DOI: 10.28919/ejma.2026.6.13
Eur. J. Math. Appl. (2026)6:13
URL: http://ejma.euap.org
© 2026 European Journal of Mathematics and Applications

RATE OF CONVERGENCE FOR THE BERNSTEIN-VON MISES THEOREM AND THE BAYES ESTIMATOR OF VOLATILITY

JAYA P. N. BISHWAL

Abstract. The paper obtains sharp rates of convergence in consistency and asymptotic normality of the quasi-posterior distributions and regular quasi-Bayes estimators, for smooth priors and loss functions, of the unknown parameter in the diffusion coefficient in a nonlinear diffusion process from n discrete observations when $n \to \infty$.

1. Introduction and Preliminaries

The Bernstein-von Mises theorem (rather called the Bernstein-von Mises phenomenon in the terminology of Le Cam and Yang [21]) or the Bayesian central limit theorem states that for large number of observations, suitably centered and normalized posterior distribution behaves like the normal distribution. In particular cases, the theorem was shown by Laplace, Bernstein and von-Mises in the early period of last century. See Le Cam [20], Le Cam and Yang [21] for a history. A complete proof in the i.i.d. case was given by Le Cam [18]. Since then many authors have extended the theorem to many dependent observations cases or refined the theorem in the i.i.d. case. Borwanker, Kallianpur and Prakasa Rao [8] extended the theorem to discrete time stationary Markov process. Strasser [30] and Hipp and Michel [14] estimated the speed of convergence in this theorem in the i.i.d. case. Prakasa Rao [24] obtained the rate of convergence in the Bernstein - von Mises theorem for Markov processes extending the work of Hipp and Michel [14] in the i.i.d. case. Note that Bernstein-von Mises theorem fails in infinite dimensional parameter case even in i.i.d. case, see Freedman [11].

In the i.i.d. case, it is well established that, for large sample size the mass of the posterior distribution concentrates in arbitrarily small neighbourhoods of the true parameter value (see Doob [10], Schwarz [27], Le Cam [19]). This phenomenon is known as *posterior consistency*. Strasser [30] estimated the rate at which the posterior distribution concentrates around the true parameter value. Recently, rates of convergence in posterior consistency in infinite dimensional parameter situation was studied by Ghoshal, Ghosh and Van der Vaart [12], Shen and Wasserman [28] and Huang [15] in the i.i.d. case. We obtain rates of convergence in posterior consistency for the nonlinear ergodic diffusion process in Theorems 2.2 and 2.3 extending Strasser [30].

Department of Mathematics and Statistics, University of North Carolina at Charlotte, 376 Fretwell Bldg, 9201 University City Blvd., Charlotte, NC 28223-0001, USA

E-mail address: J.Bishwal@charlotte.edu.

2020 *Mathematics Subject Classification.* 60F05, 60H10, 62F12, 62F15, 62M05, 91G30, 91G60.

Key words and phrases. Itô stochastic differential equation, Brownian motion, ergodic diffusion process, asymptotic analysis, posterior distribution, Bernstein-von Mises theorem, Bayes estimator, Berry-Esseen inequality, diffusion coefficient, volatility, rate of weak convergence.

The Bernstein-von Mises theorem for linear homogeneous diffusion processes was proved by Prakasa Rao [25]. Bose [9] extended the work of Prakasa Rao [25] to nonlinear homogeneous diffusion processes and Mishra [22] and Harison [13] extended the work of Bose [9] to non-homogeneous diffusion processes. Prakasa Rao [26] extended his work to linear homogeneous diffusion field. Bishwal [1] proved Bernstein-von Mises theorem for Hilbert space valued diffusions. Bishwal [3, 6]studied Bernstein-von Mises theorem and asymptotics of Bayes estimators for parabolic stochastic partial differential equations through the Fourier method.

Not much is known beyond the higher order asymptotics of posteriors for drift estimation in nonlinear diffusions. For the standard Ornstein-Uhlenbeck process, Bishwal [2] obtained $O(T^{-1/2})$ rate of convergence of the normalized posterior density to normal density and $o(T^{-1/2})$ rate for the P_θ^T probabilities of the exceptional set by using the method of Strasser [31] for the i.i.d. case. In Theorem 2.9 of Bishwal [4], we obtained same rate of convergence for posterior distributions of the drift parameter in stationary ergodic diffusion process based on continuous observation of the process. Berry-Esseen bounds of approximate Bayes estimators for the drift parameter in the discretely observed Ornstein-Uhlenbeck process was studied in Bishwal [7]. The focus of the current paper is on the Bayes estimation of the diffusion parameter or the volatility parameter based on discrete observations in a finite time interval.

We start with the preliminaries of maximum quasi-likelihood estimator and its asymptotics. Consider the SDE

$$dY_t = b_t dt + \sqrt{\theta}\sigma_t dW_t,\ t \in [0,1] \tag{1.1}$$

where b_t and σ_t are two Itô processes and our aim is to estimate θ based on discrete observations at $0 = t_0 < t_1 < t_2 \ldots < t_n = 1$. We observe the data set $\{Y_{t_i}, \sigma_{t_i}, i = 0, 1, 2, \ldots, n\}$ which is denoted by Y_0^n with $t_i = i/n,\ i = 0, 1, 2, \ldots, n$. Let $\Delta Y_j := Y_{t_j} - Y_{t_{j-1}}$.

The quasi-likelihood random field $\mathbb{H}_n(\theta)$ is defined as

$$\mathbb{H}_n(\theta) := -\frac{1}{2}\sum_{j=1}^{n}\left\{\log(\theta\sigma_{t_{j-1}}^2) + (h\theta\sigma_{t_{j-1}}^2)^{-1}(\Delta Y_j)^2\right\} \tag{1.2}$$

and the quasi maximum likelihood estimator (QMLE) of θ is given by

$$\widehat{\theta}_n := \sum_{i=1}^{n}\left(\frac{Y_{t_i} - Y_{t_{i-1}}}{\sigma_{t_{i-1}}}\right)^2.$$

In finance literature, the parameter θ is called the volatility parameter. QMLE has the following property:

$$\frac{\sqrt{n}}{\sqrt{2\theta}}(\widehat{\theta}_n - \theta) \to^{\mathcal{D}} \mathcal{N}(0,1)$$

as $n \to \infty$, see Yoshida [33]. Let $\Delta W_i := W_{t_i} - W_{t_{i-1}}$.

Lemma 1.1.

$$\frac{\sqrt{n}}{\sqrt{2\theta}}(\widehat{\theta}_n - \theta) = M_n + \frac{1}{\sqrt{n}}N_n \tag{1.3}$$

where

$$M_n := \frac{1}{\sqrt{2n}}\sum_{i=1}^{n}[(\sqrt{n}\Delta W_i)^2 - 1], \tag{1.4}$$

$$N_n := \frac{1}{\sqrt{6n}} \sum_{i=1}^{n} \frac{\sqrt{3}\sigma^{[1]}_{t_{i-1}}}{\sigma_{t_{i-1}}} [(\sqrt{n}\Delta W_i)^3 - 3\sqrt{n}\Delta W_i]$$

$$+\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\sqrt{2}\sigma^{[1]}_{t_{i-1}}}{\sigma_{t_{i-1}}} + \frac{\sqrt{2}b_{t_{i-1}}}{\sqrt{\theta}\sigma_{t_{i-1}}} \right) (\sqrt{n}\Delta W_i) + \frac{1}{n} \sum_{i=1}^{n} F_{t_{i-1}} + O_P(n^{-1/2}) \tag{1.5}$$

and

$$F_t := \frac{\sqrt{2}\sigma^{[0]}_t}{2\sigma_t} + \frac{(\sigma^{[1]}_t)^2}{2\sqrt{2}\sigma^2_t} + \frac{\sqrt{2}b^{[1]}_t}{2\sqrt{\theta}\sigma_t} + \frac{\sqrt{2}b^2_t}{2\theta\sigma^2_t}. \tag{1.6}$$

Sketch of Proof. Using differentiability in θ of the quasi-likelihood random field $\mathbb{H}_n(\theta)$ along with repeated use of Burkholder's inequality for martingales and Itô's formula, we obtain the stochastic expansion of $\widehat{\theta}_n$. Details are postponed. □

From Yoshida [33], we have

$$\sup_{x\in\mathbb{R}} |P(M_n \le x) - \Phi(x)| \le Cn^{-1/2} \tag{1.7}$$

and for any positive ϵ,

$$P\left(\sup_{t\in[0,1]} |\Delta \bar{M}_{j,n}(t)| > \epsilon \right) = O(n^{-1/2}) \tag{1.8}$$

where

$$\bar{M}_{j,n}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} G_j(t_{i-1}) H_j(\sqrt{n}\Delta W_j), \quad j = 1, 2, 3 \tag{1.9}$$

and $H_j, \quad j = 1, 2, 3$ are Hermite polynomials.

Recall that the $j-$th Hermite polynomial H_j is given by

$$H_j(x) = (-1)^j e^{x^2/2} \frac{d^j}{dx^j}(e^{-x^2/2}), \quad j \ge 1.$$

Thus $H_0(x) = 1,\ H_1(x) := x,\ H_2(x) := (x^2-1)/\sqrt{2},\ H_3(x) := (x^3-3x)/\sqrt{6}.$

Therefore, the sequence $(\bar{M}_{j,n}, n \in \mathbb{N})$ is C-tight (Jacod and Shiryayev [17, p. 315, Proposition 3.26]).

Observe that M_n has terms involving Hermite polynomial H_2 while N_n has terms involving Hermite polynomials H_1 and H_3. Here M_n is a terminal value of a continuous martingale $(M^n_t)_{t\in[0,1]}$, i.e., $M_n = M^n_1$ which converges to a mixed normal limit in the functional sense and N_n is a tight sequence of random variables.

The equations (1.4) and (1.5) can be expressed in terms of Hermite polynomials as

$$M_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_2(\sqrt{n}\Delta W_i), \tag{1.10}$$

$$N_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\sigma^{[1]}_{t_{i-1}}}{\sigma_{t_{i-1}}} H_3(\sqrt{n}\Delta W_i)$$

$$+\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\sqrt{2}\sigma^{[1]}_{t_{i-1}}}{\sigma_{t_{i-1}}} + \frac{\sqrt{2}b_{t_{i-1}}}{\sqrt{\theta}\sigma_{t_{i-1}}} \right) H_1(\sqrt{n}\Delta W_i)$$

$$+\frac{1}{n}\sum_{i=1}^{n} F_{t_{i-1}} + O_P(n^{-1/2}) \tag{1.11}$$

Alternatively,

$$M_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_2(t_{i-1})H_2(\sqrt{n}\Delta W_i),$$

$$N_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_3(t_{i-1})H_3(\sqrt{n}\Delta W_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_1(t_{i-1})H_1(\sqrt{n}\Delta W_i)$$

$$+\frac{1}{n}\sum_{i=1}^{n} F_{t_{i-1}} + O_P(n^{-1/2})$$

where

$$G_1(t) := \frac{\sqrt{2}\sigma_t^{[1]}}{\sigma_t} + \frac{\sqrt{2}b_t}{\sqrt{\theta}\sigma_t}, \quad G_2(t) := 1, \quad G_3(t) := \frac{\sqrt{3}\sigma_t^{[1]}}{\sigma_t},$$

$$\bar{F}_n(t) := \frac{1}{n}\sum_{i=1}^{[nt]} F_{t_{i-1}} \tag{1.12}$$

and F_t is defined in (1.6). Thus

$$\frac{1}{\sqrt{n}}N_n := \frac{1}{n}\sum_{i=1}^{n} G_3(t_{i-1})H_3(\sqrt{n}\Delta W_i) + \frac{1}{n}\sum_{i=1}^{n} G_1(t_{i-1})H_1(\sqrt{n}\Delta W_i)$$

$$+\frac{1}{n}\bar{F}_n(t) + O_P(n^{-1}).$$

For $j = 2$, $G_2(t) = 1$, $H_2(x) = (x^2 - 1)/\sqrt{2}$, in this case $\bar{M}_{2,n} = M_n$.

$$M_{n,t} := 2\sqrt{n}\sum_{i=1}^{n}\int_{t\wedge t_{i-1}}^{t\wedge t_i}\int_{t_{i-1}}^{s} dW_u dW_s. \tag{1.13}$$

$$\bar{M}^{\xi}_{n,t} = 4n\sqrt{n}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}(t_i - u)(W_u - W_{t_{i-1}})dW_u. \tag{1.14}$$

The sequence $(\bar{M}^{\xi}_{n,t}, n \in \mathbb{N})$ is also C-tight. It is easy to show that

$$\xi_n = \bar{M}^{\xi}_{n,1} \tag{1.15}$$

where ξ_n is defined with the original $(\widetilde{F}_t)$-bracket $\langle M_{n,\cdot}\rangle$.

Mixed CLT for generalized power variation along with its Edgeworth expansion is proved in Theorem 4.1 and 4.2 respectively in Podoloskij and Yoshida [23].

In Theorem 7 (based on continuous observation) and Theorem 14 (based on discrete observations) in Yoshida [34], asymptotic normality of simultaneous and adaptive Bayes estimators are obtained respectively. The adaptive Bayes approach is better than the simultaneous Bayes approach from a computational point of view since it can supress the dimension of the integral in the computation of the Bayes estimate. See also Theorem 2.9 in Yoshida [36].

Yoshida [35] generalized the quasi likelihood analysis to partial quasi likelihood analysis and obtained limit theorems for quasi maximum likelihood estimators (QMLEs) and quasi Bayes estimators (QBEs). The basic idea is conditioning by partial information. Conditional polynomial type large deviations (PLD) play a central role. PLD is the key to QLA. Once a

PLD is established, one can obtain a strong mode of convergence of the quasi likelihood random field and the associated estimators.

Yoshida [35] also studied *stochastic regression model* for volatility in random environment which is given by

$$dY_t = b(\gamma, X_t)dt + \sigma(\gamma, X_t, L_t, \theta)dW_t,\ t \in [0,T] \tag{1.16}$$

where $(X_t, t \geq 0)$ is a progressively measurable process, e.g., a latent diffusion process with good mixing property.

The variable γ is a measurable *random scenery* which describes the partially observed state of the random environment and $(b(\gamma,\cdot), \sigma(\gamma,\cdot))$ are regarded as *random environment* in space-time, L_t is a long-memory process, e.g., a fractional Brownian motion.

The parameter $\theta \in \Theta$ needs to be estimated based on discrete observations of $(\gamma, X_{t_j}, Y_{t_j}),\ t_j = jT/n,\ j = 0, 1, \ldots, n$.

For example, with $\sigma(\gamma, X_t, \theta) = \theta\sqrt{1+X_t^2}$,

$$dY_t = b_t dt + \theta\sqrt{1+X_t^2}dW_t,\ t \in [0,T], \tag{1.17}$$

$$X_t = \int_0^t e^{B_s^4} X_s ds + \widetilde{W}_t,\ t \in [0,T] \tag{1.18}$$

where B_t is an independent Wiener process.

The quasi-likelihood random field $\mathbb{H}_n(\theta)$ is defined as

$$\mathbb{H}_n(\theta) := -\frac{1}{2}\sum_{j=1}^{n}\left\{\log\sigma^2(\gamma, X_{t_{j-1}}, \theta) + (h\sigma^2(\gamma, X_{t_{j-1}}, \theta))^{-1}(\Delta Y_j)^2\right\}. \tag{1.19}$$

The QMLE is defined as

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta}\mathbb{H}_n(\theta). \tag{1.20}$$

The QBE (with respect to quadratic loss) is defined as

$$\widetilde{\theta}_n = \frac{\int_\Theta \theta\exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta}{\int_\Theta \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta} \tag{1.21}$$

where $\pi(\cdot)$ is the prior density which is continuous and satisfies $0 < \inf_{\theta\in\Theta}\pi(\theta) \leq \sup_{\theta\in\Theta}\pi(\theta) < \infty$.

The asymptotic statistical decision theory by Le Cam and Hajek has been given a lucid perspective by the Ibragimov-Hasminskii theory on the convergence of the likelihood ratio random field. Their scheme has been applied to stochastic processes by Kutoyants and this program is called the IHK (Ibragimov-Hasminskii-Kutoyants) program. Ibragimov and Hasminskii [16] introduced the notion of weak convergence of likelihood ratio random field, from which the asymptotic properties of likelihood estimators, i.e., the maximum likelihood estimator and Bayes estimator are induced in an unified way. L^p-boundedness of an estimator is key to develop the asymptotic theory which follows from large deviation inequality of the likelihood ratio random field. Large deviation inequality is extremely important since it quantitatively estimates the tail of the likelihood ratio random field. Ibragimov-Hasminskii [16] used the exponential large deviation inequality while Yoshida [34] relaxed this condition to polynomial type large deviation inequality.

The IHK (Ibragimov-Hasminskii-Kutoyants) program ensures that asymptotic properties of an estimator follows directly from the convergence of the likelihood ratio random field if a large deviation estimate (exponential type) exists. The quasi-likelihood analysis (QLA) proved a

polynomial type large deviation (PLD) inequality to go through a bottleneck of the program (Yoshida [34]). A conclusion of the QLA is that if the quasi-likelihood random field is locally asymptotically quadratic (LAQ) and if a key index reflecting identifiability the random field has is non-degenerate, then the PLD inequality is always valid, and as a result, the IHK program can run. Of course, exponential type large deviation estimate give strong properties of estimators such as convergence of moments.

Consider the random field

$$Z_n(u) := \exp(\mathbb{H}_n(\theta^* + a_n u) - \mathbb{H}_n(\theta^*)),\ u \in U_n \tag{1.22}$$

where

$$U_n := \{u \in \mathbb{R} : \theta_n^\dagger(u) \in \Theta\} \tag{1.23}$$

and $\theta_n^\dagger(u) := \theta^* + a_n u$. The rate a_n satisfies $a_n \to 0$ such that $n \to \infty$.

Define Δ_n and Γ_n by

$$\Delta_n := \mathbb{H}_n'(\theta^*)a_n, \quad \Gamma_n(\theta) := -\mathbb{H}_n''(\theta)a_n^2$$

We can show the large deviations and modulus of continuity of the random field $Z_n(u)$.

Large Deviations

For every $l > 0$, there exists a constant C such that

$$\sup_n \sup_r P\left(\sup_u Z_n(u) \geq \epsilon(r)\right) \leq Cr^{-l} \tag{1.24}$$

where $\epsilon(r) \to 0$ as $r \to \infty$.

Modulus of Continuity

$$\lim_{n\to\infty}\Big(\sup_n P\left\{\sup|\log Z_n(u_2) - \log Z_n(u_1)|;\ u_1, u_2 \in B_{c,n},\right.$$
$$\left.|u_2 - u_1| \leq \delta\right\} > \epsilon) \to^P 0 \tag{1.25}$$

as $\delta \to 0$ for every $\epsilon > 0,\ c > 0$ where

$$B_{c,n} := \{u \in \mathbb{R} : |u| \leq c,\ \theta_n^\dagger(u) \in \Theta\}.$$

Suppose that

$$\sup_{u\in U_n} |\Gamma_n(\theta^* + a_n u) - \Gamma| \to^P 0$$

Then the random field $Z_n(u)$ has the LAQ (local asymptotic quadratic) representation:

$$Z_n(u) = \exp\left(u\Delta_n - \frac{1}{2}\Gamma u^2 + r_n(u)\right) \tag{1.26}$$

where $r_n(u) \to^P 0$ as $n \to \infty$ for every $u \in \mathbb{R}$.

This suggest that the tail of the random field $Z_n(u)$ is like a Gaussian kernel. This fact is stated as a polynomial type large deviation (PLD) inequality. The PLD ensures the L^p-boundedness of the scaled error of the QLA estimators. A common L^p-inequality based on some kind of orthogonality such as martingale (helps to use Garcia-Rodemich-Rumsey (GRR) inequality for martingales which improves the Burkholder-Davis-Gundy (BDG) inequality) or

mixing property combined with an uniform estimate like the *Sobolev's embedding inequality* suffice to verify the regularity conditions.

Let $\mathcal{G}$ be sigma field such that $\sigma[\gamma] \subset \mathcal{G} \subset \mathcal{F}$. We obtain

$$\Delta_n \to^{\mathcal{D}_f(\mathcal{G})} \Delta.$$

Further

$$Z_n \to^{\mathcal{D}_f(\mathcal{G})} Z \tag{1.27}$$

where

$$Z(u) := \exp\left(\Delta u - \frac{1}{2}\Gamma u^2\right). \tag{1.28}$$

Let

$$\widehat{u}_n := a_n^{-1}(\widehat{\theta}_n - \theta^*) \tag{1.29}$$

be the normalized QMLE and

$$\widetilde{u}_n := a_n^{-1}(\widetilde{\theta}_n - \theta^*) \tag{1.30}$$

be the normalized QBE.

Then

$$\widehat{u}_n - \Gamma^{-1}\Delta_n \to^P 0 \tag{1.31}$$

and

$$\widetilde{u}_n - \Gamma^{-1}\Delta_n \to^P 0. \tag{1.32}$$

In particular, $\widehat{\theta}_n \to^P \theta$ and $\widetilde{\theta}_n \to^P \theta$ as $n \to \infty$.

$$E[f(\widehat{u}_n)\Phi] \to E[f(\widehat{u})\varphi]$$

and

$$E[f(\widetilde{u}_n)\Phi] \to E[f(\widehat{u})\varphi]$$

for any bounded $\mathcal{G}$ measurable random variable φ and $f \in C(\mathbb{R})$ satisfying $\limsup_{|u|\to\infty} |u|^{-1}|f(u)| < \infty$.

In particular, for all functions f of at most polynomial growth,

$$E[f(\widehat{u}_n)] \to E[f(\widehat{u})]$$

and

$$E[f(\widetilde{u}_n)] \to E[f(\widehat{u})].$$

Thus

$$\widehat{u}_n \to^d \widehat{u} \text{ as } n \to \infty. \tag{1.33}$$

where $\widehat{u} := \Gamma^{-1}\Delta$ which is the unique maximizer of the random field $Z(u)$. Observe that

$$\widetilde{u}_n := \left(\int_{\mathbb{U}_n} Z_n(u)\pi(\theta^* + a_n u)du\right)^{-1} \int_{\mathbb{U}_n} uZ_n(u)\pi(\theta^* + a_n u)du.$$

Let

$$\widetilde{u} := \left(\int_{\mathbb{R}} Z(u)du\right)^{-1} \int_{\mathbb{R}} uZ(u)du = \Gamma^{-1}\Delta.$$

We have

$$\widetilde{u}_n \to^d \widetilde{u} \text{ as } n \to \infty. \tag{1.34}$$

as $n \to \infty$

QMLE and QBE are asymptotically equivalent:

$$\widehat{u}_n - \widetilde{u}_n \to^P 0 \text{ as } n \to \infty. \tag{1.35}$$

The QMLE and QBE have limiting mixture normal distribution and are *first-order efficient*.

Using Theorem 3 in Uchida and Yoshida [32] and Theorem 1 in Yoshida [34], we have

Theorem 1.1.

$$P\left(\frac{1}{\sqrt{n}}|N_n| > \epsilon\right) \leq Cn^{-1}.$$

Using Proposition 2 (the normalized M-estimator satisfies the polynomial type LDP) in Yoshida [34], we have

Theorem 1.2.

$$\sup_{\theta\in\Theta} P\left(a_n^{-1}|\widehat{\theta}_n - \theta| > \epsilon\right) \leq Cn^{-1}.$$

In particular, this theorem gives tightness of $\{\widehat{u}_n\}_{n\geq n_0}$.

Hence, we have

Theorem 1.3.

$$\sup_{x\in\mathbb{R}}\left|P\left(\frac{\sqrt{n}}{\sqrt{2}\theta}(\widehat{\theta}_n - \theta) \leq x\right) - \Phi(x)\right| \leq Cn^{-1/2}.$$

With $\theta = 1$,

$$\sup_{x\in\mathbb{R}}\left|P\left(\frac{\sqrt{n}}{\sqrt{2}}(\widehat{\theta}_n - 1) \leq x\right) - \Phi(x)\right| \leq Cn^{-1/2}.$$

Observe that

$$\begin{aligned}\widehat{\theta}_n - \int_0^T \left(\frac{\sigma_t}{\sigma_{t_{i-1}}}\right)^2 dt &= \sum_{i=1}^n \left(\frac{Y_{t_i} - Y_{t_{i-1}}}{\sigma_{t_{i-1}}}\right)^2 - \int_0^T \left(\frac{\sigma_t}{\sigma_{t_{i-1}}}\right)^2 dt \\ &= \sum_{i=1}^n \left(\frac{Y_{t_i} - Y_{t_{i-1}}}{\sigma_{t_{i-1}}}\right)^2 - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left(\frac{\sigma_t}{\sigma_{t_{i-1}}}\right)^2 dt. \end{aligned} \tag{1.36}$$

Here is a more precise formulation of the model:

$$dY_t = b_t dt + \sqrt{\theta}\sigma_t dW_t,\ t \in [0,1] \tag{1.37}$$

where the drift process and the volatility process respectively satisfy

$$db_t = b_t^{[0]}dt + b_t^{[1]}dW_t,\ d\sigma_t = \sigma_t^{[0]}dt + \sigma_t^{[1]}dW_t,\ t \in [0,1], \tag{1.38}$$

where the drift of the drift (DoD) process and the drift of the volatility (DoV) process respectively satisfy

$$db_t^{[0]} = b_t^{[0,0]}dt + b_t^{[0,1]}dW_t,\ d\sigma_t^{[0]} = \sigma_t^{[0,0]}dt + \sigma_t^{[0,1]}dW_t,\ t \in [0,1], \tag{1.39}$$

where the volatility of the drift (VoD) process and the volatility of volatility (VoV) process respectively satisfy

$$db_t^{[1]} = b_t^{[1,0]}dt + b_t^{[1,1]}dW_t,\ d\sigma_t^{[1]} = \sigma_t^{[1,0]}dt + \sigma_t^{[1,1]}dW_t,\ t \in [0,1]. \tag{1.40}$$

Thus $b_t^{[1,1]}$ is the volatility of the volatility of the drift (VoVoD) process and and $\sigma_t^{[1,1]}$ is volatility of the volatility of the volatility (VoVoV) process.

Suppose that

$$\sup_{t\in[0,1]} \|f_t\|_{p,3} < \infty \tag{1.41}$$

for $f_t := b_t, b_t^{[0]}, b_t^{[1]}, b_t^{[1,0]}, b_t^{[1,1]}, \sigma_t, \sigma_t^{[0]}, \sigma_t^{[1]}, \sigma_t^{[0,0]}, \sigma_t^{[0,1]}, \sigma_t^{[1,0]}, \sigma_t^{[1,1]}$ where $p > 1$.

We also assume that

$$\sup_{t\in[0,1]} \|1/\sigma_t\|_{L^p} < \infty \tag{1.42}$$

for each $p > 1$. Let θ denote true value of the unknown parameter.

Proof of Lemma 1.1. We show (1.3). Let the normalized error of the QMLE be denoted by

$$\mathcal{X}_n := \frac{\sqrt{n}}{\sqrt{2}\theta}(\widehat{\theta}_n - \theta). \tag{1.43}$$

We decompose this as the sum of four terms :

$$\mathcal{X}_n = \Psi + \Phi_2 + \Phi_3 + \Phi_4 \tag{1.44}$$

where

$$\begin{aligned}
\Psi &:= \sum_{i=1}^n \frac{\sqrt{n}}{\sqrt{2}} \left\{ \frac{1}{\sigma_{t_{i-1}}^2} \left(\int_{t_{i-1}}^{t_i} \sigma_t dW_t \right)^2 - \frac{1}{n} \right\}, \\
\Phi_2 &:= \sum_{i=1}^n \frac{\sqrt{2n}}{\sqrt{\theta}\sigma_{t_{i-1}}^2} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} b_s ds \sigma_t dW_t, \qquad (1.45) \\
\Phi_3 &:= \sum_{i=1}^n \frac{\sqrt{2n}}{\sqrt{\theta}\sigma_{t_{i-1}}^2} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \sigma_s dW_s b_t dt, \\
\Phi_4 &:= \sum_{i=1}^n \frac{\sqrt{2n}}{\theta\sigma_{t_{i-1}}^2} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} b_s ds b_t dt. \qquad (1.46)
\end{aligned}$$

We say that $R_n = R(n^{-k})$ if $\|R_n\|_p = O(n^{-k})$ for any $p > 1$.

Using Burkholder's inequality for discrete martingales, we have

$$\begin{aligned}
\sqrt{n}\Phi_3 = \sum_{i=1}^n \frac{\sqrt{2}n}{\sqrt{\theta}\sigma_{t_{i-1}}^2} \Bigg\{ & \int_{t_{i-1}}^{t_i} \sigma_{t_{i-1}} b_{t_{i-1}} \int_{t_{i-1}}^{t} dW_s dt \\
& + \int_{t_{i-1}}^{t_i} b_{t_{i-1}} \left(\int_{t_{i-1}}^{s} \sigma_s^{[1]} dW_u \right) dW_s dt + \int_{t_{i-1}}^{t_i} \sigma_{t_{i-1}} \int_{t_{i-1}}^{t} b_s^{[1]} ds dt \\
& + \int_{t_{i-1}}^{t_i} \sigma_{t_{i-1}} \left[\int_{t_{i-1}}^{t} \int_{t_{i-1}}^{v} b_s^{[1]} dW_s dW_v + \int_{t_{i-1}}^{t} \int_{t_{i-1}}^{v} dW_s \cdot b_v^{[1]} dW_v \right] dt \Bigg\} + R(n^{-0.5}) \\
= \sum_{i=1}^n \frac{\sqrt{2}n}{\sqrt{\theta}\sigma_{t_{i-1}}^2} & \left\{ b_{t_{i-1}} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} dW_s dt + b_{t_{i-1}}^{[1]} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} ds dt \right\} + R(n^{-0.5}). \qquad (1.47)
\end{aligned}$$

Similarly,

$$\sqrt{n}\Phi_2 = \sum_{i=1}^n \frac{\sqrt{2}n b_{t_{i-1}}}{\sqrt{\theta}\sigma_{t_{i-1}}^2} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} ds dW_t + R(n^{-0.5}). \tag{1.48}$$

Thus

$$\sqrt{n}\Phi_2 + \sqrt{n}\Phi_3 = \sum_{i=1}^n \frac{\sqrt{2}n b_{t_{i-1}}}{\sqrt{\theta}\sigma_{t_{i-1}}^2} (W_{t_i} - W_{t_{i-1}}) + \frac{1}{n} \sum_{i=1}^n \frac{\sqrt{2}n b_{t_{i-1}}}{2\sqrt{\theta}\sigma_{t_{i-1}}^2} + R(n^{-0.5}). \tag{1.49}$$

Further

$$\sqrt{n}\Phi_4 = \frac{1}{n}\sum_{i=1}^{n}\frac{\sqrt{2}b_{t_{i-1}}^2}{2\theta\sigma_{t_{i-1}}^2} + R(n^{-0.5}). \tag{1.50}$$

On the other hand, we decompose Ψ as the sum of three terms:

$$\Psi = \Psi_1 + \Psi_2 + \Psi_3 \tag{1.51}$$

where

$$\Psi_1 := \sum_{i=1}^{n}\frac{\sqrt{n}}{\sqrt{2}}\left((\Delta W_i)^2 - \frac{1}{n}\right),\ \Psi_2 := \sum_{i=1}^{n}\frac{\sqrt{2n}}{\sigma_{t_{i-1}}}\Delta W_i\int_{t_{i-1}}^{t_i}(\sigma_t - \sigma_{t_{i-1}})dW_t,$$
$$\Psi_3 := \sum_{i=1}^{n}\frac{\sqrt{n}}{\sqrt{2}\sigma_{t_{i-1}}^2}\left(\int_{t_{i-1}}^{t_i}(\sigma_t - \sigma_{t_{i-1}})dW_t\right)^2. \tag{1.52}$$

Hence

$$\sqrt{n}\Psi_1 = \sum_{i=1}^{n}\frac{n}{\sqrt{2}}\left((\Delta W_i)^2 - \frac{1}{n}\right). \tag{1.53}$$

By repeated use of Burkholder's inequality, we have

$$\begin{aligned}\sqrt{n}\Psi_3 &= \sum_{i=1}^{n}\frac{n}{\sqrt{2}\sigma_{t_{i-1}}^2}\left(\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t}(\sigma_s^{[1]}dW_s dW_t\right)^2 + R(n^{-0.5})\\ &= \sum_{i=1}^{n}\frac{n}{\sqrt{2}\sigma_{t_{i-1}}^2}\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{t}\sigma_s^{[1]}dW_s\right)^2 dt + R(n^{-0.5})\\ &= \sum_{i=1}^{n}\frac{n}{\sqrt{2}\sigma_{t_{i-1}}^2}\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t}\left(\sigma_s^{[1]}\right)^2 dsdt + R(n^{-0.5})\\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\sqrt{2}}\frac{\left(\sigma_{t_{i-1}}^{[1]}\right)^2}{\sigma_{t_{i-1}}^2} + R(n^{-0.5}).\end{aligned} \tag{1.54}$$

Similarly, by using Itô's formula, we have

$$\sqrt{n}\Psi_2 = \sum_{i=1}^{n}\frac{\sqrt{2}}{\sigma_{t_{i-1}}}n\Delta W_i\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t}dW_s dW_t + \frac{1}{n}\sum_{i=1}^{n}\frac{\sqrt{2}\sigma_{t_{i-1}}^{[0]}}{2\sigma_{t_{i-1}}} + R(n^{-0.5}). \tag{1.55}$$

Hence we have (1.3). □

2. Rate of Convergence in the Bernstein-von Mises Theorem

Strasser [29, 30] studied the second order properties of the posterior and the regular Bayes estimator for the iid case. We extend IHK (Ibragimov-Hasminskii-Kutoyants) program to he IHKS (Ibragimov-Hasminskii-Kutoyants-Strasser) program for volatility estimation in diffusion processes.

To study rate of convergence of the posterior distribution to normal distribution and Bayesian asymptotics beyond the first order, first we make several assumptions (regularity conditions) on the prior distribution and the loss function.

Let Λ be a prior probability measure on $(\Theta, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-algebra of subsets of Θ. We assume the following regularity conditions on the prior:

(A1) For every $\delta > 0$ and every compact $K \subseteq \Theta$,

$$\inf_{\theta\in K} |\Lambda\{\phi \in \Theta : |\phi - \theta| < \delta\} > 0.$$

(A2) For every compact $K \subseteq \Theta$, there exists $\alpha \geq 0$ (depending on K) such that

$$\liminf_n \inf_{\theta\in K} n^\alpha \Lambda\{\phi \in \Theta : |\phi - \theta| < n^{-1/2}\} > 0.$$

(A3) Λ has a density $\lambda(\cdot)$ with respect to Lebesgue measure on $\mathbb{R}$ which is continuous and positive on Θ and for every $\theta \in \Theta$ there exists a neighbourhood U_θ of θ such that

$$\left|\frac{\lambda(\theta_1)}{\lambda(\theta_2)} - 1\right| \leq C|\theta_1 - \theta_2| \quad \text{for all } \theta_1, \theta_2 \in U_\theta.$$

(A4)$_m$ Λ possesses m-th order absolute moments in $\theta \in \Theta,\ m \geq 1$.

Next, we define the loss function D as follows:

A $\mathcal{B} \times \mathcal{B}$ measurable function $D : \overline{\Theta} \times \overline{\Theta} \to \overline{\mathbb{R}}$ is called a loss function if

$$D(\theta, \theta) < D(\phi, \theta) \quad \text{for all } \theta \in \Theta, \phi \in \overline{\Theta},\ \theta \neq \phi.$$

We assume the following conditions on the loss function D.

(B1) $D : \overline{\Theta} \times \Theta \to \mathbb{R}$ is continuous.

(B2) (a) The mapping $\phi \to D(\theta, \phi), \theta \in \Theta$ is extended Λ-integrable for all $\theta \in \overline{\Theta}$ and has finite Λ-expectation if $\theta \in \Theta$.

(b) For every $\theta \in \bar{\Theta}$ there exists a neighbourhood U_θ^λ of θ such that for every neighbourhood $U \subseteq U_\theta^\lambda$ of θ, the mapping $\phi \to \inf L(U, \phi)$ is Λ-integrable.

(B3) (a) For every $\phi \in \Theta$ the mapping $\theta \to D(\theta, \phi)$ is twice differentiable in Θ. We denote

$$D_{10}(\theta, \phi) := \frac{\partial}{\partial\theta} D(\theta, \phi), \quad D_{20}(\theta, \phi) := \frac{\partial^2}{\partial\theta^2} D(\theta, \phi).$$

(b) For every $\theta \in \Theta$, the mapping $\phi \to D_{10}(\theta, \phi)$ is differentiable in Θ. We denote

$$D_{11}(\theta, \phi) := \frac{\partial}{\partial\phi} D_{10}(\theta, \phi).$$

(B4) For every compact $K \subseteq \Theta$, (a) $\sup_{\theta\in K} |D_{10}(\theta, \theta)| < \infty$,
(b) $\sup_{\theta\in K} |D_{20}(\theta, \theta)| < \infty$.

(B5) For every compact $K \subseteq \Theta, \inf_{\theta\in K} D_{20}(\theta, \theta) > 0$.

(B6) For every compact $K \subseteq \Theta$, (a)$\sup_{\theta\in K} \int |D_{10}(\theta, \phi)|\Lambda(d\phi) < \infty$,
(b)$\sup_{\theta\in K} \int |D_{20}(\theta, \phi)|\Lambda(d\phi) < \infty$.

(B7) (a) For every $\theta \in \Theta$ there exists a neighbourhood $\widetilde{V}_\theta$ of θ and a constant $\widetilde{C}_\theta \geq 0$ such that

$$|D_{20}(\phi, \phi_1) - D_{20}(\phi, \phi_2)| \leq \widetilde{C}_\theta |\phi_1 - \phi_2| \quad \text{for all } \phi, \phi_1, \phi_2 \in \widetilde{V}_\theta.$$

(b) For every $\theta \in \Theta$, there exist a neighbourhood $\widetilde{\widetilde{V}}_\theta$ of θ and a constant $\widetilde{\widetilde{C}}_\theta \geq 0$ such that

$$|D_{11}(\phi, \phi_1) - D_{11}(\phi, \phi_2)| \leq \widetilde{\widetilde{C}}_\theta |\phi_1 - \phi_2| \quad \text{for all } \phi, \phi_1, \phi_2 \in \widetilde{\widetilde{V}}_\theta.$$

(B8) For every $\theta \in \Theta$ there exist a neighbourhood V_θ^λ of θ and a continuous function $k_\theta : \Theta \to \mathbb{R}$ such that

$$|D_{20}(\phi_1, \phi) - D_{20}(\phi_2, \phi)| \leq k_\theta(\phi)|\phi_1 - \phi_2| \text{ for all } \phi_1, \phi_2 \in V_\theta^\lambda,$$

and all $\theta \in \Theta$ and $\int k_\theta(\phi)\Lambda(d\phi) < \infty$ for all $\theta \in \Theta$. Obviously, the loss function $D(\theta, \phi) = F(\theta - \phi)$ where F is a twice continuously differentiable function with $F(0) = 0, F(\phi) > 0$ for $\phi \neq 0$, $F''(0) > 0$ and $|F''(\phi) - F''(\phi_1))| \leq C_\theta|\phi - \phi_1|$ for all $\phi, \phi_1 \in \Gamma_\theta$ where Γ_θ is a neighbourhood of θ and $C_\theta \geq 0$ satisfies the condition (B1), (B3) – (B5), (B7) and (B8). In particular, the quadratic loss function satisfies these conditions.

Let $K \subseteq \Theta$ be compact and $\mathcal{M}_{n,\theta} \in \mathcal{F}_n, \theta \in K$ be sets satisfying

$$\sup_{\theta \in K} P_\theta(\mathcal{M}_{n,\theta}^c) = o(n^{-1/2}). \tag{2.1}$$

The quasi-posterior probability measure of $B \in \mathcal{B}$ given Y_0^n is defined as

$$\mathcal{R}_n(B) := \frac{\int_B \mathcal{L}_n(\theta)\lambda(\theta)d\theta}{\int_\Theta \mathcal{L}_n(\theta)\lambda(\theta)d\theta} \tag{2.2}$$

where

$$\mathcal{L}_n(\theta) := \frac{dP_\theta}{dP_0}(Y_0^n)$$

$$= \exp\left\{\sum_{i=1}^n \frac{b_{t_{i-1}}}{\sigma_{t_{i-1}}}(Y_{t_i} - Y_{t_{i-1}}) - \frac{\theta}{2}\sum_{i=1}^n \left(b_{t_{i-1}}\frac{Y_{t_i} - Y_{t_{i-1}}}{\sigma_{t_{i-1}}}\right)^2\right\}. \tag{2.3}$$

A quasi-Bayes estimator (relative to a prior distribution Λ and a loss function D) is defined as

$$\widetilde{\theta}_n := \arg\min_{\theta \in \bar{\Theta}} \int_\Theta D(\theta, \phi)\mathcal{R}_n(d\phi).$$

For the quadratic loss function, $\widetilde{\theta}_T$ is the posterior mean

$$E(\theta|Y_0^n) = \frac{\int_\Theta \phi\mathcal{L}_n(\phi)\lambda(\phi)d\phi}{\int_\Theta \mathcal{L}_n(\phi)\lambda(\phi)d\phi}.$$

Let further $\Phi_m, m \in \mathbb{N}$ be signed measures on $\mathcal{B}$ defined by

$$\Phi_m(B) := \frac{1}{\sqrt{2\pi}}\int_B t^m \exp\left(-\frac{t^2}{2}\right)dt, \ B \in \mathcal{B}. \tag{2.4}$$

For every Borel set $B \in \mathcal{B}$, define

$$\mathcal{B}_n(\theta) := \left\{\phi \in \mathbb{R}^+ : (nI(\theta))^{1/2}(\phi - \widehat{\theta}_n) \in B\right\},$$

$$I(\theta) := \frac{1}{2\theta^2}, \ \mathcal{J}_n(\theta) := \frac{\mathcal{I}_n}{nI(\theta)} - 1, \ \mathcal{I}_n := \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2,$$

$$d_n(\phi) := (nI(\theta))^{1/2}(\phi - \widehat{\theta}_n), \ \phi \in \Theta,$$

$$\mathcal{V}_n(s) := \left\{\phi \in \mathbb{R}^+ : (nI(\theta))^{1/2}|\phi - \widehat{\theta}_n| \leq (s\log n)^{1/2}\right\}$$

where $s > 0, \theta \in K$, and

$$\mathcal{A}_n(\phi) := \frac{1}{\sqrt{2\pi}}(nI(\theta))^{1/2}\frac{\mathcal{L}_n(\phi)}{\mathcal{L}_n(\widehat{\theta}_n)} \tag{2.5}$$

where $\phi \in \Theta$ and $\theta \in K$.

Lemma 2.1. *(a) For every $\delta > 0$ and every compact $K \subseteq \Theta$ there exists $\epsilon_K > 0$ such that*

$$\sup_{\theta \in K} P_\theta \left\{ - \sup_{|\phi - \theta| \geq \delta} \frac{1}{n} l_n(\phi) + E_\theta \left[l_n(\theta) \right] \leq \epsilon_K \right\} \leq C_K n^{-1}.$$

(b) For every $\epsilon > 0$ and every compact $K \subseteq \Theta$ there exists $\delta_K > 0$ such that

$$\sup_{\theta \in K} P_\theta \left\{ - \inf_{|\phi - \theta| < \delta_K} \frac{1}{n} l_n(\theta) + E_\theta \left[l_n(\theta) \right] > \epsilon \right\} \leq C_K n^{-1}.$$

Proof. Similar to Lemmas 1 and 2 in Strasser [30]. Details are omitted. □

The following theorem is a consequence of Theorem 9 in Yoshida [33].

Theorem 2.1. *Under the conditions (1.41), (1.42), for every compact $K \subseteq \Theta$,*

$$\sup_{\theta \in K} \sup_{x \in \mathbb{R}} \left| P_\theta \left\{ (nI(\theta))^{1/2} (\widehat{\theta}_n - \theta) \leq x \right\} - \Phi(x) \right| = O(n^{-1/2}).$$

The next two theorems give the rates of convergence in posterior consistency.

Theorem 2.2. *Let $B_\delta(\theta) := \{\phi \in \Theta : |\phi - \theta| < \delta\}$. Then for every $\delta > 0$ and every compact $K \subseteq \Theta$ there exists $\eta_K > 0$ such that*

$$\sup_{\theta \in K} P_\theta \{ \mathcal{R}_n(B_\delta^c(\theta)) > \exp(-\eta_K n) \} \leq C_K n^{-1}.$$

Proof. From (2.2), we have

$$\mathcal{R}_n(B_\phi^c(\theta)) = \frac{\int_{|\phi - \theta| \geq \delta} \mathcal{L}_n(\phi) \lambda(\phi) d\phi}{\int_\Theta \mathcal{L}_n(\phi) \lambda(\phi) d\phi}. \tag{2.6}$$

Thus

$$\mathcal{R}_n(B_\delta^c(\theta)) > \exp(-\eta_K n) \iff \frac{1}{n} \log \mathcal{R}_n(B_\delta^c(\theta)) > -\eta_K$$

which is further equivalent to

$$\frac{1}{n} \log \int_{|\phi - \theta| \geq \delta} \mathcal{L}_n(\phi) \lambda(\phi) d\phi - \frac{1}{n} \log \int_\Theta \mathcal{L}_n(\phi) \lambda(\phi) d\phi > -\eta_K$$

which implies for arbitrary $\delta_1 > 0$ and $\theta \in K$

$$- \inf_{|\phi - \theta| < \delta_1} \frac{1}{n} \log \mathcal{L}_n(\phi) + \sup_{|\phi - \theta| \geq \delta} \frac{1}{n} \log \mathcal{L}_n(\phi) > -\eta_K + \frac{a}{n}$$

where

$$a := \inf_{\theta \in K} \log \Lambda \left\{ \phi \in \Theta : |\phi - \theta| < \delta_1 \right\}.$$

From Lemma 2.1(a), we obtain

$$\sup_{\theta \in K} P_\theta \left\{ E_\theta \left[l_n(\theta) \right] - \epsilon_K \leq \sup_{|\phi - \theta| \geq \delta} \frac{1}{n} l_n(\phi) \right\} \leq C_K n^{-1}.$$

Therefore we only have to show that

$$\sup_{\theta \in K} P_\theta \left\{ - \inf_{|\phi - \theta| < \delta_1} \frac{1}{n} l_n(\theta) + E_\theta \left(l_n(\theta) \right) > \epsilon_K - \eta_K + \frac{a}{n} \right\} \leq C_K n^{-1}.$$

But this follows from Lemma 2.1(b) choosing δ_1 sufficiently small. Combination of the last two inequalities completes the proof. □

The next theorem shows that a result similar to Theorem 2.2 is true when the radius of the neighbourhoods of the QMLE $\widehat{\theta}_n$ decreases of the order $n^{-1/2}(\log n)^{1/2}$.

Theorem 2.3. *For every $r > 0$ and every compact $K \subseteq \Theta$, there exist $s_K \geq 0$ and $C_K \geq 0$ such that*

$$P_\theta \left\{ \mathcal{R}_n \left(\mathcal{V}_n^c(s_K) \right) \geq C_K n^{-r} \right\} \leq C_K n^{-1}.$$

Proof. Using Theorem 1.2 and Theorem 2.2, we have for every $\delta > 0$ there exists $\eta_K > 0$ such that

$$\sup_{\theta \in K} P_\theta \left\{ \mathcal{R}_n \left\{ \phi \in \Theta : |\phi - \widehat{\theta}_n| \geq \delta \right\} > \exp(-\eta_K n) \right\} \leq C_K n^{-1}. \tag{2.7}$$

Hence it is sufficient to prove that for every $r > 0$ and every compact $K \subseteq \Theta$ there exists $s_K > 0, C_K$ and $\delta_K > 0$ such that

$$\sup_{\theta \in K} P_\theta \left[\mathcal{R}_n \left\{ \phi \in \Theta : \delta_K > |\phi - \widehat{\theta}_n| > n^{-1/2} \left(\frac{s_K \log n}{I(\theta)} \right)^{1/2} \right\} \geq C_K n^{-r} \right]$$
$$\leq C_K n^{-1}.$$

From Theorem 1.1, for every $\epsilon > 0$,

$$P_\theta \left\{ \left| \frac{\mathcal{I}_n}{n} - I(\theta) \right| > \epsilon \right\} \leq C n^{-1}.$$

Define the sets

$$N_n(\epsilon, \theta) := \left\{ \left| \frac{\mathcal{I}_n}{n} - I(\theta) \right| > \epsilon \right\},$$
$$S_n^\theta(s) := \left\{ \phi \in \Theta : \delta_K > |\phi - \widehat{\theta}_n| > n^{-1/2} \left(\frac{s \log n}{I(\theta)} \right)^{1/2} \right\},$$
$$H_n := \left\{ \phi \in \Theta : |\phi - \widehat{\theta}_n| < n^{-1/2} \right\}.$$

By Taylor expansion, we have

$$l_n(\phi) = l_n(\widehat{\theta}_n) + (\phi - \widehat{\theta}_n) l_n'(\widehat{\theta}_n) + \frac{(\phi - \widehat{\theta}_n)^2}{2} l_n''(\theta^*) \tag{2.8}$$

where $|\theta^* - \theta| \leq |\widehat{\theta}_n - \theta|$. Since $l_n'(\widehat{\theta}_n) = 0$, we obtain

$$\exp \left[l_n(\phi) - l_n(\widehat{\theta}_n) \right] = \exp \left\{ -\frac{n}{2} (\phi - \widehat{\theta}_n)^2 \frac{l_n''(\theta^*)}{n} \right\} =: G_n(\phi) \tag{2.9}$$

It follows that in $N_n^c(\epsilon, \theta)$, $n > n_K$ implies

$$\begin{aligned} \mathcal{R}_n(S_n^\theta(s)) &= \frac{\int_{S_n^\theta(s)} G_n(\phi) \lambda(\phi) d\phi}{\int_\Theta G_n(\phi) \lambda(\phi) d\phi} \\ &\leq \frac{\int_{S_n^\theta(s)} \exp \left\{ -\frac{n}{2} (\phi - \widehat{\theta}_n)^2 (I(\theta) - \epsilon) \right\} \lambda(\phi) d\phi}{\int_{H_n} \exp \left\{ -\frac{n}{2} (\phi - \widehat{\theta}_n)^2 (I(\theta) + \epsilon) \right\} \lambda(\phi) d\phi} \\ &\leq \frac{\exp \left\{ -\frac{1}{2} s (\log n)(1 - \epsilon I^{-1}(\theta)) \right\}}{\exp \left\{ -\frac{1}{2} (I(\theta) + \epsilon) \right\} \Lambda(H_n)}. \end{aligned}$$

Let $\rho > 0$ be such that $K_1 := \{\phi \in \Theta : dist(\phi, K) \leq \rho\}$ is a compact subset of Θ. Let

$$\widetilde{N}_n(\rho, \theta) := \left\{ |\widehat{\theta}_n - \theta| > \rho \right\}.$$

Then from Theorem 1.2,

$$\sup_{\theta\in K} P_\theta\left(\widetilde{N}_n(\rho,\theta)\right) \le Cn^{-1}.$$

Now in $(\widetilde{N}_n(\rho,\theta))^c$, by assumption (A2)

$$\liminf_n n^\alpha \Lambda(H_n) = \liminf_n n^\alpha \inf_{\theta\in K_1} \Lambda\left\{\phi\in\Theta : |\phi-\theta| < n^{-1/2}\right\} > 0.$$

Thus it follows that there exists $C_K \ge 0$ with

$$P_\theta\left\{\mathcal{R}_n(S_n^\theta(s)) > C_K n^{(2\alpha - s(1-\epsilon I^{-1}(\theta)))/2}\right\} \le C_K n^{-1}.$$

It is obvious that for every $r>0, s>0$ can be chosen such that the assertion of the Theorem holds. □

Lemma 2.4. *Let $g_\theta : \Theta \to \bar{\mathbb{R}}, \theta\in\Theta$, be a class of Λ-integrable functions satisfying*

$$\sup_{\theta\in K}\int_\Theta |g_\theta(\phi)|\Lambda(d\phi) < \infty$$

for every compact $K\subseteq\Theta$. Then for every $\delta>0$ and every compact $K\subseteq\Theta$ there exist $\eta_K>0$ such that

$$P_\theta\left\{\int_{|\phi-\theta|\ge\delta} |g_\theta(\phi)|\mathcal{R}_n(d\phi) > \exp(-\eta_K n)\right\} \le C_K n^{-1}.$$

Proof. Since g_θ is Λ-integrable for every $\theta\in\Theta$ it is also $\mathcal{R}_n$ integrable. The proof of the lemma is completely same as the proof of Theorem 2.2 if the constant a is replaced by

$$\widetilde{a} := \inf_{\theta\in K}\left\{\log\Lambda\left\{\phi\in\Theta : |\phi-\theta|<\delta_1\right\} - \log\int_\Theta |g_\theta(\phi)|\Lambda(d\phi)\right\}.$$

□

Lemma 2.5. *Let $s>0$. Let $K\subseteq\Theta$ be compact and choose $\epsilon>0$ such that $\bar{K^\epsilon}\subseteq\Theta$ where $K^\epsilon = \{\phi\in\mathbb{R}^+ : |\phi-\theta|<\epsilon \text{ if } \theta\in K\}$. Then there exist sets $\mathcal{M}_{n,\theta}\in\mathcal{F}_T, \theta\in K$ satisfying (2.2) such that $\mathcal{V}_n(s)\subseteq K^\epsilon$ on the set $\mathcal{M}_{n,\theta}$.*

Proof. The proof follows from Theorem 1.2 which implies that we can assume $|\theta_n-\theta|<\epsilon/2$ on the set $\mathcal{M}_{n,\theta}$. □

Theorem 2.4. *Under (A1) and $(A4)_m$, for every compact $K\subseteq\Theta$ there exist sets $\mathcal{M}_{n,\theta}, \theta\in K$ satisfying (2.1) such that on the set $\mathcal{M}_{n,\theta}$*

$$\sup_{B\in\mathcal{B}}\left|\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)} [d_n(\phi)]^m \mathcal{A}_n(\phi)\lambda(\phi)d\phi\right.$$

$$\left.-\lambda(\theta_n)\left[\Phi_m(B) + \frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(B)\right]\right| \le C_K n^{-1/2}.$$

Proof. According to Lemma 2.5 we can assume that $\mathcal{V}_n(s_K)\subseteq\Theta$ on the set $\mathcal{M}_{n,\theta}$. Note that

$$\exp\left[l_n(\phi) - l_n(\widehat{\theta}_n)\right] = \frac{\mathcal{L}_n(\phi)}{\mathcal{L}_n(\widehat{\theta}_n)}.$$

Hence from (2.7) and (2.9), we obtain

$$\begin{aligned}\mathcal{A}_n(\phi) &= \frac{1}{\sqrt{2\pi}}(nI(\theta))^{1/2}\frac{\mathcal{L}_n(\phi)}{\mathcal{L}_n(\widehat{\theta}_n)}\\ &= \frac{1}{\sqrt{2\pi}}(nI(\theta))^{1/2}\exp\left\{-\frac{n}{2}(\phi-\widehat{\theta}_n)^2\frac{\mathcal{I}_n}{n}\right\}\\ &= \frac{1}{\sqrt{2\pi}}(nI(\theta))^{1/2}\exp\left\{-\frac{nI(\theta)}{2}(\phi-\widehat{\theta}_n)^2(1+\mathcal{J}_n(\theta))\right\}\\ &\quad \text{(from (2.3))}.\end{aligned}$$

Define

$$Q_n := \left\{y\in\mathbb{R} : y\in B, y^2 \le s_K\log n\right\},\ B\in\mathcal{B}.$$

Then for every $B\in\mathcal{B}$, substituting $u_n(\phi)=y$, on the set $\mathcal{M}_{n,\theta}$, we have

$$\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)}[u_n(\phi))]^m\,\mathcal{A}_n(\phi)d\phi = \int_{Q_n}\exp\left(-\frac{y^2}{2}\mathcal{J}_{n,\theta}\right)\Phi_n(dy).$$

Let us now use the inequality

$$\left|e^{-\beta}-1+\beta\right| \le \beta^2 e^{|\beta|},\ \beta\in\mathbb{R}$$

and put $\beta := \frac{y^2}{2}\mathcal{J}_{n,\theta}$.

From Theorem 1.1 it follows that, on the set $\mathcal{M}_{n,\theta}$,

$$|\mathcal{J}_{n,\theta}| \le C_K n^{-1/2}(\log n)^{1/2}.$$

Thus we obtain

$$\begin{aligned}\sup_{B\in\mathcal{B}}\left|\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)}[u_n(\phi)]^m\,\mathcal{A}_n(\phi)d\phi - \int_{Q_n}\left(1-\frac{y^2}{2}\mathcal{J}_{n,\theta}\right)\Phi_m(dy)\right|\\ \le C_K n^{-1/2},\end{aligned}$$

i.e.,

$$\begin{aligned}\sup_{B\in\mathcal{B}}\left|\int_{\mathcal{B}_n(\theta)\cap V_n(s_K)}[u_n(\phi)]^m\,\mathcal{A}_n(\phi)d\phi - \Phi_m(Q_n)+\frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(Q_n)\right|\\ \le C_K n^{-1/2}.\end{aligned}$$

But it is easily seen that

$$\sup_{B\in\mathcal{B}}|\Phi_m(B)-\Phi_m(Q_n)| \le C_K n^{-1/2}$$

for sufficiently large $s_K>0$. Hence

$$\begin{aligned}\sup_{B\in\mathcal{B}}\left|\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)}[u_n(\phi)]^m\,\mathcal{A}_n(\phi)d\phi - \Phi_m(B)+\frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(B)\right|\\ \le C_K n^{-1/2}.\end{aligned}$$

According to the uniform cover theorem for every $\epsilon>0$ satisfying $\bar{K}^\epsilon\subseteq\Theta$ there exists $e_K>0$ such that $|\phi-\phi'|<e_K$ implies $|\lambda(\phi)-\lambda(\phi')|\le C_K|\phi-\phi'|$ for every $\phi\in K^\epsilon$. Hence on the set $\mathcal{M}_{n,\theta}$ we can assume that $V_n(s_K)$ is contained in an ϵ-neighbourhood of θ for every $\theta\in K$.

Then on the set $\mathcal{M}_{n,\theta}$,

$$\begin{aligned}&\left|\sup_{B\in\mathcal{B}}\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)}[u_n(\phi)]^m\,\mathcal{A}_n(\phi)\lambda(\phi)d\phi\right.\\ &\left.-\lambda(\widehat{\theta}_n)\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)}[u_n(\phi)]^m\,\mathcal{A}_n(\phi)d\phi\right|\end{aligned}$$

$$\leq C_K \lambda(\widehat{\theta}_n) \int_{\mathcal{V}_n(s_K)} |u_n(\phi)|^m \mathcal{A}_n(\phi) \left|\phi - \widehat{\theta}_n\right| d\phi.$$

Using a similar substitution argument as before we obtain on the set $\mathcal{M}_{n,\theta}$,

$$\int_{\mathcal{V}_n(s_K)} |u_n(\phi)|^m \mathcal{A}_n(\phi) |\phi - \widehat{\theta}_n| d\phi$$

$$\leq C_K n^{-1/2} \int_{y^2 \geq s_K \log n} |y|^{m+1} \exp\left(-\frac{y^2}{2}(1-\delta_K)\right) dy$$

since on the set $\mathcal{M}_{n,\theta}$ we may assume that $\mathcal{J}_{n,\theta} \geq -\delta_K$ if $y^2 \leq s_K \log n$. Now using (A1) and Theorem 1.2, we obtain $\lambda(\widehat{\theta}_n) \leq C_K < \infty$ on the set $\mathcal{M}_{n,\theta}$. Hence the assertion of the theorem follows. □

Lemma 2.6. *Under (A1) and $(A4)_m$, for $K \subseteq \Theta$ there exist $\mathcal{M}_{n,\theta} \in \mathcal{F}_T$, $\theta \in K$ satisfying (2.2) such that on the set $\mathcal{M}_{n,\theta}$*

$$\left| \int_{\Theta} |u_n(\phi)|^m \mathcal{R}_n(d\phi) - \int_{\mathcal{V}_n(s_K)} |u_T(\phi)|^m \mathcal{R}_n(d\phi) \right| \leq C_K n^{-1}.$$

Proof. Lemma 2.5 shows that we can assume $\theta_n \in K^\epsilon$ for $\epsilon > 0$ on the set $\mathcal{M}_{n,\theta}$ and $\bar{K}^\epsilon \subseteq \Theta$. From Lemma 2.4 we obtain $\delta_K > 0$ and $\eta_K > 0$ such that in $\mathcal{M}_{n,\theta}$,

$$\int_{|\phi-\theta| \geq \delta} |u_n(\phi)|^m \mathcal{R}_n(d\phi) \leq C_K n^{m/2} \exp(-\eta_K n).$$

Hence it follows that

$$\int_{\Theta \setminus \mathcal{V}_n(s_K)} |u_n(\phi)|^m \mathcal{R}_n(d\phi) \leq C_K n^{m/2} \exp(-\eta_K n) + C_K n^{m/2} \mathcal{R}_n(\mathcal{V}_n^c(s_K))$$

in $\mathcal{M}_{n,\theta}$. Now for sufficiently large $s_K > 0$ the assertion of the theorem follows from Theorem 2.3. □

Lemma 2.7. *Under (A1) and $(A4)_m$, for every $K \subseteq \Theta$ there exist sets $\mathcal{M}_{n,\theta}, \theta \in K$ satisfying (2.1) such that on the set $\mathcal{M}_{n,\theta}$*

$$\sup_{B \in \mathcal{B}} \left| \int_{\mathcal{B}_n(\theta) \cap \Theta} [u_n(\phi)]^m \mathcal{R}_n(d\phi) - \frac{\int_{\mathcal{B}_n(\theta) \cap \mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{A}_n(\phi) \lambda(\phi) d\phi}{\int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi) \lambda(\phi) d\phi} \right|$$

$$\leq C_K n^{-1/2}.$$

Proof. Lemma 2.5 shows that we may restrict our attention to the case $\mathcal{V}_n(s_K) \subseteq \Theta$. Now

$$\sup_{B \in \mathcal{B}} \left| \int_{\mathcal{B}_n(\theta) \cap \Theta} [u_n(\phi)]^m \mathcal{R}_T(d\phi) - \frac{\int_{\mathcal{B}_n(\theta) \cap \mathcal{V}_n(s_K)} [u_T(\phi)]^m \mathcal{A}_n(\phi) \lambda(\phi) d\phi}{\int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi) \lambda(\phi) d\phi} \right|$$

$$\leq \int_{\Theta \setminus \mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{R}_n(d\phi) + \frac{|\mathcal{R}_n(\mathcal{V}_n(s_K)) - 1|}{\mathcal{R}_n(\mathcal{V}_n(s_K))} \int_{\mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{R}_n(d\phi).$$

By Theorem 2.3, we have on the set $\mathcal{M}_{n,\theta}$

$$\int_{\mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{R}_n(d\phi) \leq C_K < \infty \text{ and } \mathcal{R}_n(\mathcal{V}_n(s_K)) \geq \epsilon_K > 0.$$

By Theorem 2.3 on the set $\mathcal{M}_{n,\theta}$

$$|\mathcal{R}_n(\mathcal{V}_n(s_K)) - 1| \geq C_K n^{-r}.$$

By Lemma 2.6

$$\int_{\Theta\setminus\mathcal{V}_n(s_K)} |u_n(\phi)|^m \mathcal{R}_n(d\phi) \le C_K n^{-1}.$$

This completes the proof of the lemma. □

The next theorem gives the rate of convergence in the Bernstein-von Mises theorem.

Theorem 2.5. *Under (A1) and $(A4)_m$, for every compact $K \subseteq \Theta$ there exist sets $\mathcal{M}_{n,\theta}, \theta \in K$ satisfying (2.1) such that on the set $\mathcal{M}_{n,\theta}$*

$$\sup_{B\in\mathcal{B}} \left| \int_{\mathcal{B}_n(\theta)\cap\Theta} [d_n(\phi)]^m \mathcal{R}_n(\phi)d\phi - \Phi_m(B) + \frac{1}{2}\mathcal{J}_{n,\theta}\left[\Phi_m(B) + \Phi_{m+2}(B)\right] \right| \le C_K n^{-1/2}.$$

Proof. Let us consider Lemma 2.7. The following inequality holds for $(\alpha,\beta,\gamma,\delta) \in \mathbb{R}^4, \beta,\delta \neq 0$.

$$\left|\frac{\alpha}{\beta} - \frac{\gamma}{\delta}\right| \le \frac{1}{|\delta|}|\alpha-\gamma| + \left|\frac{\gamma}{\beta\delta}\right| |\beta-\delta|. \tag{2.10}$$

For fixed $B \in \mathcal{B}$, on the set $\mathcal{M}_{n,\theta}$ define

$$\alpha := \int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{A}_n(\phi)\lambda(\phi)d\phi,\ \beta := \int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi)\lambda(\phi)d\phi,$$

$$\gamma := \lambda(\widehat{\theta}_n)\left[\Phi_m(B) - \frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(B)\right],\ \delta := \lambda(\widehat{\theta}_n)\left[1 - \frac{1}{2}\mathcal{J}_{n,\theta}\right].$$

Observe that

$$\begin{aligned}
& \sup_{B\in\mathcal{B}} \left| \int_{\mathcal{B}_n(\theta)\cap\Theta} [u_n(\phi)]^m \mathcal{R}_n(d\phi) - \frac{\Phi_m(B) - \frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(B)}{1 - \frac{1}{2}\mathcal{J}_{n,\theta}} \right| \\
\le\ & \sup_{B\in\mathcal{B}} \left| \int_{\mathcal{B}_n(\theta)\cap\Theta} [u_n(\phi)]^m \mathcal{R}_n(d\phi) - \frac{\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{A}_n(\phi)\lambda(\phi)d\phi}{\int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi)\lambda(\phi)d\phi} \right| \\
+\ & \sup_{B\in\mathcal{B}} \left| \frac{\int_{\mathcal{B}_n(\theta)\cap\mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{A}_n(\phi)\lambda(\phi)d\phi}{\int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi)\lambda(\phi)d\phi} - \frac{\lambda(\widehat{\theta}_n)\left[\Phi_m(B) - \frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}\right]}{\lambda(\widehat{\theta}_n)\left[1 - \frac{1}{2}\mathcal{J}_{n,\theta}\right]} \right| \\
=:\ & \widetilde{H}_1 + \widetilde{H}_2 \\
\le\ & C_K n^{-1/2} + \sup_{B\in\mathcal{B}} \frac{1}{\lambda(\widehat{\theta}_n)\left[1 - \frac{1}{2}\mathcal{J}_{n,\theta}\right]} \left| \int_{B_n(\theta)\cap\mathcal{V}_n(s_K)} [u_n(\phi)]^m \mathcal{A}_n(\phi)\lambda(\phi)d\phi \right. \\
& \left. -\lambda(\widehat{\theta}_n)\left[\Phi_m(B) - \frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(B)\right] \right| \\
& + \frac{|\lambda(\widehat{\theta}_n)\left[\Phi_m(B) - \frac{1}{2}\mathcal{J}_{n,\theta}\Phi_{m+2}(B)\right]|}{|\int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi)\lambda(\phi)d\phi||\lambda(\widehat{\theta}_n)\left[1 - \frac{1}{2}\mathcal{J}_{n,\theta}\right]|} \\
& \times \left| \int_{\mathcal{V}_n(s_K)} \mathcal{A}_n(\phi)\lambda(\phi)d\phi - \lambda(\widehat{\theta}_n)\left[1 - \frac{1}{2}\mathcal{J}_{n,\theta}\right] \right| \\
=:\ & C_K n^{-1/2} + \widehat{N}_1 + \widehat{N}_2 \le C_K n^{-1/2}
\end{aligned}$$

(since $\lambda(\widehat{\theta}_n) > 0$ for every $\widehat{\theta}_n \in \Theta$).

Here the bound for $\widetilde{H}_1$ comes from Lemma 2.7 and the second term $\widetilde{H}_2$ is split up by using the inequality (2.10). The bounds for $\widehat{N}_1$ and $\widehat{N}_2$ come from Theorem 2.4.

Now we shall use the following inequality which holds for $\alpha \in \mathbb{R}, |\alpha| < 1$,

$$\left| \frac{1}{1-\alpha} - (1+\alpha) \right| \leq \frac{\alpha^2}{1-|\alpha|} \tag{2.11}$$

It justifies replacing division by $1 - \frac{1}{2}\mathcal{J}_{n,\theta}$ with multiplying by $1 + \frac{1}{2}\mathcal{J}_{n,\theta}$. Another application of the inequality $\mathcal{J}_{n,\theta}^2 \leq C_K n^{-1/2}$ on the set $\mathcal{M}_{n,\theta}$ proves the assertion completely since terms containing $\mathcal{J}_{n,\theta}^2$ can be omitted. □

3. Rate of Convergence of the Bayes Estimator

We obtain $o(n^{-1/2})$ bound for the asymptotic equivalence of the Bayes estimators and the maximum likelihood estimator of the volatility. Using this bound and the Berry-Esseen bound for the maximum likelihood estimator finally we obtain $O(n^{-1/2})$ Berry-Esseen bound for the normalized Bayes estimator. Thus the maximum likelihood estimator and the Bayes estimators are *asymptotically efficient of order* $n^{-1/2}$.

To prove the main results we start with some preliminary lemmas.

Lemma 3.1. *Let conditions (B3), (B4), and (B7) (b) be satisfied. Then for every compact* $K \subseteq \Theta$ *there exist* $d_K \geq 0$ *such that*

$$\sup_{\theta \in K} \sup_{|\phi-\theta|<d_K} |D_{10}(\theta, \phi)| < \infty.$$

Proof. Similar to Lemma 2 in Strasser [29]. Details are omitted. □

Lemma 3.2. *Assume that the conditions (A1) – (A3), (B1) – (B4), (B6) (a), and (B7) (b) be satisfied. Then for every compact* $K \subseteq \Theta$ *there exist sets* $\mathcal{M}_{n,\theta}, \theta \in K$ *satisfying (3.2) such that on the set* $\mathcal{M}_{n,\theta}$

$$\left| \int_\Theta D_{10}(\widehat{\theta}_n, \phi)\mathcal{R}_n(d\phi) - \int_{\mathcal{V}_n(s_K)} D_{10}(\widehat{\theta}_n, \phi)\mathcal{R}_n(d\phi) \right| \leq C_K n^{-1}.$$

Proof. Let $\eta > 0$ and $\delta > 0$. Then the inequality

$$\int_{|\phi-\theta|\geq\delta} |D_{10}(\widehat{\theta}_n, \phi)|\mathcal{R}_n(d\phi) \geq \exp(-\eta n)$$

yields

$$\frac{\int_{|\phi-\theta|\geq\delta} |D_{10}(\widehat{\theta}_n, \phi)|\mathcal{L}_n(\phi)\Lambda(d\phi)}{\int_{|\phi-\theta|\leq\delta_1} \mathcal{L}_n(\phi)\Lambda(d\phi)} \geq \exp(-\eta n).$$

for arbitrary $\delta_1 > 0$. Hence

$$\frac{\sup_{|\phi-\theta|\geq\delta} \mathcal{L}_n(\phi) \int_\Theta |D_{10}(\widehat{\theta}_n, \phi)|\Lambda(d\phi)}{\inf_{|\phi-\theta|\geq\delta_1} \mathcal{L}_n(\phi)\Lambda\{\phi \in \Theta : |\phi-\theta| < \delta_1\}} \geq \exp(-\eta n).$$

Taking logarithms and dividing by n, we obtain

$$\sup_{|\phi-\theta|<\delta_1} \frac{1}{n} l_n(\phi) - \inf_{|\phi-\theta|>\delta} \frac{1}{n} l_n(\phi) > -\eta + \frac{a}{n}$$

where

$$a := \inf_{\theta \in K} \left[\log \Lambda\{\phi \in \Theta : |\phi-\theta| < \delta_1\} - \log \int_\Theta |D_{10}(\widehat{\theta}_n, \phi)|\Lambda(d\phi) \right].$$

Since there is a compact set $K_1 \subset \Theta$ with $K \subseteq K_1$ and $\widehat{\theta}_n \in K_1$ on the set $\mathcal{M}_{n,\theta}$, we may assume that $a \geq a_K > -\infty$ on the set $\mathcal{M}_{n,\theta}$. Then completely the same proof as for Theorem 2.3 shows that on the set $\mathcal{M}_{n,\theta}$

$$\int_{|\phi-\theta|\geq\delta} |D_{10}(\widehat{\theta}_n,\phi)|\mathcal{R}_n(d\phi) \leq \exp(-\eta_K n)$$

for sufficiently small $\eta_K > 0$ and arbitrary $\delta > 0$.

Thus

$$\begin{aligned} &\left|\int_\Theta D_{10}(\widehat{\theta}_n,\phi)\mathcal{R}_n(d\phi) - \int_{\mathcal{V}_n(s_K)} D_{10}(\theta,\phi)\mathcal{R}_n(d\phi)\right| \\ \leq\ &\int_{|\phi-\theta|\geq\delta} |D_{10}(\widehat{\theta}_n,\phi)|\mathcal{R}_n(d\phi) + \sup_{\theta\in K_1}\sup_{|\phi-\theta|<\delta} |D_{10}(\theta,\phi)|\mathcal{R}_n(\mathcal{V}_n^c(s_K)). \end{aligned}$$

Now Lemma 3.1 and Theorem 2.3 prove the theorem. □

Lemma 3.3. *Under (B7), for every compact $K \subseteq \Theta$ and every $\epsilon > 0$ with $\bar{K}_\epsilon \subseteq \Theta$ there exists $C_K > 0$ such that*

$$|D_{11}(\phi,\phi_1) - D_{11}(\phi,\phi)| \leq C_K|\phi-\phi_1|$$

Proof. Similar to the proof of Lemma 2.7. Details are omitted. □

Lemma 3.4. *Under (B2), for every compact $K \subseteq \Theta$ and every $\delta > 0$, there exists $\epsilon_K > 0$ such that $|\tau-\theta| \geq \delta, \theta \in K, \tau \in \bar{\Theta}$ implies $D(\theta,\theta) < D(\tau,\theta) - \epsilon_K$.*

Proof. Similar to Lemma 3 in Strasser [29]. Details are omitted. □

Lemma 3.5. *Let (A1)–(A3) be satisfied. Let $g_\theta : \Theta \to \mathbb{R}$, $\theta \in \Theta$ be a class of continuous Λ integrable functions satisfying the following conditions :*

(G1) *For every compact $K \subseteq \Theta$,*

$$\sup_{\theta\in K}\int_\Theta |g_\theta(\phi)|\Lambda(d\phi) < \infty.$$

(G2) *For every compact $K \subseteq \Theta$,*

$$\sup_{\theta\in K}\int_\Theta |g_\theta(\theta)| < \infty.$$

(G3) *For every compact $K \subseteq \Theta$ and every $\epsilon > 0$, there exists $\delta_K > 0$ such that $|\phi-\theta| < \delta_K, \theta \in K$ implies*

$$|g_\theta(\phi) - g_\theta(\theta)| < \epsilon.$$

Then for every $\epsilon > 0$ and every compact $K \subseteq \Theta$

$$\sup_{\theta\in K} P_\theta\left\{\left|\int_\Theta g_\theta(\phi)\mathcal{R}_n(d\phi) - g_\theta(\theta)\right| > \epsilon\right\} \leq C_K n^{-1}.$$

Proof. Similar to the proof of Lemma 4 in Strasser [29] by using Lemma 2.4 and Theorem 2.2. We omit the details. □

Lemma 3.6. *Under (A1) – (A3), for every $\epsilon > 0, \tau \in \bar{\Theta}$ there exists $\widetilde{U}_\tau \subseteq U_\tau^\lambda$ such that $U \subset \widetilde{U}_\tau$ implies*

$$\sup_{\theta\in K} P_\theta\left\{\int_\Theta \inf D(U,\phi)\mathcal{R}_n(d\phi) \leq \inf D(U,\theta) - \epsilon\right\} \leq C_K n^{-1}.$$

Proof. Similar to Lemma 5 in Strasser [29]. □

We need the following large deviations inequality for the Bayes estimator $\widetilde{\theta}_n$ proved in Yoshida [34].

Theorem 3.7. *Under (1.41), (1.42), (B1), (B2) for every $\delta > 0$ and every compact $K \subseteq \Theta$*

$$\sup_{\theta \in K} P_\theta \left\{ \left| \widetilde{\theta}_n - \theta \right| > \delta \right\} \leq C_K \exp(-Cn\delta^2).$$

Corollary 3.8. *Under (1.41), (1.42), (B1)–(B4), (B6)(b), (B7)(a) and (B8), we have for every $\epsilon > 0$,*

$$\sup_{\theta \in K} P_\theta \left\{ \left| \int_\Theta D_{20}(\theta_n^{**}, \phi) \mathcal{R}_n(d\phi) - D_{20}(\theta, \theta) \right| \geq \epsilon \right\} \leq C_K n^{-1}$$

where $|\theta_n^{**} - \theta| \leq |\widehat{\theta}_n - \theta|$.

Proof. Similar to the proof of Corollary 1 of Strasser [29]. □

Proposition 3.9. *Under (A1)–(A3), $(A4)_2$, (1.41), (1.42), (B3), (B4), (B6) (a) and (B6) (b), for every compact $K \subseteq \Theta$ there exist sets $\mathcal{M}_{n,\theta}$, $\theta \in K$ satisfying (2.1) such that on the set $\mathcal{M}_{n,\theta}$*

$$\left| \int_\Theta D_{10}(\widehat{\theta}_n, \phi) \mathcal{R}_n(d\phi) \right| \leq C_K n^{-1}.$$

Proof. By Lemma 3.2 on the set $\mathcal{M}_{n,\theta}$

$$\left| \int_\Theta D_{10}(\theta_T, \phi) \mathcal{R}_n(d\phi) - \int_{\mathcal{V}_n(s_K)} D_{10}(\widehat{\theta}_n, \phi) \mathcal{R}_n(d\phi) \right| \leq C_K n^{-1}.$$

Let $\widetilde{G}_n$ be the probability measure on $S_n := \{y \in \mathbb{R} : y^2 \leq s_K \log n\}$ which is induced by $\mathcal{R}_n$ and u_n. Since on the set $\mathcal{M}_{n,\theta}$ we have $\mathcal{V}_n(s_K) \subseteq \Theta$, the substitution $y = u_n(\phi)$ if $\phi \in \mathcal{V}_n(s_K)$ yields on the set $\mathcal{M}_{n,\theta}$

$$\int_{\mathcal{V}_n(s_K)} D_{10}(\widehat{\theta}_n, \phi) \mathcal{R}_n(d\phi) = \int_{S_n} D_{10} \left(\widehat{\theta}_n, \widehat{\theta}_n + \left(\frac{1}{nI(\theta)} \right)^{1/2} y \right) \widetilde{G}_n(dy).$$

By Taylor expansion on the set $M_{n,\theta}$

$$\int_{\mathcal{V}_n(s_K)} D_{10}(\widehat{\theta}_n, \phi) \mathcal{R}_n(d\phi) = \left(\frac{1}{nI(\theta)} \right)^{1/2} \int_{S_n} y D_{11}(\widehat{\theta}_n, \theta_n^{**}) \widetilde{G}_n(dy)$$

where $|\theta_n^{**} - \widehat{\theta}_n| \leq C_K |y| n^{-1/2}$ if $y^2 \leq \log n$.

The first term of the Taylor expansion is omitted since $D_{10}(\widehat{\theta}_n, \widehat{\theta}_n) = 0$ if $\theta_n \in \Theta$. Let $\epsilon > 0$ be such that an ϵ-neighbourhood K^ϵ of K is still contained in Θ. We may assume that on the set $\mathcal{M}_{n,\theta}, \widehat{\theta}_n \in K^\epsilon$ and $\theta_n^{**} \in K^\epsilon$ for $y^2 \leq \log n$.

Hence it follows from Lemma 3.3 that on the set $\mathcal{M}_{n,\theta}$,

$$\begin{aligned}
& \left| \int_{\mathcal{V}_n(s_K)} D_{10}(\widehat{\theta}_n, \phi) \mathcal{R}_n(d\phi) - \left(\tfrac{1}{nI(\theta)} \right)^{1/2} D_{11}(\widehat{\theta}_n, \widehat{\theta}_n) \int_{S_n} y \widetilde{G}_n(dy) \right| \\
\leq & \left(\frac{1}{nI(\theta)} \right)^{1/2} \int_{S_n} |y| \left| D_{11}(\widehat{\theta}_n, \theta_n^{**}) - D_{11}(\widehat{\theta}_n, \widehat{\theta}_n) \right| \widetilde{G}_n(dy) \\
\leq & C_K \left(\frac{1}{nI(\theta)} \right)^{1/2} \int_{S_n} |y|^2 \left| \theta_n^{**} - \theta_n \right| \widetilde{G}_n(dy) \\
\leq & C_K n^{-1} \int_{S_n} y^2 \widetilde{G}_n(dy).
\end{aligned}$$

Thus

$$\begin{aligned}&\left|\int_{\mathcal{V}_n(s_K)} D_{10}(\widehat{\theta}_n,\phi)R_n(d\phi)\right|\\ \leq\ &\left(\frac{1}{nI(\theta)}\right)^{1/2} D_{11}(\widehat{\theta}_n,\widehat{\theta}_n)\left|\int_{S_n} y\widetilde{G}_n(dy)\right| + C_K n^{-1}\int_{S_n} y^2\widetilde{G}_n(dy)\\ \leq\ &C_K n^{-1}\end{aligned}$$

since from Theorem 2.5, on the set $\mathcal{M}_{n,\theta}$

$$\left|\int_{S_n} y\widetilde{G}_n(dy)\right| \leq C_K n^{-1/2} \text{ and } \left|\int_{S_n} y^2\widetilde{G}_n(dy)\right| \leq C_K. \qquad \square$$

The next theorem provides bound on the asymptotic equivalence of the maximum likelihood and the Bayes estimators.

Theorem 3.10. *Under (A1)–(A3), $(A4)_2$, (B1)–(B8), for every compact $K \subseteq \Theta$ there exist sets $\mathcal{M}_{n,\theta}, \theta \in K$ satisfying (2.1) such that on the set $\mathcal{M}_{n,\theta}$, $\left|\widetilde{\theta}_n - \widehat{\theta}_n\right| \leq C_K n^{-1}$, i.e.,*

$$\sup_{\theta\in K} P_\theta\left\{\left|\widetilde{\theta}_n - \widehat{\theta}_n\right| > C_K n^{-1}\right\} = o(n^{-1/2}).$$

Proof. From the large deviations inequality for the Bayes estimator $\widetilde{\theta}_n$ (see Theorem 3.7) we can assume that $\widetilde{\theta}_n \in \Theta$, on the set $\mathcal{M}_{n,\theta}$. Hence by Taylor expansion

$$\begin{aligned}&\int_\Theta D_{10}(\widetilde{\theta}_n,\phi)\mathcal{R}_n(d\phi)\\ &= \int_\Theta D_{10}(\widehat{\theta}_n,\phi)\mathcal{R}_n(d\phi) + (\widetilde{\theta}_n - \widehat{\theta}_n)\int_\Theta D_{20}(\theta_n^{***},\phi)\mathcal{R}_n(d\phi)\end{aligned} \tag{3.1}$$

where $\left|\theta_n^{***} - \widehat{\theta}_n\right| \leq \left|\widetilde{\theta}_n - \widehat{\theta}_n\right|$.

But $\int_\Theta D_{10}(\widetilde{\theta}_n,\phi)\mathcal{R}_n(d\phi) = 0$ from the definition of the Bayes estimator $\widetilde{\theta}_n$.

Therefore,

$$\widetilde{\theta}_n - \widehat{\theta}_n = -\frac{\int_\Theta D_{10}(\widehat{\theta}_n,\phi)\mathcal{R}_n(d\phi)}{\int_\Theta D_{20}(\theta_n^{***},\phi)\mathcal{R}_n(d\phi)}. \tag{3.2}$$

From Proposition 3.9

$$\left|\int_\Theta D_{10}(\widehat{\theta}_n,\phi)\mathcal{R}_n(d\phi)\right| \leq C_K n^{-1}. \tag{3.3}$$

From Corollary 3.8, we have

$$\left|\int_\Theta D_{20}(\theta_n^{***},\phi)\mathcal{R}_n(d\phi) - D_{20}(\theta,\theta)\right| < \epsilon$$

on the set $\mathcal{M}_{n,\theta}$. Therefore by condition (B6)(b), we have on the set $\mathcal{M}_{n,\theta}$

$$\left|\int_\Theta D_{20}(\theta_n^{***},\phi)\mathcal{R}_n(d\phi)\right| \geq \epsilon_K \tag{3.4}$$

for suitably chosen $\epsilon_K > 0$.

Thus from (3.3) and (3.4), on the set $M_{n,\theta}$

$$\left|\widetilde{\theta}_n - \widehat{\theta}_n\right| \leq C_K n^{-1}.$$

This completes the proof of the theorem. $\square$

Theorem 3.11. *Under (A1)–(A3), $(A4)_2$, (B1)–(B8), for every compact set $K \subseteq \Theta$*

$$\sup_{\theta\in K}\sup_{x\in\mathbb{R}}\left|P_\theta\left\{((nI(\theta))^{1/2}(\widetilde{\theta}_n-\theta)\le x\right\}-\Phi(x)\right|\le C_K n^{-1/2}.$$

Proof. We will use the following elementary inequality. For any two random variables X and Y on probability space $(\Omega,\mathcal{F},P)$, for any $\epsilon>0$

$$\begin{aligned}&\sup_{x\in\mathbb{R}}|P\{X+Y\le x\}-\Phi(x)|\\ \le\ &\sup_{x\in\mathbb{R}}|P\{X\le x\}-\Phi(x)|+P\{|Y|>\epsilon\}+\epsilon.\end{aligned}$$

Hence

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|P_\theta\left\{(nI(\theta))^{1/2}(\widetilde{\theta}_n-\theta)\le x\right\}-\Phi(x)\right|\\ =\ &\sup_{x\in\mathbb{R}}\left|P_\theta\left\{(nI(\theta))^{1/2}(\widehat{\theta}_n-\theta+\widetilde{\theta}_T-\theta_n)\le x\right\}-\Phi(x)\right|\\ \le\ &\sup_{x\in\mathbb{R}}\left|P_\theta\left\{(nI(\theta))^{1/2}(\widehat{\theta}_n-\theta)\le x\right\}-\Phi(x)\right|\\ &+P_\theta\left\{(nI(\theta))^{1/2}|\widetilde{\theta}_n-\widehat{\theta}_n|>\epsilon\right\}+\epsilon.\end{aligned}\tag{3.5}$$

By Theorem 1.3, the first term is $O(n^{-1/2})$. By Theorem 3.10,

$$\sup_{\theta\in K}P_\theta\left\{(nI(\theta))^{1/2}|\widetilde{\theta}_n-\widehat{\theta}_n|>C_K n^{-1/2}\right\}=o(n^{-1/2}).$$

Thus choosing $\epsilon=C_K n^{-1/2}$ in (3.5) it is seen that the r.h.s. of (3.5) is of the order $O(n^{-1/2})$. □

Remarks

(1) Here we obtained the rate of convergence of the posterior distribution and the Bayes estimator when the diffusion process is stationary and ergodic. In the nonergodic case, the posterior distributions and the Bayes estimators converge to a mixture of normal distribution (see Bishwal [5]). It remains open to obtain rate of convergence in this case.

(2) Extension of the results of this paper to multidimensional process and multiparameter case is worth investigating.

(3) Nonuniform rates of convergence, which are more useful remains to be investigated.

(4) Based on discrete observations of an ergodic diffusion process, Bishwal [4] studied the asymptotic normality of pseudo posterior distributions and pseudo Bayes estimators of the drift parameter, for smooth prior and loss functions, based on approximate likelihood functions. Rates of convergence in this case is worth investigating.

(5) Extension problems in this paper to fractional Levy processes which include jumps and long memory is worth investigating.

Competing interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

[1] J.P.N. Bishwal, Bayes and sequential estimation in Hilbert space valued stochastic differential equations, J. Korean Stat. Soc. 28 (1999), 107–121.

[2] J.P.N. Bishwal, Rates of convergence of the posterior distributions and the Bayes estimators in the Ornstein-Uhlenbeck process, Random Oper. Stoch. Equ. 8 (2000), 51–70.

[3] J.P.N. Bishwal, The Bernstein-von Mises theorem and spectral asymptotics of Bayes estimators for parabolic SPDEs, J. Aust. Math. Soc. 72 (2002), 287–298.

[4] J.P.N. Bishwal, Parameter estimation in stochastic differential equations, Lecture Notes in Mathematics, vol. 1923, Springer, Berlin, 2008.

[5] J.P.N. Bishwal, Conditional least squares estimation for discretely sampled nonergodic diffusions, Asian Res. J. Math. 7 (2017), 1–18.

[6] J.P.N. Bishwal, Bernstein-von Mises theorem and small noise asymptotics of Bayes estimators for parabolic stochastic partial differential equations, Theory Stoch. Process. 23 (2018), 6–17.

[7] J.P.N. Bishwal, Berry-Esseen bounds of approximate Bayes estimators for the discretely observed Ornstein-Uhlenbeck process, Asian J. Stat. Sci. 1 (2021), 83–122.

[8] J. Borwanker, G. Kallianpur, B.L.S. Prakasa Rao, The Bernstein-von Mises theorem for Markov processes, Ann. Math. Stat. 42 (1971), 1241–1253.

[9] A. Bose, The Bernstein-von Mises theorem for a certain class of diffusion processes, Sankhyā Ser. A 45 (1983), 150–160.

[10] J.L. Doob, Application of the theory of martingales, Colloq. Internat. CNRS 13 (1949), 23–27.

[11] D. Freedman, Wald lecture: on the Bernstein-von Mises theorem with infinite-dimensional parameters, Ann. Stat. 27 (1999), 1119–1141.

[12] S. Ghosal, J.K. Ghosh, A.W. van der Vaart, Convergence rates of posterior distributions, Ann. Stat. 28 (2000), 500–531.

[13] V. Harison, The Bernstein-von Mises theorem for a certain class of Gaussian diffusion processes, Publ. Serv. Math. Univ. Antananarivo 6 (1992), 1–7.

[14] C. Hipp, R. Michel, On the Bernstein-v. Mises approximation of posterior distributions, Ann. Stat. 4 (1976), 972–980.

[15] T.M. Huang, Convergence rates of posterior distributions and adaptive estimation, Technical Report No. 711, Department of Statistics, Carnegie Mellon University, 2000.

[16] I.A. Ibragimov, R.Z. Has'minskii, Statistical estimation: asymptotic theory, Springer, New York, 1981.

[17] J. Jacod, A.N. Shiryaev, Limit theorems for stochastic processes, Springer, Berlin, 1987.

[18] L. Le Cam, On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, Univ. Calif. Publ. Stat. 1 (1953), 277–330.

[19] L. Le Cam, Convergence of estimates under dimensionality restrictions, Ann. Stat. 1 (1973), 38–53.

[20] L. Le Cam, Asymptotic methods in statistical decision theory, Springer, New York, 1986.

[21] L. Le Cam, G.L. Yang, Asymptotics in statistics: some basic concepts, Springer, New York, 1990.

[22] M.N. Mishra, The Bernstein-von Mises theorem for a class of non-homogeneous diffusion processes, Stat. Decis. 7 (1989), 153–165.

[23] M. Podolskij, N. Yoshida, Edgeworth expansion for functionals of continuous diffusion processes, Ann. Appl. Probab. 26 (2016), 3415–3455.

[24] B.L.S. Prakasa Rao, The rate of convergence of Bernstein-von Mises approximation for Markov processes, Serdica 4 (1978), 36–42.

[25] B.L.S. Prakasa Rao, The Bernstein-von Mises theorem for a class of diffusion processes, Teor. Sluchajnykh Protsessov 9 (1981), 95–101 (in Russian).

[26] B.L.S. Prakasa Rao, On Bayes estimation for diffusion fields, in: J.K. Ghosh, J. Roy (Eds.), Statistics: Applications and New Directions, Statistical Publishing Society, Calcutta, 1984, pp. 504–511.

[27] L. Schwartz, On Bayes procedures, Z. Wahrscheinlichkeitstheor. Verw. Geb. 4 (1965), 10–26.

[28] X. Shen, L. Wasserman, Rates of convergence of posterior distributions, Technical Report No. 678, Department of Statistics, Carnegie Mellon University, 1998.

[29] H. Strasser, The asymptotic equivalence of Bayes and maximum likelihood estimation, J. Multivariate Anal. 5 (1975), 206–226.
[30] H. Strasser, Asymptotic properties of posterior distributions, Z. Wahrscheinlichkeitstheor. Verw. Geb. 35 (1976), 269–282.
[31] H. Strasser, Improved bounds for equivalence of Bayes and maximum likelihood estimation, Theory Probab. Appl. 22 (1978), 349–361.
[32] M. Uchida, N. Yoshida, Quasi likelihood analysis of volatility and nondegeneracy of statistical random field, Stoch. Process. Appl. 123 (2013), 2851–2876.
[33] N. Yoshida, Malliavin calculus and asymptotic expansion for martingales, Probab. Theory Relat. Fields 109 (1997), 301–342.
[34] N. Yoshida, Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations, Ann. Inst. Stat. Math. 63 (2011), 431–479.
[35] N. Yoshida, Partial quasi-likelihood analysis, Jpn. J. Stat. Data Sci. 1 (2018), 157–189.
[36] N. Yoshida, Simplified quasi-likelihood analysis for a locally asymptotically quadratic random field, Ann. Inst. Stat. Math. 77 (2025), 1–24.