HYBRID ESTIMATION FOR AN ERGODIC DIFFUSION PROCESS BASED ON REDUCED DATA

Kaino, Yusuke
Graduate School of Engineering Science, Osaka University

Uchida, Masayuki
Graduate School of Engineering Science, Osaka University

Yoshida, Yuto
Science Information Systems Co., Ltd.

https://doi.org/10.5109/2232332

出版情報：Bulletin of informatics and cybernetics. 49, pp.89-118, 2017-12. Research Association of Statistical Sciences

HYBRID ESTIMATION FOR AN ERGODIC DIFFUSION PROCESS BASED ON REDUCED DATA

by

Yusuke Kaino, **Masayuki Uchida** and **Yuto Yoshida**

Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.49

FUKUOKA, JAPAN
2017

Bulletin of Informatics and Cybernetics, Vol. 49, 2017

HYBRID ESTIMATION FOR AN ERGODIC DIFFUSION PROCESS BASED ON REDUCED DATA

By

Yusuke Kaino*, **Masayuki Uchida**† and **Yuto Yoshida**‡

Abstract

We consider efficient estimation of both drift and diffusion coefficient parameters for an ergodic diffusion process from discrete observations. From the viewpoint of numerical analysis, hybrid estimators based on the initial Bayes type estimators from the reduced data are proposed and the asymptotic properties of the hybrid estimators, including convergence of moments, are shown. Furthermore, we give examples and simulation results in order to investigate the asymptotic performance of the proposed estimators.

Key Words and Phrases: Adaptive maximum likelihood type estimator, Bayes type estimator, convergence of moments, high frequency data, stochastic differential equation.

1. Introduction

We treat a d-dimensional ergodic diffusion process defined by the following stochastic differential equation

$$dX_t \quad = \quad b(X_t, \beta)dt + a(X_t, \alpha)dw_t, \quad t \geq 0, \quad X_0 = x_0, \tag{1}$$

where $\theta = (\alpha, \beta)$ is an unknown parameter, $\theta \in \Theta_\alpha \times \Theta_\beta = \Theta$, Θ_α and Θ_β are compact convex subsets of $\mathbf{R}^{m_1}$ and $\mathbf{R}^{m_2}$, respectively. $b : \mathbf{R}^d \times \Theta_\beta \to \mathbf{R}^d$ and $a : \mathbf{R}^d \times \Theta_\alpha \to \mathbf{R}^d \otimes \mathbf{R}^r$ are known functions except for parameters α and β. Furthermore, w is an r-dimensional standard Wiener process, x_0 is a deterministic initial condition. Let the true value of θ be $\theta^* = (\alpha^*, \beta^*)$ and we assume that $\theta^* \in \mathrm{Int}(\Theta)$ and the parameter spaces have locally Lipschitz boundaries, see Adams and Fournier (2003). The data are discrete observations $\mathbf{X}_n = (X_{t_i^n})_{0 \leq i \leq n}$, where $t_i^n = ih_n$. Let p be an integer and $p \geq 2$. It is assumed that $h_n \to 0$, $nh_n \to \infty$ and $nh_n^p \to 0$ as $n \to \infty$.

The statistical inference for ergodic diffusion processes has been investigated by many researchers. For statistically asymptotic theory for continuous path data, we can refer the textbooks of Kutoyants (1984, 2004). For parametric estimation based on discrete observations, see Prakasa Rao (1983, 1988), Florens-Zmirou (1989), Yoshida (1992), Bibby and Sørensen (1995), Kessler (1995, 1997), Gobet (2002), Uchida and Yoshida

* Graduate School of Engineering Science, Osaka University.
† Graduate School of Engineering Science, Osaka University, and Center for Mathematical Modeling and Data Science (MMDS), Osaka University Toyonaka, Osaka 560-8531, Japan.
‡ Science Information Systems Co., Ltd., 2-6 Kinkouchou, Kanagawa-ku Yokohama, Kanagawa 221-0056, Japan.

(2001, 2011), Uchida (2010), Fujii and Uchida (2014), Kamatani and Uchida (2015), Eguchi and Masuda (2016) and references therein. Yoshida (2011) proved the polynomial type large deviation inequality for a statistical random field and he showed the estimator has asymptotic normality and convergence of moments of both the maximum likelihood type estimator and the Bayes type estimator for discretely observed diffusion processes, see also Uchida and Yoshida (2012, 2014).

In order to explain our motivation for this paper, we consider the one-dimensional diffusion process defined by

$$dX_t = (\beta_1 - \beta_2 X_t - 2\sin(\beta_3 X_t))\,dt + \left(\frac{\alpha_2 + X_t^2}{1+\alpha_1 X_t^2}\right) dw_t, \quad t \geq 0, \quad X_0 = 2, \tag{2}$$

where $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2, \beta_3)$ are unknown parameters, and the true parameter value is $(\alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*, \beta_3^*) = (0.3, 0.5, 3, 7, 5)$. The parameter space is assumed to be $\Theta = [0.1, 50]^5$. The simulations were done for $T_n = nh_n = 250$, $h_n = 1/390$, which means that $n = 390 \times 250 = 97500$. We set $p = 4$ since $nh_n^4 \simeq 0$. Let $\Delta X_i = X_{t_i^n} - X_{t_{i-1}^n}$, $b_{i-1}(\beta) = \beta_1 - \beta_2 X_{t_{i-1}} - 2\sin(\beta_3 X_{t_{i-1}})$ and $A_{i-1}(\alpha) = \left(\frac{\alpha_2 + X_{t_{i-1}}^2}{1+\alpha_1 X_{t_{i-1}}^2}\right)^2$. For the case that $nh_n^4 \to 0$, the quasi-log likelihood functions of Kessler (1995) and Uchida and Yoshida (2012) are as follows.

$$\begin{aligned}
U_n^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^n \left\{ \frac{(\Delta X_i)^2}{h_n A_{i-1}(\alpha)} + \log(A_{i-1}(\alpha)) \right\}, \\
U_n^{(2)}(\beta \mid \bar{\alpha}) &= -\frac{1}{2}\sum_{i=1}^n \left\{ \frac{(\Delta X_i - h_n b_{i-1}(\beta))^2}{h_n A_{i-1}(\bar{\alpha})} \right\}, \\
U_n^{(3)}(\alpha \mid \bar{\theta}) &= -\frac{1}{2}\sum_{i=1}^n \left\{ \frac{(\Delta X_i)^2 - h_n^2 \bar{D}_{i-1}^{(2)}(\bar{\theta})}{h_n A_{i-1}(\alpha)} + \log A_{i-1}(\alpha) \right\}, \\
U_n^{(4)}(\beta \mid \bar{\theta}) &= -\frac{1}{2}\sum_{i=1}^n \frac{\left(\Delta X_i - h_n b_{i-1}(\beta) - h_n^2 \bar{r}_{i-1}^{(2)}(\bar{\theta})\right)^2}{h_n A_{i-1}(\bar{\alpha})}.
\end{aligned}$$

For the definition of $\bar{D}_{i-1}^{(2)}(\bar{\theta})$ and $\bar{r}_{i-1}^{(2)}(\bar{\theta})$, see Section 3 below. The adaptive maximum likelihood (ML) type estimator $(\hat{\alpha}_n^{(3)}, \hat{\beta}_n^{(4)})$ is given by

$$\begin{aligned}
\hat{\alpha}_n^{(1)} &= \arg\sup_{\alpha \in \Theta_\alpha} U_n^{(1)}(\alpha), \\
\hat{\beta}_n^{(2)} &= \arg\sup_{\beta \in \Theta_\beta} U_n^{(2)}(\beta \mid \hat{\alpha}_n^{(1)}), \\
\hat{\alpha}_n^{(3)} &= \arg\sup_{\alpha \in \Theta_\alpha} U_n^{(3)}(\alpha \mid \hat{\alpha}_n^{(1)}, \hat{\beta}_n^{(2)}), \\
\hat{\beta}_n^{(4)} &= \arg\sup_{\beta \in \Theta_\beta} U_n^{(4)}(\beta \mid \hat{\alpha}_n^{(3)}, \hat{\beta}_n^{(2)}).
\end{aligned}$$

It follows from Kessler (1995) and Uchida and Yoshida (2012) that under some regularity conditions, the adaptive ML type estimator $(\hat{\alpha}_n^{(3)}, \hat{\beta}_n^{(4)})$ has asymptotic normality and convergence of moments under $nh_n^4 \to 0$. In order to compute the ML type estimator, we used **optim()** with the "L-BFGS-B" method in the R Language. For the true model, 1000 independent sample paths are generated by the Euler-Maruyama scheme, and the mean and the standard deviation (s.d.) for the estimators are computed. Table 1 is the simulation results of the adaptive ML type estimator $(\hat{\alpha}_n^{(3)}, \hat{\beta}_n^{(4)})$ with the initial value being the true value, where the upper row is the mean of the estimator, the lower row is the s.d. of the estimator and the time means the computation time of the estimator based on a one sample path. Table 2 is the simulation results of the adaptive ML type estimator $(\hat{\alpha}_n^{(3)}, \hat{\beta}_n^{(4)})$ with the initial value being the uniform random number on Θ. As we see from Tables 1 and 2, it is quite important to choose a suitable initial value for optimization.

Table 1: adaptive ML type estimator with the initial value being the true value

$\hat{\beta}_1(3)$	$\hat{\beta}_2(7)$	$\hat{\beta}_3(5)$	$\hat{\alpha}_1(0.3)$	$\hat{\alpha}_2(0.5)$	time(sec.)
3.006	7.036	5.005	0.301	0.500	
(0.093)	(0.366)	(0.192)	(0.021)	(0.001)	20

Table 2: adaptive ML type estimator with the initial value being the uniform random number on Θ

$\hat{\beta}_1(3)$	$\hat{\beta}_2(7)$	$\hat{\beta}_3(5)$	$\hat{\alpha}_1(0.3)$	$\hat{\alpha}_2(0.5)$	time(sec.)
2.470	8.127	23.316	0.305	0.498	
(0.508)	(1.121)	(17.988)	(0.021)	(0.002)	30

Next, we consider the Bayes type estimators for α and β. We assume that the prior densities $\pi_1(\alpha)$ and $\pi_2(\beta)$ are continuous and satisfy that $0 < \inf_{\alpha\in\Theta_\alpha} \pi_1(\alpha) \le \sup_{\alpha\in\Theta_\alpha} \pi_1(\alpha) < \infty$ and $0 < \inf_{\beta\in\Theta_\beta} \pi_2(\beta) \le \sup_{\beta\in\Theta_\beta} \pi_2(\beta) < \infty$. In the same way as Uchida and Yoshida (2014), the adaptive Bayes type estimator $(\tilde{\alpha}_n^{(1)}, \tilde{\beta}_n^{(2)})$ is defined as

$$
\begin{aligned}
\tilde{\alpha}_n^{(1)} &= \frac{\int_{\Theta_\alpha} \alpha \exp\left\{\frac{1}{n^{1-\frac{2}{4}}} U_n^{(1)}(\alpha)\right\} \pi_1(\alpha) d\alpha}{\int_{\Theta_\alpha} \exp\left\{\frac{1}{n^{1-\frac{2}{4}}} U_n^{(1)}(\alpha)\right\} \pi_1(\alpha) d\alpha}, \\
\tilde{\beta}_n^{(2)} &= \frac{\int_{\Theta_\beta} \beta \exp\left\{\frac{1}{(nh_n)^{1-\frac{2}{3}}} U_n^{(2)}(\beta \mid \tilde{\alpha}_n^{(1)})\right\} \pi_2(\beta) d\beta}{\int_{\Theta_\beta} \exp\left\{\frac{1}{(nh_n)^{1-\frac{2}{3}}} U_n^{(2)}(\beta \mid \tilde{\alpha}_n^{(1)})\right\} \pi_2(\beta) d\beta}.
\end{aligned}
$$

The hybrid type estimator $(\check{\alpha}_n^{(3)}, \check{\beta}_n^{(4)})$ is given by

$$
\begin{aligned}
\check{\alpha}_n^{(3)} &= \arg\sup_{\alpha\in\Theta_\alpha} U_n^{(3)}(\alpha \mid \tilde{\alpha}_n^{(1)}, \tilde{\beta}_n^{(2)}), \\
\check{\beta}_n^{(4)} &= \arg\sup_{\beta\in\Theta_\beta} U_n^{(4)}(\beta \mid \check{\alpha}_n^{(3)}, \tilde{\beta}_n^{(2)}).
\end{aligned}
$$

It follows from Uchida and Yoshida (2012, 2014) that under some regularity conditions, the hybrid type estimator $(\check{\alpha}_n^{(3)}, \check{\beta}_n^{(4)})$ has asymptotic normality and convergence of moments under $nh_n^4 \to 0$, see also Kamatani and Uchida (2015). Table 3 is the simulation results of initial Bayes type estimator $(\tilde{\alpha}_n^{(1)}, \tilde{\beta}_n^{(2)})$ based on the full data with $n = 97500$. Table 4 is the simulation results of hybrid estimator $(\check{\alpha}_n^{(3)}, \check{\beta}_n^{(4)})$ with the initial value being the Bayes estimator based on the full data. The initial Bayes estimator of α_1 has a bias in Table 3, but the hybrid estimator in Table 4 has as good behavior as the adaptive estimator with the initial value being the true value in Table 1. The Bayes estimators are calculated with one of the MCMC methods, the mixed preconditioned Crank-Nicolson (MpCN) method proposed by Kamatani (2014) for 10^6 Markov chains and 10^5 burn-in iterations. The calculation of the Bayes estimator is essentially free from the choice of the initial value. However, it takes much time to compute the Bayes estimator with a large sample size n. Recently, Kutoyants (2017) proposed the multi-step ML type estimator with the initial estimator for a continuously observed ergodic diffusion process on $[0, T]$. Using the initial estimator obtained from the reduced continuous path data on $[0, T_0]$ for $T_0 \leq T$, he proved asymptotic efficiency of the multi-step ML type estimator as $T_0 \to \infty$.

Table 3: initial Bayes type estimator based on the full data ($n = 97500$)

$\tilde{\beta}_1(3)$	$\tilde{\beta}_2(7)$	$\tilde{\beta}_3(5)$	$\tilde{\alpha}_1(0.3)$	$\tilde{\alpha}_2(0.5)$	time(h.)
2.949	6.853	4.916	0.506	0.505	
(0.120)	(0.451)	(0.255)	(0.049)	(0.005)	3.9

Table 4: hybrid estimator with the initial value being the Bayes estimator based on the full data ($n = 97500$)

$\check{\beta}_1(3)$	$\check{\beta}_2(7)$	$\check{\beta}_3(5)$	$\check{\alpha}_1(0.3)$	$\check{\alpha}_2(0.5)$	time(sec.)
3.007	7.053	4.998	0.301	0.500	
(0.099)	(0.398)	(0.204)	(0.021)	(0.001)	30

In this paper, from the viewpoint of numerical analysis, we propose the initial Bayes type estimator based on reduced data with the sample size $n_0 \leq n$, where n is the sample size of full data. Although the estimator does not have optimal rate of convergence, the computation time of the Bayes estimator based on reduced data is much shorter than that of the Bayes estimator based on the full data with the sample size n. Furthermore, by using both the multi-step estimator in Kamatani and Uchida (2015) and the adaptive ML type estimator in Uchida and Yoshida (2012), it can be shown that under some regularity conditions, the hybrid estimator has asymptotic normality and convergence of moments. It is worth mentioning that the proposed hybrid estimator is free from the choice of the initial value for optimization of the quasi-log likelihood function since we use the Bayes type estimator as an initial value. Moreover, from the viewpoint of computational statistics, the proposed initial Bayes estimators are obtained by an MCMC method and the hybrid estimators with the initial Bayes estimators have good behavior in numerical simulations.

This paper is organized as follows. In Section 2, four kinds of the initial Bayes type estimators based on reduced data are proposed and the asymptotic properties of the estimators are stated. In Section 3, multi-step estimators with the initial Bayes type estimator based on reduced data are described. Furthermore, four kinds of hybrid type estimators are studied and their asymptotic properties, including convergence of moments, are shown. Section 4 presents numerical examples and simulation studies. We see from the simulation results that the hybrid estimator with the initial Bayes estimator is best among the competing estimators. Section 5 gives concluding remarks of this work. Section 6 is devoted to the proofs of the results presented in Sections 2 and 3.

2. Initial Bayes estimator

Let $\mathcal{F}_{\uparrow}(\mathbf{R}^d)$ be the space of all measurable functions f satisfying that $f(x)$ is an $\mathbf{R}$-valued function on $\mathbf{R}^d$ with polynomial growth in x. Let $C_{\uparrow}^{k,l}(\mathbf{R}^d \times \Theta; \mathbf{R}^d)$ denote the space of all functions f satisfying the following conditions:

(i) $f(x,\theta)$ is an $\mathbf{R}^d$-valued function on $\mathbf{R}^d \times \Theta$,

(ii) $f(x,\theta)$ is continuously differentiable with respect to x up to order k for all θ, and their derivatives up to order k are of polynomial growth in x uniformly in θ,

(iii) for $|\mathbf{n}| = 0, 1, \ldots, k$, $\partial^{\mathbf{n}} f(x,\theta)$ is continuously differentiable with respect to θ up to order l for all x. Moreover, for $|\nu| = 1, \ldots, l$ and $|\mathbf{n}| = 0, 1, \ldots, k$, $\delta^{\nu}\partial^{\mathbf{n}} f(x,\theta)$ is of polynomial growth in x uniformly in θ. Here $\mathbf{n} = (n_1, \ldots, n_d)$ and $\nu = (\nu_1, \ldots, \nu_m)$ are multi-indices, $m = \dim(\Theta)$, $|\mathbf{n}| = n_1 + \ldots + n_d$, $|\nu| = \nu_1 + \ldots + \nu_m$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x_i$, and $\delta^{\nu} = \delta_{\theta_1}^{\nu_1} \cdots \delta_{\theta_m}^{\nu_m}$, $\delta_{\theta_i} = \partial/\partial\theta_i$.

P_θ denotes the law of the process defined by the equation (1). Set $A(x,\alpha) = aa^\star(x,\alpha)$, where $\star$ denotes the transpose. Let L_θ be the infinitesimal generator of the diffusion (1): $L_\theta = \sum_{i=1}^d b_i(x,\beta)\partial_i + \frac{1}{2}\sum_{i,j=1}^d A_{ij}(x,\alpha)\partial_i\partial_j$. Set $\Delta X_i = X_{t_i^n} - X_{t_{i-1}^n}$, $A_{i-1}(\alpha) = A(X_{t_{i-1}^n}, \alpha)$ and $b_{i-1}(\beta) = b(X_{t_{i-1}^n}, \beta)$. Let $\xrightarrow{p}$ and $\xrightarrow{d}$ be the convergence in probability and the convergence in distribution, respectively. For matrices A and B of the same size, we define $A^{\otimes 2} = AA^\star$ and $B[A] = tr(BA^\star)$. Moreover, for a matrix A, $||A|| = tr(AA^*)^{1/2}$.

We make the following assumptions.

$[A1]$ (i) There exists $K > 0$ such that for all $x, y \in \mathbf{R}^d$,

$$\sup_{\beta\in\Theta_\beta} |b(x,\beta) - b(y,\beta)| + \sup_{\alpha\in\Theta_\alpha} ||a(x,\alpha) - a(y,\alpha)|| \le K|x-y|.$$

(ii) $\inf_{x,\alpha} \det(A(x,\alpha)) > 0$.
(iii) There exists a unique invariant probability measure μ_{θ^*} of X_t and for any $f \in \mathcal{F}_{\uparrow}(\mathbf{R}^d)$ satisfying $\int_{\mathbf{R}^d} |f(x)|\mu_{\theta^*}(dx) < \infty$, as $T \to \infty$,

$$\frac{1}{T}\int_0^T f(X_t)dt \xrightarrow{p} \int_{\mathbf{R}^d} f(x)\mu_{\theta^*}(dx).$$

(iv) $\sup_t E[|X_t|^M] < \infty$ for all $M > 0$.
(v) For any $g \in \mathcal{F}_{\uparrow}(\mathbf{R}^d)$ satisfying $\int_{\mathbf{R}^d} g(x)\mu_{\theta^*}(dx) = 0$, there exist $G(x)$, $\partial_{x_i} G(x) \in$

$\mathcal{F}_{\uparrow}(\mathbf{R}^d)$ $(i = 1, \ldots, d)$ such that for all x,

$$L_{\theta^*} G(x) = -g(x).$$

[*A*2](k, l) $b \in C_{\uparrow}^{k,4}(\mathbf{R}^d \times \Theta_\beta; \mathbf{R}^d)$. $a \in C_{\uparrow}^{l,4}(\mathbf{R}^d \times \Theta_\alpha; \mathbf{R}^d \otimes \mathbf{R}^r)$.

Remark 2.1. *For a sufficient condition of* [*A*1]-(v)*, see Pardoux and Veretennikov (2001), and Uchida and Yoshida (2012).*

Let $p \geq 2$. We assume that there exists $\gamma \in \left(\frac{1}{p}, 1\right)$ such that $h_n = O\left(n^{-\gamma}\right)$. Set $G \in (\gamma, 1]$ and $n_0 = [n^G]$. Let $\mathbf{Y}_{n_0} = (X_{t_i^n})_{0 \leq i \leq n_0}$ with $t_i^n = ih_n$ denote the reduced data with the sample size n_0. Moreover, we assume that there exists $\epsilon_0 \in \left(0, 1 - \frac{\gamma}{G}\right)$ such that $n_0^{\epsilon_0} \leq n_0 h_n$ for large n. Thus, we will consider the situation when $h_n \to 0$, $n_0 h_n \to \infty$ and $nh_n^p \to 0$ as $n \to \infty$, which implies that $nh_n \geq n_0 h_n \to \infty$ and $n_0 h_n^p \leq nh_n^p \to 0$ as $n \to \infty$.

Proposition 2.1. *Let $p \geq 2$, $\epsilon_1 = \epsilon_0/(2(p-1))$ and $f \in C_{\uparrow}^{1,1}(\mathbf{R}^d \times \Theta)$. Assume* [*A*1]. *Then, for all $M > 0$, as $nh_n^p \to 0$,*

$$\sup_{n \in \mathbf{N}} E_{\theta^*}\left[\sup_{\theta \in \Theta}\left(n_0^{\epsilon_1}\left|\frac{1}{n_0}\sum_{i=1}^{n_0} f(X_{t_{i-1}^n}, \theta) - \int_{\mathbf{R}^d} f(x, \theta)\mu_{\theta^*}(dx)\right|\right)^M\right] < \infty.$$

We consider four kinds of initial Bayes type estimators for α and β. Let

$$\begin{aligned}
V_{n_0}^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^{n_0}\left\{h_n^{-1}A_{i-1}^{-1}(\alpha)[(\Delta X_i)^{\otimes 2}] + \log\det(A_{i-1}(\alpha))\right\}, \\
V_{n_0}^{(2)}(\beta \mid \alpha) &= -\frac{1}{2}\sum_{i=1}^{n_0}\left\{h_n^{-1}A_{i-1}^{-1}(\alpha)[(\Delta X_i - h_n b_{i-1}(\beta))^{\otimes 2}]\right\}, \\
W_{n_0}^{(1)}(\alpha) &= -\frac{1}{2h_n^2}\sum_{i=1}^{n_0}\left|\left|(\Delta X_i)^{\otimes 2} - h_n A_{i-1}(\alpha)\right|\right|^2, \\
W_{n_0}^{(2)}(\beta) &= -\frac{1}{2h_n}\sum_{i=1}^{n_0}\left|\Delta X_i - h_n b_{i-1}(\beta)\right|^2.
\end{aligned}$$

The four kinds of quasi-log likelihood functions for α and β are as follows.

$$\begin{aligned}
(U_{1,n_0}^{(1)}(\alpha), U_{1,n_0}^{(2)}(\beta \mid \alpha)) &= (V_{n_0}^{(1)}(\alpha), V_{n_0}^{(2)}(\beta \mid \alpha)), \\
(U_{2,n_0}^{(1)}(\alpha), U_{2,n_0}^{(2)}(\beta \mid \alpha)) &= (W_{n_0}^{(1)}(\alpha), V_{n_0}^{(2)}(\beta \mid \alpha)), \\
(U_{3,n_0}^{(1)}(\alpha), U_{3,n_0}^{(2)}(\beta)) &= (V_{n_0}^{(1)}(\alpha), W_{n_0}^{(2)}(\beta)), \\
(U_{4,n_0}^{(1)}(\alpha), U_{4,n_0}^{(2)}(\beta)) &= (W_{n_0}^{(1)}(\alpha), W_{n_0}^{(2)}(\beta)).
\end{aligned}$$

Let $q = \max\{p, 2/G\}$. For $j = 1, 2$, the type j Bayes estimator $(\tilde{\alpha}_{j,n_0}^{(1)}, \tilde{\beta}_{j,n_0}^{(2)})$ is defined as

$$\tilde{\alpha}_{j,n_0}^{(1)} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{ \frac{1}{n_0^{1-\frac{2}{qG}}} U_{j,n_0}^{(1)}(\alpha) \right\} \pi_1(\alpha) d\alpha}{\int_{\Theta_\alpha} \exp\left\{ \frac{1}{n_0^{1-\frac{2}{qG}}} U_{j,n_0}^{(1)}(\alpha) \right\} \pi_1(\alpha) d\alpha},$$

$$\tilde{\beta}_{j,n_0}^{(2)} = \frac{\int_{\Theta_\beta} \beta \exp\left\{ \frac{1}{(n_0 h_n)^{1-\frac{2}{qG}}} U_{j,n_0}^{(2)}(\beta \mid \tilde{\alpha}_{j,n_0}^{(1)}) \right\} \pi_2(\beta) d\beta}{\int_{\Theta_\beta} \exp\left\{ \frac{1}{(n_0 h_n)^{1-\frac{2}{qG}}} U_{j,n_0}^{(2)}(\beta \mid \tilde{\alpha}_{j,n_0}^{(1)}) \right\} \pi_2(\beta) d\beta}.$$

For $j = 3, 4$, the type j Bayes estimator $(\tilde{\alpha}_{j,n_0}^{(1)}, \tilde{\beta}_{j,n_0}^{(2)})$ is given by

$$\tilde{\alpha}_{j,n_0}^{(1)} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{ \frac{1}{n_0^{1-\frac{2}{qG}}} U_{j,n_0}^{(1)}(\alpha) \right\} \pi_1(\alpha) d\alpha}{\int_{\Theta_\alpha} \exp\left\{ \frac{1}{n_0^{1-\frac{2}{qG}}} U_{j,n_0}^{(1)}(\alpha) \right\} \pi_1(\alpha) d\alpha},$$

$$\tilde{\beta}_{j,n_0}^{(2)} = \frac{\int_{\Theta_\beta} \beta \exp\left\{ \frac{1}{(n_0 h_n)^{1-\frac{2}{qG}}} U_{j,n_0}^{(2)}(\beta) \right\} \pi_2(\beta) d\beta}{\int_{\Theta_\beta} \exp\left\{ \frac{1}{(n_0 h_n)^{1-\frac{2}{qG}}} U_{j,n_0}^{(2)}(\beta) \right\} \pi_2(\beta) d\beta}.$$

The calculations of the above Bayes estimators are expected to be robust thanks to their normalizing terms $n_0^{1-\frac{2}{qG}}$ and $(n_0 h_n)^{1-\frac{2}{qG}}$. These normalizing terms are called temperatures, see for example, p.163 of Robert and Casella (2004). For the performance of Bayes estimator with temperature for diffusion type processes, we can refer Kamatani and Uchida (2015), Kamatani et al. (2016) and Nomura and Uchida (2016).

Let

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &= -\frac{1}{2} \int_{\mathbf{R}^d} \left\{ tr\left[A(x,\alpha)^{-1} A(x,\alpha^*) - I_d \right] + \log \frac{\det(A(x,\alpha))}{\det(A(x,\alpha^*))} \right\} \mu_{\theta^*}(dx), \\
\mathbb{Y}^{(2)}(\beta) &= -\frac{1}{2} \int_{\mathbf{R}^d} A(x,\alpha^*)^{-1} [(b(x,\beta) - b(x,\beta^*))^{\otimes 2}] \mu_{\theta^*}(dx), \\
\mathbb{W}^{(1)}(\alpha) &= -\frac{1}{2} \int_{\mathbf{R}^d} ||A(x,\alpha) - A(x,\alpha^*)||^2 \mu_{\theta^*}(dx), \\
\mathbb{W}^{(2)}(\beta) &= -\frac{1}{2} \int_{\mathbf{R}^d} |b(x,\beta) - b(x,\beta^*)|^2 \mu_{\theta^*}(dx).
\end{aligned}$$

Set

$$\begin{aligned}
(\mathbb{U}_1^{(1)}(\alpha), \mathbb{U}_1^{(2)}(\beta)) &= (\mathbb{Y}^{(1)}(\alpha), \mathbb{Y}^{(2)}(\beta)), \\
(\mathbb{U}_2^{(1)}(\alpha), \mathbb{U}_2^{(2)}(\beta)) &= (\mathbb{W}^{(1)}(\alpha), \mathbb{Y}^{(2)}(\beta)), \\
(\mathbb{U}_3^{(1)}(\alpha), \mathbb{U}_3^{(2)}(\beta)) &= (\mathbb{Y}^{(1)}(\alpha), \mathbb{W}^{(2)}(\beta)), \\
(\mathbb{U}_4^{(1)}(\alpha), \mathbb{U}_4^{(2)}(\beta)) &= (\mathbb{W}^{(1)}(\alpha), \mathbb{W}^{(2)}(\beta)).
\end{aligned}$$

We make the following assumption. Let $j = 1, 2, 3, 4$.

[*A*3]-(j)

(i) There exists a positive constant χ_j such that $\mathbb{U}_j^{(1)}(\alpha) \leq -\chi_j |\alpha - \alpha^*|^2$ for all $\alpha \in \Theta_\alpha$.

(ii) There exists a positive constant $\tilde{\chi}_j$ such that $\mathbb{U}_j^{(2)}(\beta) \leq -\tilde{\chi}_j |\beta - \beta^*|^2$ for all $\beta \in \Theta_\beta$.

Theorem 2.2. *Let* $p \geq 2$, $\gamma \in \left(\frac{1}{p}, 1\right)$, $G \in (\gamma, 1]$, $n_0 = [n^G]$, $q = \max\{p, 2/G\}$ *and* $j = 1, 2, 3, 4$. *Assume* [*A*1], [*A*2](2, 2) *and* [*A*3]-(j). *Then, for all* $M > 0$, *as* $nh_n^p \to 0$,

$$\sup_{n \in \mathbf{N}} E_{\theta^*}\left[\left| n_0^{\frac{1}{qG}} (\tilde{\alpha}_{j,n_0}^{(1)} - \alpha^*)\right|^M\right] < \infty,$$

$$\sup_{n \in \mathbf{N}} E_{\theta^*}\left[\left| (n_0 h_n)^{\frac{1}{qG}} (\tilde{\beta}_{j,n_0}^{(2)} - \beta^*)\right|^M\right] < \infty.$$

Remark 2.2. *Theorem 2.2 yields that for all* $M > 0$, *as* $nh_n^p \to 0$,

$$\sup_{n \in \mathbf{N}} E_{\theta^*}\left[\left| n^{\frac{1}{q}} (\tilde{\alpha}_{j,n_0}^{(1)} - \alpha^*)\right|^M\right] < \infty,$$

$$\sup_{n \in \mathbf{N}} E_{\theta^*}\left[\left| (nh_n)^{\frac{\epsilon_0}{q}} (\tilde{\beta}_{j,n_0}^{(2)} - \beta^*)\right|^M\right] < \infty.$$

Here we note that $h_n \to 0$ *and* $nh_n^p \to 0$ *as* $n \to \infty$, *and there exists* $\epsilon_0 \in \left(0, 1 - \frac{\gamma}{G}\right)$ *such that* $n_0^{\epsilon_0} \leq n_0 h_n$ *for large* n.

3. Hybrid estimator

Let

$$\begin{aligned}
J_n(\alpha) &:= \left\{\frac{1}{n}\partial_\alpha^2 V_n^{(1)}(\alpha) \text{ is invertible}\right\}, \\
\Gamma_n(\alpha) &:= \frac{1}{n}\partial_\alpha^2 V_n^{(1)}(\alpha) 1_{J_n(\alpha)} + E_{m_1} 1_{J_n^c(\alpha)}, \\
K_n(\beta \mid \alpha) &:= \left\{\frac{1}{nh_n}\partial_\beta^2 V_n^{(2)}(\beta \mid \alpha) \text{ is invertible}\right\}, \\
\Xi_n(\beta \mid \alpha) &:= \frac{1}{nh_n}\partial_\beta^2 V_n^{(2)}(\beta \mid \alpha) 1_{K_n(\beta \mid \alpha)} + E_{m_2} 1_{K_n^c(\beta \mid \alpha)},
\end{aligned}$$

where E_m is the $m \times m$ identity matrix, and $1_K(\omega) = 1$ if $\omega \in K$ and $1_K(\omega) = 0$ if $\omega \in K^c$.

Let $j = 1, 2, 3, 4$. Set $(\check{\alpha}_{j,n}^{(0)}, \check{\beta}_{j,n}^{(0)}) = (\tilde{\alpha}_{j,n_0}^{(1)}, \tilde{\beta}_{j,n_0}^{(2)})$ in Theorem 2.2. Let $q_1 = \max\{p - 1, 2\}$. Let $k_1 \geq \log_2(q/p)$ and $k_2 \geq \log_2(q/(\epsilon_0 q_1))$. The multi-step estimators $\check{\alpha}_{j,n}^{(k_1)}$ and $\check{\beta}_{j,n}^{(k_2)}$ of Kamatani and Uchida (2015) are defined as for $k = 1, \ldots, k_1$,

$$\check{\alpha}_{j,n}^{(k)} = \check{\alpha}_{j,n}^{(k-1)} - \Gamma_n^{-1}(\check{\alpha}_{j,n}^{(k-1)}) \frac{1}{n} \partial_\alpha V_n^{(1)}(\check{\alpha}_{j,n}^{(k-1)})$$

and for $k = 1, \dots, k_2$,

$$\check{\beta}_{j,n}^{(k)} = \check{\beta}_{j,n}^{(k-1)} - \Xi_n^{-1}(\check{\beta}_{j,n}^{(k-1)} \mid \check{\alpha}_{j,n}^{(k_1)}) \frac{1}{nh_n} \partial_\beta V_n^{(2)}(\check{\beta}_{j,n}^{(k-1)} \mid \check{\alpha}_{j,n}^{(k_1)}),$$

We assume the regularity conditions in Theorem 2.2. In an analogous way to the proofs of Kamatani and Uchida (2015), we have that for all $M > 0$, as $nh_n^p \to 0$,

$$\sup_{n \in \mathbf{N}} E_{\theta^*} \left[\left| n^{\frac{1}{p}} (\check{\alpha}_{j,n}^{(k_1)} - \alpha^*) \right|^M \right] < \infty, \tag{3}$$

$$\sup_{n \in \mathbf{N}} E_{\theta^*} \left[\left| (nh_n)^{\frac{1}{q_1}} (\check{\beta}_{j,n}^{(k_2)} - \beta^*) \right|^M \right] < \infty. \tag{4}$$

Remark 3.1. *(i) Let $(\hat{\alpha}_{n,ML}^{(1)}, \hat{\beta}_{n,ML}^{(2)})$ be the ML type estimator defined as*

$$V_n^{(1)}(\hat{\alpha}_{n,ML}^{(1)}) = \sup_\alpha V_n^{(1)}(\alpha),$$
$$V_n^{(2)}(\hat{\beta}_{n,ML}^{(2)} \mid \hat{\alpha}_{n,ML}^{(1)}) = \sup_\beta V_n^{(2)}(\beta \mid \hat{\alpha}_{n,ML}^{(1)}).$$

Let $j = 1, 2, 3, 4$. Assume the conditions in Theorem 2.2. Then, it follows from Kamatani and Uchida (2015) that as $nh_n^p \to 0$, for all $M > 0$,

$$\sup_{n \in \mathbf{N}} E_{\theta^*} \left[\left| (n^{1/p}(\hat{\alpha}_{n,ML}^{(1)} - \check{\alpha}_{j,n}^{(k_1)}), (nh_n)^{1/q_1}(\hat{\beta}_{n,ML}^{(2)} - \check{\beta}_{j,n}^{(k_2)})) \right|^M \right] < \infty.$$

(ii) We set $p = 4$, $\gamma = \frac{1}{3}$, $h_n = \frac{1}{n^{1/3}}$, $G = \frac{1}{2}$, $\epsilon_0 = \frac{1}{4}$. Then, one has that $k_1 = 0$, $k_2 = 3$ and $q_1 = 3$, and it follows from (3) and (4) that for $j = 1, 2, 3, 4$, as $nh_n^4 \to 0$,

$$\sup_{n \in \mathbf{N}} E_{\theta^*} \left[\left| n^{\frac{1}{4}} (\check{\alpha}_{j,n}^{(0)} - \alpha^*) \right|^M \right] < \infty,$$
$$\sup_{n \in \mathbf{N}} E_{\theta^*} \left[\left| (nh_n)^{\frac{1}{3}} (\check{\beta}_{j,n}^{(3)} - \beta^*) \right|^M \right] < \infty$$

for all $M > 0$.

In a similar way to Kessler (1995) and Uchida and Yoshida (2012), we use the following quasi-log likelihood functions. Let $\bar{\theta} = (\bar{\alpha}, \bar{\beta})$, $k_0 = [p/2]$ and for $k = 1, \dots, k_0$,

$$V_n^{(2k+1)}(\alpha \mid \bar{\theta}) = -\frac{1}{2} \sum_{i=1}^n \left\{ h_n^{-1} A_{i-1}^{-1}(\alpha) \left[(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - \sum_{j=2}^{k+1} h_n^j \bar{D}_{i-1}^{(j)}(\bar{\theta}) \right] + \log\det A_{i-1}(\alpha) \right\},$$

$$V_n^{(2k+2)}(\beta \mid \bar{\theta}) = -\frac{1}{2} \sum_{i=1}^n h_n^{-1} A_{i-1}^{-1}(\bar{\alpha}) \left[\left(X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta) - \sum_{j=2}^{k+1} h_n^j \bar{r}_{i-1}^{(j)}(\bar{\theta}) \right)^{\otimes 2} \right],$$

where for $l, m = 1, \dots, d$, $f_l(x) = x_l$, $h_{lm}(x) = (x - X_{t_{i-1}^n})_l (x - X_{t_{i-1}^n})_m$,

$$\bar{D}_{i-1}^{(j)}(\bar{\theta})_{lm} = \frac{1}{j!} L_{\bar{\theta}}^j h_{lm}(X_{t_{i-1}^n}), \quad \bar{r}_{i-1}^{(j)}(\bar{\theta})_l = \frac{1}{j!} L_{\bar{\theta}}^j f_l(X_{t_{i-1}^n}).$$

When $p=2$, the hybrid estimators $\hat{\alpha}_{j,n}^{(1)}$ and $\hat{\beta}_{j,n}^{(2)}$ are defined as

$$\begin{aligned}\hat{\alpha}_{j,n}^{(1)} &= \check{\alpha}_{j,n}^{(k_1)} - \Gamma_n^{-1}(\check{\alpha}_{j,n}^{(k_1)})\frac{1}{n}\partial_\alpha V_n^{(1)}(\check{\alpha}_{j,n}^{(k_1)}),\\ \hat{\beta}_{j,n}^{(2)} &= \check{\beta}_{j,n}^{(k_2)} - \Xi_n^{-1}(\check{\beta}_{j,n}^{(k_2)} \mid \hat{\alpha}_{j,n}^{(1)})\frac{1}{nh_n}\partial_\beta V_n^{(2)}(\check{\beta}_{j,n}^{(k_2)} \mid \hat{\alpha}_{j,n}^{(1)}).\end{aligned}$$

When $p=3$, the hybrid estimators $\hat{\alpha}_{j,n}^{(3)}$ and $\hat{\beta}_{j,n}^{(2)}$ are defined as

$$\begin{aligned}\hat{\beta}_{j,n}^{(2)} &= \check{\beta}_{j,n}^{(k_2)} - \Xi_n^{-1}(\check{\beta}_{j,n}^{(k_2)} \mid \check{\alpha}_{j,n}^{(k_1)})\frac{1}{nh_n}\partial_\beta V_n^{(2)}(\check{\beta}_{j,n}^{(k_2)} \mid \check{\alpha}_{j,n}^{(k_1)}),\\ \hat{\alpha}_{j,n}^{(3)} &= \arg\sup_{\alpha\in\Theta_\alpha} V_n^{(3)}(\alpha \mid \check{\alpha}_{j,n}^{(k_1)}, \hat{\beta}_{j,n}^{(2)}).\end{aligned}$$

Let $p \geq 4$. Set $(\hat{\alpha}_{j,n}^{(1)}, \hat{\beta}_{j,n}^{(2)}) = (\check{\alpha}_{j,n}^{(k_1)}, \check{\beta}_{j,n}^{(k_2)})$ for $j=1,2,3,4$, and $k_0 = [p/2]$. The hybrid estimators, $\hat{\alpha}_{j,n}^{(2k_0-1)}$, $\hat{\beta}_{j,n}^{(2k_0)}$ and $\hat{\alpha}_{j,n}^{(2k_0+1)}$ are defined as for $k=1,2,\ldots,k_0$,

$$\begin{aligned}V_n^{(2k+1)}(\hat{\alpha}_{j,n}^{(2k+1)} \mid \hat{\alpha}_{j,n}^{(2k-1)}, \hat{\beta}_{j,n}^{(2k)}) &= \sup_{\alpha\in\Theta_\alpha} V_n^{(2k+1)}(\alpha \mid \hat{\alpha}_{j,n}^{(2k-1)}, \hat{\beta}_{j,n}^{(2k)}),\\ V_n^{(2k+2)}(\hat{\beta}_{j,n}^{(2k+2)} \mid \hat{\alpha}_{j,n}^{(2k+1)}, \hat{\beta}_{j,n}^{(2k)}) &= \sup_{\beta\in\Theta_\beta} V_n^{(2k+2)}(\beta \mid \hat{\alpha}_{j,n}^{(2k+1)}, \hat{\beta}_{j,n}^{(2k)}).\end{aligned}$$

Let

$$\begin{aligned}\Gamma(\theta^*) &= \begin{pmatrix} (\Gamma_1(\alpha^*)_{ij})_{i,j=1,\ldots,m_1} & 0 \\ 0 & (\Gamma_2(\theta^*)_{kl})_{k,l=1,\ldots,m_2}\end{pmatrix},\\ \Gamma_1(\alpha^*)_{ij} &= \frac{1}{2}\int_{\mathbf{R}^d} tr\{A^{-1}(\partial_{\alpha_i}A)A^{-1}(\partial_{\alpha_j}A)(x,\alpha^*)\}\mu_{\theta^*}(dx),\\ \Gamma_2(\theta^*)_{kl} &= \int_{\mathbf{R}^d}(\partial_{\beta_k}b(x,\beta^*))^\star A(x,\alpha^*)^{-1}\partial_{\beta_k}b(x,\beta^*)\mu_{\theta^*}(dx).\end{aligned}$$

We make the assumption as follows.

[$A4$] $\Gamma(\theta^*)$ is invertible.

THEOREM 3.1. *Let $p \geq 2$, $k_0 = [p/2]$, $l_0 = [(p-1)/2]$ and $j = 1,2,3,4$. Assume* [$A1$], [$A2$]$(2k_0, 2k_0+1)$, [$A3$]-(j) *and* [$A4$]. *Then, as $nh_n^p \to 0$,*

$$(\sqrt{n}(\hat{\alpha}_{j,n}^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\hat{\beta}_{j,n}^{(2k_0)} - \beta^*)) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{m_1+m_2}(0, \Gamma(\theta^*)^{-1})$$

and

$$E_{\theta^*}[f(\sqrt{n}(\hat{\alpha}_{j,n}^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\hat{\beta}_{j,n}^{(2k_0)} - \beta^*))] \to \mathbb{E}[f(\zeta_1,\zeta_2)]$$

for all continuous functions f of at most polynomial growth.

REMARK 3.2. *We set $p=4$, $\gamma = \frac{1}{3}$, $h_n = \frac{1}{n^{1/3}}$, $G = \frac{1}{2}$, $\epsilon_0 = \frac{1}{4}$. Let $j = 1,2,3,4$. The initial Bayes type estimator $(\tilde{\alpha}_{j,n_0}^{(1)}, \tilde{\beta}_{j,n_0}^{(2)})$ with the sample size $n_0 = [\sqrt{n}]$ has that for all $M > 0$, as $nh_n^4 \to 0$,*

$$\begin{aligned}\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|n^{\frac{1}{4}}(\tilde{\alpha}_{j,n_0}^{(1)} - \alpha^*)\right|^M\right] &< \infty,\\ \sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|(nh_n)^{\frac{1}{16}}(\tilde{\beta}_{j,n_0}^{(2)} - \beta^*)\right|^M\right] &< \infty.\end{aligned}$$

Next, we obtain the multi-step estimator $(\check{\alpha}_{j,n}^{(0)}, \check{\beta}_{j,n}^{(3)})$ *based on initial Bayes estimator* $(\check{\alpha}_{j,n}^{(0)}, \check{\beta}_{j,n}^{(0)}) := (\tilde{\alpha}_{j,n_0}^{(1)}, \tilde{\beta}_{j,n_0}^{(2)})$. *By setting that* $(\hat{\alpha}_{j,n}^{(1)}, \hat{\beta}_{j,n}^{(2)}) = (\check{\alpha}_{j,n}^{(0)}, \check{\beta}_{j,n}^{(3)})$, *it follows from (3) and (4) that for all* $M > 0$, *as* $nh_n^4 \to 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|n^{\frac{1}{4}}(\hat{\alpha}_{j,n}^{(1)} - \alpha^*)\right|^M\right] < \infty,$$
$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|(nh_n)^{\frac{1}{3}}(\hat{\beta}_{j,n}^{(2)} - \beta^*)\right|^M\right] < \infty.$$

Moreover, the hybrid estimator $(\hat{\alpha}_{j,n}^{(3)}, \hat{\beta}_{j,n}^{(4)})$ *is given by*

$$\hat{\alpha}_{j,n}^{(3)} = \arg\sup_{\alpha\in\Theta_\alpha} U_n^{(3)}(\alpha \mid \hat{\alpha}_{j,n}^{(1)}, \hat{\beta}_{j,n}^{(2)}),$$
$$\hat{\beta}_{j,n}^{(4)} = \arg\sup_{\beta\in\Theta_\beta} U_n^{(4)}(\beta \mid \hat{\alpha}_{j,n}^{(3)}, \hat{\beta}_{j,n}^{(2)}),$$

and it follows from Theorem 3.1 with l_0=*1 and* $k_0 = 2$ *that as* $nh_n^4 \to 0$,

$$(\sqrt{n}(\hat{\alpha}_{j,n}^{(3)} - \alpha^*), \sqrt{nh_n}(\hat{\beta}_{j,n}^{(4)} - \beta^*)) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{m_1+m_2}(0, \Gamma(\theta^*)^{-1})$$

and

$$E_{\theta^*}[f(\sqrt{n}(\hat{\alpha}_{j,n}^{(3)} - \alpha^*), \sqrt{nh_n}(\hat{\beta}_{j,n}^{(4)} - \beta^*))] \to \mathbb{E}[f(\zeta_1, \zeta_2)]$$

for all continuous functions f *of at most polynomial growth.*

4. Examples and simulation results

Consider the following three-dimensional diffusion process defined by

$$dX_t = b(X_t, \beta)dt + a(X_t, \alpha)dw_t, \quad t \geq 0, \quad X_0 = (1, 1, 1)^\star,$$

where

$$b(X_t, \beta) = \begin{pmatrix} 1 - 3X_{t,1} - 10\sin(\beta_1 X_{t,2}^2) \\ 2 - 3X_{t,2} - 10\sin(\beta_2 X_{t,3}^2) \\ 3 - 3X_{t,3} - 10\sin(\beta_3 X_{t,1}^2 + \beta_4 X_{t,1}) \end{pmatrix},$$
$$a(X_t, \alpha) = \begin{pmatrix} \sqrt{2 + \cos(\alpha_1 X_{t,3}^2)} & 0.01 & 0 \\ 0.01 & \sqrt{2 + \cos(\alpha_2 X_{t,1}^2)} & 0 \\ 0 & 0 & \sqrt{2 + \cos(\alpha_3 X_{t,2}^2 + \alpha_4 X_{t,2})} \end{pmatrix}.$$

Furthermore, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ are unknown parameters, and the true parameter values are $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, \alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*) = (3, 6, 9, 12, 15, 18, 21, 24)$. The parameter space is assumed to be $\Theta = [0.1, 50]^8$. We note that the computation time of the initial Bayes estimator strongly depends on the parameter space Θ.

The simulations were done for $T = 250$, $h = 1/390$, which means that $n = 97500$. In this example, it is assumed that the data with $h = 1/390$ and $T = 250$ are trading data observed at every minute for one year in Japanese financial market.

Let $N_0 \geq n_0 = [n^G]$. We set $p = q = 4$, $(N_0, G) = (9000, \frac{79}{100}), (10000, \frac{80}{100}), (15000, \frac{84}{100}), (20000, \frac{86}{100})$ and $\epsilon_0 = \frac{1}{10}$. It follows from Remark 2.2 that

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|n^{\frac{1}{4}}(\tilde{\alpha}^{(1)}_{j,N_0} - \alpha^*)\right|^M\right] < \infty,$$
$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|(nh_n)^{\frac{1}{40}}(\tilde{\beta}^{(2)}_{j,N_0} - \beta^*)\right|^M\right] < \infty.$$

Let $q_1 = \max\{p-1, 2\} = 3$. Let $k_1 = \log_2(q/p) = 0$ and $k_2 = 4 \geq \log_2(q/(\epsilon_0 q_1))$. It follows from the same method as Section 3 that for $j = 1,2,3,4$, we obtain the multi-step estimator $(\check{\alpha}^{(0)}_{j,n}, \check{\beta}^{(4)}_{j,n})$ based on initial Bayes estimator $(\tilde{\alpha}^{(1)}_{j,N_0}, \tilde{\beta}^{(2)}_{j,N_0})$. Set $(\hat{\alpha}^{(1)}_{j,n}, \hat{\beta}^{(2)}_{j,n}) = (\check{\alpha}^{(0)}_{j,n}, \check{\beta}^{(4)}_{j,n})$. Moreover, the hybrid estimator $(\hat{\alpha}^{(3)}_{j,n}, \hat{\beta}^{(4)}_{j,n})$ is given by

$$\hat{\alpha}^{(3)}_{j,n} = \arg\sup_{\alpha\in\Theta_\alpha} V^{(3)}_n(\alpha \mid \hat{\alpha}^{(1)}_{j,n}, \hat{\beta}^{(2)}_{j,n}),$$
$$\hat{\beta}^{(4)}_{j,n} = \arg\sup_{\beta\in\Theta_\beta} V^{(4)}_n(\beta \mid \hat{\alpha}^{(3)}_{j,n}, \hat{\beta}^{(2)}_{j,n}),$$

where

$$V^{(3)}_n(\alpha \mid \bar{\theta}) = -\frac{1}{2}\sum_{i=1}^n \left\{h_n^{-1}A^{-1}_{i-1}(\alpha)\left[(X_{t^n_i} - X_{t^n_{i-1}})^{\otimes 2} - h_n^2 \bar{D}^{(2)}_{i-1}(\bar{\theta})\right] + \log\det A_{i-1}(\alpha)\right\},$$
$$V^{(4)}_n(\beta \mid \bar{\theta}) = -\frac{1}{2}\sum_{i=1}^n h_n^{-1}A^{-1}_{i-1}(\alpha)\left[\left(X_{t^n_i} - X_{t^n_{i-1}} - h_n b_{i-1}(\beta) - h_n^2 \bar{r}^{(2)}_{i-1}(\bar{\theta})\right)^{\otimes 2}\right].$$

For the definition of $\bar{D}^{(2)}_{i-1}(\bar{\theta})$ and $\bar{r}^{(2)}_{i-1}(\bar{\theta})$, see Section 3.

In order to compute the maximum likelihood type estimator, we used **optim()** with the "L-BFGS-B" method in the R Language. The Bayes estimators are calculated with MpCN method proposed by Kamatani (2014) for 10^6 Markov chains and 10^5 burn-in iterations. For MpCN algorithm, see Kamatani (2014) and Kaino et al. (2017).

For the true model, 100 independent sample paths are generated by the Euler-Maruyama scheme, and the mean and the standard deviation (s.d.) for the estimators in Theorems 1 and 2 are computed and shown in Tables 5-13 below. For simulations, we used the personal computer with Intel i7-5930K (3.5GHz base clock). In each table, the time means the computation time of estimator for one sample path.

Table 5 shows the simulation results of the adaptive ML type estimator $(\hat{\alpha}^{(3)}_{A,n}, \hat{\beta}^{(4)}_{A,n})$ defined as

$$\hat{\alpha}^{(1)}_{A,n} = \arg\sup_{\alpha\in\Theta_\alpha} V^{(1)}_n(\alpha),$$
$$\hat{\beta}^{(2)}_{A,n} = \arg\sup_{\beta\in\Theta_\beta} V^{(2)}_n(\beta \mid \hat{\alpha}^{(1)}_{A,n}),$$
$$\hat{\alpha}^{(3)}_{A,n} = \arg\sup_{\alpha\in\Theta_\alpha} V^{(3)}_n(\alpha \mid \hat{\alpha}^{(1)}_{A,n}, \hat{\beta}^{(2)}_{A,n}),$$
$$\hat{\beta}^{(4)}_{A,n} = \arg\sup_{\beta\in\Theta_\beta} V^{(4)}_n(\beta \mid \hat{\alpha}^{(3)}_{A,n}, \hat{\beta}^{(2)}_{A,n}),$$

where

$$\begin{aligned} V_n^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^{n}\left\{h_n^{-1}A_{i-1}^{-1}(\alpha)[(\Delta X_i)^{\otimes 2}] + \log\det(A_{i-1}(\alpha))\right\}, \\ V_n^{(2)}(\beta \mid \alpha) &= -\frac{1}{2}\sum_{i=1}^{n}\left\{h_n^{-1}A_{i-1}^{-1}(\alpha)[(\Delta X_i - h_n b_{i-1}(\beta))^{\otimes 2}]\right\}. \end{aligned}$$

The adaptive ML type estimator $(\hat{\alpha}_{A,n}^{(3)}, \hat{\beta}_{A,n}^{(4)})$ is computed by using **optim()** with the initial value being the true value. We see from Table 5 that all estimators have good behavior. Table 6 is the simulation results of the adaptive ML type estimator $(\hat{\alpha}_{A,n}^{(3)}, \hat{\beta}_{A,n}^{(4)})$ with the initial value being the uniform random number on Θ. All estimators have considerable biases, which means that the optimization fails since the initial value may be far from the true value. As we know very well, it is quite important to choose the initial value for optimization.

Table 7 shows the simulation results of four kinds of initial Bayes type estimators when the sample size of the reduced data $N_0 = 9000$. The calculation time of type 4 Bayes estimator is shortest and that of type 1 Bayes estimator is longest. For both the Bayes estimators of type 1 and type 3, $\hat{\alpha}_3$ and $\hat{\alpha}_4$ have large standard deviations. For the Bayes estimators of type 2, 3 and 4, $\hat{\beta}_3$ and $\hat{\beta}_4$ have large standard deviations. Table 8 shows the simulation results of the hybrid estimators with four initial Bayes estimators when the sample size of the reduced data $N_0 = 9000$. It does not seem that the hybrid estimators improve performance of the Bayes estimators in Table 7, which means that the initial estimator plays an important role in this example.

Table 9 shows the simulation results of four initial Bayes type estimators with $N_0 = 10000$. The calculation time of the Bayes estimator in Table 9 is longer than the one in Table 7. The standard deviations of $\hat{\alpha}_3$ and $\hat{\alpha}_4$ of the Bayes estimators of type 1 and type 3 are large. The standard deviations of $\hat{\beta}_3$ and $\hat{\beta}_4$ of the Bayes estimators of type 2 and 4 are also large. Table 10 shows the simulation results of the hybrid estimators with four initial Bayes estimators with $N_0 = 10000$. Similarly to the hybrid estimator in Table 8, the hybrid estimators do not improve the Bayes estimators in Table 9.

Tables 11 shows the simulation results of four initial Bayes type estimators with $N_0 = 15000$. The type 1, 3 and 4 Bayes estimators of β_4 have large standard deviations. The Bayes estimator of type 2 has good behavior. Table 12 shows the simulation results of the hybrid estimators with four initial Bayes estimators with $N_0 = 15000$. The hybrid estimators with the initial Bayes estimators of type 1, 3 and 4 do not improve the Bayes estimators in Table 11. On the other hand, the hybrid estimator with the initial Bayes estimator of type 2 is better than the initial Bayes estimator in Table 11. The performance of the hybrid estimator with the initial Bayes estimator of type 2 is similar to that of the estimator in Table 6.

Table 13 shows the simulation results of four initial Bayes type estimators with $N_0 = 20000$. The standard deviations of the type 3 and 4 Bayes estimators of β_3 and β_4 and the type 1 Bayes estimators of α_3 and α_4 are large. The Bayes estimator of type 2 has good behavior. Table 14 shows the simulation results of the hybrid estimators with four initial Bayes estimators with $N_0 = 20000$. The hybrid estimators with the initial Bayes estimators of type 1, 3 and 4 do not improve the Bayes estimators in Table 13. The hybrid estimator with the initial Bayes estimator of type 2 is better than the initial

Bayes estimators in Table 13. It is worth mentioning that the performance of hybrid estimator with the initial Bayes estimators of type 2 is as good as the estimator in Table 6.

In this example, we see from the simulation results that the Bayes type estimator of type 2 have good performance when the sample size of the reduced data is $N_0 \geq 15000$.

Table 5: adaptive ML type estimator with the initial value being the true value

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(sec.)
	3.002	5.997	8.984	11.981	14.988	17.981	20.972	24.008	
true	(0.014)	(0.012)	(0.033)	(0.046)	(0.045)	(0.053)	(0.089)	(0.083)	70

Table 6: adaptive ML type estimator with the initial value being the uniform random number on Θ

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(sec.)
	22.040	23.512	27.838	18.982	25.201	23.444	23.588	26.823	
unif	(16.074)	(16.379)	(14.004)	(13.446)	(16.654)	(16.702)	(17.871)	(16.628)	80

Table 7: initial Bayes type estimator ($N_0 = 9000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(h.)
	2.989	5.985	8.937	11.927	14.899	17.595	19.884	22.050	
type1	(0.047)	(0.041)	(0.123)	(0.153)	(0.311)	(0.808)	(2.946)	(3.202)	5.2
	2.991	5.984	9.406	12.680	14.988	17.956	20.873	23.903	
type2	(0.049)	(0.041)	(2.868)	(4.261)	(0.130)	(0.203)	(0.423)	(0.352)	1.6
	2.993	5.982	9.125	12.177	14.899	17.595	19.884	22.050	
type3	(0.051)	(0.047)	(1.881)	(2.464)	(0.311)	(0.808)	(2.946)	(3.202)	4.9
	2.993	5.982	9.125	12.177	14.988	17.956	20.873	23.903	
type 4	(0.051)	(0.047)	(1.881)	(2.464)	(0.130)	(0.203)	(0.423)	(0.352)	1.0

Table 8: hybrid type estimator with the initial value being the Bayes type estimator ($N_0 = 9000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(sec.)
	3.002	5.997	8.983	11.981	14.969	17.888	20.060	23.017	
type1	(0.014)	(0.012)	(0.033)	(0.044)	(0.202)	(0.671)	(2.633)	(3.396)	70
	3.002	5.998	9.444	12.721	14.989	17.981	20.961	23.998	
type2	(0.014)	(0.012)	(2.815)	(4.223)	(0.045)	(0.053)	(0.131)	(0.118)	70
	3.002	5.998	9.167	12.226	14.969	17.888	20.060	23.017	
type3	(0.014)	(0.012)	(1.837)	(2.450)	(0.202)	(0.671)	(2.633)	(3.396)	70
	3.002	5.997	9.167	12.227	14.988	17.981	20.961	23.998	
type4	(0.014)	(0.012)	(1.837)	(2.450)	(0.045)	(0.053)	(0.131)	(0.118)	70

Table 9: initial Bayes type estimator ($N_0 = 10000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(h.)
	2.995	5.981	8.946	11.943	14.866	17.664	20.071	22.327	
type1	(0.045)	(0.039)	(0.110)	(0.136)	(0.373)	(0.533)	(2.309)	(2.614)	5.8
	2.990	5.983	9.133	12.520	14.989	17.972	20.864	23.894	
type2	(0.062)	(0.041)	(1.311)	(4.088)	(0.109)	(0.177)	(0.439)	(0.351)	1.8
	2.993	5.982	9.022	12.181	14.866	17.664	20.071	22.327	
type3	(0.046)	(0.044)	(0.784)	(2.450)	(0.373)	(0.533)	(2.309)	(2.614)	5.4
	2.993	5.982	9.022	12.181	14.989	17.972	20.864	23.894	
type4	(0.046)	(0.044)	(0.784)	(2.450)	(0.109)	(0.177)	(0.439)	(0.351)	1.0

Table 10: hybrid type estimator with the initial value being the Bayes type estimator ($N_0 = 10000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(sec.)
	3.002	5.997	8.984	11.982	14.988	17.981	20.260	23.321	
type1	(0.014)	(0.012)	(0.033)	(0.044)	(0.045)	(0.053)	(2.139)	(2.530)	70
	3.002	5.997	9.161	12.546	14.988	17.981	20.948	23.985	
type2	(0.014)	(0.012)	(1.284)	(4.048)	(0.045)	(0.053)	(0.194)	(0.178)	70
	3.002	5.997	9.056	12.218	14.988	17.981	20.260	23.321	
type3	(0.014)	(0.012)	(0.718)	(2.359)	(0.045)	(0.053)	(2.139)	(2.530)	70
	3.002	5.997	9.055	12.217	14.988	17.981	20.948	23.985	
type4	(0.014)	(0.012)	(0.718)	(2.359)	(0.045)	(0.053)	(0.194)	(0.178)	70

Table 11: initial Bayes type estimator ($N_0 = 15000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(h.)
	2.988	5.982	9.207	12.753	14.891	17.489	20.388	22.568	
type1	(0.036)	(0.034)	(1.490)	(4.688)	(0.236)	(0.752)	(1.496)	(1.490)	8.7
	2.987	5.983	8.950	11.945	14.988	17.977	20.926	23.929	
type2	(0.042)	(0.033)	(0.081)	(0.105)	(0.095)	(0.136)	(0.274)	(0.252)	2.8
	2.991	5.984	9.133	12.514	14.891	17.489	20.388	22.568	
type 3	(0.035)	(0.032)	(1.308)	(4.076)	(0.236)	(0.752)	(1.496)	(1.490)	8.1
	2.991	5.984	9.133	12.514	14.988	17.977	20.926	23.929	
type4	(0.035)	(0.032)	(1.308)	(4.076)	(0.095)	(0.136)	(0.274)	(0.252)	1.6

Table 12: hybrid type estimator with the initial value being the Bayes type estimator ($N_0 = 15000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(sec.)
	3.002	5.997	9.238	12.788	14.988	17.937	20.647	23.817	
type1	(0.014)	(0.012)	(1.487)	(4.690)	(0.045)	(0.452)	(1.211)	(0.682)	70
	3.002	5.997	8.984	11.981	14.988	17.981	20.962	23.999	
type2	(0.014)	(0.012)	(0.033)	(0.046)	(0.045)	(0.053)	(0.130)	(0.117)	70
	3.002	5.997	9.165	12.552	14.988	17.937	20.647	23.817	
type3	(0.014)	(0.012)	(1.296)	(4.070)	(0.045)	(0.452)	(1.211)	(0.682)	70
	3.002	5.997	9.165	12.552	14.988	17.981	20.962	23.999	
type4	(0.014)	(0.012)	(1.296)	(4.070)	(0.045)	(0.053)	(0.130)	(0.117)	70

Table 13: initial Bayes type estimator ($N_0 = 20000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(h.)
	2.990	5.985	8.962	11.958	14.915	17.626	20.172	22.619	
type1	(0.030)	(0.031)	(0.068)	(0.083)	(0.155)	(0.570)	(1.420)	(1.572)	11.6
	2.989	5.987	8.963	11.959	14.985	17.985	20.959	23.959	
type2	(0.030)	(0.033)	(0.066)	(0.086)	(0.083)	(0.109)	(0.239)	(0.205)	3.7
	2.994	5.988	9.122	12.192	14.915	17.626	20.172	22.619	
type3	(0.037)	(0.033)	(1.553)	(2.257)	(0.155)	(0.570)	(1.420)	(1.572)	10.9
	2.994	5.988	9.122	12.192	14.985	17.985	20.959	23.959	
type4	(0.037)	(0.033)	(1.553)	(2.257)	(0.083)	(0.109)	(0.239)	(0.205)	2.1

Table 14: hybrid type estimator with the initial value being the Bayes type estimator ($N_0 = 20000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\alpha}_1(15)$	$\hat{\alpha}_2(18)$	$\hat{\alpha}_3(21)$	$\hat{\alpha}_4(24)$	time(sec.)
	3.002	5.997	8.983	11.981	14.988	17.933	20.464	23.649	
type1	(0.014)	(0.012)	(0.033)	(0.046)	(0.045)	(0.345)	(1.473)	(0.952)	70
	3.002	5.997	8.984	11.981	14.988	17.981	20.973	24.008	
type2	(0.014)	(0.012)	(0.033)	(0.046)	(0.045)	(0.053)	(0.093)	(0.086)	70
	3.002	5.997	9.138	12.204	14.988	17.933	20.464	23.649	
type3	(0.014)	(0.012)	(1.547)	(2.228)	(0.045)	(0.345)	(1.473)	(0.952)	70
	3.002	5.997	9.139	12.204	14.988	17.981	20.973	24.008	
type4	(0.014)	(0.012)	(1.546)	(2.228)	(0.045)	(0.053)	(0.093)	(0.086)	70

As another example, we treat the three-dimensional diffusion process as follows.

$$dX_t = b(X_t, \beta)dt + a(X_t, \alpha)dw_t, \quad t \geq 0, \quad X_0 = (1,1,1)^\star,$$

where

$$b(X_t, \beta) = \begin{pmatrix} \beta_1 - \beta_2 X_{t,1} - 10\sin(\beta_3 X_{t,2}^2) \\ \beta_4 - \beta_5 X_{t,2} - \beta_6 \sin(X_{t,3}^2) \\ \beta_7 - \beta_8 X_{t,3} - 10\sin(\beta_9 X_{t,1}^2) \end{pmatrix},$$

$$a(X_t, \alpha) = \begin{pmatrix} \sqrt{\alpha_1(2+\cos(X_{t,3}^2))} & 0.01 & 0 \\ 0.01 & \sqrt{\alpha_2(2+\cos(X_{t,1}^2))} & 0 \\ 0 & 0 & \sqrt{\alpha_3(2+\cos(X_{t,2}^2))} \end{pmatrix}.$$

Furthermore, $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)$ are unknown parameters, and the true parameter values are $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, \beta_5^*, \beta_6^*, \beta_7^*, \beta_8^*, \beta_9^*) = (3, 6, 9, 12, 15, 18, 21, 24, 27)$ and $(\alpha_1^*, \alpha_2^*, \alpha_3^*) = (1, 2, 3)$. The parameter space Θ is assumed to be $[0.01, 50]^{12}$. Let $p = q = 4$, $(N_0, G) = (20000, \frac{86}{100})$ and $\epsilon_0 = \frac{1}{10}$. The simulations were done in the same setting as the previous example, which means that $T = 250$, $h = 1/390$ and $n = 97500$. In this example, we will investigate the initial Bayes type estimator of type 4 when the sample size of the reduced data $N_0 = 20000$ since the computation time is shortest among the four kinds of initial Bayes type estimators. The Bayes type estimators of α and β are calculated with MpCN method proposed by Kamatani (2014) for 5×10^5 and 10^7 Markov chains and 5×10^4 and 10^6 burn-in iterations, respectively.

Tables 15 and 16 show the simulation results of the adaptive ML type estimator $(\hat{\alpha}^{(3)}_{A,n}, \hat{\beta}^{(4)}_{A,n})$ with the initial value being the true value. We can see that all estimators have good performance. Tables 17 and 18 show the simulation results of the adaptive ML type estimator $(\hat{\alpha}^{(3)}_{A,n}, \hat{\beta}^{(4)}_{A,n})$ with the initial value being the uniform random number on Θ. Similarly to the previous example, the optimization fails because of the inappropriate initial value, and several estimators of β have considerable biases.

Tables 19 and 20 show the simulation results of type 4 of initial Bayes type estimators when the sample size of the reduced data $N_0 = 20000$. The type 4 Bayes estimators of β_4 and β_6 have biases. On the other hand, the Bayes estimators of α have good performance. Tables 21 and 22 show the simulation results of the hybrid estimators for the initial Bayes estimators of type 4 with $N_0 = 20000$. The hybrid estimators improve the initial Bayes estimators of type 4 in Tables 19 and 20. We can see from Tables 21 and 22 that the performance of hybrid estimator with the initial Bayes estimators of type 4 is as good as the estimator in Tables 15 and 16.

Table 15: adaptive ML type estimator of β with the initial value being the true value

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\beta}_5(15)$	$\hat{\beta}_6(18)$	$\hat{\beta}_7(21)$	$\hat{\beta}_8(24)$	$\hat{\beta}_9(27)$
	2.996	5.973	8.883	11.992	14.970	17.975	20.940	23.913	26.723
true	(0.099)	(0.199)	(0.079)	(0.300)	(0.344)	(0.474)	(0.402)	(0.453)	(0.472)

Table 16: adaptive ML type estimator of α with the initial value being the true value

	$\hat{\alpha}_1(1)$	$\hat{\alpha}_2(2)$	$\hat{\alpha}_3(3)$	time(sec.)
	0.992	2.003	2.983	
true	(0.005)	(0.010)	(0.015)	70

Table 17: adaptive ML type estimator of β with the initial value being the uniform random number on Θ

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\beta}_5(15)$	$\hat{\beta}_6(18)$	$\hat{\beta}_7(21)$	$\hat{\beta}_8(24)$	$\hat{\beta}_9(27)$
	1.848	5.713	9.784	11.993	14.975	17.978	20.424	23.751	23.676
unif	(1.318)	(0.371)	(12.505)	(0.302)	(0.349)	(0.476)	(0.983)	(0.477)	(14.302)

Table 18: adaptive ML type estimator of α with the initial value being the uniform random number on Θ

	$\hat{\alpha}_1(1)$	$\hat{\alpha}_2(2)$	$\hat{\alpha}_3(3)$	time(sec.)
	1.001	2.002	2.988	
unif	(0.016)	(0.010)	(0.020)	80

Table 19: initial Bayes type estimator of β ($N_0 = 20000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\beta}_5(15)$	$\hat{\beta}_6(18)$	$\hat{\beta}_7(21)$	$\hat{\beta}_8(24)$	$\hat{\beta}_9(27)$	time(h.)
	3.061	6.192	8.824	11.365	14.674	16.868	20.576	23.604	26.575	
type4	(0.551)	(0.943)	(0.292)	(1.107)	(1.193)	(1.717)	(1.353)	(1.514)	(1.256)	16

Table 20: initial Bayes type estimator of α ($N_0 = 20000$)

	$\hat{\alpha}_1(1)$	$\hat{\alpha}_2(2)$	$\hat{\alpha}_3(3)$	time(h.)
	1.012	1.985	2.923	
type4	(0.010)	(0.021)	(0.030)	1.5

Table 21: hybrid type estimator of β with the initial value being the Bayes type estimator ($N_0 = 20000$)

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\beta}_5(15)$	$\hat{\beta}_6(18)$	$\hat{\beta}_7(21)$	$\hat{\beta}_8(24)$	$\hat{\beta}_9(27)$
	2.996	5.973	8.883	11.992	14.971	17.975	20.940	23.911	26.663
type4	(0.099)	(0.199)	(0.079)	(0.300)	(0.343)	(0.473)	(0.406)	(0.456)	(0.539)

Table 22: hybrid type estimator of α with the initial value being the Bayes type estimator ($N_0 = 20000$)

	$\hat{\alpha}_1(1)$	$\hat{\alpha}_2(2)$	$\hat{\alpha}_3(3)$	time(sec.)
	0.992	2.003	2.983	
type4	(0.005)	(0.010)	(0.015)	70

Next, in order to compare to the hybrid type estimator $(\hat{\alpha}_{4,n}^{(3)}, \hat{\beta}_{4,n}^{(4)})$ with the initial Bayes type estimator of type 4 based on reduced data, we consider the following two kinds of initial estimators $(\hat{\alpha}_I^{(1)}, \hat{\beta}_I^{(2)})$ and $(\hat{\alpha}_{II}^{(1)}, \hat{\beta}_{II}^{(2)})$. Let $N_0 = 20000$.

Method I. Using 27^3 uniform random numbers $\alpha_{0,m}$ $(m = 1, \ldots, 27^3)$ on $[0.01, 50]^3$, we compute

$$\hat{\alpha}_m^{(1)} = \arg\sup_{\alpha} U_{4,N_0}^{(1)}(\alpha)$$

by means of **optim()** with each initial value $\alpha_{0,m}$. The initial estimator $\hat{\alpha}_{I,N_0}^{(1)}$ is defined as

$$U_{4,N_0}^{(1)}(\hat{\alpha}_{I,N_0}^{(1)}) = \max\left\{U_{4,N_0}^{(1)}(\hat{\alpha}_1^{(1)}), U_{4,N_0}^{(1)}(\hat{\alpha}_2^{(1)}), \ldots, U_{4,N_0}^{(1)}(\hat{\alpha}_{27^3}^{(1)})\right\}.$$

Next, using 35000 uniform random numbers $\beta_{0,m}$ $(m = 1, \ldots, 35000)$ on $[0.01, 50]^9$, we compute

$$\hat{\beta}_m^{(2)} = \arg\sup_{\beta} U_{4,N_0}^{(2)}(\beta)$$

by means of **optim()** with each initial value $\beta_{0,m}$. The initial estimator $\hat{\beta}_{I,N_0}^{(2)}$ is defined as

$$U_{4,N_0}^{(2)}(\hat{\beta}_{I,N_0}^{(2)}) = \max\left\{U_{4,N_0}^{(2)}(\hat{\beta}_1^{(2)}), U_{4,N_0}^{(2)}(\hat{\beta}_2^{(2)}), \ldots, U_{4,N_0}^{(2)}(\hat{\beta}_{35000}^{(2)})\right\}.$$

Method II. For 100^3 points $\bar{\alpha}_{0,m}$ $(m = 1, \ldots, 100^3)$ with 100 equally spaced points on each axis on $[0.01, 50]^3$, the initial estimator $\hat{\alpha}_{II,N_0}^{(1)}$ is defined as

$$U_{4,N_0}^{(1)}(\hat{\alpha}_{II,N_0}^{(1)}) = \max\left\{U_{4,N_0}^{(1)}(\bar{\alpha}_{0,1}), U_{4,N_0}^{(1)}(\bar{\alpha}_{0,2}), \ldots, U_{4,N_0}^{(1)}(\bar{\alpha}_{0,100^3})\right\}.$$

Next, for 7^9 points $\bar{\beta}_{0,m}$ $(m = 1,\ldots,7^9)$ on $[0.01, 50]^9$ with 7 equally spaced points on each axis on $[0.01, 50]^9$, the initial estimator $\hat{\beta}^{(2)}_{II,N_0}$ is defined as

$$U^{(2)}_{4,N_0}(\hat{\beta}^{(2)}_{II,N_0}) = \max\left\{U^{(2)}_{4,N_0}(\beta_{0,1}), U^{(2)}_{4,N_0}(\beta_{0,2}), \ldots, U^{(2)}_{4,N_0}(\beta_{0,7^9})\right\}.$$

Let $k = I, II$. By the same method as Section 3, we obtain the multi-step estimator $(\check{\alpha}^{(0)}_{k,n}, \check{\beta}^{(4)}_{k,n})$ based on the initial estimator $(\hat{\alpha}^{(1)}_{k,N_0}, \hat{\beta}^{(2)}_{k,N_0})$. Set $(\bar{\alpha}^{(1)}_{k,n}, \bar{\beta}^{(2)}_{k,n}) = (\check{\alpha}^{(0)}_{k,n}, \check{\beta}^{(4)}_{k,n})$. Moreover, the hybrid estimator $(\bar{\alpha}^{(3)}_{k,n}, \bar{\beta}^{(4)}_{k,n})$ is given by

$$\begin{aligned}\bar{\alpha}^{(3)}_{k,n} &= \arg\sup_{\alpha\in\Theta_\alpha} V^{(3)}_n(\alpha \mid \bar{\alpha}^{(1)}_{k,n}, \bar{\beta}^{(2)}_{k,n}),\\ \bar{\beta}^{(4)}_{k,n} &= \arg\sup_{\beta\in\Theta_\beta} V^{(4)}_n(\beta \mid \bar{\alpha}^{(3)}_{k,n}, \bar{\beta}^{(2)}_{k,n}).\end{aligned}$$

Let $\hat{\theta}_B = (\hat{\alpha}_B, \hat{\beta}_B) := (\hat{\alpha}^{(3)}_{4,n}, \hat{\beta}^{(4)}_{4,n})$, $\hat{\theta}_I = (\hat{\alpha}_I, \hat{\beta}_I) := (\bar{\alpha}^{(3)}_{I,n}, \bar{\beta}^{(4)}_{I,n})$ and $\hat{\theta}_{II} = (\hat{\alpha}_{II}, \hat{\beta}_{II}) := (\bar{\alpha}^{(3)}_{II,n}, \bar{\beta}^{(4)}_{II,n})$. Tables 23 and 24 show the simulation results of the hybrid estimators $\hat{\theta}_B$, $\hat{\theta}_I$ and $\hat{\theta}_{II}$ for the initial estimator based on reduced data with $N_0 = 20000$. As seen from Tables 21 and 22, the hybrid estimator $\hat{\theta}_B$ with the initial Bayes estimator has good performance. The hybrid estimators of the method I for β_7 and β_9 have considerable biases. The hybrid estimators of the method II have bad behavior. As we know very well, it takes much time to compute $\hat{\theta}_I$ and $\hat{\theta}_{II}$ when the dimension of parameter space is large. Taking account into both accuracy and computation time of the estimator, it seems that the hybrid estimator with the initial Bayes estimator of type 4 is much better than both $\hat{\theta}_I$ and $\hat{\theta}_{II}$ in this example.

Table 23: $\hat{\beta}_B$ (hybrid), $\hat{\beta}_I$ (35000 random numbers) and $\hat{\beta}_{II}$ (7^9 lattice points) with $N_0 = 20000$

	$\hat{\beta}_1(3)$	$\hat{\beta}_2(6)$	$\hat{\beta}_3(9)$	$\hat{\beta}_4(12)$	$\hat{\beta}_5(15)$	$\hat{\beta}_6(18)$	$\hat{\beta}_7(21)$	$\hat{\beta}_8(24)$	$\hat{\beta}_9(27)$	time(h.)
	2.996	5.973	8.883	11.992	14.971	17.975	20.940	23.911	26.663	
$\hat{\beta}_B$	(0.099)	(0.199)	(0.079)	(0.300)	(0.343)	(0.473)	(0.406)	(0.456)	(0.539)	16
	2.996	5.972	8.883	11.997	14.977	17.984	19.762	23.628	12.555	
$\hat{\beta}_I$	(0.099)	(0.198)	(0.079)	(0.300)	(0.343)	(0.476)	(1.191)	(0.512)	(13.394)	16
	0.138	5.350	0.010	11.999	14.999	17.981	19.219	23.493	6.169	
$\hat{\beta}_{II}$	(0.090)	(0.1861)	(0.000)	(0.300)	(0.343)	(0.473)	(1.002)	(0.476)	(11.329)	32

Table 24: $\hat{\alpha}_B$ (hybrid), $\hat{\alpha}_I$ (27^3 random numbers) and $\hat{\alpha}_{II}$ (100^3 lattice points) with $N_0 = 20000$

	$\hat{\alpha}_1(1)$	$\hat{\alpha}_2(2)$	$\hat{\alpha}_3(3)$	time(h.)
	0.992	2.003	2.983	
$\hat{\alpha}_B$	(0.005)	(0.010)	(0.015)	1.5
	0.992	2.002	2.999	
$\hat{\alpha}_I$	(0.005)	(0.010)	(0.021)	1.5
	1.022	2.002	3.006	
$\hat{\alpha}_{II}$	(0.005)	(0.010)	(0.019)	1.5

5. Conclusion

In this paper, we have studied the ML type estimators of both drift and volatility parameters for discretely observed ergodic diffusion processes from the viewpoint of numerical analysis. In general, it is important to select a suitable initial value for optimization of the quasi likelihood function by using optim() in R language. On the other hand, the computation of the Bayes type estimator does not strongly depend on the initial value. Therefore, using the reduced data obtained from the full data, we have derived the initial Bayes type estimators of both drift and volatility parameters. Note that there is no need to use the first n_0 data as the reduced data.

Although a disadvantage of the initial Bayes type estimators with the reduced data is that they do not have optimal rates, they also have an great advantage that the computation time of them is much shorter than that of the Bayes type estimators with the full data. Furthermore, we see from the results of Kamatani and Uchida (2015) and Kutoyants (2017) that the bad convergence rates of the initial estimators do not matter at the first step for derivation of the efficient estimators. In fact, it is shown that the hybrid estimator with the initial Bayes type estimator, which means the adaptive ML type estimator with the initial Bayes type estimator, has asymptotic efficiency and convergence of moments.

We see from Tables 10 and 14 that the hybrid estimators of α_2 based on the type 1 initial Bayes estimators with $N_0 = 10,000$ and $20,000$ are unstable. Compared with the type 1, 3 and 4 initial Bayes estimators and the hybrid estimators, the type 2 initial Bayes estimator and the hybrid estimator based on the reduced data with $N_0 = 15,000$ and $20,000$ have good performance. We recommend that all the initial Bayes estimators and the hybrid estimators should be computed. The best estimator is selected by comparing the quasi likelihoods.

It follows from the numerical results in Tables 23 and 24 of Section 4 that the proposed hybrid estimators with the initial Bayes estimators are as good as the results of Tables 15 and 16 and the hybrid estimator with the initial Bayes type estimator is best among the competing estimators in the sense of both accuracy and computation time of the estimator.

6. Proofs

Proof of Proposition 2.1. In the similar way to the proof of (9) in Uchida (2010), we can show the result. For details, see Kaino et al. (2017).

Proof of Theorem 2.2. First, we will prove the case of the type 1 Bayes estimator $\tilde{\alpha}^{(1)}_{1,n_0}$. Set

$$\mathbb{H}^{(1)}_{1,n_0}(\alpha) \quad = \quad \frac{1}{n_0^{1-\frac{2}{qG}}} U^{(1)}_{1,n_0}(\alpha),$$

$$\begin{aligned}
\mathbb{Y}^{(1)}_{1,n_0}(\alpha) &= \frac{1}{n_0^{\frac{2}{qG}}}\{\mathbb{H}^{(1)}_{1,n_0}(\alpha) - \mathbb{H}^{(1)}_{1,n_0}(\alpha^*)\} = \frac{1}{n_0}\left\{U^{(1)}_{1,n_0}(\alpha) - U^{(1)}_{1,n_0}(\alpha^*)\right\},\\
\Delta^{(1)}_{1,n_0}(\alpha^*)[u_1] &= \frac{1}{n_0^{\frac{1}{qG}}}\partial_\alpha \mathbb{H}^{(1)}_{1,n_0}(\alpha^*)[u_1] = \frac{1}{n_0^{1-\frac{1}{qG}}}\partial_\alpha U^{(1)}_{1,n_0}(\alpha^*)[u_1],\\
\Gamma^{(1)}_{1,n_0}(\alpha^*)[u_1,u_1] &= -\frac{1}{n_0^{\frac{2}{qG}}}\partial^2_\alpha \mathbb{H}^{(1)}_{1,n_0}(\alpha^*)[u_1,u_1] = -\frac{1}{n_0}\partial^2_\alpha U^{(1)}_{1,n_0}(\alpha^*)[u_1,u_1],\\
\Gamma_1(\alpha^*)[u_1,u_1] &= \frac{1}{2}\int_{\mathbf{R}^d} tr\{A^{-1}(\partial_\alpha A)A^{-1}(\partial_\alpha A)(x,\alpha^*)[u_1^{\otimes 2}]\}\mu_{\theta^*}(dx)
\end{aligned}$$

for $u_1 \in \mathbf{R}^{m_1}$. Let $\mathbb{U}^{(1)}_{n_0} = \left\{u_1 \in \mathbf{R}^{m_1} \mid \alpha^* + \frac{u_1}{n_0^{\frac{1}{qG}}} \in \Theta_\alpha\right\}$ and $\mathbb{V}^{(1)}_{n_0}(r) = \{u_1 \in \mathbb{U}^{(1)}_{n_0} \mid r \leq |u_1|\}$. For $u_1 \in \mathbb{U}^{(1)}_{n_0}$, set $\mathbb{Z}^{(1)}_{1,n_0}(u_1;\alpha^*) = \exp\left\{\mathbb{H}^{(1)}_{1,n_0}\left(\alpha^* + \frac{u_1}{n_0^{\frac{1}{qG}}}\right) - \mathbb{H}^{(1)}_{1,n_0}(\alpha^*)\right\}$.

Note that $\epsilon_1 = \epsilon_0/(2(p-1))$. It is shown that for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}[|\Delta^{(1)}_{1,n_0}(\alpha^*)|^M] < \infty, \tag{5}$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\sup_{\alpha\in\Theta_1} n_0^{\epsilon_1}|\mathbb{Y}^{(1)}_{1,n_0}(\alpha) - \mathbb{Y}^{(1)}(\alpha)|\right)^M\right] < \infty, \tag{6}$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}[(n_0^{\epsilon_1}|\Gamma^{(1)}_{1,n_0}(\alpha^*) - \Gamma_1(\alpha^*)|)^M] < \infty. \tag{7}$$

Proof of (5). One has a decomposition $\partial_\alpha U^{(1)}_{1,n_0}(\alpha)[u_1] = M^{(1)}_{1,n_0}(\alpha)[u_1] + R^{(1)}_{1,n_0}(\alpha)[u_1]$, where

$$\begin{aligned}
M^{(1)}_{1,n_0}(\alpha)[u_1] &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[h_n^{-1}\left\{\partial_\alpha A^{-1}_{i-1}(\alpha)[u_1]\right\}[(\Delta X_i)^{\otimes 2} - E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}^n_{i-1}]]\right],\\
R^{(1)}_{1,n_0}(\alpha)[u_1] &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[h_n^{-1}\left\{\partial_\alpha A^{-1}_{i-1}(\alpha)[u_1]\right\}[E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}^n_{i-1}] - h_n A_{i-1}(\alpha)]\right].
\end{aligned}$$

It follows from the standard estimates together with the Burkholder inequality that for all $M > 1$,

$$\begin{aligned}
E_{\theta^*}\left[|\frac{1}{\sqrt{n_0}}M^{(1)}_{1,n_0}(\alpha^*)|^M\right] &\leq \frac{1}{n_0^{M/2}}E_{\theta^*}\left[\sum_{i=1}^{n_0}\left(h_n^{-1}\left\{\partial_\alpha A^{-1}_{i-1}(\alpha)[u_1]\right\}[(\Delta X_i)^{\otimes 2} - E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}^n_{i-1}]]\right)^2\right]^{M/2}\\
&\leq \frac{1}{n_0}E_{\theta^*}\left[\sum_{i=1}^{n_0}\left(h_n^{-1}\left\{\partial_\alpha A^{-1}_{i-1}(\alpha)[u_1]\right\}[(\Delta X_i)^{\otimes 2} - E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}^n_{i-1}]]\right)^M\right]\\
&\leq C,\\
E_{\theta^*}\left[|\frac{1}{\sqrt{n_0}}R^{(1)}_{1,n_0}(\alpha^*)|^M\right] &\leq C(\sqrt{n_0}h_n)^M.
\end{aligned}$$

Noting that $1 - 1/(qG) \geq 1/2$ and

$$\frac{n_0 h_n}{n_0^{1-\frac{1}{qG}}} = n_0^{\frac{1}{qG}}h_n \leq n^{\frac{1}{q}}h_n = (nh_n^q)^{\frac{1}{q}} \leq (nh_n^p)^{\frac{1}{q}},$$

one has that as $nh_n^p \to 0$, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|\frac{1}{n_0^{1-\frac{1}{qG}}}\partial_\alpha U_{1,n_0}^{(1)}(\alpha^*)\right|^M\right] < \infty$, which completes the proof of (5).

Proof of (6). A decomposition is given by $U_{1,n_0}^{(1)}(\alpha) - U_{1,n_0}^{(1)}(\alpha^*) = \mathcal{M}_{1,n_0}^{(1)}(\alpha) + \mathcal{R}_{1,n_0}^{(1)}(\alpha) + \bar{\mathbb{Y}}_{1,n_0}^{(1)}(\alpha)$, where

$$\begin{aligned}
\mathcal{M}_{1,n_0}^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[h_n^{-1}\left\{(A_{i-1}^{-1}(\alpha) - A_{i-1}^{-1}(\alpha^*))\right\}[(\Delta X_i)^{\otimes 2} - E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}_{i-1}^n]]\right], \\
\mathcal{R}_{1,n_0}^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[h_n^{-1}\left\{(A_{i-1}^{-1}(\alpha) - A_{i-1}^{-1}(\alpha^*))\right\}[E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}_{i-1}^n] - h_n A_{i-1}(\alpha)]\right], \\
\bar{\mathbb{Y}}_{1,n_0}^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[\left\{A_{i-1}^{-1}(\alpha) - A_{i-1}^{-1}(\alpha^*)\right\}[A_{i-1}(\alpha)] + \log\frac{\det A_{i-1}(\alpha)}{\det A_{i-1}(\alpha^*)}\right].
\end{aligned}$$

The standard estimates yield that for $\epsilon_1 = \epsilon_0/(2(p-1)) \in (0, 1/(2p))$,

$$\sup_\alpha E_{\theta^*}\left[\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] \le C({n_0}^{\epsilon_1}\frac{1}{\sqrt{n_0}})^M, \qquad \sup_\alpha E_{\theta^*}\left[\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\mathcal{R}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] \le C({n_0}^{\epsilon_1}h_n)^M.$$

In the same way,

$$\sup_{n\in\mathbf{N}}\sup_\alpha E_{\theta^*}\left[\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\partial_\alpha\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] < \infty, \qquad \sup_{n\in\mathbf{N}}\sup_\alpha E_{\theta^*}\left[\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\partial_\alpha\mathcal{R}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] < \infty.$$

The Sobolev inequality (Lemma 4.65 of Adams and Fournier (2003)) implies that for $M > m_1$,

$$\begin{aligned}
& E_{\theta^*}\left[\left({n_0}^{\epsilon_1}\sup_\alpha|\frac{1}{n_0}\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] \\
\le\ & E_{\theta^*}\left[C\int_{\Theta_\alpha}\left\{\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M + \left({n_0}^{\epsilon_1}|\frac{1}{n_0}\partial_\alpha\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right\}d\alpha\right] \\
\le\ & C_{\Theta_\alpha}\left\{\sup_\alpha E_{\theta^*}\left[\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] + \sup_\alpha E_{\theta^*}\left[\left({n_0}^{\epsilon_1}|\frac{1}{n_0}\partial_\alpha\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right]\right\}.
\end{aligned}$$

Therefore, for all $M > 0$, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left({n_0}^{\epsilon_1}\sup_\alpha|\frac{1}{n_0}\mathcal{M}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] < \infty$, and in a similar way, for all $M > 0$, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left({n_0}^{\epsilon_1}\sup_\alpha|\frac{1}{n_0}\mathcal{R}_{1,n_0}^{(1)}(\alpha)|\right)^M\right] < \infty$. It follows from Proposition 2.1 that for all $M > 0$, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left({n_0}^{\epsilon_1}\sup_\alpha|\frac{1}{n_0}\bar{\mathbb{Y}}_{1,n_0}^{(1)}(\alpha) - \mathbb{Y}^{(1)}(\alpha)|\right)^M\right] < \infty$, which completes the proof of (6).

Proof of (7). We obtain a decomposition $\partial_\alpha^2 U_{1,n_0}^{(1)}(\alpha)[u_1, u_1] = \mathbf{M}_{1,n_0}^{(1)}(\alpha)[u_1, u_1] +$

$\mathbf{R}^{(1)}_{1,n_0}(\alpha)[u_1,u_1] - \bar{\Gamma}^{(1)}_{1,n_0}(\alpha)[u_1,u_1]$, where

$$\begin{aligned}
\mathbf{M}^{(1)}_{1,n_0}(\alpha)[u_1,u_1] &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[h_n^{-1}\left\{\partial_\alpha^2 A_{i-1}^{-1}(\alpha)[u_1,u_1]\right\}\left[(\Delta X_i)^{\otimes 2} - E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}^n_{i-1}]\right]\right],\\
\mathbf{R}^{(1)}_{1,n_0}(\alpha)[u_1,u_1] &= -\frac{1}{2}\sum_{i=1}^{n_0}\left[h_n^{-1}\left\{\partial_\alpha^2 A_{i-1}^{-1}(\alpha)[u_1,u_1]\right\}\left[E_{\theta^*}[(\Delta X_i)^{\otimes 2}|\mathcal{G}^n_{i-1}] - h_n A_{i-1}(\alpha)\right]\right],\\
\bar{\Gamma}^{(1)}_{1,n_0}(\alpha)[u_1,u_1] &= \frac{1}{2}\sum_{i=1}^{n_0}\left[\left\{\partial_\alpha^2 A_{i-1}^{-1}(\alpha)[u_1,u_1]\right\}[A_{i-1}(\alpha)] + \partial_\alpha^2 \log\det A_{i-1}(\alpha)[u_1,u_1]\right].
\end{aligned}$$

Similarly to the proof of (6), one has that for all $M>0$,
$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(n_0{}^{\epsilon_1}|\frac{1}{n_0}\bar{\Gamma}^{(1)}_{1,n_0}(\alpha^*) - \Gamma_1(\alpha^*)|\right)^M\right] < \infty$,

$$E_{\theta^*}\left[\left(n_0{}^{\epsilon_1}|\frac{1}{n_0}\mathbf{M}^{(1)}_{1,n_0}(\alpha^*)|\right)^M\right] \le C\left(\frac{n_0{}^{\epsilon_1}}{\sqrt{n_0}}\right)^M, \qquad E_{\theta^*}\left[\left(n_0{}^{\epsilon_1}|\frac{1}{n_0}\mathbf{R}^{(1)}_{1,n_0}(\alpha^*)|\right)^M\right] \le C(n_0{}^{\epsilon_1}h_n)^M,$$

which completes the proof of (7).

Moreover, we obtain that for all $M>0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(n_0^{-1}\sup_{\alpha\in\Theta_\alpha}|\partial_\alpha^3 U^{(1)}_{1,n_0}(\alpha)|\right)^M\right] < \infty. \tag{8}$$

Theorem 3 of Yoshida (2011) together with (5)-(8) implies that for any $L>0$, there exists $C_L>0$ such that for all $n\in\mathbf{N}$ and $r>0$,

$$P_{\theta^*}\left[\sup_{u_1\in\mathbb{V}^{(1)}_{n_0}(r)} \mathbb{Z}^{(1)}_{1,n_0}(u_1;\alpha^*) \ge e^{-r}\right] \le \frac{C_L}{r^L}. \tag{9}$$

Note that

$$n_0^{\frac{1}{qG}}(\tilde{\alpha}^{(1)}_{1,n_0} - \alpha^*) = \frac{\int_{\mathbb{U}^{(1)}_{n_0}} u_1 \mathbb{Z}^{(1)}_{1,n_0}(u_1;\alpha^*)\pi_1\left(\alpha^* + \frac{u_1}{n_0^{\frac{1}{qG}}}\right)du_1}{\int_{\mathbb{U}^{(1)}_{n_0}} \mathbb{Z}^{(1)}_{1,n_0}(u_1;\alpha^*)\pi_1\left(\alpha^* + \frac{u_1}{n_0^{\frac{1}{qG}}}\right)du_1}.$$

We can show that

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(1)}_{n_0}} \mathbb{Z}^{(1)}_{1,n_0}(u_1;\alpha^*)du_1\right)^{-1}\right] < \infty. \tag{10}$$

In an analogous way to the proof of Theorem 8 of Yoshida (2011), it follows from (9) and (10) that for all $M>0$, as $nh_n^p \to 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|n_0^{\frac{1}{qG}}(\tilde{\alpha}^{(1)}_{1,n_0} - \alpha^*)\right|^M\right] < \infty. \tag{11}$$

Next, we will prove the case of the type 1 Bayes estimator $\tilde{\beta}^{(2)}_{1,n_0}$. Set $U^{(2)}_{1,n_0}(\alpha,\beta) = U^{(2)}_{1,n_0}(\beta \mid \alpha)$ and

$$\begin{aligned}
\mathbb{H}^{(2)}_{1,n_0}(\alpha,\beta) &= \frac{1}{(n_0h_n)^{1-\frac{2}{qG}}} U^{(2)}_{1,n_0}(\alpha,\beta), \\
\mathbb{Y}^{(2)}_{1,n_0}(\beta) &= \frac{1}{(n_0h_n)^{\frac{2}{qG}}} \left\{ \mathbb{H}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta) - \mathbb{H}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*) \right\} \\
&= \frac{1}{n_0h_n} \left\{ U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta) - U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*) \right\}, \\
\Delta^{(2)}_{1,n_0}(\beta^*)[u_2] &= \frac{1}{(n_0h_n)^{\frac{1}{qG}}} \partial_\beta \mathbb{H}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)[u_2] = \frac{1}{(n_0h_n)^{1-\frac{1}{qG}}} \partial_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)[u_2],
\end{aligned}$$

$$\begin{aligned}
\Gamma^{(2)}_{1,n_0}(\beta^*)[u_2,u_2] &= -\frac{1}{(n_0h_n)^{\frac{2}{qG}}} \partial^2_\beta \mathbb{H}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)[u_2,u_2] = -\frac{1}{n_0h_n} \partial^2_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)[u_2,u_2], \\
\Gamma^{(2)}_1(\theta^*)[u_2,u_2] &= \int_{\mathbf{R}^d} A(x,\alpha^*)^{-1}[\partial_\beta b(x,\beta^*)[u_2], \partial_\beta b(x,\beta^*)[u_2]] \mu_{\theta^*}(dx)
\end{aligned}$$

for $u_2 \in \mathbf{R}^{m_2}$. Let $\mathbb{U}^{(2)}_{n_0} = \left\{ u_2 \in \mathbf{R}^{m_2} \mid \beta^* + \frac{u_2}{(n_0h_n)^{\frac{1}{qG}}} \in \Theta_\beta \right\}$ and $\mathbb{V}^{(2)}_{n_0}(r) = \{u_2 \in \mathbb{U}^{(2)}_{n_0} \mid r \le |u_2|\}$.

For $u_2 \in \mathbb{U}^{(2)}_{n_0}$, set $\mathbb{Z}^{(2)}_{1,n_0}(u_2;\beta^*) = \exp\left\{ \mathbb{H}^{(2)}_{1,n_0}\left(\tilde{\alpha}^{(1)}_{1,n_0}, \beta^* + \frac{u_2}{(n_0h_n)^{\frac{1}{qG}}} \right) - \mathbb{H}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*) \right\}$.

It follows that for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[|\Delta^{(2)}_{1,n_0}(\beta^*)|^M \right] < \infty, \tag{12}$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\sup_{\beta\in\Theta_2} (n_0h_n)^{\epsilon_1} |\mathbb{Y}^{(2)}_{1,n_0}(\beta) - \mathbb{Y}^{(2)}(\beta)| \right)^M \right] < \infty, \tag{13}$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1} |\Gamma^{(2)}_{1,n_0}(\beta^*) - \Gamma^{(2)}_1(\alpha^*,\beta^*)| \right)^M \right] < \infty. \tag{14}$$

Proof of (12). We have that

$$\begin{aligned}
\Delta^{(2)}_{1,n_0}(\beta^*)[u_2] &= \frac{1}{(n_0h_n)^{1-\frac{1}{qG}}} \partial_\beta U^{(2)}_{1,n_0}(\alpha^*,\beta^*)[u_2] \\
&\quad + \frac{1}{(n_0h_n)^{1-\frac{1}{qG}}} \frac{1}{n_0^{\frac{1}{qG}}} \int_0^1 \partial_\alpha \partial_\beta U^{(2)}_{1,n_0}(\alpha^* + t(\tilde{\alpha}^{(1)}_{1,n_0} - \alpha^*), \beta^*) dt [u_2, n_0^{\frac{1}{qG}}(\tilde{\alpha}^{(1)}_{1,n_0} - \alpha^*)].
\end{aligned}$$

One has a decomposition $\partial_\beta U^{(2)}_{1,n_0}(\alpha,\beta)[u_2] = M^{(2)}_{1,n_0}(\theta)[u_2] + R^{(2)}_{1,n_0}(\theta)[u_2]$, where

$$\begin{aligned}
M^{(2)}_{1,n_0}(\theta)[u_2] &= \sum_{i=1}^{n_0} A^{-1}_{i-1}(\alpha)[\partial_\beta b_{i-1}(\beta)[u_2], X_{t^n_i} - E_{\theta^*}[X_{t^n_i} | \mathcal{G}^n_{i-1}]], \\
R^{(2)}_{1,n_0}(\theta)[u_2] &= \sum_{i=1}^{n_0} A^{-1}_{i-1}(\alpha)[\partial_\beta b_{i-1}(\beta)[u_2], E_{\theta^*}[X_{t^n_i} | \mathcal{G}^n_{i-1}] - X_{i-1} - h_n b_{i-1}(\beta)].
\end{aligned}$$

By the Burkholder inequality, one has that for all $M > 1$,

$$E_{\theta^*}\left[|\frac{1}{\sqrt{n_0 h_n}} M^{(2)}_{1,n_0}(\theta^*)|^M\right] \quad \leq \quad C. \tag{15}$$

Moreover, one has that for all $M > 1$,

$$E_{\theta^*}\left[|\frac{1}{\sqrt{n_0 h_n}} R^{(2)}_{1,n_0}(\theta^*)|^M\right] \quad \leq \quad C\left(n_0 h_n^3\right)^{M/2}. \tag{16}$$

Hence, as $nh_n^p \to 0$, for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[|\frac{1}{(n_0 h_n)^{1-\frac{1}{qG}}} \partial_\beta U^{(2)}_{1,n_0}(\alpha^*, \beta^*)|^M\right] < \infty. \tag{17}$$

Since $1 - 1/(qG) \geq 1/2$, we obtain that $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[|\frac{1}{(n_0 h_n)^{1-\frac{1}{qG}}} \partial_\beta U^{(2)}_{1,n_0}(\alpha^*, \beta^*)|^M\right] < \infty$. Moreover, for all $M > 0$, $\sup_{n\in\mathbf{N}} \sup_\alpha E_{\theta^*}\left[|\frac{1}{n_0 h_n} \partial_\alpha \partial_\beta U^{(2)}_{1,n_0}(\alpha, \beta^*)|^M\right] < \infty$, $\sup_{n\in\mathbf{N}} \sup_\alpha E_{\theta^*}\left[|\frac{1}{n_0 h_n} \partial^2_\alpha \partial_\beta U^{(2)}_{1,n_0}(\alpha, \beta^*)|^M\right] < \infty$. The Sobolev inequality implies that

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\sup_\alpha |\frac{1}{n_0 h_n} \partial_\alpha \partial_\beta U^{(2)}_{1,n_0}(\alpha, \beta^*)|\right)^M\right] < \infty. \tag{18}$$

Noting that $\frac{1}{(n_0 h_n)^{1-\frac{1}{qG}} n_0^{\frac{1}{qG}}} = \frac{h_n^{\frac{1}{qG}}}{n_0 h_n}$, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\sup_\alpha |\frac{1}{(n_0 h_n)^{1-\frac{1}{qG}} n_0^{\frac{1}{qG}}} \partial_\alpha \partial_\beta U^{(2)}_{1,n_0}(\alpha, \beta^*)|\right)^M\right] < \infty$ as $nh_n^p \to 0$, which completes the proof of (12).

Proof of (13). A decomposition is given by $U^{(2)}_{1,n_0}(\alpha,\beta) - U^{(2)}_{1,n_0}(\alpha,\beta^*) = \mathcal{M}^{(2)}_{1,n_0}(\alpha,\beta) + \mathcal{R}^{(2)}_{1,n_0}(\alpha,\beta) + \bar{\mathbb{Y}}^{(2)}_{1,n_0}(\alpha,\beta)$, where

$$\begin{aligned}
\mathcal{M}^{(2)}_{1,n_0}(\alpha,\beta) &= \sum_{i=1}^{n_0} A^{-1}_{i-1}(\alpha)[b_{i-1}(\beta) - b_{i-1}(\beta^*), (X_{t_i^n} - E_{\theta^*}[X_{t_i^n}|\mathcal{G}^n_{i-1}])], \\
\mathcal{R}^{(2)}_{1,n_0}(\alpha,\beta) &= \sum_{i=1}^{n_0} A^{-1}_{i-1}(\alpha)[b_{i-1}(\beta) - b_{i-1}(\beta^*), E_{\theta^*}[X_{t_i^n}|\mathcal{G}^n_{i-1}] - X_{t_{i-1}^n} - h_n b_{i-1}(\beta^*))], \\
\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\alpha,\beta) &= -\frac{h_n}{2}\sum_{i=1}^{n_0} A^{-1}_{i-1}(\alpha)[(b_{i-1}(\beta) - b_{i-1}(\beta^*))^{\otimes 2}].
\end{aligned}$$

It follows that for $\epsilon_1 \in (0, 1/(2p))$, $\sup_\theta E_{\theta^*}\left[\left((n_0 h_n)^{\epsilon_1} |\frac{1}{(n_0 h_n)} \mathcal{M}^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right] \leq C\left((n_0 h_n)^{\epsilon_1} \frac{1}{\sqrt{n_0 h_n}}\right)^M$ and $\sup_\theta E_{\theta^*}\left[\left((n_0 h_n)^{\epsilon_1} |\frac{1}{(n_0 h_n)} \mathcal{R}^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right] \leq C((n_0 h_n)^{\epsilon_1} h_n)^M$ for all $M > 1$. In an analogous argument, we obtain that for all $M > 1$, $\sup_{n\in\mathbf{N}} \sup_\theta E_{\theta^*}\left[\left((n_0 h_n)^{\epsilon_1} |\frac{1}{(n_0 h_n)} \partial_\theta \mathcal{M}^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right] < \infty$ and

$\sup_{n\in\mathbf{N}}\sup_\theta E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}|\frac{1}{(n_0h_n)}\partial_\theta\mathcal{R}^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right] < \infty$. Hence, using the Sobolev inequality, we have that for all $M > m_1+m_2$, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\theta|\frac{1}{(n_0h_n)}\mathcal{M}^{(2)}_{1,n_0}(\theta_1,\beta)|\right)^M\right] < \infty$ and $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\theta|\frac{1}{(n_0h_n)}\mathcal{R}^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right] < \infty$. Noting that

$$\begin{aligned}
&\frac{1}{n_0h_n}\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta)-\mathbb{Y}^{(2)}(\beta)\\
=\;&\frac{1}{n_0^{1/qG}}\frac{1}{n_0h_n}\int_0^1\partial_\alpha\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\alpha^*+t(\tilde{\alpha}^{(1)}_{1,n_0}-\alpha^*),\beta)dt[n_0^{1/(qG)}(\tilde{\alpha}^{(1)}_{1,n_0}-\alpha^*)]+\frac{1}{n_0h_n}\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\alpha^*,\beta)-\mathbb{Y}^{(2)}(\beta),
\end{aligned}$$

we have that for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\beta|\frac{1}{n_0h_n}\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\alpha^*,\beta)-\mathbb{Y}^{(2)}(\beta)|\right)^M\right]<\infty,$$
$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\theta\frac{1}{n_0^{1/(qG)}}|\frac{1}{n_0h_n}\partial_\alpha\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right]<\infty,$$

where the last estimate is derived from the fact that $\epsilon_1=\epsilon_0/(2(p-1))<1/(2p)<1/(qG)$. Therefore, $\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\beta|\frac{1}{n_0h_n}\bar{\mathbb{Y}}^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta)-\mathbb{Y}^{(2)}(\beta)|\right)^M\right]<\infty$ for all $M>0$, which completes the proof of (13).

For the proof of (14), we obtain a decomposition

$$\partial^2_\beta U^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]=\mathbf{M}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]+\mathbf{R}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]-\bar{\Gamma}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2],$$

where $\mathbf{M}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]=\partial^2_\beta\mathcal{M}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]$, $\mathbf{R}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]=\partial^2_\beta\mathcal{R}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]$,

$$\bar{\Gamma}^{(2)}_{1,n_0}(\alpha,\beta)[u_2,u_2]\;=\;h_n\sum_{i=1}^{n_0}A^{-1}_{i-1}(\alpha)[\partial_\beta b_{i-1}(\beta)[u_2],\partial_\beta b_{i-1}(\beta)[u_2]].$$

Since $\sup_\alpha E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}|\frac{1}{n_0h_n}\mathbf{M}^{(2)}_{1,n_0}(\alpha,\beta^*)|\right)^M+\left((n_0h_n)^{\epsilon_1}|\frac{1}{n_0h_n}\partial_\alpha\mathbf{M}^{(2)}_{1,n_0}(\alpha,\beta^*)|\right)^M\right]\leq C\left(\frac{(n_0h_n)^{\epsilon_1}}{\sqrt{n_0h_n}}\right)^M$ and $\sup_\alpha E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}|\frac{1}{n_0h_n}\mathbf{R}^{(2)}_{1,n_0}(\alpha,\beta^*)|\right)^M+\left((n_0h_n)^{\epsilon_1}|\frac{1}{n_0h_n}\partial_\alpha\mathbf{R}^{(2)}_{1,n_0}(\alpha,\beta^*)|\right)^M\right]\leq C((n_0h_n)^{\epsilon_1}h_n)^M$, it follows from Sobolev's inequality that

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\alpha|\frac{1}{n_0h_n}\mathbf{M}^{(2)}_{1,n_0}(\alpha,\beta^*)|\right)^M\right]<\infty,\quad \sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\alpha|\frac{1}{n_0h_n}\mathbf{R}^{(2)}_{1,n_0}(\alpha,\beta^*)|\right)^M\right]<\infty.$$

One has that

$$\begin{aligned}
&\frac{1}{n_0h_n}\bar{\Gamma}^{(2)}_{n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)-\Gamma_2(\alpha^*,\beta^*)\\
=\ &\frac{1}{n_0^{1/(qG)}}\frac{1}{n_0h_n}\int_0^1\partial_\alpha\bar{\Gamma}^{(2)}_{n_0}(\alpha^*+t(\tilde{\alpha}^{(1)}_{1,n_0}-\alpha^*),\beta^*)dt[n_0^{1/(qG)}(\tilde{\alpha}^{(1)}_{1,n_0}-\alpha^*)]\\
&+\frac{1}{n_0h_n}\bar{\Gamma}^{(2)}_{n_0}(\alpha^*,\beta^*)-\Gamma_2(\alpha^*,\beta^*),
\end{aligned}$$

and it is shown that for all $M>0$,

$$\sup_{n\in\mathbf{N}}E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}\sup_\alpha\frac{1}{n_0^{1/qG}}|\frac{1}{n_0h_n}\partial_\alpha\bar{\Gamma}^{(2)}_{n_0}(\alpha,\beta^*)|\right)^M+\left((n_0h_n)^{\epsilon_1}|\frac{1}{n_0h_n}\bar{\Gamma}^{(2)}_{n_0}(\alpha^*,\beta^*)-\Gamma_2(\alpha^*,\beta^*)|\right)^M\right]<\infty.$$

By (9) and the above estimates, $\sup_{n\in\mathbf{N}}E_{\theta^*}\left[\left((n_0h_n)^{\epsilon_1}|\frac{1}{n_0h_n}\bar{\Gamma}^{(2)}_{n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)-\Gamma_2(\alpha^*,\beta^*)|\right)^M\right]<\infty$, which completes the proof of (14).

Furthermore, one can show that for all $M>0$,

$$\sup_{n\in\mathbf{N}}E_{\theta^*}\left[\left(\frac{1}{n_0h_n}\sup_{\theta\in\Theta}|\partial^3_\beta U^{(2)}_{1,n_0}(\alpha,\beta)|\right)^M\right]<\infty. \tag{19}$$

By (12)-(14), (19) and Theorem 3 of Yoshida (2011), one has that for any $L>0$, there exists $C_L>0$ such that for all $n\in\mathbf{N}$ and $r>0$,

$$P_{\theta^*}\left[\sup_{u_2\in\mathbb{V}^{(2)}_{n_0}(r)}\mathbb{Z}^{(2)}_{1,n_0}(u_2;\beta^*)\geq e^{-r}\right]\leq\frac{C_L}{r^L}. \tag{20}$$

Note that

$$(n_0h_n)^{\frac{1}{qG}}(\tilde{\beta}^{(2)}_{1,n_0}-\beta^*)=\frac{\int_{\mathbb{U}^{(2)}_{n_0}}u_2\mathbb{Z}^{(2)}_{1,n_0}(u_2;\beta^*)\pi_2\left(\beta^*+\frac{u_2}{(n_0h_n)^{\frac{1}{qG}}}\right)du_2}{\int_{\mathbb{U}^{(2)}_{n_0}}\mathbb{Z}^{(2)}_{1,n_0}(u_2;\beta^*)\pi_2\left(\beta^*+\frac{u_2}{(n_0h_n)^{\frac{1}{qG}}}\right)du_2}.$$

Furthermore,

$$\begin{aligned}
K^{(2)}_{1,n_0}(u_2)\ &=\ \log\mathbb{Z}^{(2)}_{1,n_0}(u_2;\beta^*)\\
&=\ \frac{1}{(n_0h_n)^{1-\frac{2}{qG}}}\left\{\partial_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)[u_2]\frac{1}{(n_0h_n)^{\frac{1}{qG}}}+\frac{1}{2}\partial^2_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)[u_2^{\otimes 2}]\frac{1}{(n_0h_n)^{\frac{2}{qG}}}\right.\\
&\quad\left.+\frac{1}{2}\int_0^1(1-t)^2\partial^3_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*+\frac{tu_2}{(n_0h_n)^{\frac{1}{qG}}})dt[u_2^{\otimes 3}]\frac{1}{(n_0h_n)^{\frac{3}{qG}}}\right\}.
\end{aligned}$$

By (12), (18) and the estimate in the proof of (14), for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|\frac{1}{(n_0h_n)^{1-\frac{1}{qG}}}\partial_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)\right|^M\right] < \infty,$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|\frac{1}{n_0h_n}\partial^2_\beta U^{(2)}_{1,n_0}(\tilde{\alpha}^{(1)}_{1,n_0},\beta^*)\right|^M\right] < \infty,$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\sup_{\theta\in\Theta}\left|\frac{1}{n_0h_n}\partial^3_\beta U^{(2)}_{1,n_0}(\theta)\right|^M\right] < \infty.$$

Hence, for some $M > m_2$, $\delta > 0$ and $C_0 > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}[|K^{(2)}_{1,n_0}(u_2)|^M] \le C_0|u_2|^M$$

for all $u_2 \in \mathbb{U}^{(2)}_{n_0}(\delta) := \{u_2 \in \mathbb{U}^{(2)}_{n_0} \mid |u_2| \le \delta\}$. By Lemma 2 of Yoshida (2011),

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(2)}_{n_0}}\mathbb{Z}^{(2)}_{1,n_0}(u_2;\beta^*)du_2\right)^{-1}\right] < \infty. \tag{21}$$

In the same manner as the proof of (11), it follows from (20) and (21) that for all $M > 0$, as $nh_n^p \to 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left|(n_0h_n)^{\frac{1}{qG}}(\tilde{\beta}^{(2)}_{1,n_0}-\beta^*)\right|^M\right] < \infty.$$

This completes the proof of the case of the type 1 Bayes estimator.

In a similar way to the proof of the type 1 Bayes estimators, we can show the moment estimates of the type 2, type 3 and type 4 Bayes estimators, see Kaino et al. (2017). This completes the proof.

Proof of Theorem 3.1. In the analogous way to the proofs of Theorem 1 of Kamatani and Uchida (2015) and Theorem 3 of Uchida and Yoshida (2012), we can prove the result. For details, see Kaino et al. (2017).

Acknowledgement

The authors wish to thank two anonymous referees for their valuable comments. Masayuki Uchida's research was partially supported by JST CREST, JSPS KAKENHI Grant Numbers JP24300107, JP17H01100 and Cooperative Research Program of the Institute of Statistical Mathematics.

References

Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev spaces.* Second edition. Pure and Applied Mathematics (Amsterdam), 140. Elsevier/Academic Press, Amsterdam.

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

Eguchi, S. and Masuda, H. (2016). Schwarz type model comparison for LAQ models. To appear in Bernoulli.

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

Fujii, T. and Uchida, M. (2014). AIC type statistics for discretely observed ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, **17**, 267–282.

Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **38**, 711–737.

Kaino, Y., Uchida, M. and Yoshida, Y. (2017). Hybrid type estimators for a diffusion process with ergodicity from reduced data. Preprint.
(http://www.sigmath.es.osaka-u.ac.jp/~uchida/Paper/ergodic_reduced_KUY.pdf)

Kamatani, K. (2014). Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. arXiv:1412.6231. To appear in Bernoulli.

Kamatani, K. and Uchida, M. (2015). Hybrid multi-step estimators for ergodic diffusion processes from discrete observations. *Statistical Inference for Stochastic Processes*, **18**, 177-204.

Kamatani, K., Nogita, A. and Uchida, M. (2016). Hybrid multi-step estimation of the volatility for stochastic regression models. *Bulletin of Informatics and Cybernetics*, **48**, 19-35.

Kessler, M. (1995). Estimation des paramètres d'une diffusion par des contrastes corrigés. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, **320**, 359–362.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, **24**, 211–229.

Kutoyants, Yu. A. (1984). Parameter estimation for stochastic processes. Prakasa Rao, B.L.S. (ed.) Heldermann, Berlin.

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

Kutoyants, Yu. A. (2017). On the multi-step MLE-process for ergodic diffusion. *Stochastic Processes and their Applications*, **127**, 2243–2261.

Nomura, R. and Uchida, M. (2016). Adaptive Bayes estimators and hybrid estimators for small diffusion processes based on sampled data. *Journal of The Japan Statistical Society*, **46**, 129–154.

Pardoux, E. and Veretennikov, A.Y. (2001). On the Poisson equation and diffusion approximation 1. *Annals of Probability*, **29**, 1061–1085.

Prakasa Rao, B. L. S. (1983). Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Math. Operationsforsch. Statist. Ser. Statist.*, **14**, 195–209.

Prakasa Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. *Contemporary Mathematics*, **80**, 249–284. Amer. Math. Soc., Providence, RI.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods.*, 3rd ed. Springer Verlag, New York.

Uchida, M. (2010). Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Annals of the Institute of Statistical Mathematics*, **62**, 161–187.

Uchida, M. and Yoshida, N. (2001). Information criteria in model selection for mixing processes. *Statistical Inference for Stochastic Processes*, **4**, 73–98.

Uchida, M. and Yoshida, N. (2011). Estimation for misspecified ergodic diffusion processes from discrete observations. *European Series in Applied and Industrial Mathematics: Probability and Statistics*, **15**, 270–290.

Uchida, M. and Yoshida, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Processes and their Applications*, **122**, 2885–2924.

Uchida, M. and Yoshida, N. (2014). Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. *Statistical Inference for Stochastic Processes*, **17**, 181-219.

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

Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, **63**, 431–479.

Received September 7, 2017
Revised October 26, 2017