AIC for ergodic diffusion processes from discrete observations

Uchida, Masayuki
Faculty of Mathematics, Kyushu University

Yoshida, Nakahiro
Graduate School of Mathematical Sciences, University of Tokyo

出版情報：MHF Preprint Series. MHF2005-12, 2005-03-08. 九州大学大学院数理学研究院
バージョン：
権利関係：

MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

AIC for ergodic diffusion processes from discrete observations

M. Uchida & N. Yoshida

MHF 2005-12

(Received March 8, 2005)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

AIC for ergodic diffusion processes from discrete observations

[1]Masayuki Uchida and [2]Nakahiro Yoshida
[1]Faculty of Mathematics, Kyushu University
Ropponmatsu, Fukuoka 810-8560, Japan
[2]Graduate School of Mathematical Sciences, University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan

Abstract. Akaike's information criterion (AIC) is proposed for evaluating statistical models constructed by the maximum likelihood estimators under the situation where the parametric models contain the true model. In order to obtain AIC, it suffices to get a log likelihood function and the maximum likelihood estimator. However, we can not generally derive AIC for discretely observed diffusion processes since the transition densities of diffusion processes do not commonly have explicit forms. This paper presents AIC type of information criterion for discretely observed ergodic diffusion processes. The information criterion is constructed by using an approximate log likelihood function and an asymptotically efficient estimator. The approximate log likelihood function is based on a result of Dacunha-Castelle and Florens-Zmirou (1986). The asymptotically efficient estimator is derived from a contrast function based on a locally Gaussian approximation.

AMS 2000 subject classifications: Primary 62B10, 62M05; Secondary 62F12, 60J60.

Key words and phrases: Akaike's information criteria, model selection, maximum contrast estimator, approximate log likelihood function, discrete time observation.

Abbreviated Title: AIC for diffusion processes.

1 Introduction

In this paper we consider a family of one-dimensional diffusion processes defined by the stochastic differential equations

$$\begin{aligned} dX_t &= b(X_t,\alpha)dt + \sigma(X_t,\beta)dw_t,\ t\in[0,T], \\ X_0 &= x_0, \end{aligned} \tag{1}$$

where $\theta=(\alpha,\beta)\in\Theta_\alpha\times\Theta_\beta=\Theta$ with Θ_α and Θ_β being compact convex subsets of $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively. Furthermore, b is an $\mathbf{R}$-valued function defined on $\mathbf{R}\times\Theta_\alpha$, σ is an $\mathbf{R}$-valued function defined on $\mathbf{R}\times\Theta_\beta$, and w is a one-dimensional standard Wiener process. We assume that the drift b and the diffusion coefficient σ are known apart from the parameters α and β. Moreover, it is assumed that the process X is ergodic for every θ with invariant probability measure μ_θ. For details of ergodic diffusion processes and the invariant probability measures, see Kutoyants (2004). The data we treat are discrete observations $\mathbf{X}_n=(X_{t_k^n})_{0\le k\le n}$ with $t_k^n=kh_n$, where h_n is the discretization step. The type of asymptotics we consider is when $h_n\to 0$, $nh_n\to\infty$ and $nh_n^2\to 0$ as $n\to\infty$.

Using an information criterion, we discuss the problem of choosing a statistical model among specified parametric models which include the true model. In order to explain the concepts of information criteria, we treat the following simple situation for the moment. Based on the information contained in the observations $\mathbf{X}_n = x_n$, we choose a parametric model which consists of a family of probability densities $\{f(x_n, \theta); \theta \in \Theta\}$. We assume that this specified family of probability densities contain the true density $g(x_n) := f(x_n, \theta_0)$. The adopted parametric model is estimated by replacing the unknown parameter vector θ with an estimator $\hat{\theta}(\mathbf{X}_n)$, for example the maximum likelihood estimator. Then a future observation $\mathbf{Z}_n = z_n$ derived from the true density $g(z_n)$ is predicted by using the statistical model $f(z_n, \hat{\theta}(\mathbf{X}_n))$. We can also recognize that $g(z_n)$ is predicted by the statistical model $f(z_n, \hat{\theta}(\mathbf{X}_n))$. After fitting a parametric model to the observations $\mathbf{X}_n$, we would like to assess the closeness of the statistical model $f(z_n, \hat{\theta}(\mathbf{X}_n))$ to the true density $g(z_n)$. The estimated Kullback-Leibler information

$$I\{g(z_n); f(z_n, \hat{\theta}(\mathbf{X}_n))\} = E_{\mathbf{Z}_n}\left[\log \frac{g(\mathbf{Z}_n)}{f(\mathbf{Z}_n, \hat{\theta}(\mathbf{X}_n))}\right]$$

is used as an overall measure of the divergence of the statistical model $f(z_n, \hat{\theta}(\mathbf{X}_n))$ from the true density $g(z_n)$, conditional on the observations $\mathbf{X}_n$. It can be expressed as

$$I\{g(z_n); f(z_n, \hat{\theta}(\mathbf{X}_n))\} = \int g(z_n) \log g(z_n) dz_n - \int g(z_n) \log f(z_n, \hat{\theta}(\mathbf{X}_n)) dz_n. \tag{2}$$

In the same way as in Akaike (1973,1974), we use here the concept of the model selection based on minimizing the estimated Kullback-Leibler information (2). We see that the first term in the right hand side of (2) does not depend on the statistical model $f(z_n, \hat{\theta}(\mathbf{X}_n))$ while the second term $\eta(\hat{\theta}(\mathbf{X}_n)) := \int g(z_n) \log f(z_n, \hat{\theta}(\mathbf{X}_n)) dz_n$ depends on it, where $\eta(\hat{\theta}(\mathbf{X}_n))$ is called the expected log likelihood. Taking account of it, our selection rule is to choose a model which is maximizing the expected log likelihood $\eta(\hat{\theta}(\mathbf{X}_n))$ among parametric models. However, since the expected log likelihood $\eta(\hat{\theta}(\mathbf{X}_n))$ depends on the true density $g(z_n)$, we need to estimate it. The simple estimator of the expected log likelihood $\eta(\hat{\theta}(\mathbf{X}_n))$ is the estimated log likelihood $l_n(\mathbf{X}_n, \hat{\theta}(\mathbf{X}_n)) := \log f(\mathbf{X}_n, \hat{\theta}(\mathbf{X}_n))$. Usually the estimated log likelihood $l_n(\mathbf{X}_n, \hat{\theta}(\mathbf{X}_n))$ provides an optimistic assessment (overestimation) of the expected log likelihood $\eta(\hat{\theta}(\mathbf{X}_n))$ because the same observations $\mathbf{X}_n$ are used both to estimate the unknown parameter vector θ and to evaluate the expected log likelihood $\eta(\hat{\theta}(\mathbf{X}_n))$. The bias of the estimated log likelihood appearing in estimating the expected log likelihood is given by

$$\text{bias} = E_{\mathbf{X}_n}\left[\log f(\mathbf{X}_n, \hat{\theta}(\mathbf{X}_n)) - \int g(z_n) \log f(z_n, \hat{\theta}(\mathbf{X}_n)) dz_n\right].$$

If the maximum likelihood estimator $\hat{\theta}_n^{(ML)}$ can be used, one has that under some regularity conditions, bias $= \dim(\Theta) + o(1)$ as $n \to \infty$, where $\dim(\Theta)$ denotes the dimension of a parameter space Θ. The bias corrected log likelihood is given by $\log f(\mathbf{X}_n, \hat{\theta}_n^{(ML)}) - \dim(\Theta)$. Thus, Akaike (1973, 1974) proposed Akaike's information criterion (AIC) as follows:

$$\text{AIC}(\mathbf{X}_n, \hat{\theta}_n^{(ML)}) = -2 \log f(\mathbf{X}_n, \hat{\theta}_n^{(ML)}) + 2\dim(\Theta). \tag{3}$$

Consequently, we choose a statistical model which minimizes the value of AIC among a set of competing models. Note that AIC should be used under the situation where statistical model f contains the true model g, that is, f is a specified parametric model. For information criteria for misspecified models, we can refer Takeuchi (1976) and Konishi and Kitagawa (1996, 2003). For applications of model selection by information criteria, see, for example, Shibata (1976), Hall (1990), Burman and Nolan (1995), Burnham and Anderson (1998), Hurvich et al. (1998), Shimodaira (1998) and references therein.

As seen above, in order to construct AIC, it is enough to obtain the log likelihood function and the maximum likelihood estimator. For that reason, there seems no difficulty to derive AIC even if we consider diffusion processes. As a positive fact, Yoshida and Uchida (2001, 2004) obtained several types of information criteria including AIC for continuously observed diffusion processes. Unfortunately, as for AIC for discretely observed diffusion models, there are two serious problems. First, we can not explicitly obtain the log likelihood functions since the transition densities of diffusion processes do not generally have explicit forms. Because of the first difficulty, the maximum likelihood estimators can not be derived. Therefore, it is not a trivial problem to obtain AIC for diffusion models.

In order to obtain AIC type of information criteria for diffusion processes, we consider two kinds of functions. One is an approximate log likelihood function u_n based on a result of Dacunha-Castelle and Florens-Zmirou (1986). The other is a contrast function g_n based on a locally Gaussian approximation. The approximate log likelihood function u_n is used as an approximation of the log likelihood function and an asymptotically efficient estimator is derived from the contrast function g_n. The essential point is that in general we can not use the contrast function g_n as an approximation of the log likelihood function.

The rest of this paper is organized as follows. In section 2, using an approximate log likelihood function u_n and a contrast function g_n based on a locally Gaussian approximation, we propose AIC type of information criterion for discretely observed ergodic diffusion processes. In order to check that we can not generally use the contrast function g_n as an approximation of the log likelihood function, simulation studies for both the approximate log likelihood function u_n and the contrast function g_n are considered in section 3. In section 4, we study an example of model selection based on AIC including simulation results of the number of models selected by AIC. Section 5 gives conclusion of this paper and discussion on several possibilities of both an approximate log likelihood function and an asymptotically efficient estimator. Moreover, there are two directions of information criteria for discretely observed diffusion processes as future works. The results presented in section 2 are proved in section 6.

2 AIC type of information criterion

We introduce the notation used in this paper.
1. α_0, β_0 and θ_0 denote the true values of α, β and θ, respectively.
2. For a function $f(x,\theta)$, define that $\delta_{\theta_i} f(x,\theta) = \frac{\partial}{\partial\theta_i} f(x,\theta)$, $f'(x,\theta) = \frac{\partial}{\partial x} f(x,\theta)$, $\delta_\theta f(x,\theta) = (\delta_{\theta_i} f(x,\theta))_{i=1,\dots,p}$ and $\delta_\theta^2 f(x,\theta) = \left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x,\theta)\right)_{i,j=1,\dots,p+q}$.
3. E denotes the state space of X, $E \subseteq \mathbf{R}$.
4. When the distribution of X_t given $X_0 = x$ has a strictly positive density with respect to the Lebesgue measure on the state space E, we denote it by $y \longmapsto p(t,x,y,\theta)$, $y \in E$.
5. Let $\xrightarrow{p}$ be the convergence in probability and $\xrightarrow{d}$ be the convergence in distribution.

Moreover, we define the following functions.

$$\begin{aligned}
s(x,\beta) &= \int_0^x \frac{du}{\sigma(u,\beta)}, \\
B(x,\theta) &= \frac{b(x,\alpha)}{\sigma(x,\beta)} - \frac{1}{2}\sigma'(x,\beta), \\
\tilde{B}(x,\theta) &= B(s^{-1}(x,\beta),\theta), \\
\tilde{h}(x,\theta) &= \tilde{B}^2(x,\theta) + \tilde{B}'(x,\theta).
\end{aligned}$$

We make three assumptions as follows.

Assumption 1 *(i) Equation (1) has a unique strong solution on* $[0, T]$.
(ii) $\inf_{x,\beta} \sigma^2(x, \beta) > 0$.
(iii) The process X *is ergodic for every* θ *with invariant probability measure* μ_θ. *All polynomial moments of* μ_θ *are finite.*
(iv) For all $m \geq 0$ *and for all* θ, $\sup_t E_\theta[|X_t|^m] < \infty$.
(v) For every θ, *the functions* $b(x, \alpha)$ *and* $\sigma(x, \beta)$ *are twice continuously differentiable with respect to* x *and the derivatives are of polynomial growth in* x, *uniformly in* θ.
(vi) The functions $b(x, \alpha)$ *and* $\sigma(x, \beta)$ *and all their partial* x*-derivatives up to order 2 are three times differentiable with respect to* θ *for all* x *in the state space. All these derivatives with respect to* θ *are of polynomial growth in* x, *uniformly in* θ.

Assumption 2 *(i)* $\tilde{h}(x, \theta) = O(|x|^2)$ *as* $x \to \infty$.
(ii) $\inf_x \tilde{h}(x, \theta) > -\infty$ *for all* θ.
(iii) $\sup_\theta \sup_x |\tilde{h}^3(x, \theta)| \leq M < \infty$.
(iv) There exists $\gamma > 0$ *such that for every* θ *and* $j = 1, 2$, $|\tilde{B}^j(x, \theta)| = O(|\tilde{B}|^\gamma(x, \theta))$ *as* $|x| \to \infty$.

Assumption 3

$$\begin{aligned} b(x, \alpha) = b(x, \alpha_0) \quad &\text{for } \mu_{\theta_0} \text{ a.s. all } x \quad \Rightarrow \quad \alpha = \alpha_0, \\ \sigma(x, \beta) = \sigma(x, \beta_0) \quad &\text{for } \mu_{\theta_0} \text{ a.s. all } x \quad \Rightarrow \quad \beta = \beta_0. \end{aligned}$$

Remark 1 *Assumptions 1 and 3 are made in order to estimate an unknown parameter* θ. *For more details, see the conditions in Kessler (1997). It follows from assumptions 1 and 2 that we can obtain an approximate log likelihood function based on lemma 2 in Dacunha-Castelle and Florens-Zmirou (1986).*

The log likelihood function of $\mathbf{X}_n$ is

$$l_n(\mathbf{X}_n, \theta) = \sum_{k=1}^n l(h_n, X_{t_{k-1}^n}, X_{t_k^n}, \theta),$$

where $l(t, x, y, \theta) = \log p(t, x, y, \theta)$. Define the maximum likelihood estimator

$$\hat{\theta}_n^{(ML)} = \arg\sup_\theta l_n(\mathbf{X}_n, \theta).$$

Then, Akaike's information criterion is as follows:

$$AIC = -2l_n(\mathbf{X}_n, \hat{\theta}_n^{(ML)}) + 2dim(\Theta).$$

However, since the transition density p of the diffusion process X does not generally have an explicit form, we can not directly obtain the log likelihood function l_n and the maximum likelihood estimator $\hat{\theta}_n^{(ML)}$. That is why we need to obtain both an approximation of the log-likelihood function l_n and an asymptotically efficient estimator $\hat{\theta}_n$ in order to construct AIC type of information criteria for diffusion processes.

As an approximation of log likelihood function l_n, we use the following approximate log likelihood function based on lemma 2 in Dacunha-Castelle and Florens-Zmirou (1986).

$$u_n(\mathbf{X}_n, \theta) = \sum_{k=1}^n u(h_n, X_{t_{k-1}^n}, X_{t_k^n}, \theta),$$

where

$$u(t,x,y,\theta) \quad = \quad -\frac{1}{2}\log(2\pi t) - \log\sigma(y,\beta) - \frac{[S(x,y,\beta)]^2}{2t} + H(x,y,\theta) + t\tilde{g}(x,y,\theta). \qquad (4)$$

Here,

$$\begin{aligned}
S(x,y,\beta) &= \int_x^y \frac{du}{\sigma(u,\beta)}, \\
H(x,y,\theta) &= \int_x^y \left\{ \frac{b(u,\alpha)}{\sigma^2(u,\beta)} - \frac{1}{2}\frac{\sigma'(u,\beta)}{\sigma(u,\beta)} \right\} du, \\
\tilde{g}(x,y,\theta) &= -\frac{1}{2}\left\{ C(x,\theta) + C(y,\theta) + \frac{1}{3}B(x,\theta)B(y,\theta) \right\}, \\
C(x,\theta) &= \frac{1}{3}[B(x,\theta)]^2 + \frac{1}{2}[B(x,\theta)]'\sigma(x,\beta).
\end{aligned}$$

Next, in order to derive an asymptotically efficient estimator, we use the contrast function based on locally Gaussian approximation as follows:

$$g_n(\mathbf{X}_n,\theta) = \sum_{k=1}^{n} g(h_n, X_{t_{k-1}^n}, X_{t_k^n}, \theta),$$

where

$$g(t,x,y,\theta) = -\frac{1}{2}\log(2\pi t) - \log\sigma(x,\beta) - \frac{[y-x-tb(x,\alpha)]^2}{2t\sigma^2(x,\beta)}. \qquad (5)$$

We then define the maximum contrast estimator as

$$\hat{\theta}_n^{(C)} = \arg\sup_{\theta} g_n(\mathbf{X}_n,\theta).$$

For a process $\mathbf{Z}_n$ which is independent of (but has the same distribution as) the observed process $\mathbf{X}_n$,

$$\begin{aligned}
& u_n(\mathbf{X}_n,\hat{\theta}_n^{(C)}(\mathbf{X}_n)) - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n,\hat{\theta}_n^{(C)}(\mathbf{X}_n))] \\
= \; & u_n(\mathbf{X}_n,\hat{\theta}_n^{(C)}(\mathbf{X}_n)) - u_n(\mathbf{X}_n,\theta_0) && (6) \\
& + u_n(\mathbf{X}_n,\theta_0) - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n,\theta_0)] && (7) \\
& + E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n,\theta_0)] - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n,\hat{\theta}_n^{(C)}(\mathbf{X}_n))]. && (8)
\end{aligned}$$

Under the regularity conditions, one has

$$\begin{aligned}
(6) \;=\; & [D^{1/2}\delta_\theta u_n(\mathbf{X}_n,\theta_0)]^T D^{-1/2}(\hat{\theta}_n^{(C)}(\mathbf{X}_n) - \theta_0) \\
& + \frac{1}{2}[D^{-1/2}(\hat{\theta}_n^{(C)}(\mathbf{X}_n) - \theta_0)]^T [D^{1/2}\delta_\theta^2 u_n(\mathbf{X}_n,\theta_0) D^{1/2}] \\
& \times D^{-1/2}(\hat{\theta}_n^{(C)}(\mathbf{X}_n) - \theta_0) + o_p(1), \\
(8) \;=\; & -\left[D^{1/2}E_{\mathbf{Z}_n}[\delta_\theta l_n(\mathbf{Z}_n,\theta_0)]\right]^T D^{-1/2}(\hat{\theta}_n^{(C)}(\mathbf{X}_n) - \theta_0) \\
& - \frac{1}{2}[D^{-1/2}(\hat{\theta}_n^{(C)}(\mathbf{X}_n) - \theta_0)]^T D^{1/2} E_{\mathbf{Z}_n}[\delta_\theta^2 l_n(\mathbf{Z}_n,\theta_0)] D^{1/2} \\
& \times D^{-1/2}(\hat{\theta}_n^{(C)}(\mathbf{X}_n) - \theta_0) + o_p(1),
\end{aligned}$$

where A^T is the transpose of A for a vector A, D is the following $(p+q)\times(p+q)$ matrix

$$D = \begin{pmatrix} \frac{1}{nh_n} I_p & 0 \\ 0 & \frac{1}{n} I_q \end{pmatrix},$$

and I_p is the $p \times p$ identity matrix.

Let $I(\theta_0)$ denote the Fisher information matrix as follows:

$$I(\theta_0) = \begin{pmatrix} (I_b^{ij}(\theta_0))_{i,j=1,\dots,p} & 0 \\ 0 & (I_\sigma^{ij}(\theta_0))_{i,j=1,\dots,q} \end{pmatrix},$$

where

$$\begin{aligned} I_b^{ij}(\theta_0) &= \int_{\mathbf{R}} \frac{\delta_{\alpha_i} b(x,\alpha_0)\delta_{\alpha_j} b(x,\alpha_0)}{\sigma^2(x,\beta_0)} \mu_{\theta_0}(dx), \\ I_\sigma^{ij}(\theta_0) &= 2\int_{\mathbf{R}} \frac{\delta_{\beta_i} \sigma(x,\beta_0)\delta_{\beta_j} \sigma(x,\beta_0)}{\sigma^2(x,\beta_0)} \mu_{\theta_0}(dx). \end{aligned}$$

In order to obtain our main result, we need the following four lemmas.

Lemma 1 *Suppose that assumptions 1 and 2 hold true. Then, as* $nh_n^2 \to 0$,

$$E[u_n(\mathbf{X}_n,\theta_0) - l_n(\mathbf{X}_n,\theta_0)] = o(1).$$

Lemma 2 *Suppose that assumptions 1 and 2 hold true. Then, as* $nh_n^2 \to 0$,

$$D^{1/2}[\delta_\theta u_n(\mathbf{X}_n,\theta_0) - \delta_\theta g_n(\mathbf{X}_n,\theta_0)] = o_p(1).$$

Lemma 3 (Kessler (1997)) *Suppose that assumptions 1 and 3 hold true. Then, as* $nh_n^2 \to 0$,
(i) $D^{-1/2}(\hat{\theta}_n^{(C)} - \theta_0) = I^{-1}(\theta_0) D^{1/2}(\delta_\theta g_n)(\mathbf{X}_n,\theta_0) + o_p(1)$,
(ii) $D^{1/2}(\delta_\theta g_n)(\mathbf{X}_n,\theta_0) \xrightarrow{d} N(0, I(\theta_0))$.

Lemma 4 *Suppose that assumptions 1 and 2 hold true. Then, as* $nh_n^2 \to 0$,

$$D^{1/2}(\delta_\theta^2 u_n)(\mathbf{X}_n,\theta_0) D^{1/2} \xrightarrow{p} -I(\theta_0).$$

The main result is as follows.

Theorem 1 *Suppose that assumptions 1, 2 and 3 hold true. Then, as* $nh_n^2 \to 0$,

$$E_{\mathbf{X}_n}\left[u_n(\mathbf{X}_n,\hat{\theta}_n^{(C)}) - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n,\hat{\theta}_n^{(C)}(\mathbf{X}_n))]\right] = dim(\Theta) + o(1).$$

Remark 2 *(i) By theorem 1, AIC type of information criterion for diffusion processes is*

$$AIC = -2u_n(\mathbf{X}_n,\hat{\theta}_n^{(C)}) + 2dim(\Theta).$$

(ii) Instead of the maximum contrast estimator $\hat{\theta}_n^{(C)}$, *we can also use the approximate maximum likelihood estimator* $\hat{\theta}_n^{(AML)}$ *derived from the approximate log likelihood function* u_n. *Under assumptions 1-3,* $\hat{\theta}_n^{(AML)}$ *has the same properties as lemma 3, that is,* $\hat{\theta}_n^{(AML)}$ *is asymptotically efficient. Therefore, even if* $\hat{\theta}_n^{(AML)}$ *is used, we can make the same assertion as theorem 1 and*

$$AIC = -2u_n(\mathbf{X}_n,\hat{\theta}_n^{(AML)}) + 2dim(\Theta).$$

However, as seen in several examples put later, u_n *has a complicated expression while* g_n *is a simple form. Thus, as for an asymptotically efficient estimator, it is better to use* g_n *than* u_n.

3 Simulation studies on approximate log likelihoods

After getting theorem 1, we immediately have the following question. "Is it possible to use the contrast function g_n based on a locally Gaussian approximation as an approximation of the log likelihood function l_n?" The answer is negative. In order to understand this fact, we examine the asymptotic behaviours of $E_{\theta_0}[g_n(\mathbf{X}_n, \theta_0) - l_n(\mathbf{X}_n, \theta_0)]$ and $E_{\theta_0}[u_n(\mathbf{X}_n, \theta_0) - l_n(\mathbf{X}_n, \theta_0)]$ through simulations, which are done for each $T = nh_n = 10, 30, 50$ and $h_n = 1/100, 1/1000$. Three models we simulate are the Ornstein-Uhlenbeck process, the Radial Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross process, which have explicit transition densities. For details of the transition densities of the three models, see Karlin and Taylor (1981). For a true parameter value θ_0 and an initial value x_0, 5000 independent sample paths are generated by the Milstein scheme. For the Milstein scheme, see Kloeden and Platen (1992).

3.1 The Ornstein-Uhlenbeck process

Consider the Ornstein-Uhlenbeck process defined by the following stochastic differential equation

$$dX_t = -\alpha X_t dt + \beta dw_t, \quad X_0 = x_0,$$

where $\alpha > 0$ and $\beta > 0$ are unknown parameters.

It follows from (4) and (5) that

$$\begin{aligned} g(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \frac{1}{2}\log(\beta^2) - \frac{(y-x+t\alpha x)^2}{2t\beta^2}, \\ u(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \frac{1}{2}\log(\beta^2) - \frac{(y-x)^2}{2t\beta^2} - \frac{\alpha}{2\beta^2}(y^2-x^2) \\ &\quad -\frac{t}{2}\left\{\frac{\alpha^2}{3\beta^2}(x^2+y^2+xy) - \alpha\right\}. \end{aligned}$$

The log likelihood function l_n is obtained from

$$\begin{aligned} l(t,x,y,\theta) &= \log p(t,x,y,\theta), \\ p(t,x,y,\theta) &= \frac{1}{\sqrt{\pi\beta^2(1-\exp\{-2\alpha t\})/\alpha}} \exp\left[\frac{-(y-\exp\{-\alpha t\}x)^2}{\beta^2(1-\exp\{-2\alpha t\})/\alpha)}\right]. \end{aligned}$$

In tables 1 and 2, both u_n and g_n have good approximations of l_n for all cases. It is worth mentioning that u_n is better than g_n. For the the Ornstein-Uhlenbeck model, it seems that g_n can be substituted as an approximation of l_n instead of u_n.

Table 1: *The Ornstein-Uhlenbeck process. Means of $g_n - l_n$ and $u_n - l_n$ for 5000 independent simulated sample paths with $\alpha_0 = 1$, $\beta_0 = 2$ and $x_0 = 10$.*

T	h_n	$E[g_n - l_n]$	$E[u_n - l_n]$
10	1/100	-0.02436	0.00826
	1/1000	-0.00227	0.00083
30	1/100	-0.06979	0.02488
	1/1000	-0.00790	0.00249
50	1/100	-0.11241	0.04151
	1/1000	-0.01186	0.00416

Table 2: *The Ornstein-Uhlenbeck process. Means of $g_n - l_n$ and $u_n - l_n$ for* 5000 *independent simulated sample paths with $\alpha_0 = 2$, $\beta_0 = 5$ and $x_0 = 10$.*

T	h_n	$E[g_n - l_n]$	$E[u_n - l_n]$
10	1/100	-0.10091	0.03310
	1/1000	-0.00922	0.00333
30	1/100	-0.28569	0.09944
	1/1000	-0.02453	0.00999
50	1/100	-0.49459	0.16578
	1/1000	-0.04084	0.01665

3.2 The Radial Ornstein-Uhlenbeck process

We consider the Radial Ornstein-Uhlenbeck process defined by

$$dX_t = (\theta X_t^{-1} - X_t)dt + dw_t, \quad X_0 = x_0,$$

where $\theta > 0$ is an unknown parameter.

The contrast function g_n, the approximate log likelihood function u_n and the log likelihood function l_n are constructed by

$$\begin{aligned} g(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \frac{(y - x - t(\theta x^{-1} - x))^2}{2t}, \\ u(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \frac{(y-x)^2}{2t} + \theta \log\left(\frac{y}{x}\right) - \frac{1}{2}(y^2 - x^2) \\ &\quad -\frac{t}{2}\left\{\frac{1}{3}\left(\frac{\theta}{x} - x\right)^2 - \frac{1}{2}\left(\frac{\theta}{x^2} + 1\right) + \frac{1}{3}\left(\frac{\theta}{y} - y\right)^2 - \frac{1}{2}\left(\frac{\theta}{y^2} + 1\right)\right. \\ &\quad \left. + \frac{1}{3}\left(\frac{\theta}{x} - x\right)\left(\frac{\theta}{y} - y\right)\right\}, \\ l(t,x,y,\theta) &= \log p(t,x,y), \end{aligned}$$

respectively, where

$$p(t,x,y,\theta) = \frac{(y/x)^\theta \sqrt{xy}\exp\{-y^2 + (\theta + \frac{1}{2})t\}}{\sinh(t)} \exp\left[\frac{-(x^2+y^2)}{\exp\{2t\} - 1}\right] I_{\theta - \frac{1}{2}}\left(\frac{xy}{\sinh(t)}\right)$$

and I_ν is a modified Bessel function with index ν.

In tables 3 and 4, g_n has a small bias when $T = 30, 50$ and $h = 1/100$, while u_n has a good approximation for all cases. It follows from these tables that g_n is not suitable for an approximation of l_n when $T = 30, 50$ and $h = 1/100$.

Table 3: *The Radial Ornstein-Uhlenbeck process. Means of $g_n - l_n$ and $u_n - l_n$ for* 5000 *independent simulated sample paths with $\theta_0 = 2$ and $x_0 = 10$.*

T	h_n	$E[g_n - l_n]$	$E[u_n - l_n]$
10	1/100	-0.24315	0.05307
	1/1000	-0.10851	0.00544
30	1/100	-0.86191	0.18446
	1/1000	-0.37434	0.01892
50	1/100	-1.45814	0.31463
	1/1000	-0.64074	0.03222

Table 4: *The Radial Ornstein-Uhlenbeck process. Means of* $g_n - l_n$ *and* $u_n - l_n$ *for* 5000 *independent simulated sample paths with* $\theta_0 = 100$ *and* $x_0 = 10$.

T	h_n	$E[g_n - l_n]$	$E[u_n - l_n]$
10	1/100	-0.09708	0.03347
	1/1000	-0.01881	0.00334
30	1/100	-0.29819	0.10038
	1/1000	-0.05279	0.01002
50	1/100	-0.52674	0.16733
	1/1000	-0.09156	0.01671

3.3 The Cox-Ingersoll-Ross process

Consider the Cox-Ingersoll-Ross process defined by the following stochastic differential equation

$$dX_t = -\alpha_1(X_t - \alpha_2)dt + \beta\sqrt{X_t}dw_t, \quad X_0 = x_0,$$

where $\alpha_1 > 0$, $\alpha_2 > 0$ and $\beta > 0$ are unknown parameters.

The contrast function g_n and the approximate log likelihood function u_n are obtained from

$$g(t,x,y,\theta) = -\frac{1}{2}\log(2\pi t) - \frac{1}{2}\log(\beta^2 x) - \frac{(y - x + t\alpha_1(x-\alpha_2))^2}{2t\beta^2 x}, \tag{9}$$

$$\begin{aligned} u(t,x,y,\theta) = & -\frac{1}{2}\log(2\pi t) - \frac{1}{2}\log(\beta^2 y) - \frac{2(\sqrt{y}-\sqrt{x})^2}{t\beta^2} - \frac{\alpha_1(y-x)}{\beta^2} + \left(\frac{\alpha_1\alpha_2}{\beta^2} - \frac{1}{4}\right)\log\left(\frac{y}{x}\right) \\ & -\frac{t}{2}\left[\frac{1}{3}\left\{-\frac{\alpha_1}{\beta}\sqrt{x} + \left(\frac{\alpha_1\alpha_2}{\beta} - \frac{\beta}{4}\right)\frac{1}{\sqrt{x}}\right\}^2 + \frac{1}{2}\left\{-\frac{\alpha_1}{2} - \frac{1}{2}\left(\alpha_1\alpha_2 - \frac{\beta^2}{4}\right)\frac{1}{x}\right\}\right. \\ & +\frac{1}{3}\left\{-\frac{\alpha_1}{\beta}\sqrt{y} + \left(\frac{\alpha_1\alpha_2}{\beta} - \frac{\beta}{4}\right)\frac{1}{\sqrt{y}}\right\}^2 + \frac{1}{2}\left\{-\frac{\alpha_1}{2} - \frac{1}{2}\left(\alpha_1\alpha_2 - \frac{\beta^2}{4}\right)\frac{1}{y}\right\} \\ & \left. +\frac{1}{3}\left\{-\frac{\alpha_1}{\beta}\sqrt{x} + \left(\frac{\alpha_1\alpha_2}{\beta} - \frac{\beta}{4}\right)\frac{1}{\sqrt{x}}\right\}\left\{-\frac{\alpha_1}{\beta}\sqrt{y} + \left(\frac{\alpha_1\alpha_2}{\beta} - \frac{\beta}{4}\right)\frac{1}{\sqrt{y}}\right\}\right], \end{aligned} \tag{10}$$

respectively. The log likelihood function l_n is constructed by

$$\begin{aligned} l(t,x,y,\theta) &= \log p(t,x,y,\theta), \\ p(t,x,y,\theta) &= \frac{\gamma(y/x)^{\frac{1}{2}\nu}\exp\{\frac{1}{2}\alpha_1\nu t - \gamma y\}}{1 - \exp\{-\alpha_1 t\}}\exp\left[\frac{-\gamma(x+y)}{\exp\{\alpha_1 t\} - 1}\right] I_\nu\left(\frac{\gamma\sqrt{xy}}{\sinh(\frac{1}{2}\alpha_1 t)}\right), \end{aligned}$$

where $\gamma = 2\alpha_1\beta^{-2}$, $\nu = \gamma\alpha_2 - 1$ and I_ν is a modified Bessel function with index ν.

In tables 5 and 6, u_n is a good approximation for all cases, while g_n is a considerable bias for all cases. These results show that we can not use g_n instead of u_n.

Table 5: *The Cox-Ingersoll-Ross process. Means of* $g_n - l_n$ *and* $u_n - l_n$ *for* 5000 *independent simulated sample paths with* $\alpha_{1,0} = 1$, $\alpha_{2,0} = 5$, $\beta_0 = 2$ *and* $x_0 = 10$.

T	h_n	$E[g_n - l_n]$	$E[u_n - l_n]$
10	1/100	-9.23405	0.01440
	1/1000	-31.57794	0.00145
30	1/100	-28.95825	0.04840
	1/1000	-98.57695	0.00489
50	1/100	-48.62306	0.08193
	1/1000	-165.32443	0.00830

Table 6: *The Cox-Ingersoll-Ross process. Means of $g_n - l_n$ and $u_n - l_n$ for* 5000 *independent simulated sample paths with $\alpha_{1,0} = 1$, $\alpha_{2,0} = 10$, $\beta_0 = 2$ and $x_0 = 10$.*

T	h_n	$E[g_n - l_n]$	$E[u_n - l_n]$
10	1/100	-5.63288	0.01009
	1/1000	-20.92659	0.00100
30	1/100	-17.06086	0.03066
	1/1000	-63.01867	0.00305
50	1/100	-28.42232	0.05122
	1/1000	-105.04559	0.00510

4 Example of model selection based on AIC

As an example of model selection based on AIC, we treat the following setting. The true model is

$$dX_t = -(X_t - 10)dt + 2\sqrt{X_t}dw_t,$$

where $X_0 = 10$ and $t \in [0, T]$. We consider the following three statistical models:

$$\begin{aligned} dX_t &= -\alpha_1(X_t - \alpha_2)dt + \beta\sqrt{X_t}dw_t, && (11)\\ dX_t &= -\alpha_1(X_t - \alpha_2)dt + \sqrt{\beta_1 + \beta_2 X_t}dw_t, && (12)\\ dX_t &= -\alpha_1(X_t - \alpha_2)dt + (\beta_1 + \beta_2 X_t)^{\beta_3}dw_t, && (13)\end{aligned}$$

where $\alpha_1 > 0$, $\alpha_2 > 0$, $\beta > 0$, $\beta_1 \geq 0$, $\beta_2 > 0$ and $\beta_3 \geq 0$.

As seen in section 3, the contrast function $g^{(1)}(t, x, y, \theta)$ of the model (11) can be used (9). It follows from (5) that the contrast functions for the models (12) and (13) are

$$\begin{aligned} g^{(2)}(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \frac{1}{2}\log(\beta_1 + \beta_2 x) - \frac{(y - x + t\alpha_1(x - \alpha_2))^2}{2t(\beta_1 x + \beta_2)}, && (14)\\ g^{(3)}(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \beta_3\log(\beta_1 + \beta_2 x) - \frac{(y - x + t\alpha_1(x - \alpha_2))^2}{2t(\beta_1 x + \beta_2)^{2\beta_3}}, && (15)\end{aligned}$$

respectively. For the approximate log-likelihood function $u^{(1)}(t, x, y, \theta)$ of the model (11), we have already obtained (10) in section 3. Moreover, by (4), the approximate log-likelihood functions of the models (12) and (13) are

$$\begin{aligned} u^{(2)}(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \frac{1}{2}\log(\beta_1 + \beta_2 y) - \frac{2(\sqrt{\beta_1 + \beta_2 y} - \sqrt{\beta_1 + \beta_2 x})^2}{t\beta_2^2}\\ &\quad -\frac{\alpha_1(y - x)}{\beta_2} + \left(\frac{\alpha_1\alpha_2}{\beta_2} + \frac{\alpha_1\beta_1}{\beta_2^2} - \frac{1}{4}\right)\log\left(\frac{\beta_1 + \beta_2 y}{\beta_1 + \beta_2 x}\right)\\ &\quad -\frac{t}{2}\left[C(x,\theta) + C(y,\theta) + \frac{1}{3}B(x,\theta)B(y,\theta)\right], && (16)\end{aligned}$$

where

$$\begin{aligned} B(x,\theta) &= \frac{-\alpha_1 x + \alpha_1\alpha_2 - \beta_2/4}{\sqrt{\beta_1 + \beta_2 x}},\\ C(x,\theta) &= \frac{1}{3}\left\{\frac{-\alpha_1 x + \alpha_1\alpha_2 - \beta_2/4}{\sqrt{\beta_1 + \beta_2 x}}\right\}^2 + \frac{1}{2}\left\{\frac{-\alpha_1\beta_2 x/2 - \alpha_1\beta_1 - \alpha_1\alpha_2\beta_2/2 + \beta_2^2/8}{\beta_1 + \beta_2 x}\right\},\end{aligned}$$

and

$$
\begin{aligned}
u^{(3)}(t,x,y,\theta) &= -\frac{1}{2}\log(2\pi t) - \beta_3\log(\beta_1+\beta_2 y) - \frac{\left\{(\beta_1+\beta_2 y)^{1-\beta_3} - (\beta_1+\beta_2 x)^{1-\beta_3}\right\}^2}{2t(1-\beta_3)^2\beta_2^2} \\
&\quad -\alpha_1\left[\frac{(-\beta_1+\beta_2 y(1-2\beta_3) - 2\alpha_2\beta_2(1-\beta_3))}{2\beta_2^2(1-\beta_3)(1-2\beta_3)(\beta_1+\beta_2 y)^{2\beta_3-1}}\right. \\
&\quad \left. -\frac{(-\beta_1+\beta_2 x(1-2\beta_3) - 2\alpha_2\beta_2(1-\beta_3))}{2\beta_2^2(1-\beta_3)(1-2\beta_3)(\beta_1+\beta_2 x)^{2\beta_3-1}}\right] - \frac{\beta_3}{2}\log\left(\frac{\beta_1+\beta_2 y}{\beta_1+\beta_2 x}\right) \\
&\quad -\frac{t}{2}\left[C(x,\theta)+C(y,\theta)+\frac{1}{3}B(x,\theta)B(y,\theta)\right], \qquad (17)
\end{aligned}
$$

where

$$
\begin{aligned}
B(x,\theta) &= \frac{-\alpha_1(x-\alpha_2)}{(\beta_1+\beta_2 x)^{\beta_3}} - \frac{\beta_2\beta_3}{2(\beta_1+\beta_2 x)^{1-\beta_3}}, \\
C(x,\theta) &= \frac{1}{3}\left\{\frac{-\alpha_1(x-\alpha_2)}{(\beta_1+\beta_2 x)^{\beta_3}} - \frac{\beta_2\beta_3}{2(\beta_1+\beta_2 x)^{1-\beta_3}}\right\}^2 \\
&\quad +\frac{1}{2}\left\{-\alpha_1 + \frac{\alpha_1(x-\alpha_2)\beta_2\beta_3}{\beta_1+\beta_2 x} - \frac{\beta_2^2\beta_3(\beta_3-1)}{2(\beta_1+\beta_2 x)^{2-2\beta_3}}\right\},
\end{aligned}
$$

respectively. Note that $u^{(3)}(t,x,y,\theta)$ is obtained under the assumption $\beta_3 \neq 1/2$. When $\beta_3 = 1/2$, it suffices to consider the model (12).

Therefore, AIC for each model (11), (12), (13) is as follows.

$$
\begin{aligned}
AIC_1(\mathbf{X}_n, \hat{\theta}_n^{(1)}) &= -2u^{(1)}(\mathbf{X}_n, \hat{\theta}_n^{(1)}) + 2\times 3, \\
AIC_2(\mathbf{X}_n, \hat{\theta}_n^{(2)}) &= -2u^{(2)}(\mathbf{X}_n, \hat{\theta}_n^{(2)}) + 2\times 4, \\
AIC_3(\mathbf{X}_n, \hat{\theta}_n^{(3)}) &= -2u^{(3)}(\mathbf{X}_n, \hat{\theta}_n^{(3)}) + 2\times 5,
\end{aligned}
$$

where $\hat{\theta}_n^{(i)}$ is obtained from the contrast function $g_n^{(i)}$ for $i = 1,2,3$.

We examine the number of models selected by AIC among competing models (11), (12), (13) for 1000 independent sample paths generated by the Milstein scheme through simulations. The simulations are done for each $T = 10, 30, 50$ and $h_n = 1/100, 1/500$.

By table 7, we see that model 1 is selected with high frequency as the best model for all cases. However, we must note that model 2 is selected in a significant probability. This fact implies that AIC is not a tool for estimating the true model. Note that AIC is a tool to choose the best model among competing models from the aspect of both model-fitting and prediction.

Table 7: *The number of models selected by AIC for* 1000 *independent simulated sample paths.*

T	h_n	model 1	model 2	model 3
10	1/100	761	185	54
	1/500	744	194	62
30	1/100	803	185	12
	1/500	673	238	89
50	1/100	769	186	45
	1/500	610	275	115

5 Conclusion and discussion

In order to get AIC type of information criterion, we first use the contrast function g_n based on a locally Gaussian approximation to obtain an efficient estimator. Next, for an approximation of log likelihood l_n, it is better to use the approximate log likelihood u_n based on Dacunha-Castelle and Florens-Zmirou (1986) than the locally Gaussian approximation g_n. As for the CIR model, we should not use g_n as an approximation of l_n because g_n has a considerable bias.

In this paper, we proposed AIC type of information criterion under assumptions 1-3. In order to obtain the information criterion, however, the most important point is to show the lemmas 1-4 presented in section 2. For this reason, it is possible to consider AIC type of information criterion by replacing assumptions 1-3 with the following assumption.

Assumption 4 *There exist functions u_n and g_n satisfying lemmas 1-4.*

Under assumption 4 and appropriate regularity conditions, it is possible to make the same assertion as theorem 1. Therefore, there is a possibility of having both other approximate log likelihood functions and other asymptotically efficient estimators. For approximate log likelihood functions, the essential point is to satisfy lemma 1. Nicolau (2002) considered an approximate likelihood function by means of a simulation-based technique based on a results of Dacunha-Castelle and Florens-Zmirou (1986). Aït-Sahalia (2002) presented an explicit sequence of closed-form transition densities by using Hermite expansion. It seems that under appropriate regularity conditions, their approximate log likelihood functions satisfy lemma 1. Moreover, although the contrast function g_n generally has a considerable bias for an approximation of l_n, the bias corrected contrast function $\tilde{g}_n$ may be available to an approximation of l_n. For asymptotically efficient estimators satisfying lemma 3, there are a number of works for various diffusion models, see Florens-Zmirou (1989), Yoshida (1992), Bibby and Sørensen (1995), Aït-Sahalia and Mykland (2003, 2004), Bibby et al. (2004) and reference therein.

With regard to future projects, there are three directions. One is to obtain AIC by replacing the assumption that $nh_n^2 \to 0$ with a weaker asymptotics, for example $nh_n^3 \to 0$. Another is to extend the results of this paper to a multi-dimensional diffusion model. The third is to consider information criteria for a misspecified diffusion model, that is, Takeuchi's information criterion (TIC) presented in Takeuchi (1976) and the generalized information criterion (GIC) considered in Konishi and Kitagawa (1996). For the first objective, it seems that there is no difficulty under the situation when $nh_n^3 \to 0$. Using the contrast function $l_{3,n}$ presented in Kessler (1997) and the third order approximate log likelihood function $u_{3,n}$ based on a results of Dacunha-Castelle and Florens-Zmirou (1986), we will be able to show that $l_{3,n}$ and $u_{3,n}$ satisfy assumption 4 under $nh_n^3 \to 0$. However, there is no doubt that $l_{3,n}$ and $u_{3,n}$ have very complicated expressions. For a multi-dimensional diffusion model, an asymptotically efficient estimator can be obtained from a multi-dimensional version of the contrast function based on a locally Gaussian approximation in the same way as the case of a one-dimensional diffusion model. Therefore, it is important to consider an approximate log likelihood function satisfying lemma 1 for a multi-dimensional diffusion model. One possibility is to use the approximate log likelihood function based on a result of Aït-Sahalia (2003). As for information criteria for misspecified diffusion models, we will need to compose a statistically asymptotic theory of parametric estimation for misspecified diffusion models from discrete observations. Using the misspecified version of estimators together with non-parametric estimators of both drift and diffusion coefficient functions for discretely observed diffusion models, we will be able to obtain TIC and GIC for misspecified diffusion models. However, it seems that there are a lot of difficulties in order to prove the desired result.

6 Proof

Let R denote a function $(0,1]\times\mathbf{R}$ for which there exists a $C>0$ such that $|R(h_n,x)|\le h_n(1+|x|)^C$ for all n,x.

Proof of Lemma 1.

$$E[u_n(\mathbf{X}_n,\theta_0)-l_n(\mathbf{X}_n,\theta_0)]=\sum_{k=1}^n E\left[E[u(X_{t_{k-1}^n},X_{t_k^n},\theta_0)-l(X_{t_{k-1}^n},X_{t_k^n},\theta_0)|X_{t_{k-1}^n}]\right].$$

It follows from (3.7) and lemma 2 in Dacunha-Castelle and Florens-Zmirou (1986) that

$$E[u(X_{t_{k-1}^n},X_{t_k^n},\theta_0)-l(X_{t_{k-1}^n},X_{t_k^n},\theta_0)|X_{t_{k-1}^n}]=R(h_n^2,X_{t_{k-1}^n}).$$

Thus, as $n\to\infty$, $h_n\to 0$ and $nh_n^2\to 0$,

$$\begin{aligned}E[u_n(\mathbf{X}_n,\theta_0)-l_n(\mathbf{X}_n,\theta_0)] &= \sum_{k=1}^n h_n^2 E[R(1,X_{t_{k-1}^n})]\\ &\le nh_n^2\frac{1}{n}\sum_{k=1}^n C\longrightarrow 0.\end{aligned}$$

This completes the proof.

Proof of Lemma 2. Set that for $i=1,\dots,p+q$,

$$\xi_k^i \;:=\; d_i[\delta_{\theta_i}u(h_n,X_{t_{k-1}^n},X_{t_k^n},\theta_0)-\delta_{\theta_i}g_n(h_n,X_{t_{k-1}^n},X_{t_k^n},\theta_0)],$$

where $d_i=1/\sqrt{nh_n}$ for $i=1,\dots,p$ and $d_i=1/\sqrt{n}$ for $i=p+1,\dots,p+q$. By lemma 9 in Genon-Catalot and Jacod (1993), it is enough to prove that for $i,j=1,\dots,p+q$,

$$\begin{aligned}\sum_{k=1}^n E[\xi_k^i|X_{t_{k-1}^n}] &\xrightarrow{p} 0,\\ \sum_{k=1}^n E[(\xi_k^i)^2|X_{t_{k-1}^n}] &\xrightarrow{p} 0\end{aligned}$$

as $n\to\infty$, $h_n\to 0$ and $nh_n^2\to 0$.

For $i=1,\dots,p$, one has that

$$\begin{aligned}\xi_k^i &= \frac{1}{\sqrt{nh_n}}\Big[\delta_{\alpha_i}H(X_{t_{k-1}^n},X_{t_k^n},\theta_0)+h_n\delta_{\alpha_i}\tilde g(X_{t_{k-1}^n},X_{t_k^n},\theta_0)\\ &\quad -(\delta_{\alpha_i}b)(X_{t_{k-1}^n},\alpha_0)\frac{X_{t_k^n}-X_{t_{k-1}^n}-h_nb(X_{t_{k-1}^n},\alpha_0)}{\sigma^2(X_{t_{k-1}^n},\beta_0)}\Big].\end{aligned}$$

By the Ito-Taylor expansion based on the generator $L_{\theta_0}=b(x,\alpha_0)(\partial/\partial x)+\frac{1}{2}\sigma(x,\beta_0)(\partial/\partial x)^2$,

$$\begin{aligned}E[\xi_k^i|X_{t_{k-1}^n}] &= \frac{1}{\sqrt{nh_n}}\Big[\delta_{\alpha_i}H(X_{t_{k-1}^n},X_{t_{k-1}^n},\theta_0)+h_nL_{\theta_0}\delta_{\alpha_i}H(X_{t_{k-1}^n},X_{t_{k-1}^n},\theta_0)\\ &\quad +h_n\delta_{\alpha_i}\tilde g(X_{t_{k-1}^n},X_{t_{k-1}^n},\theta_0)+R(h_n^2,X_{t_{k-1}^n})\\ &\quad -\frac{(\delta_{\alpha_i}b)(X_{t_{k-1}^n},\alpha_0)}{\sigma^2(X_{t_{k-1}^n},\beta_0)}E[X_{t_k^n}-X_{t_{k-1}^n}-h_nb(X_{t_{k-1}^n},\alpha_0)|X_{t_{k-1}^n}]\Big].\end{aligned}$$

For details of the Ito-Taylor expansion based on the generator L_{θ_0}, see Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989) and Kessler (1997). Therefore,

$$\begin{aligned}\sum_{k=1}^{n} E[\xi_k^i | X_{t_{k-1}^n}] &= \frac{1}{\sqrt{nh_n}} \sum_{k=1}^{n} R(h_n^2, X_{t_{k-1}^n}) \\ &= \sqrt{nh_n^3}\frac{1}{n}\sum_{k=1}^{n} R(1, X_{t_{k-1}^n}) \xrightarrow{p} 0.\end{aligned}$$

Next, we obtain that for $i = 1, \ldots, p$,

$$\begin{aligned}\sum_{k=1}^{n} E[(\xi_k^i)^2 | X_{t_{k-1}^n}] &= \frac{1}{nh_n} \sum_{k=1}^{n} R(h_n^2, X_{t_{k-1}^n}) \\ &= h_n\frac{1}{n}\sum_{k=1}^{n} R(1, X_{t_{k-1}^n}) \xrightarrow{p} 0.\end{aligned}$$

In the same way as the proof for $i = 1, \ldots, p$, we can obtain the results for $i = p+1, \ldots, p+q$. This completes the proof.

Proof of Lemma 3. See the proof of theorem 1 in Kessler (1997).

Proof of Lemma 4. Set that for $i, j = 1, \ldots, p+q$,

$$\eta_k^{ij} \;:=\; d_i d_j (\delta_{\theta_i}\delta_{\theta_j} u_n)(h_n, X_{t_{k-1}^n}, X_{t_k^n}, \theta_0),$$

where d_i is defined in the proof of lemma 2. In the same way as in lemma 2, we can show that

$$\begin{aligned}\sum_{k=1}^{n} E[\eta_k^{ij} | X_{t_{k-1}^n}] &\xrightarrow{p} -I^{ij}(\theta_0), \\ \sum_{k=1}^{n} E[(\eta_k^{ij})^2 | X_{t_{k-1}^n}] &\xrightarrow{p} 0\end{aligned}$$

as $n \to \infty$, $h_n \to 0$ and $nh_n^2 \to 0$. This completes the proof.

Proof of Theorem 1.

$$\begin{aligned}& E_{\mathbf{X}_n}\left[u_n(\mathbf{X}_n, \hat{\theta}_n^{(C)}(\mathbf{X}_n)) - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n, \hat{\theta}_n^{(C)}(\mathbf{X}_n))]\right] \\ =\; & E_{\mathbf{X}_n}\left[u_n(\mathbf{X}_n, \hat{\theta}_n^{(C)}(\mathbf{X}_n)) - u_n(\mathbf{X}_n, \theta_0)\right] \qquad (18) \\ & + E_{\mathbf{X}_n}\left[u_n(\mathbf{X}_n, \theta_0) - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n, \theta_0)]\right] \qquad (19) \\ & + E_{\mathbf{X}_n}\left[E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n, \theta_0)] - E_{\mathbf{Z}_n}[l_n(\mathbf{Z}_n, \hat{\theta}_n^{(C)}(\mathbf{X}_n))]\right]. \qquad (20)\end{aligned}$$

Lemma 1 implies that (19) $= o(1)$. By (6) and lemmas 2-4, one has that

$$(18) \;=\; \frac{1}{2}\mathrm{tr}[I(\theta_0)I^{-1}(\theta_0)] + o(1).$$

It follows from (8) and lemmas 2-4 that

$$(20) \;=\; \frac{1}{2}\mathrm{tr}[I(\theta_0)I^{-1}(\theta_0)] + o(1).$$

This completes the proof.

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science under Grants-in-Aid for Scientific Researches, and by Cooperative Research Program of the Institute of Statistical Mathematics. The first author was also supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan under Kyushu University 21st Century COE Program, Development of Dynamic Mathematics with High Functionality.

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach, *Econometrica* **70**, 223-262.

Aït-Sahalia, Y. (2003). Closed-form likelihood expansions for multivariate diffusions, Tech. rep., Princeton University. http://www.princeton.edu/ yacine/multivarmle.pdf

Aït-Sahalia, Y. and Mykland, P. A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* **71**, 483-549.

Aït-Sahalia, Y. and Mykland, P. A. (2004). Estimators of diffusions with discrete observations: a general theory. *Ann. Statist.* **32**, 2186-2222.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium in Information Theory*, Petrov, B.N. & Csaki, F. eds., Akademiai Kaido, Budapest, 267-281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Auto. Control* **AC-19**, 716-723.

Bibby, B. M. and Sørensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.

Bibby, B. M., Jacobsen, M. and Sørensen, M. (2004) Estimating functions for discretely sampled diffusion-type models. Preprint No. 2004-4, Department of Applied Mathematics and Statistics, University of Copenhagen. (to appear in Aït-Sahalia, Y. and Hansen, L. P. (eds.): Handbook of Financial Econometrics, North-Holland, Amsterdam).

Burman, P. and Nolan, D. (1995). A general Akaike-type criterion for model selection in robust regression. *Biometrika* **82**, 877–886.

Burnham, K. P. and Anderson, D. R. (1998). *Model selection and Inference.* Springer-Verlag, New York.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19**, 263–284.

Florens-Zmirou, D. (1989). Approximate discrete time schemes for statistics of diffusion processes. *Statistics* **20**, 547–557.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Statist.* **29**, 119–151.

Hurvich, C. M., Simonoff, J. S. and Tsai, C. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**, 271–293.

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes.* Academic Press, New York.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations.* Springer-Verlag, New York.

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

Konishi, S. and Kitagawa, G. (2003). Asymptotic theory for information criteria in model selection —functional approach. C. R. Rao 80th birthday felicitation volume, Part IV. *J. Statist. Plann. Inference* **114**, 45–61.

Kutoyants, Yu. A. (2004). *Statistical inference for ergodic diffusion processes.* Springer-Verlag, London.

Nicolau, J. (2002). A new technique for simulating the likelihood of stochastic differential equations. *Econom. J.* **5**, 91–103.

Hall, P. (1990). Akaike's information criterion and Kullback-Leibler loss for histogram density estimation. *Probab. Theory Related Fields* **85**, 449–467.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**, 117–126.

Shimodaira, H. (1998). An application of multiple comparison techniques to model selection. *Ann. Inst. Stat. Math.* **50**, 1-13.

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Sciences* **153**, 12-18. (in Japanese).

Uchida, M. and Yoshida, N. (2001). Information criteria in model selection for mixing processes. *Statist. Infer. Stochast. Process.* **4**, 73–98.

Uchida, M. and Yoshida, N. (2004). Information criteria for small diffusions via the theory of Malliavin-Watanabe. *Statist. Infer. Stochast. Process.* **7**, 35–67.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41**, 220–242.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

MHF2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

MHF2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

MHF2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

MHF2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

MHF2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

MHF2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

MHF2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

MHF2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model

MHF2003-10 Atsushi YOSHIKAWA
Maple and wave-front tracking — an experiment

MHF2003-11 Masanobu KANEKO
On the local factor of the zeta function of quadratic orders

MHF2003-12 Hidefumi KAWASAKI
Conjugate-set game for a nonlinear programming problem

MHF2004-1 Koji YONEMOTO & Takashi YANAGAWA
Estimating the Lyapunov exponent from chaotic time series with dynamic noise

MHF2004-2 Rui YAMAGUCHI, Eiko TSUCHIYA & Tomoyuki HIGUCHI
State space modeling approach to decompose daily sales of a restaurant into time-dependent multi-factors

MHF2004-3 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Cubic pencils and Painlevé Hamiltonians

MHF2004-4 Atsushi KAWAGUCHI, Koji YONEMOTO & Takashi YANAGAWA
Estimating the correlation dimension from a chaotic system with dynamic noise

MHF2004-5 Atsushi KAWAGUCHI, Kentarou KITAMURA, Koji YONEMOTO, Takashi YANAGAWA & Kiyofumi YUMOTO
Detection of auroral breakups using the correlation dimension

MHF2004-6 Ryo IKOTA, Masayasu MIMURA & Tatsuyuki NAKAKI
A methodology for numerical simulations to a singular limit

MHF2004-7 Ryo IKOTA & Eiji YANAGIDA
Stability of stationary interfaces of binary-tree type

MHF2004-8 Yuko ARAKI, Sadanori KONISHI & Seiya IMOTO
Functional discriminant analysis for gene expression data via radial basis expansion

MHF2004-9 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Hypergeometric solutions to the q - Painlevé equations

MHF2004-10 Raimundas VIDŪNAS
Expressions for values of the gamma function

MHF2004-11 Raimundas VIDŪNAS
Transformations of Gauss hypergeometric functions

MHF2004-12 Koji NAKAGAWA & Masakazu SUZUKI
Mathematical knowledge browser

MHF2004-13 Ken-ichi MARUNO, Wen-Xiu MA & Masayuki OIKAWA
Generalized Casorati determinant and Positon-Negaton-Type solutions of the Toda lattice equation

MHF2004-14 Nalini JOSHI, Kenji KAJIWARA & Marta MAZZOCCO
Generating function associated with the determinant formula for the solutions of the Painlevé II equation

MHF2004-15 Kouji HASHIMOTO, Ryohei ABE, Mitsuhiro T. NAKAO & Yoshitaka WATANABE
Numerical verification methods of solutions for nonlinear singularly perturbed problem

MHF2004-16 Ken-ichi MARUNO & Gino BIONDINI
Resonance and web structure in discrete soliton systems: the two-dimensional Toda lattice and its fully discrete and ultra-discrete versions

MHF2004-17 Ryuei NISHII & Shinto EGUCHI
Supervised image classification in Markov random field models with Jeffreys divergence

MHF2004-18 Kouji HASHIMOTO, Kenta KOBAYASHI & Mitsuhiro T. NAKAO
Numerical verification methods of solutions for the free boundary problem

MHF2004-19 Hiroki MASUDA
Ergodicity and exponential β-mixing bounds for a strong solution of Lévy-driven stochastic differential equations

MHF2004-20 Setsuo TANIGUCHI
The Brownian sheet and the reflectionless potentials

MHF2004-21 Ryuei NISHII & Shinto EGUCHI
Supervised image classification based on AdaBoost with contextual weak classifiers

MHF2004-22 Hideki KOSAKI
On intersections of domains of unbounded positive operators

MHF2004-23 Masahisa TABATA & Shoichi FUJIMA
Robustness of a characteristic finite element scheme of second order in time increment

MHF2004-24 Ken-ichi MARUNO, Adrian ANKIEWICZ & Nail AKHMEDIEV
Dissipative solitons of the discrete complex cubic-quintic Ginzburg-Landau equation

MHF2004-25 Raimundas VIDŪNAS
Degenerate Gauss hypergeometric functions

MHF2004-26 Ryo IKOTA
The boundedness of propagation speeds of disturbances for reaction-diffusion systems

MHF2004-27 Ryusuke KON
Convex dominates concave: an exclusion principle in discrete-time Kolmogorov systems

MHF2004-28 Ryusuke KON
Multiple attractors in host-parasitoid interactions: coexistence and extinction

MHF2004-29 Kentaro IHARA, Masanobu KANEKO & Don ZAGIER
Derivation and double shuffle relations for multiple zeta values

MHF2004-30 Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Generalized partitioned quantum cellular automata and quantization of classical CA

MHF2005-1 Hideki KOSAKI
Matrix trace inequalities related to uncertainty principle

MHF2005-2 Masahisa TABATA
Discrepancy between theory and real computation on the stability of some finite element schemes

MHF2005-3 Yuko ARAKI & Sadanori KONISHI
Functional regression modeling via regularized basis expansions and model selection

MHF2005-4 Yuko ARAKI & Sadanori KONISHI
Functional discriminant analysis via regularized basis expansions

MHF2005-5 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Point configurations, Cremona transformations and the elliptic difference Painlevé equations

MHF2005-6 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Construction of hypergeometric solutions to the q - Painlevé equations

MHF2005-7 Hiroki MASUDA
Simple estimators for non-linear Markovian trend from sampled data:
I. ergodic cases

MHF2005-8 Hiroki MASUDA & Nakahiro YOSHIDA
Edgeworth expansion for a class of Ornstein-Uhlenbeck-based models

MHF2005-9 Masayuki UCHIDA
Approximate martingale estimating functions under small perturbations of dynamical systems

MHF2005-10 Ryo MATSUZAKI & Masayuki UCHIDA
One-step estimators for diffusion processes with small dispersion parameters from discrete observations

MHF2005-11 Junichi MATSUKUBO, Ryo MATSUZAKI & Masayuki UCHIDA
Estimation for a discretely observed small diffusion process with a linear drift

MHF2005-12 Masayuki UCHIDA & Nakahiro YOSHIDA
AIC for ergodic diffusion processes from discrete observations