HYBRID MULTI-STEP ESTIMATION OF THE VOLATILITY FOR STOCHASTIC REGRESSION MODELS

Kamatani, Kengo
Graduate School of Engineering Science and MMDS, Osaka University

Nogita, Akihiro
MMDS, Osaka University

Uchida, Masayuki
Graduate School of Engineering Science and MMDS, Osaka University

https://doi.org/10.5109/2203025

出版情報：Bulletin of informatics and cybernetics. 48, pp.19-35, 2016-12. 統計科学研究会
バージョン：
権利関係：

Bulletin of Informatics and Cybernetics, Vol. 48, 2016

HYBRID MULTI-STEP ESTIMATION OF THE VOLATILITY FOR STOCHASTIC REGRESSION MODELS

By

Kengo Kamatani[*] **Akihiro Nogita**[†] and **Masayuki Uchida**[‡]

Abstract

We deal with an estimation problem of a volatility parameter for stochastic regression models based on high frequency data. Hybrid multi-step estimators are proposed and their asymptotic properties, including convergence of moments, are obtained.

Key Words and Phrases: Asymptotic mixed normality, Bayes type estimator, convergence of moments, diffusion process, discrete time observations, maximum likelihood type estimator.
We treat parametric estimation of the volatility for a stochastic regression model specified by the stochastic integral equation

$$Y_t \quad = \quad Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s, \quad t \in [0, T], \tag{1}$$

where w is an r-dimensional standard Wiener process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, b and X are progressively measurable processes with values in $\mathbb{R}^m$ and $\mathbb{R}^d$, respectively, Y_0 is an $\mathbb{R}^m$-valued initial condition, σ is an $\mathbb{R}^m \otimes \mathbb{R}^r$-valued function defined on $\mathbb{R}^d \times \Theta$, and Θ is a bounded domain in $\mathbb{R}^p$. The data are discrete observations $\mathbf{Z}_n = (X_{t_k}, Y_{t_k})_{0\le k\le n}$ with $t_k = kh$ for $h = h_n = T/n$. Note that the process b is completely unknown. The asymptotics will be considered in the situation where $n \to \infty$, which means that $\mathbf{Z}_n$ are high frequency data.

Statistical inference for the stochastic differential equation from discrete observations has been developed by many researchers, see for example, Prakasa Rao (1983,1988), Yoshida (1992, 2011), Kessler (1995, 1997), Gobet (2002), Uchida and Yoshida (2011, 2012, 2014) for ergodic diffusions, Shimizu and Yoshida (2006), Shimizu (2006), Ogihara and Yoshida (2011), Masuda (2013a, 2013b) for jump diffusion processes and Lévy type processes, Dohnal (1987), Genon-Catalot and Jacod (1993, 1994), Gobet (2001) for non-ergodic diffusions. Uchida and Yoshida (2013) showed that both the maximum likelihood (ML) and Bayes type estimators have asymptotic normality with convergence of moments for the stochastic regression models. However, from a computational point of view, numerical optimization is necessary to get the ML-type estimators and it is important to choose appropriate initial values for optimization. Furthermore, it takes much time to compute the Bayes type estimators. Although the one-step estimator is very

[*] Graduate School of Engineering Science and MMDS, Osaka University
[†] MMDS, Osaka University
[‡] Graduate School of Engineering Science and MMDS, Osaka University

efficient, it is difficult to implement the one-step estimation for diffusion type processes since it is not easy to find the initial estimator with $\sqrt{n}$-consistency. For the details of one-step estimator, see Lehmann (1999). Recently, Kamatani and Uchida (2015) considered the multi-step estimation of both drift and volatility parameters for ergodic diffusion processes based on sampled data. The method can be applied to parametric inference of non-ergodic diffusion type processes from the high frequency data observed on the fixed interval.

In order to illustrate the multi-step estimator, we consider the case of I.I.D. model, see also Kamatani and Uchida (2015). Let $l_n(\theta)$ be a smooth log-likelihood function for I.I.D. model. Let $q \in (0, 1/2]$ and $J = [-\log_2 q]$, which yields that $2^{J-1}q \le 1/2 < 2^J q$. We assume that for $M > 0$, the initial estimator $\hat{\theta}^{(0)}$ satisfies a moment condition $\sup_n E_{\theta^*}\left[\left|n^q(\hat{\theta}^{(0)} - \theta^*)\right|^M\right] < \infty$. For $k = 1, \ldots, J$, we define the k-step estimator $\hat{\theta}^{(k)}$ as

$$\hat{\theta}^{(k)} = \hat{\theta}^{(k-1)} - \left[\partial_\theta^2 l_n(\hat{\theta}^{(k-1)})\right]^{-1}\left[\partial_\theta l_n(\hat{\theta}^{(k-1)})\right].$$

Since

$$\begin{aligned}
\partial_\theta l_n(\theta^*) &= \partial_\theta l_n(\hat{\theta}^{(k-1)}) + \partial_\theta^2 l_n(\hat{\theta}^{(k-1)})[\theta^* - \hat{\theta}^{(k-1)}] + R_n[(\theta^* - \hat{\theta}^{(k-1)})^{\otimes 2}], \\
R_n &= \int_0^1 (1-t)\partial_\theta^3 l_n(\hat{\theta}^{(k-1)} + t(\theta^* - \hat{\theta}^{(k-1)}))dt,
\end{aligned}$$

one has that

$$\begin{aligned}
&\hat{\theta}^{(k)} \\
= \; &\hat{\theta}^{(k-1)} - \left[\partial_\theta^2 l_n(\hat{\theta}^{(k-1)})\right]^{-1}\left[\partial_\theta l_n(\theta^*) - \partial_\theta^2 l_n(\hat{\theta}^{(k-1)})[\theta^* - \hat{\theta}^{(k-1)}] - R_n[(\theta^* - \hat{\theta}^{(k-1)})^{\otimes 2}]\right] \\
= \; &\hat{\theta}^{(k-1)} - \left[\partial_\theta^2 l_n(\hat{\theta}^{(k-1)})\right]^{-1}[\partial_\theta l_n(\theta^*)] + (\theta^* - \hat{\theta}^{(k-1)}) \\
&+ \left[\partial_\theta^2 l_n(\hat{\theta}^{(k-1)})\right]^{-1} R_n[(\theta^* - \hat{\theta}^{(k-1)})^{\otimes 2}].
\end{aligned}$$

Therefore,

$$\hat{\theta}^{(k)} - \theta^* = -\left[\partial_\theta^2 l_n(\hat{\theta}^{(k-1)})\right]^{-1}[\partial_\theta l_n(\theta^*)] + \left[\partial_\theta^2 l_n(\hat{\theta}^{(k-1)})\right]^{-1} R_n[(\hat{\theta}^{(k-1)} - \theta^*)^{\otimes 2}].$$

In particular, when $2q \le 1/2$,

$$\begin{aligned}
&n^{2q}(\hat{\theta}^{(1)} - \theta^*) \\
= \; &-\left[\frac{1}{n}\partial_\theta^2 l_n(\hat{\theta}^{(0)})\right]^{-1}\left[\frac{n^{2q}}{n}\partial_\theta l_n(\theta^*)\right] + \left[\frac{1}{n}\partial_\theta^2 l_n(\hat{\theta}^{(0)})\right]^{-1}\frac{1}{n}R_n[(n^q(\hat{\theta}^{(0)} - \theta^*))^{\otimes 2}].
\end{aligned}$$

Hence, under some regularity conditions,

$$\sup_n E_{\theta^*}\left[\left|n^q(\hat{\theta}^{(0)} - \theta^*)\right|^M\right] < \infty \quad \Longrightarrow \quad \sup_n E_{\theta^*}\left[\left|n^{2q}(\hat{\theta}^{(1)} - \theta^*)\right|^M\right] < \infty.$$

By obtaining the multi-step estimator recursively, one has that under some regularity conditions,

$$\sup_n E_{\theta^*}\left[\left|n^{2q}(\hat{\theta}^{(1)} - \theta^*)\right|^M\right] < \infty \quad \Longrightarrow \quad \sup_n E_{\theta^*}\left[\left|n^{2^2 q}(\hat{\theta}^{(2)} - \theta^*)\right|^M\right] < \infty,$$

$$\vdots$$

$$\sup_n E_{\theta^*}\left[\left|n^{2^{J-2}q}(\hat{\theta}^{(J-2)}-\theta^*)\right|^M\right]<\infty \quad \Longrightarrow \quad \sup_n E_{\theta^*}\left[\left|n^{2^{J-1}q}(\hat{\theta}^{(J-1)}-\theta^*)\right|^M\right]<\infty,$$
$$\sup_n E_{\theta^*}\left[\left|n^{2^{J-1}q}(\hat{\theta}^{(J-1)}-\theta^*)\right|^M\right]<\infty \quad \Longrightarrow \quad \sup_n E_{\theta^*}\left[\left|\sqrt{n}(\hat{\theta}^{(J)}-\theta^*)\right|^M\right]<\infty$$

and the J-step estimator $\hat{\theta}^{(J)}$ is asymptotically efficient. Here we note that the initial estimator does not have the optimal rate $\sqrt{n}$, but the multi-step estimator has the optimal rate of convergence. By using the similar property to multi-step estimator, Kutoyants (2015) studied one-step and two-step maximum likelihood estimator (MLE)-processes of a drift parameter of an ergodic diffusion process based on the initial estimator obtained from a learning time interval.

Based on the above procedure of Kamatani and Uchida (2015), in this paper, we propose the hybrid multi-step estimator with the initial Bayes type estimator for a stochastic regression model and show that the multi-step estimator has asymptotic mixed normality and convergence of moments. Needless to say, limiting distribution of the estimator is essential for asymptotic statistical decision theory and it is worth mentioning that it is indispensable to show the convergence of moments for estimators in order to validate the information criteria in model selection problems. Among many researches on statistically asymptotic decision theory, we refer the readers to Ibragimov and Has'minskii (1981), Kutoyants (1984, 2004) and Yoshida (2011). Moreover, for model selection for diffusion type processes, see Uchida and Yoshida (2001, 2004, 2006, 2016) and Uchida (2010).

This paper is organized as follows. In Section 2, we state the main results. After the notation and assumptions are stated, the multi-step estimator is proposed and the asymptotic properties, including convergence of moments, are shown. Section 3 gives an example and simulation studies. Section 4 is devoted to the proofs of the results presented in Section 2.

1. Multi-step estimator

Let θ^* denote the true value of θ. We assume that Θ is a bounded domain in $\mathbb{R}^p$ with a locally Lipschitz boundary, which means that Θ has the strong local Lipschitz condition and satisfies Sobolev's inequality, see Adams and Fournier (2003). The convergence in probability and the $\mathcal{F}$-stable convergence in distribution are denoted by $\to^p$ and $\to^{d_s(\mathcal{F})}$, respectively. Set $A^{\otimes 2}=AA^\star$ and $A[B]=\mathrm{Tr}(AB^\star)$ for matrices A and B of the same size. Here $\star$ means the transpose. Set $S(x,\theta)=\sigma(x,\theta)^{\otimes 2}$ and $\Delta_k Y=Y_{t_k}-Y_{t_{k-1}}$. Let $C_\uparrow^{k,l}(\mathbb{R}^d\times\Theta;\mathbb{R}^m)$ be the space of all functions f such that (i) $f(x,\theta)$ is an $\mathbb{R}^m$-valued function on $\mathbb{R}^d\times\Theta$, (ii) $f(x,\theta)$ is continuously differentiable with respect to x up to order k for all θ. (iii) for $|\mathbf{n}|=0,1,\ldots,k$, $\partial_x^{\mathbf{n}}f(x,\theta)$ is continuously differentiable with respect to θ up to order l for all x. Moreover, for $|\nu|=0,1,\ldots,l$ and $|\mathbf{n}|=0,1,\ldots,k$, $\partial_\theta^\nu\partial_x^{\mathbf{n}}f(x,\theta)$ is of polynomial growth in x uniformly in θ. Here $\mathbf{n}=(n_1,\ldots,n_d)$ and $\nu=(\nu_1,\ldots,\nu_p)$ are multi-indices, $p=\dim(\Theta)$, $|\mathbf{n}|=n_1+\ldots+n_d$, $|\nu|=\nu_1+\ldots+\nu_p$, $\partial_x^{\mathbf{n}}=\partial_{x_1}^{n_1}\cdots\partial_{x_d}^{n_d}$, $\partial_{x_i}=\partial/\partial x_i$, and $\partial_\theta^\nu=\partial_{\theta_1}^{\nu_1}\cdots\partial_{\theta_p}^{\nu_p}$, $\partial_{\theta_i}=\partial/\partial\theta_i$. Suppose that σ admits a continuous extension over $\mathbb{R}^d\times\bar{\Theta}$, and also denotes it by σ. For $f\in L^p(P)$, set $||f||_p=(E\left[|f|^p\right])^{1/p}$ for $p>1$.

We make the assumptions as follows.

[A1] **(i)** For every $p > 1$, $\sup_{0\le t\le T} \|b_t\|_p < \infty$.

(ii) $\inf_{x,\theta} \det S(x,\theta) > 0$ and $\sigma \in C^{2,4}_{\uparrow}(\mathbb{R}^d \times \Theta; \mathbb{R}^m \otimes \mathbb{R}^r)$.

[A2] The process X has the following form

$$X_t \quad = \quad X_0 + \int_0^t \tilde{b}_s ds + \int_0^t a_s dw_s + \int_0^t \tilde{a}_s d\tilde{w}_s,$$

where $\tilde{b}$, a and $\tilde{a}$ are progressively measurable processes and take values in $\mathbb{R}^d$, $\mathbb{R}^d \otimes \mathbb{R}^r$ and $\mathbb{R}^d \otimes \mathbb{R}^{r_1}$, respectively, satisfying

$$\|X_0\|_p + \sup_{t\in[0,T]} (\|\tilde{b}_t\|_p + \|a_t\|_p + \|\tilde{a}_t\|_p) < \infty$$

for all $p > 1$, and $\tilde{w}$ is an r_1-dimensional Wiener process independent of w.

The quasi-log likelihood function $\mathbb{H}_n(\theta)$ is given by

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{k=1}^{n} \left\{\log\det S(X_{t_{k-1}},\theta) + h^{-1}S^{-1}(X_{t_{k-1}},\theta)[(\Delta_k Y)^{\otimes 2}]\right\}.$$

Let $\mathbb{Y}_n(\theta) = \frac{1}{n}\{\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta^*)\}$, which converges in probability to

$$\mathbb{Y}(\theta) = -\frac{1}{2T}\int_0^T \left\{\log\left(\frac{\det S(X_t,\theta)}{\det S(X_t,\theta^*)}\right) + \mathrm{Tr}\left(S^{-1}(X_t,\theta)S(X_t,\theta^*) - I_d\right)\right\} dt$$

uniformly in $\theta \in \Theta$ under [A1] and [A2]. Set

$$\chi_0 = \inf_{\theta\neq\theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta-\theta^*|^2}.$$

The following condition is about nondegeneracy of the index χ_0.

[A3] For every $L > 0$, there exists $c_L > 0$ such that $P\left[\chi_0 \le r^{-1}\right] \le \frac{c_L}{r^L}$ for all $r > 0$.

Note that [A3] is the same condition as $1/\chi_0$ has finite moments of all order. For sufficient conditions for [A3], see Uchida and Yoshida (2013).

The assumption on the initial estimators is made as follows.

[B] Let $q \in (0, 1/2]$. $\hat{\theta}_n^{(0)}$ is an initial estimator of θ satisfying that as $n \to \infty$,

$$\sup_n E_{\theta^*}\left[\left|n^q(\hat{\theta}_n^{(0)} - \theta^*)\right|^{M_1}\right] < \infty$$

for all $M_1 > 0$.

We can obtain the initial estimators satisfying $[B]$ as follows. Let $q \in (0, 1/2]$. The initial Bayes type estimator $\tilde{\theta}^{(0)}_{q,n}$ for a prior density $\pi : \Theta \to \mathbb{R}_+$ with respect to the quadratic loss is defined by

$$\tilde{\theta}^{(0)}_{q,n} = \frac{\int_\Theta \theta \exp\left\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\right\}\pi(\theta)d\theta}{\int_\Theta \exp\left\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\right\}\pi(\theta)d\theta}.$$

We assume that π is continuous and $0 < \inf_{\theta\in\Theta}\pi(\theta) \le \sup_{\theta\in\Theta}\pi(\theta) < \infty$.

Let $\mathbb{U}_{q,n} = \left\{u \in \mathbb{R}^p \ ; \ \theta^* + \frac{1}{n^q}u \in \Theta\right\}$ and $\mathbb{V}_{q,n}(r) = \{u \in \mathbb{U}_{q,n} \ ; \ r \le |u|\}$. We define the random field $\mathbb{Z}_{q,n}$ on $\mathbb{U}_{q,n}$ by

$$\mathbb{Z}_{q,n}(u) = \exp\left\{\frac{1}{n^{1-2q}}\mathbb{H}_n\left(\theta^* + \frac{1}{n^q}u\right) - \frac{1}{n^{1-2q}}\mathbb{H}_n(\theta^*)\right\} \tag{2}$$

for $u \in \mathbb{U}_{q,n}$.

Proposition 1.1. *Let $q \in (0, 1/2]$. Assume [A1], [A2] and [A3]. Then, for every $L > 0$, there exists a positive constant C_L such that*

$$P\left[\sup_{u\in\mathbb{V}_{q,n}(r)} \mathbb{Z}_{q,n}(u) \ge e^{-r}\right] \le \frac{C_L}{r^L}$$

for all $r > 0$ and $n \in \mathbf{N}$.

Proposition 1.2. *Let $q \in (0, 1/2]$. Assume [A1], [A2] and [A3]. Then, as $n \to \infty$,*

$$\sup_n E_{\theta^*}\left[\left|n^q(\tilde{\theta}^{(0)}_{q,n} - \theta^*)\right|^M\right] < \infty$$

for all $M > 0$.

We consider the multi-step estimators. Set

$$\begin{aligned} \Gamma_n(\theta) &:= \frac{1}{n}\partial^2_\theta \mathbb{H}_n(\theta), \\ K_n(\theta) &:= \{\Gamma_n(\theta) \text{ is invertible }\}, \\ \bar{\Gamma}_n(\theta) &:= \Gamma_n(\theta)1_{K_n(\theta)} + E_p 1_{K^c_n(\theta)}, \end{aligned}$$

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

Let $J = [-\log_2 q]$ for $q \in (0, 1/2]$. The multi-step estimator $\hat{\theta}^{(J)}_n$ is defined as for $k = 1, \ldots, J$,

$$\hat{\theta}^{(k)}_n = \hat{\theta}^{(k-1)}_n - \bar{\Gamma}^{-1}_n(\hat{\theta}^{(k-1)}_n)\frac{1}{n}\partial_\theta \mathbb{H}_n(\hat{\theta}^{(k-1)}_n).$$

Let $\Gamma(\theta^*) = (\Gamma^{ij}(\theta^*))_{i,j=1,\ldots,p}$ with

$$\Gamma^{ij}(\theta^*) = \frac{1}{2T}\int_0^T \mathrm{Tr}\left((\partial_{\theta_i}S)S^{-1}(\partial_{\theta_j}S)S^{-1}(X_t, \theta^*)\right)dt$$

and let ζ be a p-dimensional standard normal random variable independent of $\Gamma(\theta^*)$.

Lemma 1.3. *Let* $q \in (0, 1/2]$ *and* $J = [-\log_2 q]$. *Assume* [A1], [A2], [A3] *and* [B]. *Then, for* $k = 0, 1, \ldots, J-1$, *as* $n \to \infty$,

$$\sup_n E_{\theta^*}\left[\left|n^{2^k q}(\hat{\theta}_n^{(k)} - \theta^*)\right|^M\right] < \infty$$

for all $M > 0$.

Theorem 1.4. *Let* $q \in (0, 1/2]$ *and* $J = [-\log_2 q]$. *Assume* [A1], [A2], [A3] *and* [B]. *Then, as* $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*) \to^{d_s(\mathcal{F})} \Gamma(\theta^*)^{-1/2}\zeta$$

and

$$E\left[f(\sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*))\right] \to \mathbb{E}\left[f(\Gamma(\theta^*)^{-1/2}\zeta)\right]$$

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

2. Examples and simulations

Consider the one-dimensional diffusion process defined by

$$\begin{cases} dX_t = -(X_t - 1)dt + [\theta_1 + \theta_2\{1 + \sin(\theta_3 X_t)\}]dW_t, & t \in [0, 1], \\ X_0 = 1, \end{cases}$$

where the true value is $\theta^* = (1,\ 4,\ 8)$, the parameter space is $\Theta = [0.01, 20] \times [0, 20] \times [0, 20]$. The data are $(X_{t_i})_{i=0,1,\ldots,n}$ with $t_i = ih$, $h = 1/10000$, $t_n = nh = T = 1$, and the sample size n is 10000.

We do simulations for the maximum likelihood type estimator $\hat{\theta}_{M,n}$ (Genon-Catalot and Jacod (1993)), the Bayes type estimator $\hat{\theta}_{B,n}$ (Uchida and Yoshida (2013)) and the HMS estimator proposed in this paper. The ML type estimator $\hat{\theta}_{M,n}$ is defined by

$$\mathbb{H}_n(\hat{\theta}_{M,n}) = \sup_{\theta\in\Theta} \mathbb{H}_n(\theta).$$

The Bayes type estimator $\hat{\theta}_{B,n}$ with uniform prior is defined as

$$\hat{\theta}_{B,n} := \frac{\int_\Theta \theta \exp(\mathbb{H}_n(\theta))d\theta}{\int_\Theta \exp(\mathbb{H}_n(\theta))d\theta}.$$

Let $q \in (0, 1/2]$. The initial Bayes type estimator $\tilde{\theta}_{q,n}^{(0)}$ for uniform prior is defined as

$$\tilde{\theta}_{q,n}^{(0)} = \frac{\int_\Theta \theta \exp\left\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\right\} d\theta}{\int_\Theta \exp\left\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\right\} d\theta}.$$

In order to maximize $\mathbb{H}_n(\theta)$, we use the optim() with the method being "L-BFGS-B" in R Language. The Bayes type estimator is calculated with the Markov chain Monte Carlo (MCMC) method defined below.

For the target distribution $p(\theta)d\theta$ in $\mathbb{R}^d$, we run the following Markov chain Monte Carlo method, which is a version of the mixed preconditioned Crank-Nicolson (MpCN) method studied in Kamatani (2014) and Kamatani (2017). Fix $h \in (0, 1)$ and $\nu \geq 0$, and set $g(\theta) = (1 + |\theta|^2/\nu)^{-(\nu+d)/2}$ for $\nu > 0$ and set $g(\theta) = |\theta|^{-d}$ for $\nu = 0$.

- For $m = 0$. Initialize θ.
- For $m \geq 1$, iterate
  - Generate r from the inverse gamma distribution with the shape parameter $\nu/2 + d/2$ and the rate parameter $\nu/2 + |\theta|^2/2$.
  - Set $\theta^* = h^{1/2}\theta + (1-h)^{1/2} r^{1/2} w$ where w follows the standard normal distribution.
  - Accept θ^* as θ with probability $\min\left\{1, \frac{p(\theta^*)g(\theta)}{p(\theta)g(\theta^*)}\right\}$. Otherwise, discard θ^*.

In this paper, we set $h = 0.8$ and $\nu = 2$. This is one of Metropolis-Hastings algorithms. The Markov kernel associated with the transition from θ to θ^* admits an invariant distribution $g(\theta)d\theta$. Thanks to accept/reject process, the resulting Markov kernel is $p(\theta)d\theta$-invariant. The choice of h has little effect and the choice of ν has moderate effect in practice. This MCMC method is efficient for complicated target distribution. For details of this MCMC method, see Kamatani (2014). To apply this method, set $p(\theta) \propto \exp\{\mathbb{H}_n(\theta)\}$ for $\hat{\theta}_{B,n}$, and set $p(\theta) \propto \exp\left\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\right\}$ for $\tilde{\theta}^{(0)}_{q,n}$.

For the true model, 1000 independent sample paths are generated, and the mean and the standard deviation for the estimators are computed and shown in Tables 1-3. Table 1 is the simulation result of the ML type estimator $\hat{\theta}_{M,n}$ with two different initial values. The maximum likelihood estimator derived by using optim() with the initial value being the true value has a good performance. On the other hand, the optimization fails since the initial value derived from the uniform distribution on Θ can be far from the true value.

Table 2 is the simulation result on the Bayes type estimator $\hat{\theta}_{B,n}$ with uniform prior. The simulation was done by using the MCMC method for $M = 5\times10^4$, 5×10^5 and 10^7 with $Bi = 5\times10^3$, 5×10^4 and 10^6, respectively. Here M is the number of Markov chains and Bi is the number of burn-in iteration. The Bayes type estimator with $M = 10^7$ and $Bi = 10^6$ has a good behavior, but under the situations where $(M, Bi) = (5\times10^4, 5\times10^3)$ and $(5\times10^5, 5\times10^4)$, the computation of the Bayes type estimator fails because the Markov chains generated by the MCMC method does not converge to the theoretical ones.

Table 3 is the simulation results of the initial Bayes type estimator $\tilde{\theta}^{(0)}_{q,n}$ with the uniform prior, $M = 5\times10^4$ and $Bi = 5\times10^3$, and the HMS estimators $\hat{\theta}^{(J)}_{q,n}$ for $q = 0.5, 0.45, 0.4, \ldots, 0.05$, where $J = [-\log_2 q]$. Note that $\sup_n E_{\theta^*}\left[\left|n^q(\tilde{\theta}^{(0)}_{q,n} - \theta^*)\right|^M\right] < \infty$ for all $M > 0$. We can see that in this example, the HMS estimator with $q = 0.2$ is the best among the HMS estimators with $q = 0.5, 0.45, 0.4, \ldots, 0.1, 0.05$. It is a difficult problem to choose the optimal q from the theoretical point of view. In practice, however, we can obtain the best estimator among the competing HMS estimators with various values of q, where the best estimator $\hat{\theta}^*_n$ satisfies that $\mathbb{H}_n(\hat{\theta}^*_n) = \max_{q\in K}\mathbb{H}_n(\hat{\theta}^{(J)}_{q,n})$ and K is a set of values of q, e.g., $K = \{0.05, 0.1, 0.15, \ldots, 0.5\}$.

Next, we focus on the computation time for obtaining the estimators. The personal computer with Intel i7 4930K (3.4GHz base clock/3.9GHz Turbo, 12MB cache) was used for simulations. The average times of computation for $\hat{\theta}_{M,n}$, $\hat{\theta}_{B,n}$ and $\hat{\theta}^{(J)}_{0.2,n}$ are 0.9, 1733 and 1667 seconds, respectively, where $\hat{\theta}_{M,n}$ is the ML type estimator with the initial value derived from the uniform distribution on Θ, $\hat{\theta}_{B,n}$ is the Bayes type estimator with uniform prior and $M = 5 \times 10^4$, and $\hat{\theta}^{(J)}_{0.2,n}$ is the HMS estimator obtained from the initial Bayes type estimator $\tilde{\theta}^{(0)}_{q,n}$ with $q = 0.2$, uniform prior and $M = 5 \times 10^4$. The average time of computation for the Bayes type estimator $\hat{\theta}_{B,n}$ with uniform prior and $M = 5 \times 10^5$ is 281 minutes. There is almost no difference of computation time for the HMS estimators based on the initial Bayes type estimator with uniform prior, $M = 5 \times 10^4$ and all q in Table 3. From the computational point of view, obtaining the estimator in a short time is extremely important but the most important thing is to obtain the estimator precisely. In that sense, although we need much time to obtain the HMS estimator with the initial Bayes estimator compared with the ML type estimator by using optim(), the HMS estimator is much better than the ML type estimator in this model.

3. Proofs

Proof of Proposition 1. Set

$$
\begin{aligned}
\mathbb{L}_{q,n}(\theta) &= \frac{1}{n^{1-2q}}\mathbb{H}_n(\theta), \\
\mathbb{Y}_n(\theta) &= \frac{1}{n^{2q}}\{\mathbb{L}_{q,n}(\theta) - \mathbb{L}_{q,n}(\theta^*)\} = \frac{1}{n}\left\{\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta^*)\right\}, \\
\Delta_n(\theta^*)[u] &= \frac{1}{n^q}\partial_\theta \mathbb{L}_{q,n}(\theta^*)[u] = \frac{1}{n^{1-q}}\partial_\theta \mathbb{H}_n(\theta^*)[u], \\
\Gamma_n(\theta^*)[u,u] &= \frac{1}{n^{2q}}\partial^2_\theta \mathbb{L}_{q,n}(\theta^*)[u,u] = \frac{1}{n}\partial^2_\theta \mathbb{H}_n(\theta^*)[u,u]
\end{aligned}
$$

for $u \in \mathbf{R}^p$. Note that $\mathbb{Z}_{q,n}(u;\theta^*) = \exp\left\{\mathbb{L}_{q,n}\left(\theta^* + \frac{u}{n^q}\right) - \mathbb{L}_{q,n}(\theta^*)\right\}$ for $u \in \mathbb{U}_{q,n}$.

Let $\epsilon_1 \in (0, 1/2)$. By Lemma 6 in Uchida and Yoshida (2013), one has that for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}[|\Delta_n(\theta^*)|^M] < \infty, \tag{3}$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(\sup_{\theta\in\Theta} n^{\epsilon_1}|\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta)|\right)^M\right] < \infty. \tag{4}$$

It follows from Lemma 7 in Uchida and Yoshida (2013) that for all $M > 0$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}[(n^{\epsilon_1}|\Gamma_n(\theta^*) - \Gamma_1(\theta^*)|)^M] < \infty, \tag{5}$$

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\left(n^{-1}\sup_{\theta\in\Theta}|\partial_{\theta_i}\partial_{\theta_j}\partial_{\theta_k}\mathbb{H}_n(\theta)|\right)^M\right] < \infty \tag{6}$$

for $i, j, k = 1, \ldots, p$. We can check the regularity conditions $[A1'']$, $[A2]$, $[A3]$, $[A4']$, $[A5]$ and $[A6]$ in Theorem 2 of Yoshida (2011). Indeed, by (5) and (6), we show $[A1'']$. The assumption $[A3]$ in this paper implies $[A2]$, $[A3]$ with $\rho = 2$ and $[A5]$ in Theorem 2 of Yoshida (2011). One can take appropriate parameters satisfying $[A4']$. It follows from (3) and (4) that [A6] holds for every $L > 0$. This completes the proof.

Proof of Proposition 2. Since

$$n^q(\tilde{\theta}_n^{(0)}-\theta^*)=\frac{\int_\Theta n^q(\theta-\theta^*)\exp\Big\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\Big\}\pi(\theta)d\theta}{\int_\Theta \exp\Big\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta)\Big\}\pi(\theta)d\theta}$$

Table 1: ML type estimator with $n = 1 \times 10^4$

initial value	mean (1, 4, 8)	s.d.
true value	1.008, 4.078, 7.998	0.010, 0.102, 0.007
random number	1.622, 1.487, 9.659	0.638, 1.622, 6.412

Table 2: Bayes type estimator with $n = 1 \times 10^4$.

M (Numbers of MCMC)	Bi (Burn-in)	mean (1, 4, 8)	s.d.
5×10^4	5×10^3	1.075, 3.445, 6.644	0.290, 1.242, 2.698
5×10^5	5×10^4	1.020, 3.812, 7.383	0.165, 0.812, 1.868
1×10^7	1×10^6	1.006, 4.066, 7.962	0.027, 0.152, 0.275

Table 3: HMS estimator with $n = 1 \times 10^4$, $M = 5 \times 10^4$, $Bi = 5 \times 10^3$.

q	J	initial Bayes estimator: mean (1, 4, 8), (s.d.) HMS estimator: mean (1, 4, 8), (s.d.)
0.5	1	1.075, 3.445, 6.644 (0.290, 1.242, 2.698) 1.077, 3.673, 6.637 (0.305, 2.034, 2.710)
0.45	1	1.074, 3.552, 6.876 (0.278, 1.163, 2.429) 1.069, 3.774, 6.877 (0.286, 1.841, 2.438)
0.4	1	1.073, 3.599, 7.005 (0.248, 1.113, 2.298) 1.060, 3.902, 7.001 (0.275, 2.042, 2.324)
0.35	1	1.050, 3.747, 7.270 (0.187, 0.954, 1.977) 1.041, 3.933, 7.260 (0.200, 1.710, 2.005)
0.3	1	1.029, 3.993, 7.712 (0.095, 0.562, 1.149) 1.012, 4.194, 7.713 (0.122, 1.575, 1.163)
0.25	2	1.033, 4.208, 7.951 (0.033, 0.268, 0.309) 1.009, 4.135, 7.956 (0.058, 0.774, 0.321)
0.2	2	1.064, 4.488, 7.998 (0.021, 0.250, 0.026) 1.008, 4.058, 7.999 (0.012, 0.157, 0.024)
0.15	2	1.196, 5.072, 8.085 (0.039, 0.497, 0.186) 1.014, 4.166, 8.060 (0.159, 1.845, 0.237)
0.1	3	1.933, 4.587, 8.978 (0.146, 0.729, 0.476) 2.109, 7.799, 9.111 (0.920, 4.439, 0.721)
0.05	4	4.748, 5.968, 9.830 (0.253, 0.323, 0.390) 11.680, 8.632, 9.746 (4.544, 3.355, 0.599)

$$
\begin{aligned}
&= \frac{\int_{\mathbb{U}_{q,n}} u \exp\Big\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta^* + \frac{u}{n^q})\Big\}\pi\Big(\theta^* + \frac{u}{n^q}\Big)du}{\int_{\mathbb{U}_{q,n}} \exp\Big\{\frac{1}{n^{1-2q}}\mathbb{H}_n(\theta^* + \frac{u}{n^q})\Big\}\pi\Big(\theta^* + \frac{u}{n^q}\Big)du} \\
&= \frac{\int_{\mathbb{U}_{q,n}} u\mathbb{Z}_{q,n}(u:\theta^*)\pi\Big(\theta^* + \frac{u}{n^q}\Big)du}{\int_{\mathbb{U}_{q,n}} \mathbb{Z}_{q,n}(u:\theta^*)\pi\Big(\theta^* + \frac{u}{n^q}\Big)du},
\end{aligned}
$$

we obtain that

$$
\begin{aligned}
&E_{\theta^*}[|n^q(\tilde{\theta}_n^{(0)} - \theta^*)|^M] \\
\le& E_{\theta^*}\Big[\Big\{\int_{\mathbb{U}_{q,n}} \mathbb{Z}_{q,n}(u:\theta^*)\pi\Big(\theta^* + \frac{u}{n^q}\Big)du\Big\}^{-1}\int_{\mathbb{U}_{q,n}} |u|^M\mathbb{Z}_{q,n}(u:\theta^*)\pi\Big(\theta^* + \frac{u}{n^q}\Big)du\Big] \\
\le& C\sum_{r=0}^{\infty}(r+1)^M E_{\theta^*}\Big[\Big\{\int_{\mathbb{U}_{q,n}} \mathbb{Z}_{q,n}(u:\theta^*)du\Big\}^{-1}\int_{\{u|r<|u|\le r+1\}\cap\mathbb{U}_{q,n}} \mathbb{Z}_{q,n}(u:\theta^*)du\Big] \\
\le& C\sum_{r=0}^{\infty}(r+1)^M\Big\{P_{\theta^*}\Big[\sup_{u\in\mathbb{V}_{q,n}(r)} \mathbb{Z}_{q,n}(u;\theta^*) \ge e^{-r}\Big] \\
&\qquad + e^{-r}\Big(\int_{\{u|r<|u|\le r+1\}} du\Big)E_{\theta^*}\Big[\Big(\int_{\mathbb{U}_{q,n}} \mathbb{Z}_{q,n}(u:\theta^*)du\Big)^{-1}\Big]\Big\}.
\end{aligned}
$$

Next one has that

$$
\sup_{n\in\mathbb{N}} E_{\theta^*}\Big[\Big(\int_{\mathbb{U}_{q,n}} \mathbb{Z}_{q,n}(u:\theta^*)du\Big)^{-1}\Big] < \infty. \tag{7}
$$

Proof of (7). Note that

$$
\begin{aligned}
\log\mathbb{Z}_{q,n}(u;\theta^*) \;=\; & \frac{1}{n^{1-2q}}\Big\{\partial_\theta\mathbb{H}_n(\theta^*)[u]\frac{1}{n^q} + \frac{1}{2}\partial_\theta^2\mathbb{H}_n(\theta^*)[u^{\otimes 2}]\frac{1}{n^{2q}} \\
& + \frac{1}{2}\int_0^1 (1-t)^2\partial_\theta^3\mathbb{H}_n(\theta^* + \frac{tu}{n^q})dt[u^{\otimes 3}]\frac{1}{n^{3q}}\Big\}.
\end{aligned}
$$

By Lemmas 6 and 7 of Uchida and Yoshida (2013), for every $M > p$, $\delta > 0$, there exists $C_0 > 0$ such that

$$
\sup_{n\in\mathbb{N}} E_{\theta^*}\Big[|\log\mathbb{Z}_{q,n}(u;\theta^*)|^M\Big] \le C_0|u|^M
$$

for all $u \in \{u \in \mathbb{U}_{q,n}; |u| \le \delta\}$. It follows from Lemma 2 of Yoshida (2011) that

$$
\sup_{n\in\mathbb{N}} E_{\theta^*}\Big[\Big(\int_{\{u\in\mathbb{U}_{q,n};|u|\le\delta\}} e^{\log\mathbb{Z}_{q,n}(u:\theta^*)}du\Big)^{-1}\Big] < \infty.
$$

This completes the proof.

Proof of Lemma 1. We will show the result by mathematical induction.

When $k = 0$, it follows from assumption [B] that the statement holds.

Assume that the statement holds for some $k = l$. On $K_{l,n} := K_n(\hat{\theta}_n^{(l)})$, one has that

$$\hat{\theta}_n^{(l+1)} = \hat{\theta}_n^{(l)} - \Big[\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(l)})\Big]^{-1}\frac{1}{n}\partial_\theta H_n(\hat{\theta}_n^{(l)}),$$

and

$$\partial_\theta \mathbb{H}_n(\theta^*) = \partial_\theta \mathbb{H}_n(\hat{\theta}_n^{(l)}) + \partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(l)})(\theta^* - \hat{\theta}_n^{(l)}) + R_n^{(l)}[(\theta^* - \hat{\theta}_n^{(l)})^{\otimes 2}],$$

where $R_n^{(l)} = \int_0^1 (1-t)\partial_\theta^3 \mathbb{H}_n(\hat{\theta}_n^{(l)} + t(\theta^* - \hat{\theta}_n^{(l)}))dt$. Therefore,

$$\begin{aligned} n^{2^{l+1}q}(\hat{\theta}_n^{(l+1)} - \theta^*) = &- \Big[\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(l)})\Big]^{-1}\Big[\frac{n^{2^{l+1}q}}{n}\partial_\theta \mathbb{H}_n(\theta^*)\Big] \\ &+ \Big[\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(l)})\Big]^{-1}\frac{1}{n}R_n^{(l)}[(n^{2^l q}(\theta^* - \hat{\theta}_n^{(l)}))^{\otimes 2}]. \end{aligned}$$

By using the standard estimates and Lemmas 6 and 7 of Uchida and Yoshida (2013), one has that

$$\begin{aligned} &\sup_n E_{\theta^*}\Big[\Big|\Big[\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(l)})\Big]^{-1}\Big|^M 1_{K_{l,n}}\Big] < \infty, \\ &\sup_n E_{\theta^*}\Big[\Big|\frac{1}{\sqrt{n}}\partial_\theta \mathbb{H}_n(\theta^*)\Big|^M\Big] < \infty, \\ &\sup_n E_{\theta^*}\Big[\Big|\frac{1}{n}R_n^{(l)}\Big|^M\Big] < \infty. \end{aligned}$$

Hence,

$$\sup_n E_{\theta^*}[|n^{2^{l+1}q}(\hat{\theta}_n^{(l+1)} - \theta^*)|^M 1_{K_{l,n}}] < \infty. \tag{8}$$

Next, we note that for all $\theta \neq \theta^*$,

$$\begin{aligned} \chi_0 &= \inf_{\theta\neq\theta^*}\frac{-\mathbb{Y}(\theta)}{|\theta-\theta^*|^2} \\ &\leq \frac{-\mathbb{Y}(\theta)}{|\theta-\theta^*|^2} \\ &= \frac{-\mathbb{Y}(\theta^*)}{|\theta-\theta^*|^2} + \frac{-\partial_\theta \mathbb{Y}(\theta^*)[\theta-\theta^*]}{|\theta-\theta^*|^2} + \frac{1}{2}\Gamma(\theta^*)\left[\left(\frac{\theta-\theta^*}{|\theta-\theta^*|}\right)^{\otimes 2}\right] \\ &\quad + \frac{1}{2}\sum_{i,j,k=1}^p \int_0^1 (1-u)^2 \partial_{\theta_i}\partial_{\theta_j}\partial_{\theta_k}\mathbb{Y}(\theta^* + u(\theta-\theta^*))du \\ &\qquad\qquad \times \frac{(\theta-\theta^*)_i}{|\theta-\theta^*|}\frac{(\theta-\theta^*)_j}{|\theta-\theta^*|}\frac{(\theta-\theta^*)_k}{|\theta-\theta^*|}|\theta-\theta^*| \\ &\leq \frac{1}{2}\Gamma(\theta^*)\left[\left(\frac{\theta-\theta^*}{|\theta-\theta^*|}\right)^{\otimes 2}\right] + \sum_{i,j,k=1}^p \frac{1}{6}\left(\sup_\theta |\partial_{\theta_i}\partial_{\theta_j}\partial_{\theta_k}\mathbb{Y}(\theta)|\right)|\theta-\theta^*|. \end{aligned}$$

Let $\lambda_{min} = \inf_{\theta\neq\theta^*}\Gamma(\theta^*)\left[\left(\frac{\theta-\theta^*}{|\theta-\theta^*|}\right)^{\otimes 2}\right]$ and $\Xi = \left\{\theta \neq \theta^* \mid \Gamma(\theta^*)\left(\frac{\theta-\theta^*}{|\theta-\theta^*|}\right) = \lambda_{min}\left(\frac{\theta-\theta^*}{|\theta-\theta^*|}\right)\right\}$. For every $\theta \in \Xi$, one has that

$$\chi_0 \leq \frac{1}{2}\lambda_{min} + \sum_{i,j,k=1}^p \frac{1}{6}\left(\sup_\theta |\partial_{\theta_i}\partial_{\theta_j}\partial_{\theta_k}\mathbb{Y}(\theta)|\right)|\theta-\theta^*|.$$

Therefore,

$$\chi_0 \leq \inf_{\theta\in\Xi}\left\{\frac{1}{2}\lambda_{min}+\sum_{i,j,k=1}^{p}\frac{1}{6}\left(\sup_{\theta}|\partial_{\theta_i}\partial_{\theta_j}\partial_{\theta_k}\mathbb{Y}(\theta)|\right)|\theta-\theta^*|\right\}$$
$$= \frac{1}{2}\lambda_{min}=\frac{1}{2}\inf_{|x|=1}|\Gamma(\theta^*)x|. \tag{9}$$

It follows from [A3] and (9) that for all $r>0$ and $L>0$,

$$1-\frac{C_L}{r^L} \leq P\left[\chi_0>\frac{1}{r}\right]\leq P\left[\inf_{|x|=1}|\Gamma(\theta^*)x|>\frac{2}{r}\right]. \tag{10}$$

Set

$$A_n:=\left\{\left|\Gamma_n(\theta_n^{(l)})-\Gamma_n(\theta^*)\right|<\frac{1}{2r},\ \left|\Gamma_n(\theta^*)+\Gamma(\theta^*)\right|<\frac{1}{2r},\ \inf_{|x|=1}|\Gamma(\theta^*)x|>\frac{2}{r}\right\}.$$

Since

$$\inf_{|x|=1}|\Gamma_n(\hat{\theta}_n^{(l)})x|\geq\inf_{|x|=1}|\Gamma(\theta^*)x|-\sup_{|x|=1}|(\Gamma_n(\hat{\theta}_n^{(l)})+\Gamma(\theta^*))x|,$$

we have that

$$P[A_n] \leq P\left[\left|\Gamma_n(\hat{\theta}_n^{(l)})+\Gamma(\theta^*)\right|<\frac{1}{r},\ \inf_{|x|=1}|\Gamma(\theta^*)x|>\frac{2}{r}\right]$$
$$\leq P\left[\inf_{|x|=1}|\Gamma_n(\hat{\theta}_n^{(l)})x|>\frac{1}{r}\right].$$

Noting that for all $r>0$,

$$P[A_n^c] \leq P\left[\left|\Gamma_n(\hat{\theta}_n^{(l)})-\Gamma_n(\theta^*)\right|\geq\frac{1}{2r}\right]+P\left[\left|\Gamma_n(\theta^*)+\Gamma(\theta^*)\right|\geq\frac{1}{2r}\right]$$
$$+P\left[\inf_{|x|=1}|\Gamma(\theta^*)x|\leq\frac{2}{r}\right]$$
$$\leq (2r)^L\frac{E\left[\left(\frac{1}{n}\sup_\theta\left|\partial_\theta^3\mathbb{H}_n(\theta)\right|\left|n^{2^lq}(\hat{\theta}^{(l)}-\theta^*)\right|\right)^L\right]}{n^{2^lqL}}$$
$$+(2r)^L\frac{E\left[\left|n^{1/2}(\Gamma_n(\theta^*)+\Gamma(\theta^*))\right|^L\right]}{n^{L/2}}+\frac{C_L}{r^L},$$

and

$$P\left[K_{l,n}^c\right] \leq P\left[\inf_{|x|=1}|\Gamma_n(\hat{\theta}_n^{(l)})x|=0\right]\leq P\left[\inf_{|x|=1}|\Gamma_n(\hat{\theta}_n^{(l)})x|\leq\frac{1}{r}\right]\leq P[A_n^c],$$

and setting $r=n^{2^{l-1}q}$ and $L_0=2^{l-1}qL$, we obtain that

$$P\left[K_{l,n}^c\right] \leq \frac{C_{L_0}}{n^{L_0}}. \tag{11}$$

Here we notice that $2^{l-1}q \in (0, 1/4]$ since $0 < 2^l q \le 1/2$.

It follows from (11) that for every $M > 0$,

$$\begin{aligned}
&\sup_n E_{\theta^*}[|n^{2^{l+1}q}(\hat{\theta}_n^{(l+1)} - \theta^*)|^M 1_{K_{l,n}^c}] \\
\le&\sup_n E_{\theta^*}[|n^{2^{l+1}q}(\hat{\theta}_n^{(l+1)} - \hat{\theta}_n^{(l)})|^M 1_{K_{l,n}^c}] + \sup_n E_{\theta^*}[|n^{2^{l+1}q}(\hat{\theta}_n^{(l)} - \theta^*)|^M 1_{K_{l,n}^c}] \\
\le&\sup_n E_{\theta^*}\left[\left|\frac{1}{n}\partial_\theta \mathbb{H}(\hat{\theta}_n^{(l)})\right|^M n^{2^{l+1}qM} 1_{K_{l,n}^c}\right] + \sup_n E_{\theta^*}\left[\left|n^{2^l q}(\hat{\theta}_n^{(l)} - \theta^*)\right|^M n^{2^l qM} 1_{K_{l,n}^c}\right] \\
<&\infty,
\end{aligned}$$

which together with (8) completes the proof.

Proof of Theorem 1. On $K_{J-1,n}$,

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*) =& -\left[\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(J-1)})\right]^{-1}\left[\frac{1}{\sqrt{n}}\partial_\theta \mathbb{H}_n(\theta^*)\right] \\
&+ n^{-2^J q + \frac{1}{2}}\left[\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{(J-1)})\right]^{-1}\frac{1}{n}R_n^{(j-1)}[(n^{2^{J-1}q}(\theta^* - \hat{\theta}_n^{(J-1)}))^{\otimes 2}].
\end{aligned}$$

Since Lemma 1 yields that for all $M > 0$,

$$\sup_n E_{\theta^*}[|n^{2^{J-1}q}(\hat{\theta}_n^{(J-1)} - \theta^*)|^M] < \infty,$$

one has that

$$\sup_n E_{\theta^*}[|n^{1/2}(\hat{\theta}_n^{(J)} - \theta^*)|^M 1_{K_{J-1,n}}] < \infty.$$

It follows from (11) with $l = J - 1$ that for every $M > 0$,

$$\begin{aligned}
&\sup_n E_{\theta^*}[|n^{1/2}(\hat{\theta}_n^{(J)} - \theta^*)|^M 1_{K_{J-1,n}^c}] \\
\le&\sup_n E_{\theta^*}[|n^{1/2}(\hat{\theta}_n^{(J)} - \hat{\theta}_n^{(J-1)})|^M 1_{K_{J-1,n}^c}] + \sup_n E_{\theta^*}[|n^{1/2}(\hat{\theta}_n^{(J-1)} - \theta^*)|^M 1_{K_{J-1,n}^c}] \\
\le&\sup_n E_{\theta^*}\left[\left|\frac{1}{n}\partial_\theta \mathbb{H}(\hat{\theta}_n^{(J-1)})\right|^M n^{1/2} 1_{K_{J-1,n}^c}\right] \\
&+ \sup_n E_{\theta^*}\left[\left|n^{2^{J-1}q}(\hat{\theta}_n^{(J-1)} - \theta^*)\right|^M n^{(1/2 - 2^{J-1}q)M} 1_{K_{J-1,n}^c}\right] \\
<&\infty.
\end{aligned}$$

Thus, for all $M > 0$,

$$\sup_n E_{\theta^*}[|\sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*)|^M] < \infty. \tag{12}$$

Set

$$\begin{aligned}
S_n &:= \sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*), \quad \Delta_n := \frac{1}{\sqrt{n}}\partial_\theta \mathbb{H}_n(\theta^*), \\
\bar{\Gamma}_{J-1,n} &:= \Gamma_{J-1,n} 1_{K_{J-1,n}} + E_p 1_{K_{J-1,n}^c},
\end{aligned}$$

$$\bar{A}_n \quad := \quad n^{-2^J q+\frac{1}{2}}\bar{\Gamma}_{J-1,n}^{-1}\frac{1}{n}R_n^{(l)}[(n^{2^{J-1}q}(\theta^* - \hat{\theta}_n^{(J-1)}))^{\otimes 2}]$$

and

$$A_n \quad := \quad n^{-2^J q+\frac{1}{2}}\Gamma_{J-1,n}^{-1}\frac{1}{n}R_n^{(l)}[(n^{2^{J-1}q}(\theta^* - \hat{\theta}_n^{(J-1)}))^{\otimes 2}]$$

on $K_{J-1,n}$. Here we note that $\frac{1}{2} < 2^J q \leq 1$ since $J = [-\log_2 q]$.

For any closed set $C \subset \mathbb{R}^{p+1}$ and any $\mathcal{F}$-measurable random variable Y,

$$\begin{aligned}
P[(Y,S_n) \in C] &= P[\{(Y,S_n) \in C\} \cap K_{J-1,n}] + P[\{(Y,S_n) \in C\} \cap K_{J-1,n}^c] \\
&\leq P[\{(Y, -\Gamma_{J-1,n}^{-1}\Delta_n + A_n) \in C\} \cap K_{J-1,n}] + P[K_{J-1,n}^c] \\
&= P[\{(Y, -\bar{\Gamma}_{J-1,n}^{-1}\Delta_n + \bar{A}_n) \in C\} \cap K_{J-1,n}] + P[K_{J-1,n}^c] \\
&\leq P[(Y, -\bar{\Gamma}_{J-1,n}^{-1}\Delta_n + \bar{A}_n) \in C] + o(1).
\end{aligned}$$

Note that

$$-\bar{\Gamma}_{J-1,n}^{-1} \xrightarrow{p} \Gamma(\theta^*)^{-1}, \quad \Delta_n \xrightarrow{d_s(\mathcal{F})} \Gamma(\theta^*)^{1/2}\zeta, \quad \bar{A}_n \xrightarrow{p} 0,$$

where in a similar way to the proof of Lemma 9 in Uchida and Yoshida (2013), we can show the stable convergence of Δ_n. Furthermore, it follows from Proposition 5.33 in Chapter VIII of Jacod and Shiryaev (2002), (2.2.5) in Jacod and Protter (2012) and the continuous mapping theorem that

$$\begin{aligned}
\limsup_{n\to\infty} P[(Y,S_n) \in C] &\leq \limsup_{n\to\infty} P[(Y, \bar{\Gamma}_{J-1,n}^{-1}\Delta_n + A_n) \in C] \\
&\leq P[(Y, \Gamma(\theta^*)^{-1/2}\zeta) \in C].
\end{aligned}$$

Hence, $\sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*) \xrightarrow{d_s(\mathcal{F})} \Gamma(\theta^*)^{-1/2}\zeta$, which together with (12) implies that

$$E[f(\sqrt{n}(\hat{\theta}_n^{(J)} - \theta^*))] \to E[f(\Gamma(\theta^*)^{-1/2}\zeta)]$$

for all continuous functions f of at most polynomial growth. This completes the proof.

Acknowledgement

The authors wish to thank the referee for valuable comments. The research of Kamatani was partially supported by JSPS KAKENHI Grant Number 24740062, CREST, JST, and by Cooperative Research Program of the Institute of Statistical Mathematics. Uchida's research was partially supported by JSPS KAKENHI Grant Numbers 24300107, 25245034, CREST, JST, and by 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.

Brouste, A., Fukasawa, M., Hino, H., Iacus, S., Kamatani, K., Koike, Y., Masuda, H., Nomura, R., Shimuzu, Y., Uchida, M., and Yoshida, N. (2014). The YUIMA Project : a Computational Framework for Simulation and Inference of Stochastic Differential Equations, *Journal of Statistical Software*, Vol. 57, Issue 4, Apr 2014.

Dohnal, G. (1987). On Estimating the Diffusion Coefficient. *Journal of Applied Probability*, **24**, 105–114.

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. *Annales de l' Institut Henri Poincaré Probabilités et Statistiques,* **29**, 119–151.

Genon-Catalot, V. and Jacod, J. (1994). Estimation of the Diffusion Coefficient for Diffusion Processes: Random Sampling *Scandinavian Journal of Statistics*, **21**, 193–221.

Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli,* **7**, 899–912.

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

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

Jacod, J. and Protter, P. E. (2012). *Discretization of Processes.* Springer.

Jacod, J. and Shiryaev, A. (2002). *Limit Theorems for Stochastic Processes, 2nd. Edition.* Springer.

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

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. (2017). Ergodicity of Markov chain Monte Carlo with reversible proposal. *Journal of Applied Probability* **54**(2), to appear.

Kessler, M. (1995). Estimation des paramètres d'une diffusion par des contrastes corrigés. *Comptes Rendus de l'Acadmie 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. (2015). On multi-step MLE-process for ergoodic diffusion. (arXiv 1504.01869). Preprint.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory.* Springer-Verlag, New York.

Masuda, H. (2013a). Asymptotics for functionals of self-normalized residuals of discretely observed stochastic processes. *Stochastic Processes and their Applications*, bf 123, 2752–2778

Masuda, H. (2013b). Convergence of Gaussian quasi-likelihood random fields for ergodic Levy driven SDE observed at high frequency. *The Annals of Statistics*, **41**, 1593–1641.

Ogihara, T. and Yoshida, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes*, **14**, 189-229.

Prakasa Rao, B. L. S. (1983). Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Mathematische Operationsforschung und Statistik, Series Optimization,* **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.

R Development Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Shimizu, Y. (2006). M-estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Statistical Inference for Stochastic Processes*, **9**, 179–225.

Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes*, **9**, 227–277.

Sørensen, M. (2008). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint, Department of Mathematical Sciences, University of Copenhagen.

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. (2004). Information criteria for small diffusions via the theory of Malliavin-Watanabe. *Statistical Inference for Stochastic Processes*, **7**, 35–67.

Uchida, M. and Yoshida, N. (2006). Asymptotic expansion and information criteria. Dedicated to Professor Minoru Siotani on his 80th birthday. *SUT Journal of Mathematics*, **42**, 31–58.

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. (2013). Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. *Stochastic Processes and their Applications*, **123**, 2851-2876.

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.

Uchida, M. and Yoshida, N. (2016). Model selection for volatility prediction. In Mark Podolskij, Robert Stelzer, Steen Thorbjrnsen, Almut E. D. Veraart (eds.) The Fascination of Probability, Statistics and their Applications. In Honour of Ole E. Barndorff-Nielsen, Springer, 343–360.

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 January 14, 2016
Revised March 3, 2016