

Article

Berry–Esseen Bounds of the Quasi Maximum Likelihood Estimators for the Discretely Observed Diffusions

Jaya P. N. Bishwal

Department of Mathematics and Statistics, University of North Carolina at Charlotte, 376 Fretwell Bldg., 9201 University City Blvd., Charlotte, NC 28223-0001, USA; J.Bishwal@unc.edu

Abstract: For stationary ergodic diffusions satisfying nonlinear homogeneous Itô stochastic differential equations, this paper obtains the Berry–Esseen bounds on the rates of convergence to normality of the distributions of the quasi maximum likelihood estimators based on stochastic Taylor approximation, under some regularity conditions, when the diffusion is observed at equally spaced dense time points over a long time interval, the high-frequency regime. It shows that the higher-order stochastic Taylor approximation-based estimators perform better than the basic Euler approximation in the sense of having smaller asymptotic variance.

Keywords: Itô stochastic differential equation; diffusion process; Malliavin calculus; discrete observations; almost slowly increasing experimental design; quasi maximum likelihood estimators; conditional least squares estimator; Monte Carlo methods

MSC: 60F05; 60F10; 60H10; 60J60; 62F12; 62F15; 62M05



Citation: Bishwal, J.P.N.

Berry–Esseen Bounds of the Quasi Maximum Likelihood Estimators for the Discretely Observed Diffusions. *AppliedMath* **2022**, *2*, 39–53. <https://doi.org/10.3390/appliedmath2010003>

Academic Editor: Tommi Sottinen

Received: 2 December 2021

Accepted: 7 January 2022

Published: 8 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Preliminaries

Parameter estimation in diffusion processes based on discrete observations is the recent trend of investigation in financial econometrics and mathematical biology since the data available in finance and biology are high-frequency discrete, though the model is continuous. For a treatise on this subject, see Bishwal (2008, 2021) [1,2].

Consider the Itô stochastic differential equation

$$\begin{aligned} dX_t &= f(\theta, X_t)dt + dW_t, \quad t \geq 0 \\ X_0 &= X^0 \end{aligned} \quad (1)$$

where $\{W_t, t \geq 0\}$ is a one-dimensional standard Wiener process, $\theta \in \Theta$, Θ is a compact subset of \mathbb{R} , f is a known real valued function defined on $\Theta \times \mathbb{R}$, the unknown parameter θ is to be estimated on the basis of observation of the process $\{X_t, t \geq 0\}$. Let θ_0 be the true value of the parameter that is in the interior of Θ . We assume that the process $\{X_t, t \geq 0\}$ is observed at $0 = t_0 < t_1 < \dots < t_n = T$ with $\Delta t_i := t_i - t_{i-1} = \frac{T}{n} = h$ and $T = dn^{1/2}$ for some fixed real number $d > 0$. We estimate θ from the observations $\{X_{t_0}, X_{t_1}, \dots, X_{t_n}\}$.

The conditional least squares estimator (CLSE) of θ is defined as

$$\begin{aligned} \theta_{n,T} &:= \arg \min_{\theta \in \Theta} Q_{n,T}(\theta) \\ \text{where } Q_{n,T}(\theta) &= \sum_{i=1}^n \frac{[X_{t_i} - X_{t_{i-1}} - f(\theta, X_{t_{i-1}})h]^2}{\Delta t_i} \end{aligned} \quad (2)$$

This estimator was first studied by Dorogovcev (1976) [3], who obtained its weak consistency under some regularity conditions as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$. Kasonga (1988) [4] obtained the strong consistency of the CLSE under some regularity conditions as $n \rightarrow \infty$

assuming that $T = dn^{1/2}$ for some fixed real number $d > 0$. Prakasa Rao (1983) [5] obtained asymptotic normality of the CLSE as $T \rightarrow \infty$ and $\frac{T}{n^{1/2}} \rightarrow 0$.

Florens-Zmirou (1989) [6] studied the minimum contrast estimator, based on a Euler–Maruyama-type first-order approximate discrete time scheme of the SDE (1), which is given by

$$Z_{t_i} - Z_{t_{i-1}} = f(\theta, Z_{t_{i-1}})(t_i - t_{i-1}) + W_{t_i} - W_{t_{i-1}}, \quad i \geq 1, \quad Z_0 = X^0. \tag{3}$$

The log-likelihood function of $\{Z_{t_i}, 0 \leq i \leq n\}$ is given by

$$L_{n,T} = C \sum_{i=1}^n \frac{[Z_{t_i} - Z_{t_{i-1}} - f(\theta, Z_{t_{i-1}})h]^2}{\Delta t_i} \tag{4}$$

where C is a constant independent of θ . A contrast for the estimation of θ is derived from the above log-likelihood by substituting $\{Z_{t_i}, 0 \leq i \leq n\}$ with $\{X_{t_i}, 0 \leq i \leq n\}$. The resulting contrast is

$$H_{n,T} = C \sum_{i=1}^n \frac{[X_{t_i} - X_{t_{i-1}} - f(\theta, X_{t_{i-1}})h]^2}{\Delta t_i} \tag{5}$$

and the resulting minimum contrast estimator, called the Euler–Maruyama estimator, is given by

$$\check{\theta}_{n,T} := \arg \min_{\theta \in \Theta} H_{n,T}(\theta)$$

Florens-Zmirou (1989) [6] showed the L_2 -consistency of the estimator as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$ and asymptotic normality as $T \rightarrow \infty$ and $\frac{T}{n^{2/3}} \rightarrow 0$.

Notice that the contrast $H_{n,T}$ would be the log-likelihood of $(X_{t_i}, 0 \leq i \leq n)$ if the transition probability was $\mathcal{N}(f(\theta, x)h, h)$. This led Kessler (1997) [7] to consider Gaussian approximation of the transition density. The most natural one is achieved through choosing its mean and variance to be the mean and variance of the transition density. Thus, the transition density is approximated by $\mathcal{N}(E(X_{t_i}|X_{t_{i-1}}), h)$, which produces the contrast

$$K_{n,T} = C \sum_{i=1}^n \frac{[X_{t_i} - E(X_{t_i}|X_{t_{i-1}})]^2}{\Delta t_i} \tag{6}$$

Since the transition density is unknown, in general, there is no closed-form expression for $E(X_{t_i}|X_{t_{i-1}})$. Using the stochastic Taylor formula obtained in Florens-Zmirou (1989) [6], he obtained a closed-form expression of $E(X_{t_i}|X_{t_{i-1}})$. The contrast $H_{n,T}$ is an example of such an approximation when $E(X_{t_i}|X_{t_{i-1}}) \approx X_{t_{i-1}} + hf(\theta, X_{t_{i-1}})$.

The resulting minimum contrast estimator, which is also the quasi-maximum likelihood estimator (QMLE), is given by

$$\theta_{n,T} := \arg \min_{\theta \in \Theta} K_{n,T}(\theta)$$

Kessler (1997) [7] showed the L_2 -consistency of the estimator as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$ and asymptotic normality as $T \rightarrow \infty$ and $\frac{T}{n^{(p-1)/p}} \rightarrow 0$ for an arbitrary integer p .

Denote

$$\mu(\theta, X_{t_{i-1}}) := E(X_{t_i}|X_{t_{i-1}}), \quad \mu(\theta, x) := E(X_{t_i}|X_{t_{i-1}} = x) \tag{7}$$

which is the mean function of the transition probability distribution. Hence, the contrast is given by

$$K_{n,T} = C \sum_{i=1}^n \frac{[X_{t_i} - \mu(\theta, X_{t_{i-1}})]^2}{\Delta t_i} \tag{8}$$

If continuous observation of $\{X_t\}$ on the interval $[0, T]$ were available, then the likelihood function of θ would be

$$L_T(\theta) = \exp \left\{ \int_0^T f(\theta, X_t) dX_t - \frac{1}{2} \int_0^T f^2(\theta, X_t) dt \right\} \tag{9}$$

(see Liptser and Shiriyayev (1977) [8]). Since we have discrete data, we have to approximate the likelihood to obtain the MLE. Taking Itô-type approximation of the stochastic integral and rectangle rule approximation of the ordinary integral in (9), we obtain the approximate likelihood function

$$\hat{L}_{n,T}(\theta) := \exp \left\{ \sum_{i=1}^n f(\theta, X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) - \frac{h}{2} \sum_{i=1}^n f^2(\theta, X_{t_{i-1}}) \right\}. \tag{10}$$

The Itô approximate maximum likelihood estimate (IAMLE) based on $\hat{L}_{n,T}$ is defined as

$$\hat{\theta}_{n,T} := \arg \max_{\theta \in \Theta} \hat{L}_{n,T}(\theta).$$

Weak consistency and asymptotic normality of this estimator were obtained by Yoshida (1992) [9] as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$.

Note that the CLSE, the Euler–Maruyama estimator and the IAMLE are the same estimator (see Shoji (1997) [10]). For the Ornstein–Uhlenbeck process, Bishwal and Bose (2001) [11] studied the rates of weak convergence of approximate maximum likelihood estimators, which are of conditional least squares type. For the Ornstein–Uhlenbeck process, Bishwal (2010) [12] studied the uniform rate of weak convergence for the minimum contrast estimator, which has a close connection to the Stratonovich–Milstein scheme. Bishwal (2009) [13] studied Berry–Esseen inequalities for conditional least squares estimator in discretely observed nonlinear diffusions. Bishwal (2009) [14] studied the Stratonovich-based approximate M-estimator of discretely sampled nonlinear diffusions. Bishwal (2011) [15] studied Milstein approximation of the posterior density of diffusions. Bishwal (2010) [16] studied conditional least squares estimation in nonlinear diffusion processes based on Poisson sampling. Bishwal (2011) [17] obtained some new estimators of integrated volatility using the stochastic Taylor-type schemes, which could be useful for option pricing in stochastic volatility models; see also Bishwal (2021) [2].

Prime denotes the derivative with respect to θ , dot denotes the derivative with respect to x and \vee denotes the max symbol throughout the paper. In order to obtain a better estimator in terms of lowering variance in Monte Carlo simulation, which may have a faster rate of convergence, first, we use the algorithm proposed in Bishwal (2008) [1]. Note that the Itô integral and the Fisk–Stratonovich (FS, henceforth; Fisk, while introducing the concept of quasimartingale, had the trapezoidal approximation and Stratonovich had the midpoint approximation, converging to the same mean square limit) integral are connected by

$$\int_0^T f(\theta, X_t) dX_t = \int_0^T f(\theta, X_t) \circ dX_t - \frac{1}{2} \int_0^T \dot{f}(\theta, X_t) dt, \tag{11}$$

where \circ is the Itô’s circle for the FS integral. We transform the Itô integral (the limit of the rectangular approximation to preserve the martingale property) in (9) to the FS integral and apply FS-type trapezoidal approximation of the stochastic integral and rectangular rule-type approximation of the Lebesgue integrals and obtain the approximate likelihood

$$\tilde{L}_{n,T}(\theta) := \exp \left\{ \frac{1}{2} \sum_{i=1}^n [f(\theta, X_{t_{i-1}}) + f(\theta, X_{t_i})](X_{t_i} - X_{t_{i-1}}) - \frac{h}{2} \sum_{i=1}^n \dot{f}(\theta, X_{t_{i-1}}) - \frac{h}{2} \sum_{i=1}^n f^2(\theta, X_{t_{i-1}}) \right\} \tag{12}$$

The Fisk–Stratonovich approximate maximum likelihood estimator (FSAMLE) based on $\tilde{L}_{n,T}$ is defined as

$$\tilde{\theta}_{n,T} := \arg \max_{\theta \in \Theta} \tilde{L}_{n,T}(\theta).$$

Weak consistency as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$ and asymptotic normality as $T \rightarrow \infty$ and $\frac{T}{n^{2/3}} \rightarrow 0$ of the FSAMLE were shown in Bishwal (2008) [1]. Berry–Esseen bounds for the IAMLE and the FSAMLE for the Ornstein–Uhlenbeck processes were obtained in Bishwal and Bose (2001) [11].

We shall use the following notations: $\Delta X_i = X_{t_i} - X_{t_{i-1}}$, $\Delta W_i = W_{t_i} - W_{t_{i-1}}$, C is a generic constant independent of h, n and other variables (it may depend on θ). Throughout the paper, \dot{f} denotes the derivative with respect to x and f' denotes the derivative with respect to θ of the function $f(\theta, x)$. Suppose that θ_0 denotes the true value of the parameter and $\theta_0 \in \Theta$. We assume the following conditions:

Assumption 1.

- (A1) $|f(\theta, x)| \leq a(\theta)(1 + |x|)$,
 $|f(\theta, x) - f(\theta, y)| \leq a(\theta)|x - y|$.
- (A2) $|f(\theta, x) - f(\phi, y)| \leq b(x)|\theta - \phi|$ for all $\theta, \phi \in \Theta, x, y \in \mathbb{R}$
 where $\sup_{\theta \in \Theta} |a(\theta)| = a < \infty, E|b(X^0)|^r < \infty$ for any integer r .

(A3) The diffusion process X is stationary and ergodic with invariant measure ν , i.e., for any g with $E[g(\cdot)] < \infty$,

$$\frac{1}{n} \sum_{i=1}^n g(X_{t_i}) \rightarrow E_\nu[g(X_0)] \text{ a.s. as } T \rightarrow \infty \text{ and } h \rightarrow 0.$$

- (A4) $\sup_{t \geq 0} E|X_t|^q < \infty$ for all $q \geq 0$.
- (A5) $E|f(\theta, X^0) - f(\theta_0, X^0)|^2 = 0$ iff $\theta = \theta_0$.
- (A6) f is continuously differentiable function in x up to order p for all θ .
- (A7) $f(\cdot, x)$ and all its derivatives are three times continuously differentiable with respect to θ for all $x \in \mathbb{R}$. Moreover, these derivatives up to third order with respect to θ are of polynomial growth in x uniformly in θ .

The Fisher information is given by

$$0 < I(\theta) := \int_{-\infty}^{\infty} (f'(\theta, x))^2 d\nu(x) < \infty$$

and for any $\delta > 0$, or any compact $\bar{\Theta} \subset \Theta$,

$$\inf_{\theta_0 \in \bar{\Theta}} \sup_{|\theta - \theta_0| > \delta} E_{\theta_0} |f'(\theta, X_0) - f'(\theta_0, X_0)|^2 > 0.$$

(A8) The Malliavin covariance of the process is nondegenerate.

The Malliavin covariance matrix of a smooth random variable S is defined as $\gamma_T = \int_0^T D_t S [D_t S]^* dt$, where D_t is the Malliavin derivative. The Malliavin covariance is nondegenerate if $\det(\gamma_T)$ is almost surely positive and, for any $m \geq 1$, one has $\|1/\det(\gamma_T)\|_{L^m} < \infty$. This, associated with the functional $\omega \rightarrow X(t, \omega)$, is given by $0 < \sigma^2(t) = Y_t^2 \int_0^t f^2(\theta, X_s) Z_s^2 ds < \infty$ where Y_t and Z_t , respectively, satisfy

$$dY_t = \dot{f}(\theta, X_t) Y_t dt + Y_t dW_t, Y_0 = 1, dZ_t = -f(\theta, X_t) Z_t dt - Z_t dW_t, Z_0 = 1.$$

In the case of independent observations, in order to prove the validity of asymptotic expansion, one usually needs a certain regularity condition for the underlying distribution, such as the Cramér condition; see Bhattacharya and Ranga Rao (1976) [18]. This type of condition then ensures the regularity of the distribution and hence the smoothness assumption of the functional under the expectation whose martingale expansion is desired can be removed. This type of condition for dependent observations leads to the regularity of

the distribution of a functional with nondegenerate Malliavin covariance, which is known in Malliavin calculus; see Ikeda and Watanabe (1989) [19] and Nualart (1995) [20]. Malliavin covariance is connected to the Hörmander condition, which is a sufficient condition for a second-order differential operator to be hypoelliptic; see Bally (1991) [21]. For operators with analytic coefficients, this condition turns out to be also necessary, but this is not true for general smooth coefficients.

More precisely, let X be a differentiable \mathbb{R} -valued Wiener functional defined on a Wiener space. Assume that there exists a functional ψ such that

$$\sup_{u \in \mathbb{R}} |u|^j E[e^{iuX} X^k \psi] < \infty, \quad j, k \in \mathbb{Z}^+.$$

Thus, it is a regularity condition of the characteristic function, which is a consequence of the nondegeneracy of the Malliavin covariance in the case of Wiener functionals. The functional ψ , which is a random variable satisfying $0 \leq \psi \leq 1$, is a truncation functional extracting from the Wiener space, the portion on which the distribution is regular. If X is almost regular, one may take ψ nearly equal to one. Uniform degeneracy of the Malliavin covariance of the functional $T^{-1/2} \int_0^T f(\theta_0, X_t) dW_t$ can be shown under (A8); see Yoshida (1997) [22].

Bishwal (2009) [13] obtained the rate of convergence to normality of the Itô AMLE and the Fisk–Stratonovich AMLE of the order $O\left(T^{-1/2} \sqrt{\frac{T^2}{n}}\right)$ and $O\left(T^{-1/2} \sqrt{\frac{T^3}{n^2}}\right)$, respectively, under the regularity conditions given above with $q > 16$ for (A4). We obtain the rate of convergence to normality, i.e., Berry–Esseen bound of the order $O\left(T^{-1/2} \sqrt{\frac{T^{p+1}}{n^p}}\right)$ for the QMLE $\theta_{n,T}$ for arbitrary integer p .

We need the following lemma from Michel and Pfanzagl (1971) [23] to prove our main results.

Lemma 1. *Let ξ, ζ and η be any three random variables on a probability space (Ω, \mathcal{F}, P) with $P(\eta > 0) = 1$. Then, for any $\epsilon > 0$, we have*

- (a) $\sup_{x \in \mathbb{R}} |P\{\xi + \zeta \leq x\} - \Phi(x)| \leq \sup_{x \in \mathbb{R}} |P\{\xi \leq x\} - \Phi(x)| + P(|\zeta| > \epsilon) + \epsilon,$
- (b) $\sup_{x \in \mathbb{R}} |P\{\frac{\xi}{\eta} \leq x\} - \Phi(x)| \leq \sup_{x \in \mathbb{R}} |P\{\xi \leq x\} - \Phi(x)| + P\{|\eta - 1| > \epsilon\} + \epsilon.$

2. Main Results

We start with some preliminary lemmas. Let L denote the generator of the diffusion process, $g \in C^2(\mathbb{R})$

$$Lg(x) := f(\theta, x)\dot{g}(x) + \frac{1}{2}\ddot{g}(x).$$

The k -th iterate of L is denoted as L^k . Its domain is $C^{2k}(\mathbb{R})$. We set $L^0 = I$.

Stochastic Taylor formula (Kloeden and Platen (1992) [24]): For a $p + 1$ times continuously differentiable function $g : \mathbb{R} \rightarrow \mathbb{R}$, we have for $t \in [0, T]$ and $p = 1, 2, 3, \dots$

$$g(X_t) = g(X_0) + \sum_{k=1}^p \frac{t^k}{k!} L^k g(X_0) + \int_0^t \dots \int_0^{s_2} L^{p+1} g(X_{s_1}) ds_1 \dots ds_{p+1}.$$

Lemma 2. *With $f(x) = x$, the stochastic Taylor expansion of $\mu(\theta, x)$ is given by*

$$\mu(\theta, X_{t_{i-1}}) := E(X_{t_i} | X_{t_{i-1}}) = \sum_{k=0}^p \frac{h^k}{k!} L^k f(X_{t_{i-1}}) + R(\theta, h^{p+1}, X_{t_{i-1}})$$

where R denotes a function for which there exists a constant C such that

$$R(\theta, h^{p+1}, X_{t_{i-1}}) \leq h^{p+1}C(1 + |X_{t_{i-1}}|)^C.$$

Proof. Applying the stochastic Taylor formula of Florens-Zmirou (1989, Lemma 1) [6], one obtains the result. See also Kloeden and Platen (1992) [24].

Consider the following special cases:

Euler Scheme: For $p = 1$,

$$\mu(\theta, x) = L^0 f(x) + hL^1 f(x) + R(\theta, h^2, x).$$

Milstein Scheme: For $p = 2$,

$$\mu(\theta, x) = L^0 f(x) + hL^1 f(x) + \frac{h^2}{2!} L^2 f(x) + R(\theta, h^3, x).$$

Simpson Scheme: For $p = 4$,

$$\mu(\theta, x) = L^0 f(x) + hL^1 f(x) + \frac{h^2}{2!} L^2 f(x) + \frac{h^3}{3!} L^3 f(x) + \frac{h^4}{4!} L^4 f(x) + R(\theta, h^5, x).$$

Boole Scheme: For $p = 6$,

$$\mu(\theta, x) = L^0 f(x) + hL^1 f(x) + \frac{h^2}{2!} L^2 f(x) + \frac{h^3}{3!} L^3 f(x) + \frac{h^4}{4!} L^4 f(x) + \frac{h^5}{5!} L^5 f(x) + \frac{h^6}{6!} L^6 f(x) + R(\theta, h^7, x). \quad \square$$

Remark 1. For $p = 1$, $\mu(\theta, X_{t_{i-1}}) \approx X_{t_{i-1}} + hf(\theta, X_{t_{i-1}})$. This produces the CLSE. This estimator has been very well studied in the literature (see Shoji (1997) [10]).

Remark 2. Note that the Milstein scheme is equivalent to Stratonovich approximation of the stochastic integral after converting the Itô integral to the Stratonovich integral.

Lemma 3. For all $p \geq 2$, we have

$$E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \leq C \left(\frac{T^{p+1}}{n^{p-1}} \right).$$

Proof. First, we show that, for $p = 2$,

$$E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \leq C \frac{T^4}{n^2}.$$

We emphasize that the Itô formula is a stochastic Taylor formula of order 2. By the Itô formula, we have

$$\begin{aligned} & f(\theta_0, X_t) - f(\theta_0, X_{t_{i-1}}) \\ &= \int_{t_{i-1}}^t \dot{f}(\theta_0, X_u) dX_u + \frac{1}{2} \int_{t_{i-1}}^t \ddot{f}(\theta_0, X_u) du \\ &= \int_{t_{i-1}}^t \dot{f}(\theta_0, X_u) dW_u + \int_{t_{i-1}}^t [\dot{f}(\theta_0, X_u) f(\theta_0, X_u) + \frac{1}{2} \ddot{f}(\theta_0, X_u)] du \\ &=: \int_{t_{i-1}}^t \dot{f}(\theta_0, X_u) dW_u + \int_{t_{i-1}}^t F(\theta_0, X_u) du \end{aligned}$$

where

$$F(\theta_0, X_u) =: \dot{f}(\theta_0, X_u) f(\theta_0, X_u) + \frac{1}{2} \ddot{f}(\theta_0, X_u).$$

We employ Taylor expansion in the local neighborhood of θ_0 . Let $\theta = \theta_0 + T^{-1/2}u, u \in \mathbb{R}$. Then, we have

$$\begin{aligned}
 & E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \\
 = & E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f(\theta, X_t) - f(\theta, X_{t_{i-1}})] dt \right|^2 \\
 = & E \sup_{u \in \mathbb{R}} \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f(\theta_0 + T^{-1/2}u, X_t) - f(\theta_0 + T^{-1/2}u, X_{t_{i-1}})] dt \right|^2 \\
 = & E \sup_{u \in \mathbb{R}} \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f(\theta_0, X_t) - f(\theta_0, X_{t_{i-1}})] dt \right. \\
 & \left. + T^{-1/2}u \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f'(\bar{\theta}, X_t) - f'(\bar{\theta}, X_{t_{i-1}})] dt \right|^2 \\
 = & 2E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f(\theta_0, X_t) - f(\theta_0, X_{t_{i-1}})] dt \right|^2 \\
 & + 2E \sup_{u \in \mathbb{R}} \left| T^{-1/2}u \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f'(\bar{\theta}, X_t) - f'(\bar{\theta}, X_{t_{i-1}})] dt \right|^2 \\
 =: & 2E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f(\theta_0, X_t) - f(\theta_0, X_{t_{i-1}})] dt \right|^2 + 2G_1 \\
 = & E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left[\int_{t_{i-1}}^t \dot{f}(\theta_0, X_u) dW_u + \int_{t_{i-1}}^t f'(\theta_0, X_{t_{i-1}}) F(\theta_0, X_u) du \right] dt \right|^2 + 2G_1 \\
 \leq & 2E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^t \dot{f}(\theta_0, X_u) dW_u dt \right|^2 \\
 & + 2E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^t f'(\theta_0, X_{t_{i-1}}) F(\theta_0, X_u) dudt \right|^2 + 2G_1 \\
 =: & 2(J_1 + J_2) + 2G_1
 \end{aligned}$$

where

$$\begin{aligned}
 J_1 & =: E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^t \dot{f}(\theta_0, X_u) dW_u dt \right|^2, \\
 J_2 & =: E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^t f'(\theta_0, X_{t_{i-1}}) F(\theta_0, X_u) dudt \right|^2, \\
 G_1 & =: E \sup_{u \in \mathbb{R}} \left| T^{-1/2}u \sum_{i=1}^n \int_{t_{i-1}}^{t_i} [f'(\bar{\theta}, X_t) - f'(\bar{\theta}, X_{t_{i-1}})] dt \right|^2
 \end{aligned}$$

and $|\bar{\theta} - \theta_0| \leq |\theta - \theta_0|$. Further

$$\begin{aligned} & E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \\ & \leq 2E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \\ & \quad + 2E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i \right|^2. \end{aligned}$$

By Lemma 2, we have

$$\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}} - hf(\theta, X_{t_{i-1}}) = \sum_{k=2}^p \frac{h^k}{k!} L^k f(\theta, X_{t_{i-1}}) + R(\theta, h^{p+1}, X_{t_{i-1}}).$$

Further

$$\begin{aligned} & \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \int_0^T f(\theta, X_t) dt \\ & = \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i + \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i - \int_0^T f(\theta, X_t) dt \\ & = \sum_{i=1}^n \left[\sum_{k=2}^p \frac{h^k}{k!} L^k f(\theta, X_{t_{i-1}}) + R(\theta, h^{p+1}, X_{t_{i-1}}) \right] + \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i - \int_0^T f(\theta, X_t) dt. \end{aligned}$$

Hence

$$\begin{aligned} & E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \\ & \leq 4(J_1 + J_2) + 2E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n \left[\sum_{k=2}^p \frac{h^k}{k!} L^k f(\theta, X_{t_{i-1}}) + R(\theta, h^{p+1}, X_{t_{i-1}}) \right] \right|^2. \end{aligned}$$

Observe that, with $B_{i,t} := \int_{t_{i-1}}^t f'(\theta_0, X_{t_{i-1}}) \dot{f}(\theta_0, X_u) dW_u$, $1 \leq i \leq n$, we have

$$\begin{aligned} J_1 & = \sum_{i=1}^n E \left(\int_{t_{i-1}}^{t_i} B_{i,t} dt \right)^2 + \sum_{j \neq i=1}^n E \left(\int_{t_{i-1}}^{t_i} B_{i,t} dt \right) \left(\int_{t_{j-1}}^{t_j} B_{j,t} dt \right) \\ & \leq \sum_{i=1}^n (t_i - t_{i-1}) \int_{t_{i-1}}^{t_i} E(B_{i,t}^2) dt \\ & \quad \text{(the last term being zero due to the orthogonality of the integrals)} \\ & \leq \sum_{i=1}^n (t_i - t_{i-1}) \int_{t_{i-1}}^{t_i} \left\{ \int_{t_{i-1}}^t E [f'(\theta_0, X_{t_{i-1}}) \dot{f}(\theta_0, X_u)]^2 du \right\} dt \\ & \leq C \frac{T}{n} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (t - t_{i-1}) dt \text{ (by (A4) and (A3))} \\ & \leq C \frac{T}{n} \sum_{i=1}^n (t_i - t_{i-1})^2 \\ & = C \frac{T^3}{n^2}. \end{aligned}$$

On the other hand, with $A_{i,t} := \int_{t_{i-1}}^t f'(\theta_0, X_{t_{i-1}})F(\theta_0, X_u)du, 1 \leq i \leq n$, we have

$$\begin{aligned}
 J_2 &= E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^t f'(\theta_0, X_{t_{i-1}})F(\theta_0, X_u)dudt \right|^2 \\
 &= E \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} A_{i,t}dt \right|^2 \\
 &= \sum_{i=1}^n E \left(\int_{t_{i-1}}^{t_i} A_{i,t}dt \right)^2 + \sum_{j \neq i=1}^n E \left(\int_{t_{i-1}}^{t_i} A_{i,t}dt \right) \left(\int_{t_{j-1}}^{t_j} A_{j,t}dt \right) \\
 &\leq \sum_{i=1}^n (t_i - t_{i-1}) E \left(\int_{t_{i-1}}^{t_i} A_{i,t}^2 dt \right) + \sum_{j \neq i=1}^n \left\{ E \left(\int_{t_{i-1}}^{t_i} A_{i,t}dt \right)^2 E \left(\int_{t_{j-1}}^{t_j} A_{j,t}dt \right)^2 \right\}^{1/2} \\
 &\leq \sum_{i=1}^n (t_i - t_{i-1}) \int_{t_{i-1}}^{t_i} E(A_{i,t}^2)dt \\
 &\quad + \sum_{j \neq i=1}^n \left\{ (t_i - t_{i-1}) \int_{t_{i-1}}^{t_i} E(A_{i,t}^2)dt (t_j - t_{j-1}) \int_{t_{j-1}}^{t_j} E(A_{j,t}^2)dt \right\}^{1/2}.
 \end{aligned}$$

However, $E(A_{i,t}^2) \leq C(t - t_{i-1})^2$ using (A4) and (A3). On substitution, the last term is dominated by

$$\begin{aligned}
 &C \sum_{i=1}^n (t_i - t_{i-1})^4 + C \sum_{j \neq i=1}^n (t_i - t_{i-1})^2 (t_j - t_{j-1})^2 \\
 &= C \frac{T^4}{n^3} + C \frac{n(n-1)T^4}{2n^4} \leq C \frac{T^4}{n^2}.
 \end{aligned}$$

Thus

$$J_1 + J_2 \leq C \frac{T^4}{n^2}.$$

By the same method, we have

$$G_1 \leq C \frac{T^3}{n^2}.$$

Hence

$$\begin{aligned}
 &E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}})\Delta t_i - \int_0^T f(\theta, X_t)dt \right|^2 \\
 &\leq 2(J_1 + J_2) + 2G_1 \\
 &\leq C \frac{T^4}{n^2}.
 \end{aligned}$$

Thus, the proof for $p = 2$ is complete. Next, we consider the general case $p \geq 3$. Denote

$$J_3 := E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n \left[\sum_{k=2}^p \frac{h^k}{k!} L^k f(\theta, X_{t_{i-1}}) + R(\theta, h^{p+1}, X_{t_{i-1}}) \right] \right|^2.$$

We have

$$\begin{aligned}
 & E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \\
 & \leq 2E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i \right|^2 \\
 & \quad + 2E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n f(\theta, X_{t_{i-1}}) \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \\
 & \leq 2J_3 + 4(J_1 + J_2) + 2G_1.
 \end{aligned}$$

Observe that, by Lemma 2, we have

$$J_3 \leq C \frac{T^{p+1}}{n^{p-1}}.$$

Thus, by combining the bounds for J_1, J_2, J_3 and G_1 , we have

$$E \sup_{\theta \in \Theta} \left| \sum_{i=1}^n [\mu(\theta, X_{t_{i-1}}) - X_{t_{i-1}}] \Delta t_i - \int_0^T f(\theta, X_t) dt \right|^2 \leq C \left(\frac{T^{p+1}}{n^{p-1}} \right).$$

□

The following lemma is from Bishwal (2008) [1].

Lemma 4. *Let*

$$I_T := \frac{1}{TI(\theta_0)} \int_0^T f^2(\theta_0, X_t) dt.$$

Then, under the conditions (A1)–(A8),

$$\sup_{\theta \in \Theta} E[I_T(\theta) - 1]^2 \leq CT^{-1}.$$

The following lemma follows from Theorem 7 in Yoshida (1997) [22].

Lemma 5. *Let*

$$M_T := \frac{1}{\sqrt{TI(\theta_0)}} \int_0^T f(\theta_0, X_t) dW_t.$$

Then, under the conditions (A1)–(A8),

$$\sup_{x \in \mathbb{R}} |P_{\theta_0}\{M_T \leq x\} - \Phi(x)| \leq CT^{-1/2}.$$

Our main result is the following theorem.

Theorem 1. *Under the conditions (A1)–(A8), for any $p \geq 1$, we have*

$$\sup_{x \in \mathbb{R}} \left| P_{\theta} \left\{ \sqrt{TI(\theta)}(\theta_{n,T} - \theta) \leq x \right\} - \Phi(x) \right| = O \left(T^{-1/2} \sqrt{\frac{T^{p+1}}{n^p}} \right).$$

Proof. We start with $p = 1$ and $p = 2$. Let

$$\hat{l}_{n,T}(\theta) := \log \hat{L}_{n,T}(\theta), \text{ and } \tilde{l}_{n,T}(\theta) := \log \tilde{L}_{n,T}(\theta).$$

By Taylor expansion, we have

$$\hat{l}'_{n,T}(\hat{\theta}_{n,T}) = \hat{l}'_{n,T}(\theta_0) + (\hat{\theta}_{n,T} - \theta_0)\hat{l}''_{n,T}(\bar{\theta}_{n,T})$$

where $|\bar{\theta}_{n,T} - \theta| \leq |\hat{\theta}_{n,T} - \theta_0|$. Since $\hat{l}'_{n,T}(\hat{\theta}_{n,T}) = 0$, hence we have

$$\sqrt{TI(\theta_0)}(\hat{\theta}_{n,T} - \theta_0) = -\frac{\frac{1}{\sqrt{TI(\theta_0)}}\hat{l}'_{n,T}(\theta_0)}{\frac{1}{TI(\theta_0)}\hat{l}''_{n,T}(\bar{\theta}_{n,T})} = -\frac{\frac{1}{\sqrt{TI(\theta_0)}}\sum_{i=1}^n f'(\theta_0, X_{t_{i-1}})\Delta W_i}{\frac{1}{TI(\theta_0)}\sum_{i=1}^n f''(\bar{\theta}_{n,T}, X_{t_{i-1}})\Delta t_i} =: \frac{M_{n,T}}{V_{n,T}}$$

Note that

$$V_{n,T} = \frac{1}{TI(\theta_0)} \sum_{i=1}^n f''(\bar{\theta}_{n,T}, X_{t_{i-1}})\Delta t_i = \frac{1}{TI(\theta_0)} \sum_{i=1}^n f''(\bar{\theta}_{n,T}, X_{t_{i-1}})^2 \Delta t_i.$$

However, $E(I_T - 1)^2 \leq CT^{-1}$ from Lemma 4 (see also Pardoux and Veretennikov (2001) [25] and Yoshida (2011) [26]). It can be shown that $E(V_{n,T} - I_T)^2 \leq C\frac{T}{n}$ (see Altmeyer and Chorowski (2018) [27]). Hence

$$E(V_{n,T} - 1)^2 = E[(V_{n,T} - I_T) + (I_T - 1)]^2 \leq C(T^{-1} \sqrt{\frac{T}{n}}).$$

Further, by Lemma 1 (b), we have

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ \sqrt{TI(\theta)}(\hat{\theta}_{n,T} - \theta) \leq x \right\} - \Phi(x) \right| \\ &= \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ \frac{M_{n,T}}{V_{n,T}} \leq x \right\} - \Phi(x) \right| \\ &= \sup_{x \in \mathbb{R}} |P_\theta\{M_{n,T} \leq x\} - \Phi(x)| + P_\theta\{|V_{n,T} - 1| \geq \epsilon\} + \epsilon \\ &\leq C(T^{-1/2} \sqrt{\frac{T^2}{n}}) + \epsilon^{-2}C(T^{-1} \sqrt{\frac{T}{n}}) + \epsilon. \end{aligned}$$

since, by Lemmas 1 (a) and 5, we have

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |P_\theta\{M_{n,T} \leq x\} - \Phi(x)| \\ &\leq \sup_{x \in \mathbb{R}} |P_\theta\{M_T \leq x\} - \Phi(x)| + P_\theta\{|M_{n,T} - M_T| \geq \epsilon\} + \epsilon \\ &\leq CT^{-1/2} + \epsilon^{-2}E|M_{n,T} - M_T|^2 + \epsilon \\ &\leq C(T^{-1/2} \sqrt{\frac{T^2}{n}}) + \epsilon^{-2}C\frac{T}{n} + \epsilon. \end{aligned}$$

Choosing $\epsilon = T^{-1/2}$, we have the result.

On the other hand, by Taylor expansion, we have

$$\tilde{l}'_{n,T}(\tilde{\theta}_{n,T}) = \tilde{l}'_{n,T}(\theta_0) + (\tilde{\theta}_{n,T} - \theta_0)\tilde{l}''_{n,T}(\bar{\tilde{\theta}}_{n,T})$$

where $|\bar{\theta}_{n,T} - \theta| \leq |\tilde{\theta}_{n,T} - \theta|$. Since $\tilde{l}'_{n,T}(\tilde{\theta}_{n,T}) = 0$, hence we have

$$\begin{aligned} & \sqrt{TI(\theta_0)}(\tilde{\theta}_{n,T} - \theta_0) \\ = & -\frac{\frac{1}{\sqrt{TI(\theta_0)}}\tilde{l}'_{n,T}(\theta_0)}{\frac{1}{TI(\theta_0)}\tilde{l}''_{n,T}(\bar{\theta}_{n,T})} \\ = & -\left\{ \frac{1}{\sqrt{TI(\theta_0)}} \left\{ \frac{1}{2} \sum_{i=1}^n [f'(\theta_0, X_{t_{i-1}}) + f'(\theta_0, X_{t_i})] \Delta W_i \right. \right. \\ & + \frac{1}{2} \sum_{i=1}^n [f'(\theta_0, X_{t_{i-1}}) + f'(\theta_0, X_{t_i})] \int_{t_{i-1}}^{t_i} f'(\theta_0, X_t) dt \\ & \left. \left. - \frac{h}{2} \sum_{i=1}^n \dot{f}'(\theta_0, X_{t_{i-1}}) - h \sum_{i=1}^n f(\theta_0, X_{t_{i-1}}) f'(\theta_0, X_{t_{i-1}}) \right\} \right\} \\ & \times \left\{ \frac{1}{TI(\theta_0)} \left\{ \frac{1}{2} \sum_{i=1}^n [f''(\bar{\theta}_{n,T}, X_{t_{i-1}}) + f''(\bar{\theta}_{n,T}, X_{t_i})] \Delta W_i \right. \right. \\ & - \frac{1}{2} \sum_{i=1}^n [f'(\bar{\theta}_{n,T}, X_{t_{i-1}}) + f'(\bar{\theta}_{n,T}, X_{t_i})] \int_{t_{i-1}}^{t_i} f'(\bar{\theta}_{n,T}, X_t) dt \\ & \left. \left. - \frac{h}{2} \sum_{i=1}^n \dot{f}''(\bar{\theta}_{n,T}, X_{t_{i-1}}) - h \sum_{i=1}^n f(\bar{\theta}_{n,T}, X_{t_{i-1}}) f''(\bar{\theta}_{n,T}, X_{t_{i-1}}) \right. \right. \\ & \left. \left. - h \sum_{i=1}^n f'^2(\bar{\theta}_{n,T}, X_{t_{i-1}}) \right\} \right\}^{-1} \\ =: & \{R_{n,T}\} \{S_{n,T}\}^{-1}. \end{aligned}$$

Let $\lim S_{n,T} = S_T$ in L_2 as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$. Similar to Lemma 4, it can be shown that $E(S_T - 1)^2 \leq CT^{-1}$ (see also Pardoux and Veretennikov (2001) [25] and Yoshida (2011) [26]). It can be shown that $E[(S_{n,T} - S_T)^2] \leq C\frac{T}{n}$ (see Altmeyer and Chorowski (2018) [27]). Hence

$$E(S_{n,T} - 1)^2 = E[(S_{n,T} - S_T) + (S_T - 1)]^2 \leq C(T^{-1} \sqrt{\frac{T}{n}}).$$

Thus, by Lemma 1 (b), we have

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ \sqrt{TI(\theta)}(\tilde{\theta}_{n,T} - \theta) \leq x \right\} - \Phi(x) \right| \\ = & \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ \frac{R_{n,T}}{S_{n,T}} \leq x \right\} - \Phi(x) \right| \\ = & \sup_{x \in \mathbb{R}} |P_\theta \{R_{n,T} \leq x\} - \Phi(x)| + P_\theta \{|S_{n,T} - 1| \geq \epsilon\} + \epsilon \\ \leq & C(T^{-1/2} \sqrt{\frac{T^3}{n^2}}) + \epsilon^{-2} C(T^{-1} \sqrt{\frac{T}{n}}) + \epsilon. \end{aligned}$$

since, by Lemmas 1 (a) and 5, we have

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |P_\theta\{R_{n,T} \leq x\} - \Phi(x)| \\ & \leq \sup_{x \in \mathbb{R}} |P_\theta\{M_T \leq x\} - \Phi(x)| + P_\theta\{|R_{n,T} - M_T| \geq \epsilon\} + \epsilon \\ & \leq CT^{-1/2} + \epsilon^{-2}E|R_{n,T} - M_T|^2 + \epsilon \\ & \leq CT^{-1/2} + \epsilon^{-2}C\frac{T^3}{n^2} + \epsilon. \end{aligned}$$

Choosing $\epsilon = T^{-1/2}$, we have the result.

Now, we study the general case for arbitrary p . By Taylor expansion, we have

$$K'_{n,T}(\theta_{n,T}) = K'_{n,T}(\theta_0) + (\theta_{n,T} - \theta_0)K''_{n,T}(\bar{\theta}_{n,T})$$

where $|\bar{\theta}_{n,T} - \theta| \leq |\theta_{n,T} - \theta_0|$. Since $K'_{n,T}(\theta_{n,T}) = 0$, hence we have

$$\sqrt{TI(\theta_0)}(\theta_{n,T} - \theta_0) = -\frac{\frac{1}{\sqrt{TI(\theta_0)}}K'_{n,T}(\theta_0)}{\frac{1}{TI(\theta_0)}K''_{n,T}(\bar{\theta}_{n,T})} = -\frac{\frac{1}{\sqrt{TI(\theta_0)}}\sum_{i=1}^n m'(\theta_0, X_{t_{i-1}})\Delta W_i}{\frac{1}{TI(\theta_0)}\sum_{i=1}^n m''(\bar{\theta}_{n,T}, X_{t_{i-1}})\Delta t_i} =: \frac{N_{n,T}}{U_{n,T}}$$

Note that

$$U_{n,T} = \frac{1}{TI(\theta_0)} \sum_{i=1}^n m''(\bar{\theta}_{n,T}, X_{t_{i-1}})\Delta t_i = \frac{1}{TI(\theta_0)} \sum_{i=1}^n m'(\bar{\theta}_{n,T}, X_{t_{i-1}})^2 \Delta t_i.$$

Let $\lim U_{n,T} = U_T$ in L_2 as $T \rightarrow \infty$ and $\frac{T}{n} \rightarrow 0$. Similar to Lemma 4, it can be shown that $E(U_T - 1)^2 \leq CT^{-1}$ (see also Pardoux and Veretennikov (2001) [25] and Yoshida (2011) [26]). It can be shown that $E(U_{n,T} - U_T)^2 \leq C\frac{T}{n}$ (see Altmeyer and Chorowski (2018) [27]). Hence

$$E(U_{n,T} - 1)^2 = E[(U_{n,T} - U_T) + (U_T - 1)]^2 \leq C(T^{-1} \sqrt{\frac{T}{n}}).$$

Further, by Lemma 1 (b), we have

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ \sqrt{TI(\theta)}(\theta_{n,T} - \theta) \leq x \right\} - \Phi(x) \right| \\ & = \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ \frac{N_{n,T}}{U_{n,T}} \leq x \right\} - \Phi(x) \right| \\ & = \sup_{x \in \mathbb{R}} |P_\theta\{N_{n,T} \leq x\} - \Phi(x)| + P_\theta\{|U_{n,T} - 1| \geq \epsilon\} + \epsilon \\ & \leq C(T^{-1/2} \sqrt{\frac{T^{p+1}}{n^p}}) + \epsilon^{-2}C(T^{-1} \sqrt{\frac{T}{n}}) + \epsilon. \end{aligned}$$

since, by Lemmas 1 (a) and 5, we have

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |P_\theta\{N_{n,T} \leq x\} - \Phi(x)| \\ & \leq \sup_{x \in \mathbb{R}} |P_\theta\{M_T \leq x\} - \Phi(x)| + P_\theta\{|N_{n,T} - M_T| \geq \epsilon\} + \epsilon \\ & \leq CT^{-1/2} + \epsilon^{-2}E|N_{n,T} - M_T|^2 + \epsilon \\ & \leq CT^{-1/2} + \epsilon^{-2}C\frac{T^{p+1}}{n^p} + \epsilon. \end{aligned}$$

Choosing $\epsilon = T^{-1/2}$, we have the result. \square

Remark 3. With $p = 1$, for the Euler scheme, which produces the conditional least squares estimator, one obtains the rate $O\left(T^{-1/2} \sqrt{\frac{T^2}{n}}\right)$. With $p = 2$, for the Milstein scheme, one obtains the rate $O\left(T^{-1/2} \sqrt{\frac{T^3}{n^2}}\right)$. With $p = 4$, for the Simpson scheme, one obtains the rate $O\left(T^{-1/2} \sqrt{\frac{T^5}{n^4}}\right)$. With $p = 6$, for the Boole scheme, one obtains the rate $O\left(T^{-1/2} \sqrt{\frac{T^7}{n^6}}\right)$. Thus, the higher the p , the sharper the bound. Thus, the Itô/Euler scheme gives the first-order QMLE, the Milstein/Stratonovich scheme produces the second-order QMLE, the Simpson scheme produces the fourth-order QMLE and the Boole scheme produces the sixth-order QMLE. See Bishwal (2011) [28] for a connection of this area to the stochastic moment problem and hedging of generalized Black–Scholes options.

3. Example

Consider the stochastic differential equation

$$\begin{aligned} dX_t &= \theta \frac{X_t}{\sqrt{1 + X_t^2}} dt + dW_t, \quad t \geq 0 \\ X_0 &= x_0 \end{aligned}$$

The solution to the above SDE is called the hyperbolic diffusion process because it has a hyperbolic stationary distribution when $\theta < 0$. The process has nonlinear drift and the process is stationary and ergodic, which distinguishes this from a linear drift case, such as the Ornstein–Uhlenbeck process and the Cox–Ingersoll–Ross process, which have linear drift. This model verifies assumption (A3). In fact, the stationary density is proportional to $\exp(\theta \sqrt{1 + X_t^2})$. It is not possible to calculate the conditional expectation for the hyperbolic diffusion process and hence one needs a higher-order Taylor expansion approach.

Remark 4 (Concluding Remark). *It would be interesting to extend the results of the paper to diffusions with jumps using the strong stochastic Taylor expansion with jumps results in Chapter 6 of Kloeden and Bruti-Liberati (2010) [29].*

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bishwal, J.P.N. *Parameter Estimation in Stochastic Differential Equations*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 2008.
2. Bishwal, J.P.N. *Parameter Estimation in Stochastic Volatility Models*; Springer Nature Switzerland AG: Cham, Switzerland, 2021; (forthcoming).
3. Dorogovcev, A.J. The consistency of an estimate of a parameter of a stochastic differential equation. *Theory Prob. Math. Stat.* **1976**, *10*, 73–82.
4. Kasonga, R.A. The consistency of a nonlinear least squares estimator from diffusion processes. *Stoch. Proc. Appl.* **1988**, *30*, 263–275. [[CrossRef](#)]
5. Prakasa Rao, B.L.S. Asymptotic theory for non-linear least squares estimator for diffusion processes. *Math. Oper. Stat. Ser. Stat.* **1983**, *14*, 195–209. [[CrossRef](#)]
6. Florens-Zmirou, D. Approximate discrete time schemes for statistics of diffusion processes. *Statistics* **1989**, *20*, 547–557. [[CrossRef](#)]
7. Kessler, M. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **1997**, *24*, 211–229. [[CrossRef](#)]
8. Liptser, R.S.; Shiriyayev, A.N. *Statistics of Random Processes I*; Springer: New York, NY, USA, 1977.
9. Yoshida, N. Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.* **1992**, *41*, 220–242. [[CrossRef](#)]
10. Shoji, I. A note on asymptotic properties of estimator derived from the Euler method for diffusion processes at discrete times. *Stat. Probab. Lett.* **1997**, *36*, 153–159. [[CrossRef](#)]

11. Bishwal, J.P.N.; Bose, A. Rates of convergence of approximate maximum likelihood estimators in the Ornstein-Uhlenbeck process. *Comput. Math. Appl.* **2001**, *42*, 23–38. [[CrossRef](#)]
12. Bishwal, J.P.N. Uniform rate of weak convergence for the minimum contrast estimator in the Ornstein-Uhlenbeck process. *Methodol. Comput. Appl. Probab.* **2010**, *12*, 323–334. [[CrossRef](#)]
13. Bishwal, J.P.N. Berry-Esseen inequalities for discretely observed diffusions. *Monte Carlo Methods Appl.* **2009**, *15*, 229–239. [[CrossRef](#)]
14. Bishwal, J.P.N. M-Estimation for discretely sampled diffusions. *Theory Stoch. Process.* **2009**, *15*, 62–83.
15. Bishwal, J.P.N. Milstein approximation of posterior density of diffusions. *Int. J. Pure Appl. Math.* **2011**, *68*, 403–414.
16. Bishwal, J.P.N. Conditional least squares estimation in diffusion processes based on Poisson sampling. *J. Appl. Probab. Stat.* **2010**, *5*, 169–180.
17. Bishwal, J.P.N. Some new estimators of integrated volatility. *Am. Open J. Stat.* **2011**, *1*, 74–80. [[CrossRef](#)]
18. Bhattacharya, R.N.; Ranga Rao, R. *Normal Approximation and Asymptotic Expansion*; Wiley: New York, NY, USA, 1976.
19. Ikeda, N.; Watanabe, S. *Stochastic Differential Equations and Diffusion Processes*, 2nd ed.; North-Holland: Amsterdam, The Netherlands; Kodansha Ltd.: Tokyo, Japan, 1989.
20. Nualart, D. *Malliavin Calculus and Related Topics*; Springer: Berlin, Germany, 1995.
21. Bally, V. On the connection between the Malliavin covariance matrix and Hörmander condition. *J. Funct. Anal.* **1991**, *96*, 219–255. [[CrossRef](#)]
22. Yoshida, M. Malliavin calculus and asymptotic expansion for martingales. *Probab. Theory Relat. Fields* **1997**, *109*, 301–342. [[CrossRef](#)]
23. Michel, R.; Pfanzagl, J. The accuracy of the normal approximation for minimum contrast estimate. *Zeit. Wahr. Verw. Gebiete* **1971**, *18*, 73–84. [[CrossRef](#)]
24. Kloeden, P.E.; Platen, E. *Numerical Solution of Stochastic Differential Equations*; Springer: Berlin, Germany, 1992.
25. Pardoux, E.; Veretennikov, A.Y. On the Poisson equation and diffusion equation I. *Ann. Probab.* **2001**, *29*, 1061–1085. [[CrossRef](#)]
26. Yoshida, N. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Stat. Math.* **2011**, *63*, 431–479. [[CrossRef](#)]
27. Altmeyer, R.; Chorowski, J. Estimation error for occupation functionals of stationary Markov processes. *Stoch. Proc. Appl.* **2018**, *128*, 1830–1848. [[CrossRef](#)]
28. Bishwal, J.P.N. Stochastic moment problem and hedging of generalized Black-Scholes options. *Appl. Numer. Math.* **2011**, *61*, 1271–1280. [[CrossRef](#)]
29. Kloeden, P.E.; Bruti-Liberati, N. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*; Springer: Berlin, Germany, 2010.