On Bayesian Asymptotics in Stochastic Differential Equations with Random Effects

Trisha Maitra and Sourabh Bhattacharya

Abstract

[Delattre et al.](2012) investigated asymptotic properties of the maximum likelihood estimator of the hyperparameters of the random effect parameters associated with \( n \) independent stochastic differential equations (SDE’s) assuming that the SDE’s are independent and identical (iid).

In this article, we consider the Bayesian approach to learning about the hyperparameters, and prove consistency and asymptotic normality of the posterior distribution of the hyperparameters in the iid set-up as well as when the SDE’s are independent but non-identical.

Keywords: Asymptotic normality; Maximum likelihood estimator; Posterior consistency; Posterior normality; Random effects; Stochastic differential equations.

1 Introduction

[Delattre et al.](2012) consider classical inference in the context of mixed-effects stochastic differential equations (SDE’s) having the following form: for \( i = 1, \ldots, n \),

\[
dX_i(t) = b(X_i(t), \phi_i)dt + \sigma(X_i(t))dW_i(t),
\]

where, for \( i = 1, \ldots, n \), \( X_i(0) = x^i \) is the initial value of the stochastic process \( X_i(t) \), which is assumed to be continuously observed on the time interval \([0, T_i] ; T_i > 0\) assumed to be known. In the context of statistical modelling, \( X_i(\cdot) \) models the \( i \)-th individual. The SDE’s given by (1.1) are driven by independent standard Weiner processes \( \{W_i(\cdot) ; i = 1, \ldots, n\} \), and \( \{\phi_i ; i = 1, \ldots, n\} \), which are to be interpreted as the random effect parameters associated with the \( n \) individuals, are assumed to be independent of the Brownian motions and independently and identically distributed (iid) random variables with common distribution \( g(\varphi, \theta)dv(\varphi) \). Here \( g(\varphi, \theta) \) is a density with respect to a dominating measure on \( \mathbb{R}^d \) (\( \mathbb{R} \) is the real line and \( d \) is the dimension) for all \( \theta \), where \( \theta \in \Omega \subset \mathbb{R}^d \) is the unknown parameter of interest, which is to be estimated. [Delattre et al.](2012) impose regularity conditions that ensure existence of solutions of (1.1). We adopt their assumptions, which they denote by (H1), (H2) and (H3). As in [Delattre et al.](2012), for statistical inference we let \( b(x, \phi_i) = \phi_i b(x) \), and assume that \( b(\cdot) \) and \( \sigma(\cdot) \) are real, continuous functions, having linear growth.

Assuming that \( g(\varphi, \theta)dv(\varphi) \equiv N(\mu, \omega^2) \), [Delattre et al.](2012) obtain the likelihood \( L(\theta) \) as the product of the following:

\[
\lambda_i(X_i, \theta) = \frac{1}{(1 + \omega^2 V_i)^{1/2}} \exp \left[ - \frac{V_i}{2(1 + \omega^2 V_i)} \left( \frac{\mu - U_i V_i}{V_i} \right)^2 \right] \exp \left( \frac{U_i^2}{2V_i} \right),
\]

where \( \theta = (\mu, \omega^2) \in \mathbb{R} \times \mathbb{R}^+ \) \((R^+ = (0, \infty))\), and

\[
U_i = \int_0^{T_i} \frac{b(X_i(s))}{\sigma^2(X_i(s))} dX_i(s), \quad V_i = \int_0^{T_i} \frac{b^2(X_i(s))}{\sigma^2(X_i(s))} ds; \quad i = 1, \ldots, n,
\]

are sufficient statistics. We adopt assumption (H4) of [Delattre et al.](2012), that the function \( b(\cdot)/\sigma(\cdot) \) is not constant, and that, for \( i \geq 1 \), \((U_i, V_i)\) admits a density \( \varphi_i(u, v) \) with respect to the Lebesgue measure on \( \mathbb{R} \times \mathbb{R}^+ \) which is jointly continuous and positive on an open ball of \( \mathbb{R} \times \mathbb{R}^+ \). In the iid situation, \( \varphi_i = \varphi_1 \) for \( i \geq 1 \).

*Trisha Maitra is a PhD student and Sourabh Bhattacharya is an Assistant Professor in Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108. Corresponding e-mail: sourabh@isical.ac.in.
Delattre et al. (2012) consider the iid set-up by setting $x^i = x$ and $T_i = T$ for $i = 1, \ldots, n$, and independently investigate asymptotic properties of the MLE of $\theta$, without invoking the general results already existing in the literature. As an alternative, Maitra and Bhattacharya (2014) verify the regularity conditions of the existing results to prove asymptotic properties of the MLE in this SDE set-up. Not only in the iid situation, Maitra and Bhattacharya (2014) prove asymptotic results related to the MLE even in the independent but non-identical (we refer to this as non-iid) case.

In this article, we consider the Bayesian framework, for both iid and non-iid set-ups, and prove consistency and asymptotic normality of the Bayesian posterior distribution of $\theta = (\mu, \omega^2)$. In what follows, in Section 2 we investigate asymptotic properties of the posterior in the iid context. In Section 3 we investigate Bayesian asymptotics in the non-iid set-up.

Notationally, “$\xrightarrow{d}$”, “$P_x$” and “$\xrightarrow{L}$” denote convergence “almost surely”, “in probability” and “in distribution”, respectively.

2 Consistency and asymptotic normality of the Bayesian posterior in the iid set-up

2.1 Consistency of the Bayesian posterior distribution

Theorem 7.80 presented in Schervish (1995) provides easy-to-verify sufficient conditions that ensure posterior consistency. We state the theorem below, using which we prove posterior consistency in our case.

**Theorem 1** (Schervish (1995)) Let $\{X_n\}_{n=1}^{\infty}$ be conditionally iid given $\theta$ with density $f_1(x|\theta)$ with respect to a measure $\nu$ on a space $(X^1, B^1)$. Fix $\theta_0 \in \Omega$, and define, for each $M \subseteq \Omega$ and $x \in X^1$,

$$Z(M, x) = \inf_{\psi \in M} \log \frac{f_1(x|\theta_0)}{f_1(x|\psi)}.$$  

Assume that for each $\theta \neq \theta_0$, there is an open set $N_0$ such that $\theta \in N_0$ and that $E_{\theta_0} Z(N_0, X_1) > -\infty$. Also assume that $f_1(x|\cdot)$ is continuous at $\theta$ for every $\theta$, a.s. $[P_{\theta_0}]$. For $\epsilon > 0$, define $C_\epsilon = \{\theta : K_1(\theta_0, \theta) < \epsilon\}$, where

$$K_1(\theta_0, \theta) = E_{\theta_0} \left( \log \frac{f_1(X_1|\theta_0)}{f_1(X_1|\theta)} \right)$$  \hspace{1cm} (2.1)

is the Kullback-Leibler divergence measure associated with observation $X_1$. Let $\pi$ be a prior distribution such that $\pi(C_\epsilon) > 0$, for every $\epsilon > 0$. Then, for every $\epsilon > 0$ and open set $N_0$ containing $C_\epsilon$, the posterior satisfies

$$\lim_{n \to \infty} \pi(N_0|X_1, \ldots, X_n) = 1, \text{ a.s. } [P_{\theta_0}].$$  \hspace{1cm} (2.2)

2.1.1 Verification of posterior consistency

The conditions of Theorem 1 above are verified in the context of Theorem 1 in Maitra and Bhattacharya (2014). Here, all we need to ensure is that there exists a prior $\pi$ which gives positive probability to $C_\epsilon$ for every $\epsilon > 0$. Since $K_1(\theta_0, \theta) = 0$ if and only if $\theta = \theta_0$, for any $\epsilon > 0$, the set $C_\epsilon$ is non-empty provided that $\Omega \setminus \{\theta_0\}$ is non-empty.

Let $d\theta = g(\theta) d\theta$ almost everywhere on $\Omega$, where $g(\theta)$ is any positive, continuous density on $\Omega$ with respect to the Lebesgue measure $\nu$.

Since Delattre et al. (2012) show that $K_1(\theta_0, \theta)$ is continuous in $\theta$, and since the parameter space $\Omega$ is compact, it follows that $K_1(\theta_0, \theta)$ is uniformly continuous on $\Omega$. Hence, for any $\epsilon > 0$, there exists $\delta_\epsilon$ which is independent of $\theta$, such that $||\theta - \theta_0|| \leq \delta_\epsilon$ implies $K_1(\theta_0, \theta) < \epsilon$.
Hence,

\[ \pi(C_\epsilon) \geq \pi(\{\theta : \|\theta - \theta_0\| \leq \delta_\epsilon\}) \geq \left[ \inf_{\{\theta : \|\theta - \theta_0\| \leq \delta_\epsilon\}} g(\theta) \right] \times \nu(\{\theta : \|\theta - \theta_0\| \leq \delta_\epsilon\}) > 0. \]  

Hence, (2.2) holds in our case with any prior with continuous density with respect to the Lebesgue measure.

### 2.2 Asymptotic normality of the Bayesian posterior distribution

We now investigate asymptotic normality of posterior distributions in our SDE set-up. For our purpose, we make use of Theorem 7.102 in conjunction with Theorem 7.89 provided in Schervish (1995). These theorems make use of seven regularity conditions, of which only the first four will be required for the iid set-up. Hence, in this iid context we state the four requisite conditions.

#### 2.2.1 Regularity conditions – iid case

1. The parameter space is \( \Omega \subseteq \mathbb{R}^d \) for some finite \( d \).

2. \( \theta_0 \) is a point interior to \( \Omega \).

3. The prior distribution of \( \Theta \) has a density with respect to Lebesgue measure that is positive and continuous at \( \theta_0 \).

4. There exists a neighborhood \( \mathcal{N}_0 \subseteq \Omega \) of \( \theta_0 \) on which \( \ell_n(\theta) = \log f(X_1, \ldots, X_n|\theta) \) is twice continuously differentiable with respect to all co-ordinates of \( \theta \), a.s. \([P_{\theta_0}]\).

Before proceeding to justify asymptotic normality of our posterior, we furnish the relevant theorem below (Theorem 7.102 of Schervish (1995)).

**Theorem 2 (Schervish (1995))** Let \( \{X_n\}_{n=1}^{\infty} \) be conditionally iid given \( \theta \). Assume the above four regularity conditions; also assume that there exists \( H_r(x, \theta_0) \) such that, for each \( \theta_0 \in \text{int}(\Omega) \) and each \( k, j \),

\[ \sup_{\|\theta - \theta_0\| \leq r} \left| \frac{\partial^2}{\partial \theta_k \partial \theta_j} \log f(X_1|\theta_0) - \frac{\partial^2}{\partial \theta_k \partial \theta_j} \log f(X_1|\theta_0) \right| \leq H_r(x, \theta_0), \]  

with

\[ \lim_{r \to 0} \mathbb{E}_{\theta_0} H_r(X, \theta_0) = 0. \]  

Further suppose that the conditions of Theorem 1 hold, and that the Fisher’s information matrix \( I(\theta_0) \) is positive definite. Now let

\[ \Sigma_n^{-1} = \begin{cases} -\ell_n''(\hat{\theta}_n) & \text{if the inverse and } \hat{\theta}_n \text{ exist} \\ \mathbb{1}_d & \text{if not}, \end{cases} \]  

where, for any \( t \),

\[ \ell_n'(t) = \left( \left( \frac{\partial}{\partial \theta_k} \ell_N(\theta) \right) \right)_{\theta = t}, \]  

and \( \mathbb{1}_d \) is the identity matrix of order \( d \). Thus, \( \Sigma_n^{-1} \) is the observed Fisher’s information matrix.

Letting \( \Psi_n = \Sigma_n^{-1/2} (\theta - \hat{\theta}_n) \), it follows that for each compact subset \( B \) of \( \mathbb{R}^d \) and each \( \epsilon > 0 \), it holds that

\[ \lim_{n \to \infty} P_{\theta_0} \left( \sup_{\psi \in B} \left| \pi(\psi|X_1, \ldots, X_n) - \tilde{\phi}(\psi) \right| > \epsilon \right) = 0, \]  

\[ \text{(2.8)} \]
where \( \tilde{\phi}(\cdot) \) denotes the density of the standard normal distribution.

### 2.2.2 Verification of posterior normality

The first three regularity conditions in Section 2.2.1 trivially hold. The remaining conditions of Theorem 2 are verified in the context of Theorem 2 of Maitra and Bhattacharya (2014). Hence, (2.3) holds in our SDE set-up.

### 3 Consistency and asymptotic normality of the Bayesian posterior in the non-iid set-up

In this section, as in Maitra and Bhattacharya (2014), we do not enforce the restrictions \( T_i = T \) and \( x^i = x \) for \( i = 1, \ldots, n \). Consequently, here we deal with the set-up where the processes \( X_i(\cdot); \ i = 1, \ldots, n \), are independently, but not identically distributed. Following Maitra and Bhattacharya (2014), we assume that the sequences \( \{T_1, T_2, \ldots\} \) and \( \{x^1, x^2, \ldots\} \) are sequences in compact sets \( \mathcal{I} \) and \( \mathcal{X} \), respectively, so that there exist convergent subsequences with limits in \( \mathcal{I} \) and \( \mathcal{X} \). For notational convenience, we continue to denote the convergent subsequences as \( \{T_1, T_2, \ldots\} \) and \( \{x^1, x^2, \ldots\} \). Let us denote the limits by \( T^\infty \) and \( x^\infty \), where \( T^\infty \in \mathcal{I} \) and \( x^\infty \in \mathcal{X} \).

Under mild assumptions, Maitra and Bhattacharya (2014) prove continuity of the moments as functions of \( x, T \) and \( \theta \). In particular, the Kullback-Leibler distance and the information matrix, which we denote by \( K_{x,T}(\theta_0, \theta) \) (or, \( K_{x,T}(\theta, \theta_0) \)) and \( I_{x,T}(\theta) \) to emphasize dependence on the initial values \( x \) and \( T \), are continuous in \( x, T \) and \( \theta \). For \( x = x^k \) and \( T = T_k \), if we denote the Kullback-Leibler distance and the Fisher’s information as \( K_k(\theta_0, \theta) \) (or, \( K_k(\theta, \theta_0) \)) and \( I_k(\theta) \), respectively, then continuity of \( K_{x,T}(\theta_0, \theta) \) (or, \( K_{x,T}(\theta, \theta_0) \)) and \( I_{x,T}(\theta_0) \) with respect to \( x \) and \( T \) ensures that as \( x^k \to x^\infty \) and \( T_k \to T^\infty \), \( K_{x^k,T_k}(\theta_0, \theta) \to K_{x^\infty,T^\infty}(\theta_0, \theta) = K(\theta_0, \theta) \), say. Similarly, \( K_{x^k,T_k}(\theta, \theta_0) \to K(\theta, \theta_0) \) and \( I_{x^k,T_k}(\theta) \to I_{x^\infty,T^\infty}(\theta) = I(\theta) \), say. Thanks to compactness, the limits \( K(\theta_0, \theta) \), \( K(\theta, \theta_0) \) and \( I(\theta) \) are well-defined Kullback-Leibler divergences and Fisher’s information, respectively. Consequently (see Maitra and Bhattacharya (2014)), the following hold for any \( \theta \in \Omega \),

\[
\begin{align*}
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} K_k(\theta_0, \theta) &= K(\theta_0, \theta); \\
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} K_k(\theta, \theta_0) &= K(\theta, \theta_0); \\
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} I_k(\theta) &= I(\theta).
\end{align*}
\]

The above results will be seen to have important roles as we proceed with the non-iid Bayesian set-up. For consistency in the Bayesian framework we utilize the theorem of Choi and Schervish (2007), and for asymptotic normality of the posterior we make use of Theorem 7.89 of Schervish (1995).

### 3.1 Posterior consistency in the non-iid set-up

Before we proceed, we need to state an extra assumption which will be necessary in this context. Recall our assumption that \( (U_i, V_i) \) admits a density with respect to the Lebesgue measure and that the density is continuous and positive on an open ball of \( \mathbb{R} \times \mathbb{R}^+ \). Here we additionally assume that the density decays sufficiently fast outside the open ball. More formally, we assume that there exists a strictly positive function \( \alpha^*(x, T, \theta) \), continuous in \( (x, T, \theta) \), such that for any \( (x, T, \theta) \),

\[
E_{\theta} \left[ \exp \left\{ \alpha^*(x, T, \theta)U^2(x, T) \right\} \right] < \infty, \tag{3.4}
\]

Let

\[
\alpha^*_{\min} = \inf_{x \in \mathcal{X}, T \in \mathcal{I}, \theta \in \Omega} \alpha^*(x, T, \theta), \tag{3.5}
\]
and
\[ \alpha = \min \{ \alpha_{\text{min}}^*, e^* \}, \tag{3.6} \]
where \(0 < c^* < 1/16\). Compactness ensures that \(\alpha_{\text{min}}^* > 0\), so that \(0 < \alpha < 1/16\). It also holds due to compactness that for \(\theta \in \Omega\),
\[ \sup_{x \in \mathcal{X}, T \in \mathcal{T}} E_\theta \left[ \exp \left\{ \alpha U^2(x, T) \right\} \right] < \infty. \tag{3.7} \]
This choice of \(\alpha\) ensuring (3.7) will be useful in verification of the conditions of Theorem 3, which we next state.

**Theorem 3 (Choi and Schervish (2007))** Let \(\{X_i\}_{i=1}^\infty\) be independently distributed with densities \(\{f_i(\cdot | \theta)\}_{i=1}^\infty\), with respect to a common \(\sigma\)-finite measure, where \(\theta \in \Omega\), a measurable space. The densities \(f_i(\cdot | \theta)\) are assumed to be jointly measurable. Let \(\theta_0 \in \Omega\) and let \(P_{\theta_0}\) be the joint distribution of \(\{X_i\}_{i=1}^\infty\) when \(\theta_0\) is the true value of \(\theta\). Let \(\{U_n\}_{n=1}^\infty\) be a sequence of subsets of \(\Omega\). Let \(\theta\) have prior \(\pi\) on \(\Omega\). Define the following:
\[ \Lambda_i(\theta_0, \theta) = \log \frac{f_i(X_i|\theta_0)}{f_i(X_i|\theta)}, \]
\[ K_i(\theta_0, \theta) = E_{\theta_0} \left( \Lambda_i(\theta_0, \theta) \right), \]
\[ W_i(\theta_0, \theta) = \text{Var}_{\theta_0} \left( \Lambda_i(\theta_0, \theta) \right). \]

Make the following assumptions:

1. Suppose that there exists a set \(B\) with \(\pi(B) > 0\) such that
   \[ (i) \sum_{i=1}^\infty \frac{W_i(\theta_0, \theta)}{i^2} < \infty, \quad \forall \theta \in B, \]
   \[ (ii) \text{For all } \epsilon > 0, \pi \left( B \cap \{ \theta : K_i(\theta_0, \theta) < \epsilon, \forall i \} \right) > 0. \]

2. Suppose that there exist test functions \(\{\Phi_n\}_{n=1}^\infty\), sets \(\{\Omega_n\}_{n=1}^\infty\) and constants \(C_1, C_2, c_1, c_2 > 0\) such that
   \[ (i) \sum_{i=1}^\infty E_{\theta_0} \Phi_n < \infty, \]
   \[ (ii) \sup_{\theta \in U_i \cap \Omega_n} E_\theta \left( 1 - \Phi_n \right) \leq C_1 e^{-c_1 n}, \]
   \[ (iii) \pi \left( \Omega_n \right) \leq C_2 e^{-c_2 n}. \]

Then,
\[ \pi \left( \theta \in U_n^c | X_1, \ldots, X_n \right) \rightarrow 0 \quad \text{a.s.} \quad [P_{\theta_0}] \tag{3.8} \]

### 3.1.1 Validation of posterior consistency

From the proof of Proposition 7 of [Delattre et al. (2012)] it follows that \(\left| \log \frac{f_i(X_i|\theta_0)}{f_i(X_i|\theta)} \right|\) has an upper bound which has finite expectation and square of expectation under \(\theta_0\), and is uniform for all \(\theta \in B\), where \(B\) is of the form \([\mu_{\theta}, \overline{\mu}] \times [\overline{\omega}^2, \omega^2]\), say, with \(\mu < \overline{\mu}\) and \(0 < \overline{\omega}^2 < \omega^2\). Hence, for each \(i\), \(W_i(\theta_0, \theta)\) is finite. Moreover, since the sequences \(\{T_1, T_2, \ldots\}\) and \(\{x^1, x^2, \ldots\}\) belong to compact spaces \(\mathcal{X}\) and \(\mathcal{X}\), and the variance function \(W_{x,T}(\theta_0, \theta)\) viewed as a function of \(x\) and \(T\), is bounded by a function continuous in \(x\) and \(T\), \(W_i(\theta_0, \theta) < \kappa\), for some \(0 < \kappa < \infty\), uniformly in \(i\). Continuity of \(W_{x,T}(\theta_0, \theta)\) follows as an application of Theorem 3 of [Maitra and Bhattacharya (2014)] where the required uniform integrability is assured by finiteness of the moments (see [Delattre et al. (2012)]) for every \(x \in \mathcal{X}, T \in \mathcal{T}\) and compactness of \(\mathcal{X}\) and \(\mathcal{T}\). Hence, choosing a prior that gives positive probability to the set \(B\), it follows that for all \(\theta \in B\),
\[ \sum_{i=1}^\infty \frac{W_i(\theta_0, \theta)}{i^2} < \kappa \sum_{i=1}^\infty \frac{1}{i^2} < \infty. \]
Hence, condition (1)(i) holds.

To verify (1)(ii) note that because of compactness of $B$, $K_i(\theta_0, \theta)$, which is continuous in $\theta$, is uniformly continuous in $B$. Hence, for every $\varepsilon > 0$, there exists $\delta_i(\varepsilon)$ independent of $\theta$ such that $\|\theta - \theta_0\| < \delta_i(\varepsilon)$ implies $K_i(\theta_0, \theta) < \varepsilon$. Let us define

$$
\delta(\varepsilon) = \inf \{ \delta_{x,T}(\varepsilon) : x \in \mathcal{X}, T \in \mathcal{T} \},
$$

where $\delta_{x,T}(\varepsilon)$ is any strictly positive continuous function of $x$ and $T$, depending upon $\varepsilon$ such that $\delta_{x,T}(\varepsilon) = \delta_i(\varepsilon)$, for every $i = 1, 2, \ldots$. Compactness of $\mathcal{X}$ and $\mathcal{T}$ ensures that $\delta(\varepsilon) > 0$. So, for any $\varepsilon > 0$,

$$
\{ \theta \in B : K_i(\theta_0, \theta) < \varepsilon, \ \forall \ i \} \supseteq \{ \theta \in B : \|\theta - \theta_0\| < \delta(\varepsilon) \}.
$$

It follows that

$$
\pi(B \cap \{ \theta : K_i(\theta_0, \theta) < \varepsilon, \ \forall \ i \}) \geq \pi(\{ \theta \in B : \|\theta - \theta_0\| < \delta(\varepsilon) \}). \tag{3.11}
$$

The remaining part of the proof that the right hand side of (3.11) is strictly positive, follows exactly in the same way as the proof of strict positivity (2.3) in Section 2.1.1 with a continuous prior density on $\Omega$ with respect to the Lebesgue measure.

We now verify conditions (2)(i), (2)(ii) and (2)(iii). We let $\Omega_n = (\Omega_{1n} \times \mathbb{R}^+)$, where $\Omega_{1n} = \{ \mu : |\mu| < M_n \}$, where $M_n = O(e^n)$. Note that

$$
\pi(\Omega_n^\varepsilon) = \pi(\Omega_{1n}^\varepsilon) = \pi(|\mu| > M_n) < E_\pi(|\mu|) M_n^{-1}, \tag{3.12}
$$

so that (2)(iii) holds, assuming that the prior $\pi$ is such that the expectation $E_\pi(|\mu|)$ is finite.

Fixing $\delta > 0$, we construct the tests $\Phi_n$ as follows.

$$
\Phi_n = \begin{cases} 
1 & \text{if } \beta_n < \sqrt{e^{-n\delta}}, \\
0 & \text{otherwise},
\end{cases} \tag{3.13}
$$

where

$$
\beta_n = \frac{L_n(\theta_0)}{\sup_{\theta \in \Omega_n} L_n(\theta)} = \frac{L_n(\theta_0)}{L_n(\hat{\theta}_n)} \tag{3.14}
$$

is the likelihood ratio test statistic under $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Here $L_n(\theta) = \prod_{i=1}^n f_i(X_i|\theta)$, and $\hat{\theta}_n$ is the MLE associated with $n$ observations. Now, denoting $-2 \log \beta_n$ by $Z_n^2$, we obtain for $\alpha$ given by (3.6),

$$
E_{\theta_0} \Phi_n = P_{\theta_0} \left( \beta_n < \sqrt{e^{-n\delta}} \right) = P_{\theta_0} \left( Z_n^2 > n\delta \right) < e^{-\alpha n\delta} E_{\theta_0} \left( e^{\alpha Z_n^2} \right). \tag{3.15}
$$

Note that $Z_n^2 = -n \left( \hat{\theta}_n - \theta_0 \right)^T \ell''_{n,i}(\theta_0)^{\ast\ast} \left( \hat{\theta}_n - \theta_0 \right)$, where $\ell_n(\theta) = \sum_{i=1}^n \log f_i(X_i|\theta)$, and $\theta_n^\ast$ lies between $\theta_0$ and $\hat{\theta}_n$. Also,

$$
\ell''_{n,ij}(\theta_n^\ast) = \ell''_{n,ij}(\theta_0^\ast) + \left( \theta_n^\ast - \theta_0 \right)^T \ell''_{n,ij}(\theta_0^\ast) \tag{3.16}
$$

where $\ell''_{n,ij}$ is the $(i, j)$-th element of $\ell''_n$ and $\ell''_{n,ij}$ is its derivative, and $\theta_n^\ast$ lies between $\theta_0$ and $\theta_n^\ast$. Using Kolmogorov’s strong law of large numbers for the non-iid case (see, for example, [3.3] in Section 2.1.1), which holds in our problem due to finiteness of the moments for every $x$ and $T$ belonging to the compact spaces $\mathcal{X}$ and $\mathcal{T}$, respectively, yields, in conjunction with (3.3), that

$$
\ell''_{n,ij}(\theta_0^\ast) \xrightarrow{a.s.} \mathcal{I}(\theta_0). \tag{3.17}
$$
Also, by Cauchy-Schwartz,
\[
\left| (\hat{\theta}_n^* - \theta_0)^T \frac{\ell'''_{n,i,j}(\theta_n^{**})}{n} \right| \leq \|\hat{\theta}_n^* - \theta_0\| \cdot \left\| \frac{\ell'''_{n,i,j}(\theta_n^{**})}{n} \right\|. \tag{3.18}
\]
In (3.18), due to boundedness of the third derivative (see the proof of Proposition 8 of Delattre et al. (2012)), and due to continuity of the moments with respect to \( x \) and \( T \) (which follows from Theorem 3 of Maitra and Bhattacharya (2014) where uniform integrability is ensured by finiteness of the moments for every \( x, T \) belonging to compact sets \( \mathcal{X} \) and \( \mathcal{T} \), and then finally applying Kolmogorov’s strong law of large numbers for the non-iid case, it can be easily shown that \( \|\hat{\theta}_n - \theta_0\| = O_P(1) \). Since \( \|\hat{\theta}_n - \theta_0\| = o_P(1) \), it follows that \( \|\hat{\theta}_n^* - \theta_0\| = o_P(1) \) as well. Hence,
\[
\left| (\hat{\theta}_n^* - \theta_0)^T \frac{\ell'''_{n,i,j}(\theta_n^{**})}{n} \right| = o_P(1),
\]
implicating that
\[
\frac{\ell'''_{n,i,j}(\theta_n^{**})}{n} \xrightarrow{P} \mathcal{I}(\theta_0). \tag{3.20}
\]
Hence, due to (3.20) and due to asymptotic normality of MLE in our non-iid set-up addressed in Maitra and Bhattacharya (2014), under \( P_{\theta_0} \),
\[
Z_n^2 = -n (\hat{\theta}_n - \theta_0)^T \frac{\ell''_{n,i,j}}{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_1,
\]
and so, by the continuous mapping theorem, \( e^{\alpha Z_n^2} \xrightarrow{d} e^{\alpha \chi^2_1} \). Moreover, using the form \( Z_n^2 = -2 \log \beta_n = -2 \sum_{i=1}^{n} \left( \log f(X_i | \theta_0) - \log f(X_i | \hat{\theta}_n) \right) \), we can write
\[
E_{\theta_0} \left( e^{\alpha Z_n^2} \right)^2 \tag{3.22}
= E_{\theta_0} \left[ \exp \left( -4 \alpha \sum_{i=1}^{n} \left( \log f(X_i | \theta_0) - \log f(X_i | \hat{\theta}_n) \right) \right) \right]
= E_{\theta_0} E_{X_1, \ldots, X_n | \theta_0} \left[ \exp \left( -4 \alpha \sum_{i=1}^{n} \left( \log f(X_i | \theta_0) - \log f(X_i | \hat{\theta}_n) \right) \right) \right]
= E_{\theta_0} \prod_{i=1}^{n} E_{X_i | \theta_0} \left[ \exp \left( -4 \alpha \left( \log f(X_i | \theta_0) - \log f(X_i | \hat{\theta}_n) \right) \right) \right]. \tag{3.23}
\]
Note that \( \hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, X_n) = \hat{\theta}_n(U_1, \ldots, U_n, V_1, \ldots, V_n) \) is a many-to-one function of \( (X_1, \ldots, X_n) \); in particular, it is a many-to-one function of \( X_i = (U_i, V_i) \) (see the expressions yielding MLE in page 327 of Delattre et al. (2012)). Consequently, the conditional distribution of \( X_i \) given \( \hat{\theta}_n \) is non-degenerate.
It follows from the lower bound obtained in the proof of Proposition 7 of Delattre et al. (2012), that conditional on \( \hat{\theta}_n = \varphi = (\mu, \omega^2), \log f(X_i | \theta_0) - \log f(X_i | \hat{\theta}_n) \geq C_3(U_i, V_i, \varphi) \), where
\[
C_3(U_i, V_i, \varphi) = \frac{1}{2} \left\{ \log \left( 1 + \frac{\omega^2}{\omega_0^2} \right) + \frac{\omega^2 - \omega_0^2}{\omega_0^2} \right\} - \frac{1}{2} \left( \frac{\omega^2}{1 + \omega_0^2 V_i} \right)^2 - \frac{1}{2} \left( \frac{\omega_0^2 V_i}{1 + \omega_0^2 V_i} \right)^2 \tag{3.24}
= |\mu| \left| \frac{U_i}{1 + \omega_0^2 V_i} \right| \left( 1 + \frac{\omega_0^2 - \omega^2}{\omega_0^2} \right) - \frac{1}{2} \left| \frac{\omega_0^2 V_i}{1 + \omega_0^2 V_i} \right| - \frac{\mu_0 U_i}{1 + \omega_0^2 V_i}.
\]
Hence, for every given \( n \geq 1 \), due to the lower bound (3.24) and the moment existence assumption
where \( K(2)(i) \) holds. Using this in conjunction with summation over (3.15), it is easily seen that condition proving uniform integrability of \( \alpha \) exists, so that

\[
E_{\theta_0} \left( e^{\alpha Z_n^2} \right)^2 = E_{\theta_n}[E_{\theta_0}(\varphi)] \leq \sup_{\vartheta \in \Omega} E_n(\vartheta) < \infty, \text{ for any given } n. \text{ So, for } n \text{ at most finite,}
\]

\[
\sup_{n \text{ at most finite}} E_{\theta_0} \left( e^{\alpha Z_n^2} \right)^2 < \infty. \tag{3.25}
\]

In our problem, for large enough \( n \), at most the following case can occur: for any given \( \epsilon > 0 \), there exists \( N_0(\epsilon) \) such that

\[
\left| E_{\theta_0} \left( e^{\alpha Z_n^2} \right)^2 - E_{\theta_0} \left( e^{\alpha \chi_n^2} \right)^2 \right| < \epsilon \text{ for } n \geq N_0(\epsilon), \text{ where } E_{\theta_0} \left( e^{\alpha \chi_n^2} \right)^2 < \infty.
\]

Combining this with (3.25) it follows that

\[
\sup_{n \geq 1} E_{\theta_0} \left( e^{\alpha Z_n^2} \right)^2 < \infty, \tag{3.26}
\]

proving uniform integrability of \( \left\{ e^{\alpha Z_n^2} \right\}_{n=1}^{\infty} \). Consequently, \( E_{\theta_0} \left( e^{\alpha Z_n^2} \right) \to E_{\theta_0} \left( e^{\alpha \chi_n^2} \right) \), which is a finite quantity. Using this in conjunction with summation over (3.15), it is easily seen that condition (2)(ii) holds.

Let us now verify condition (2)(ii). For our purpose, let us define \( U_n = U_{\delta} = \{ (\mu, \omega^2) : K(\theta, \theta_0) < \delta \} \), where \( K(\theta, \theta_0) \), defined as in (3.2), is the proper Kullback-Leibler divergence. Thus, \( K(\theta, \theta_0) \to 0 \) if and only if \( \theta \neq \theta_0 \). Now,

\[
E_{\theta} (1 - \Phi_n)
= P_\theta \left( \beta_n > \sqrt{e^{-n\delta}} \right) = P_\theta \left( -2 \log \beta_n < n\delta \right)
= P_\theta \left( -n \left( \hat{\theta}_n - \theta \right)^T \frac{\ell''(\theta_n)}{n} \left( \hat{\theta}_n - \theta \right) + 2 \ell_n(\theta_0) - 2 \ell_n(\theta) - 2 \left( \hat{\theta}_n - \theta \right)^T \ell'_n(\theta) > -n\delta \right)
\]

(here \( \theta_n^* \) lies between \( \theta \) and \( \hat{\theta}_n \))

\[
< e^{\alpha n \delta} E_{\theta} \left( \exp \left\{ -n \left( \hat{\theta}_n - \theta \right)^T \frac{\ell''(\theta_n)}{n} \left( \hat{\theta}_n - \theta \right) + 2 \ell_n(\theta_0) - 2 \ell_n(\theta) - 2 \left( \hat{\theta}_n - \theta \right)^T \ell'_n(\theta) \right\} \right),
\]

\[
\leq e^{\alpha n \delta} E_{\theta} \left( \exp \left\{ -n \left( \hat{\theta}_n - \theta \right)^T \frac{\ell''(\theta_n)}{n} \left( \hat{\theta}_n - \theta \right) + 2 \ell_n(\theta_0) - 2 \ell_n(\theta) + 2 \left( \hat{\theta}_n - \theta \right)^T \ell'_n(\theta) \right\} \right)
\]

\[
\leq e^{\alpha n \delta} \sqrt{E_{\theta} \left( \exp \left\{ -2n \left( \hat{\theta}_n - \theta \right)^T \frac{\ell''(\theta_n)}{n} \left( \hat{\theta}_n - \theta \right) + 4 \ell_n(\theta_0) - 4 \ell_n(\theta) \right\} \right)} \times \sqrt{E_{\theta} \left( \exp \left\{ 4n \left( \hat{\theta}_n - \theta \right)^T \ell'_n(\theta) \right\} \right)} \tag{3.27}
\]

(3.28)

using Cauchy-Schwartz inequality).

where \( \alpha \) is given by (3.6). Now observe that

\[
- n \left( \hat{\theta}_n - \theta \right)^T \frac{\ell''(\theta_n)}{n} \left( \hat{\theta}_n - \theta \right) \xrightarrow{\mathcal{L}} \chi_1^2, \tag{3.29}
\]

8
The Cauchy-Schwartz inequality entails
\[
\frac{\ell_n(\theta_0) - \ell_n(\theta)}{n} \overset{a.s.}{\rightarrow} -K(\theta, \theta_0),
\]
the latter convergence (3.30) being possible due to Kolmogorov’s strong law of large numbers in the non-
\textit{iid} case, using finiteness of our moments for all \(x, T\) belonging to compact spaces \(\mathcal{X}\) and \(\mathcal{\bar{X}}\), respectively. The Cauchy-Schwartz inequality entails
\[
\left| \left( \hat{\theta}_n - \theta \right)^T \ell'_n(\theta) \right| = \left| \left( \hat{\theta}_n - \theta \right)^T \mathcal{I}(\theta)^{1/2} \mathcal{I}(\theta)^{-1/2} \ell'_n(\theta) \right| \\
\leq \sqrt{n \left( \hat{\theta}_n - \theta \right)^T \mathcal{I}(\theta) \left( \hat{\theta}_n - \theta \right)} \times \sqrt{n^{-1} \{ \ell'_n(\theta) \}^T \mathcal{I}^{-1}(\theta) \ell'_n(\theta)},
\]
where
\[
n \left( \hat{\theta}_n - \theta \right)^T \mathcal{I}(\theta) \left( \hat{\theta}_n - \theta \right) \overset{\mathcal{L}}{\rightarrow} \chi^2_1
\]
and
\[
n^{-1} \{ \ell'_n(\theta) \}^T \mathcal{I}^{-1}(\theta) \ell'_n(\theta) = n^{-1} \text{tr} \mathcal{I}^{-1}(\theta) \ell'_n(\theta) \{ \ell'_n(\theta) \}^T \overset{a.s.}{\rightarrow} \text{tr} (\mathcal{I}^{-1}(\theta) \mathcal{I}(\theta)) = 2,
\]
where, for any matrix \(A\), \(\text{tr} (A)\) denotes trace of the matrix \(A\).

Hence, combining the asymptotic inequalities we obtain that for \(\theta \in U_n \cap \Omega_n\), where \(n\) is sufficiently large,
\[
E_\theta (1 - \Phi_n) < e^{\alpha n \delta} \times e^{-2\alpha n K(\theta, \theta_0)} \times \sqrt{E_\theta \left( e^{2\alpha \chi^2_1} \right)} \times E_\theta \left( e^{4\alpha \sqrt{2}\chi^2_1} \right) \\
< e^{\alpha n \delta} \times e^{-2\alpha n \delta} \times \sqrt{E_\theta \left( e^{2\alpha \chi^2_1} \right)} \times E_\theta \left( e^{4\alpha \sqrt{2}\chi^2_1} \right) \\
= e^{-\alpha n \delta} \times \sqrt{E_\theta \left( e^{2\alpha \chi^2_1} \right)} \times E_\theta \left( e^{4\alpha \sqrt{2}\chi^2_1} \right).
\]
For our choice of \(\alpha\), the expectations in (3.34) are finite. Also since the right hand side of (3.34) does not depend upon \(\theta\), (2)(ii) is proved in our case. That is, finally, posterior consistency (3.8) holds in our non-\textit{iid} SDE set-up.

A clear advantage of this theorem is that compactness of the parameter space \(\Omega\) is not required, unlike in all the previous results.

### 3.2 Asymptotic normality of the posterior distribution in the non-\textit{iid} set-up

For asymptotic normality of the posterior in the \textit{iid} situation, four regularity conditions, stated in Section 2.2.1 were necessary. In the non-\textit{iid} framework, three more are necessary, in addition to the already presented four conditions. They are as follows (see Schervish (1995) for details).

#### 3.2.1 Extra regularity conditions in the non-\textit{iid} set-up

(5) The largest eigenvalue of \(\Sigma_n\) goes to zero in probability.

(6) For \(\delta > 0\), define \(\mathcal{N}_\delta(\delta)\) to be the open ball of radius \(\delta\) around \(\theta_0\). Let \(\rho_n\) be the smallest eigenvalue of \(\Sigma_n\). If \(\mathcal{N}_\delta(\delta) \subseteq \Omega\), there exists \(K(\delta) > 0\) such that
\[
\lim_{n \to \infty} P_{\theta_0} \left( \sup_{\theta \in \Omega \cap \mathcal{N}_\delta(\delta)} \rho_n [\ell_n(\theta) - \ell_n(\theta_0)] < -K(\delta) \right) = 1.
\]
(7) For each $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that

$$\lim_{n \to \infty} P_{\theta_0} \left( \sup_{\theta \in \mathcal{N}_0(\delta(\epsilon), \|\gamma\| = 1} \left| 1 + \gamma^T \Sigma_n^\frac{1}{2} \ell''_n(\theta) \Sigma_n^\frac{1}{2} \gamma \right| < \epsilon \right) = 1. \quad (3.36)$$

In the non-iid case, the four regularity conditions presented in Section 2.2.1 and additional three provided above, are sufficient to guarantee (2.8).

### 3.2.2 Verification of the regularity conditions

For $i = 1, 2$ and $j = 1, 2$, let the $(i, j)$-th element of $\ell''_n(\hat{\theta}_n)$ be denoted by $\ell''_{n,ij}(\hat{\theta}_n)$. Then $\ell''_{n,ij}(\hat{\theta}_n)/n$ admits the following Taylor’s series expansion around $\theta_0$:

$$\frac{\ell''_{n,ij}(\hat{\theta}_n)}{n} = \frac{\ell''_{n,ij}(\theta_0)}{n} + \frac{(\hat{\theta}_n - \theta_0)^T \ell''_{n,ij}(\theta^*_n)}{n}, \quad (3.37)$$

where $\theta^*_n$ lies between $\theta_0$ and $\hat{\theta}_n$. Now let us consider

$$\left| -\frac{\ell''_{n,ij}(\hat{\theta}_n)}{nI_{ij}(\theta_0)} - 1 \right| = \left| -\frac{\ell''_{n,ij}(\theta_0)}{nI_{ij}(\theta_0)} - \frac{(\hat{\theta}_n - \theta_0)^T \ell''_{n,ij}(\theta^*_n)}{nI_{ij}(\theta_0)} - 1 \right| \leq \left| -\frac{\ell''_{n,ij}(\theta_0)}{nI_{ij}(\theta_0)} - 1 \right| + \left| \frac{(\hat{\theta}_n - \theta_0)^T \ell''_{n,ij}(\theta^*_n)}{nI_{ij}(\theta_0)} \right|. \quad (3.38)$$

Since (3.17) guarantees that $\ell''_{n,ij}(\theta_0)/n \xrightarrow{a.s.} -I_{ij}(\theta_0)$, it follows that $\left| -\frac{\ell''_{n,ij}(\theta_0)}{nI_{ij}(\theta_0)} - 1 \right| \xrightarrow{a.s.} 0$. That the second term is $o_P(1)$ follows in the same way as (3.19). That is, (3.38), and hence $\left| -\frac{\ell''_{n,ij}(\theta_0)}{nI_{ij}(\theta_0)} - 1 \right|$, is $o_P(1)$. In other words, $-\ell''_{n,ij}(\hat{\theta}_n)$ and $nI(\theta_0)$ are asymptotically equivalent (in probability). Since the maximum eigenvalue of $n^{-1}I^{-1}(\theta_0)$ goes to zero in probability as $n \to \infty$, so does the maximum eigenvalue of $\Sigma_n$. Hence, condition (5) holds.

To verify condition (6), note that again by Kolmogorov’s strong law of large numbers,

$$\frac{1}{n} (\ell_n(\theta) - \ell_n(\theta_0)) \xrightarrow{a.s.} -\mathcal{K}(\theta_0, \theta), \quad (3.39)$$

where, $\mathcal{K}(\theta_0, \theta)$ is given by (3.1). Now, writing $\rho_n [\ell_n(\theta) - \ell_n(\theta_0)]$ as $n \rho_n \left[ \ell_n(\theta) - \ell_n(\theta_0) \right]$ and noting that $\Sigma_n = O_P(n^{-1})$ implies $n \rho_n \xrightarrow{P} c$, where $c > 0$, it follows from (3.39) that $\rho_n [\ell_n(\theta) - \ell_n(\theta_0)] \xrightarrow{P} -c\mathcal{K}(\theta_0, \theta) < 0$. Hence, condition (6) holds.

For condition (7) note that for $\theta \in \mathcal{N}_0(\delta(\epsilon))$, $\theta = \theta_0 + \delta_2 \frac{\theta}{\|\theta\|}$, where $0 < \delta_2 \leq \delta(\epsilon)$. So, using Taylor’s series expansion around $\theta_0$, the $(i, j)$-th element of $\ell''_n(\theta)/n$ can be written as

$$\frac{\ell''_{n,ij}(\theta)}{n} = \frac{\ell''_{n,ij}(\theta_0)}{n} + \frac{\theta^T \ell''_{n,ij}(\theta^*)}{n\|\theta_0\|}, \quad (3.40)$$

where $\theta^*$ lies between $\theta_0$ and $\theta$. As $n \to \infty$, $\frac{\ell''_{n,ij}(\theta_0)}{n}$ tends, in probability, to the $(i, j)$-th element of $-I(\theta_0)$. Now notice that

$$\left| \frac{\theta^T \ell''_{n,ij}(\theta^*)}{n\|\theta_0\|} \right| \leq \frac{\|\theta^T \ell''_{n,ij}(\theta^*)\|}{n},$$

so that $\frac{|\theta^T \ell''_{n,ij}(\theta^*)|}{n\|\theta_0\|} = O_P(1)$ since $\frac{\ell''_{n,ij}(\theta_0)}{n\|\theta_0\|} = O_P(1)$ as before. Considering the inequality analogous
to (3.38) it follows that $\ell''_n(\theta)$ is asymptotically equivalent (in probability) to $-nI(\theta_0) + O_P(\delta_2)$. Since $\Sigma_1$ is asymptotically equivalent (in probability) to $n^{-\frac{1}{2}}I^{-\frac{1}{2}}(\theta_0)$, condition (7) holds.

References

Choi, T. and Schervish, M. J. (2007). On Posterior Consistency in Nonparametric Regression Problems. *Journal of Multivariate Analysis*, 98, 1969–1987.

Delattre, M., Genon-Catalot, V., and Samson, A. (2012). Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects. *Scandinavian Journal of Statistics*, 40, 322–343.

Maitra, T. and Bhattacharya, S. (2014). On Asymptotics Related to Classical Inference in Stochastic Differential Equations with Random Effects. Submitted.

Schervish, M. J. (1995). *Theory of Statistics*. Springer-Verlag, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York.