)}$.

Again,

$$\begin{aligned} E(L_n(m^u)1_{(m^a \le \lambda(\theta))}) &= E\left\{ \left(\sum_{1}^{m^a} (y_i - \theta_i)^2 + \sum_{m^a+1}^{\lambda(\theta)} \theta_i^{\ 2} + \sum_{\lambda(\theta)+1}^{\infty} \theta_i^2 \right) 1_{(m^a \le \lambda(\theta))} \right\} \\ &< (B + C + 1)n^{-\frac{2q}{2q+1}} + E\left\{ \left(\sum_{m^a+1}^{\lambda(\theta)} \theta_i^2 \right) 1_{(m^a \le \lambda(\theta))} \right\}. \end{aligned} \tag{29}$$

Now consider any number $K > \sqrt{2}$. The second expression in Eq. (29) equals

$$E\left\{\left(\sum_{m^a+1}^{\lambda(\theta)} \theta_i^2\right) 1_{\left(m^a \le \lambda(\theta), \sum_{m^a+1}^{\lambda(\theta)} \theta_i^2 \le K^2 n^{-2q/(2q+1)}\right)}\right\}$$
$$+E\left\{\left(\sum_{m^a+1}^{\lambda(\theta)} \theta_i^2\right) 1_{\left(m^a \le \lambda(\theta), \sum_{m^a+1}^{\lambda(\theta)} \theta_i^2 > K^2 n^{-2q/(2q+1)}\right)}\right\}$$
$$\le K^2 n^{-\frac{2q}{2q+1}} + E\left\{\left(\sum_{m^a+1}^{\lambda(\theta)} \theta_i^2\right) 1_{\left(m^a < \lambda(\theta), \sum_{m^a+1}^{\lambda(\theta)} \theta_i^2 > K^2 n^{-2q/(2q+1)}\right)}\right\}$$
$$\le K^2 n^{-\frac{2q}{2q+1}} + \sum_{m=1}^{\lambda(\theta)-1} B\,P(m^u = m)\,I\left(\sum_{m+1}^{\lambda(\theta)} \theta_i^2 > K^2 n^{-2q/(2q+1)}\right),$$
as $\theta \in E_q(B)$.

Now for any $m < \lambda(\theta)$,

$$P(m^u = m) \le P\{S(\lambda(\theta)) > S(m)\} = P\left\{n \sum_{m+1}^{\lambda(\theta)} y_i^2 < 2(\lambda(\theta) - m)\right\}. \quad (30)$$

Noting that $n \sum_{m+1}^{\lambda(\theta)} y_i^2 \overset{D}{=} \chi^2_{(\lambda(\theta)-m-1)} + W^2$, where $W \sim N\left(\sqrt{n \sum_{m+1}^{\lambda(\theta)} \theta_i^2}, 1\right)$, the expression in Eq. (30) is less than

$$P\left\{W < \sqrt{2}\sqrt{\lambda(\theta) - m}\right\} \le P\left\{Z < (\sqrt{2} - K) n^{\frac{1}{2(1+2q)}}\right\},$$

where $Z \sim N(0, 1)$, using the fact that $\sum_{m+1}^{\lambda(\theta)} \theta_i^2 > K^2 n^{-2q/(2q+1)}$ and $\lambda(\theta) \le n^{1/(2q+1)}$. But,

$$P\left\{Z < (\sqrt{2} - K) n^{1/[2(1+2q)]}\right\}$$
$$\le \frac{1/\sqrt{2\pi} \exp\{-1/2(K-\sqrt{2})^2 n^{1/(1+2q)}\}}{(K-\sqrt{2}) n^{1/2(1+2q)}}.$$

So,

$$\sum_{m=1}^{\lambda(\theta)-1} B\,P(m^u = m) I\left(\sum_{m+1}^{\lambda(\theta)} \theta_i^2 > K^2 n^{-2q/(2q+1)}\right)$$
$$\le \frac{B n^{1/[2(1+2q)]}}{\sqrt{2\pi}(K - \sqrt{2}) \exp\{1/2(K - \sqrt{2})^2 n^{1/(1+2q)}\}} \le C^1 n^{-2q/(2q+1)},$$

for some constant C^1. Hence Theorem 2.3 is proved. □

4 Discussion

A counter-example to show the nonoptimal behaviour of BIC

Consider $q = 1$, $B = 1$. Define $H(n) = \left[\frac{n^{1/3}}{(\log n)^{(1-\delta)/2}}\right]$, where $0 < \delta < 1$. Consider a sequence θ^n as follows.

$$\theta = \theta^n = \left\{\theta_j = \frac{(\log n)^{1-\delta/2}}{\sqrt{n}},\ j = 1, 2, \dots, H(n) \text{ and } \theta_j = 0 \text{ if } j > H(n)\right\}.$$

Fact 1. Easy to check that for large enough n, $\theta^n \in E_1(1)$.

Fact 2. For $\theta = \theta^n$ as defined above $\sum_{i=2}^{H(n)} \theta_i{}^2 = \{H(n) - 1\}\frac{(\log n)^{1-\delta}}{n}$.

Consider now the upper truncated BIC defined as follows:

$$m^u = \underset{1 \le m \le [n^{1/3}]}{\operatorname{argmin}} \left\{\sum_{i=m+1}^{n} y_i{}^2 + \frac{m}{n}\log n\right\}.$$

Then the estimate of this rule followed by least squares estimates is $\hat{\theta}_i = y_i$ for $i \le m^u$ and $\hat{\theta}_i = 0$ otherwise. It is easy to see that the expected squared error loss for this estimate is greater than or equal to

$$E\left(\sum_{i=m^u+1}^{H(n)} \theta_i{}^2\right) \ge E\left(\left(\sum_{i=m^u+1}^{H(n)} \theta_i{}^2\right) I_{(m^u=1)}\right) = \left\{\sum_{i=2}^{H(n)} \theta_i{}^2\right\}\left\{1 - \sum_{j=2}^{[n^{1/3}]} P(m^u = j)\right\}.$$

Now evaluating at $\theta = \theta^n$, the last expression is

$$\{H(n) - 1\}\frac{(\log n)^{1-\delta}}{n}\left\{1 - \sum_{j=2}^{[n^{1/3}]} P(m^u = j)\right\}.$$

Note that $P(m^u = j) \le P(Y_j^2 \ge \log n/n)$.

Now $P(Y_j^2 \ge \log n/n) = P(Y_j \ge \sqrt{\log n/n}) + P(Y_j \le -\sqrt{\log n/n})$. Using tails of standard normal probabilities, it is easy to show that for each $2 \le j \le H(n)$, under $\theta = \theta^n$, the above probability is less than $4n\frac{1/(\log n)^{\delta/2}}{\sqrt{\log n}\sqrt{n}}$. So

$$\sum_{j=2}^{[n^{1/3}]} P(m^u = j) \le 4n\frac{1/(\log n)^{\delta/2}}{\sqrt{\log n} \times (n^{1/6})}.$$

Then it immediately follows by a simple algebra that

$$\left(1 - \sum_{j=2}^{[n^{1/3}]} P(m^u = j)\right) > \frac{1}{2}$$

for all sufficiently large n.

But note that

$$n^{2/3}\{H(n)-1\}\frac{(\log n)^{1-\delta}}{n} \to \infty \text{ as } n \to \infty.$$

Using the above facts it follows that $\lim_{n\to\infty} \sup_{\theta\in E_1(1)} n^{2/3} E\left(\sum_{i=1}^{\infty}(\hat{\theta}_i - \theta_i)^2\right) = \infty$. So the upper truncated BIC does not achieve the minimax rate of convergence.

In fact, a careful inspection reveals that this same sequence θ^n can be used to show that BIC does not attain the minimax rate of convergence for any kind of upper truncation. More importantly, the same sequence can be used to show unrestricted BIC followed by least squares also does not achieve the minimax rate, even in the sense of convergence in probability as shown to be true for AIC. But we do not present those arguments in the present paper.

We explore below the connection between the problem studied in our paper and nonparametric regression, vide Eq. (7). Define

$$\begin{aligned}\bar{f}(t) &= f\left(\frac{i}{n+1}\right), \quad \text{if } \frac{i-1}{n} \le t < \frac{i}{n} \text{ for } i = 1, 2, \ldots, n-1\\ &= f\left(\frac{n}{n+1}\right) \quad \text{if } \frac{n-1}{n} \le t \le 1.\end{aligned}$$

It is shown in Brown and Low (1996) that under certain conditions, estimating $\{f(t)\}$ in problem (1.1) through $\{Z(t)\}$ is asymptotically equivalent to estimating $\{\bar{f}(t)\}$ through $\{\bar{Z}(t)\}$, where

$$d\bar{Z}(t) = \bar{f}(t)dt + \frac{dB(t)}{\sqrt{n}},\ 0 \le t \le 1.$$

They also observe that $\{S_i = n\left(\bar{Z}_{i/n} - \bar{Z}_{i-1/n}\right), \text{i=1,2,}\ldots, n\}$ are sufficient for $\bar{Z}(t)$. Now note that $n(\bar{Z}_{i/n} - \bar{Z}_{i-1/n}) \overset{D}{=} Y_i$, $i = 1, 2, \ldots, n$, and the Y_i's are trivially sufficient for the problem (Eq. 7). So any decision rule based on S_i's can be replaced by the same decision rule based on the Y_i's and both will have the same properties. It is also easy to verify, at least heuristically, that the minimization criterion for AIC studied in our theorems is close (up to $O_p(1/n)$) in distribution to the minimization criterion for AIC based on the Y_is and so the models selected by AIC in these two problems are also expected to be close.

We briefly explain below how the theoretical results about the rate optimality of AIC obtained for continuous path data can be applied to the situation when one observes the process $\{Z(t)\}$ only at points $\{t_K = K/N : K = 0, 1, \ldots, N\}$, where $N = N_n$; i.e, N depends on n.

Let $\{\phi_i : i \ge 1\}$ be the usual Fourier basis of $L_2[0, 1]$. Analogous to the y_is, let us define

$$y_i' = \sum_{K=1}^{N} \phi_i\left(\frac{K}{N}\right)\left(z\left(\frac{K}{N}\right) - z\left(\frac{K-1}{N}\right)\right),$$

$i = 1, \ldots, n$; which can be rewritten as

$$y_i' = \theta_i' + \epsilon_i';\ i = 1, \ldots, n$$

where $\theta_i' = \sum_{K=1}^{N} \int_{K-1/N}^{K/N} \phi_i(K/N) f(u)\,du$ and $\epsilon_i' = \frac{1}{\sqrt{n}} \sum_{K=1}^{N} \phi_i(K/N)$ $(B(K/N) - B(K-1/N))$; where $B(.)$ is the Brownian motion. It follows that $|\theta_i - \theta_i'| \leq \frac{ci}{N}$ for some positive constant c which does not depend on i, θ's and θ''s and $\epsilon_i' \sim N(0, 1/n \cdot 1/N \sum_{K=1}^{N} \phi_i^2(K/N))$. So, if N is large compared to n as $n \to \infty$; it is expected that θ_i' and θ_i will be close; ϵ_i' will be approximately $N(0, \frac{1}{n})$ i.e., distributionally close to ϵ_i and then any result/procedure obtained from using the continuous data will be expected to be asymptotically close to the corresponding analogous one based on the discretized version of the problem. Towards that, let us define

$$S_1(m) = \sum_{i=m+1}^{n} y_i'^2 + \frac{2m}{n};$$

the criterion function based on the y_i's corresponding to $S(m)$. Let us heuristically define the "Akaike Information Criterion" for the discrete problem as

$$m_1^{\text{AIC}} = \underset{1 \leq m \leq n}{\arg\min}\; S_1(m)$$

and the loss $L_n'(m_1^{\text{AIC}}) = \sum_{i=1}^{m_1^{\text{AIC}}} (y_i' - \theta_i)^2 + \sum_{i=m_1^{\text{AIC}}+1}^{n} \theta_i^2 + \sum_{i=n+1}^{\infty} \theta_i^2$, which is defined in a manner analogous to $L_n(m^{\text{AIC}})$ (as in Sect. 2). Note that $L_n'(m)$ is the loss in estimating θ_i by y_i' if $i \leq m$ and by 0 otherwise and $L_n'(m) = ||f - \hat{f}||^2$, where $\hat{f}(t) = \sum_{i=1}^{m} y_i' \phi_i(t)$.

We have a rigorous argument (not presented here, so as to not increase the length of the paper) which shows, that

$$L_n(m^{\text{AIC}}) - L_n'(m_1^{\text{AIC}}) = O_p\left(n^{-2q/(2q+1)}\right); \tag{31}$$

uniformly over $\theta \in E_q(B)$; provided $\frac{n^{3+2q/(2q+1)}}{N} = O(1)$ as $n \to \infty$. This implies that $L_n'(m_1^{\text{AIC}}) = O_p(n^{-2q/(2q+1)})$, uniformly in $\theta \in E_q(B)$ using our previous result (Theorem 2.2). (The heart of the argument in proving Eq. (31) lies in showing that with probability tending to 1 (uniformly over $\theta \in E_q(B)$) as $n \to \infty$; $S(m) - S_1(m)$ and $L_n(m) - L_n'(m)$ are uniformly small in magnitute for $1 \leq m \leq n$ (upto the minimax rate $n^{-2q/2q+1}$)).

In summary, we heuristically apply the analogous definition of AIC based on the y_i''s and define the natural loss function based on the same observations. We are able to establish that this transformed version of the AIC for the discretized process does as good a job as the AIC based on the original continuous process, in terms of minimax rate of convergence, provided we observe the discrete process at enough number of equally spaced points. So the results proved for the continuous path data are adaptive to the need for adjustment for discrete data.

Acknowledgements This research was partly supported by Purdue Research Foundation Grant. We would like to thank the referee for the valuable comments and suggestions which led to significant improvement in the content and exposition of the paper.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov & J.O. Berger (Eds.), *Second international symposium on information theory*. Budapest: Akademia Kiado. pp 267–281.

Akaike, H. (1978). A Bayesian analysis of the minimum aic procedure. *Annals of the Institute of Statistical Mathematics*, *30*, 9–14.

Brown, L.D., Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, *24*, 2384–2398.

Dharmadhikari, S.W., Fabian, V., Jogdeo, K. (1968). Bounds on the moments of martingales. *The Annals of Mathematical Statistics*, *39*, 1719–1723.

Ibragimov, I.A., Has'minskii, R.Z. (1981). *Statistical estimation: asymptotic theory*. New York: Springer-Verlag.

Li, K.C. (1987). Asymptotic optimality for c_p, c_l, cross validation and generalized cross validation: discrete index set. *The Annals of Statistics*, *15*, 958–975.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, *7*, 221–264.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika*, *68*, 45–54.

Shibata, R. (1983). Asymptotic mean efficiency of a selection of regression variables. *Annals of the Institute of Statistical Mathematics*, *35*, 415–423.

Zhao, L.H. (2000). Bayesian aspects of some nonparametric problems. *The Annals of Statistics*, *28*, 532–552.