On Asymptotics Related to Classical Inference in Stochastic Differential Equations with Random Effects

Trisha Maitra and Sourabh Bhattacharya

Abstract

Delattre et al. (2012) considered \( n \) independent stochastic differential equations (SDE's), where in each case the drift term is associated with a random effect, the distribution of which depends upon unknown parameters. Assuming the independent and identical (iid) situation the authors provide independent proofs of consistency and asymptotic normality of the maximum likelihood estimators (MLE's) of the hyper-parameters of their random effects parameters.

In this article, as an alternative route to proving consistency and asymptotic normality in the SDE set-up involving random effects, we verify the regularity conditions required by existing relevant theorems. But much more importantly, we further consider the independent, but non-identical set-up associated with the random effects based SDE framework, and prove asymptotic results associated with the MLE's.

Keywords: Asymptotic normality; Burkholder-Davis-Gundy inequality; Itô isometry; Maximum likelihood estimator; Random effects; Stochastic differential equations.

1 Introduction

Delattre et al. (2012) study mixed-effects stochastic differential equations (SDE's) of the following form:

\[ dX_i(t) = b(X_i(t), \phi_i) dt + \sigma(X_i(t)) dW_i(t), \quad \text{with} \quad X_i(0) = x^i, \quad i = 1, \ldots, n. \] (1.1)

Here, for \( i = 1, \ldots, n \), the stochastic process \( X_i(t) \) is assumed to be continuously observed on the time interval \([0, T_i]\) with \( T_i > 0 \) known, and \( \{x^i; \ i = 1, \ldots, n\} \) are the known initial values of the \( i \)-th process. The processes \( \{W_i(\cdot); \ i = 1, \ldots, n\} \) are independent standard Brownian motions, and \( \{\phi_i; \ i = 1, \ldots, n\} \) are independently and identically distributed (iid) random variables with common distribution \( \nu(\cdot) \) on \( \mathbb{R}^d \), which are independent of the Brownian motions. Here \( \theta \in \Omega \subset \mathbb{R}^d \) is an unknown parameter to be estimated. The functions \( b: \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R} \) and \( \sigma: \mathbb{R} \rightarrow \mathbb{R}^d \) are the drift function and the diffusion coefficient, respectively, both assumed to be known.

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).

Statistically, the \( i \)-th process \( X_i(\cdot) \) can be thought of as modelling the \( i \)-th individual and the corresponding random variable \( \phi_i \) denotes the random effect of individual \( i \). For statistical inference, we follow Delattre et al. (2012) who consider the special case where \( b(x, \phi) = \phi_i b(x) \) such that \( b(\cdot) \) and \( \sigma(\cdot) \) are real, continuous functions having linear growth. Delattre et al. (2012) show that the likelihood, depending upon \( \theta \), admits a relatively simple form composed of the following sufficient statistics:

\[ U_i = \int_0^{T_i} \frac{b(X_i(s))}{\sigma^2(X_i(s))} dX_i(s), \quad V_i = \int_0^{T_i} \frac{b^2(X_i(s))}{\sigma^2(X_i(s))} ds, \quad i = 1, \ldots, n. \] (1.2)

Delattre et al. (2012) assume that

\[ V_i < \infty \] (1.3)

almost surely for every \( i \geq 1 \). Adopting assumption (H4) of Delattre et al. (2012), we also assume that the function \( b(\cdot)/\sigma(\cdot) \) is not constant, and that, for \( i \geq 1 \), \( (U_i, V_i) \) has a density \( g_i(u, v) \) with respect to the Lebesgue measure on \( \mathbb{R} \times \mathbb{R}^+ \), which is jointly continuous and positive on an open ball of \( \mathbb{R} \times \mathbb{R}^+ \), where \( \mathbb{R}^+ = (0, \infty) \). In the iid set-up, \( g_1 = g_i \) 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.
The exact likelihood is given by
\[ L(\theta) = \prod_{i=1}^{n} \lambda_i(X_i, \theta), \tag{1.4} \]
where
\[ \lambda_i(X_i, \theta) = \int_{\mathbb{R}} g(\varphi, \theta) \exp \left( \varphi U_i - \frac{\varphi^2}{2} V_i \right) d\nu(\varphi). \tag{1.5} \]

Assuming that \( g(\varphi, \theta)d\nu(\varphi) \equiv N(\mu, \omega^2), \) Delattre et al. (2012) obtain the following form of \( \lambda_i(X_i, \theta): \)
\[ \lambda_i(X_i, \theta) = \frac{1}{(1 + \omega^2 V_i)^{1/2}} \exp \left[ -\frac{V_i}{2(1 + \omega^2 V_i)} \left( \mu - \frac{U_i}{V_i} \right)^2 \right] \exp \left( \frac{U_i^2}{2V_i} \right), \tag{1.6} \]
where \( \theta = (\mu, \omega^2) \in \mathbb{R} \times \mathbb{R}^+. \)

Delattre et al. (2012) consider \( x^i = x \) and \( T_i = T \) for \( i = 1, \ldots, n, \) so that the set-up boils down to the iid situation, and investigate asymptotic properties of the MLE of \( \theta, \) providing proofs of consistency and asymptotic normality independently, without invoking the general results already existing in the literature. In this article, as an alternative, we prove asymptotic properties of the MLE in this SDE set-up by verifying the regularity conditions of relevant theorems already existing in the literature.

But far more importantly, we consider the independent but non-identical case (we refer to the latter as non-\textit{iid}), and prove consistency and asymptotic normality of the MLE in this set-up. Identifiability of the likelihood in the \textit{iid} set-up follows from the proof of Proposition 7 of Delattre et al. (2012); this continues to hold in the non-\textit{iid} set-up as well. In what follows, in Section 2 we investigate asymptotic properties of MLE in the \textit{iid} context. In Section 3 we investigate classical asymptotics in the non-\textit{iid} set-up.

Notationally, "\( \xrightarrow{a.s.} \)", "\( \xrightarrow{P} \)" and "\( \xrightarrow{d} \)" denote convergence “almost surely”, “in probability” and “in distribution”, respectively.

2 Consistency and asymptotic normality of MLE in the \textit{iid} set-up

2.1 Strong consistency of MLE

Consistency of the MLE under the \textit{iid} set-up can be verified by validating the regularity conditions of the following theorem (Theorems 7.49 and 7.54 of Schervish (1995)); for our purpose we present the version for compact \( \Omega. \)

\textbf{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 \((\mathcal{X}, \mathcal{B})\). Fix \( \theta_0 \in \Omega, \) and define, for each \( M \subseteq \Omega \) and \( x \in \mathcal{X}, \)
\[ Z(M, x) = \inf_{\psi \in M} \log \frac{f_1(x|\theta_0)}{f_1(x|\psi)}. \]

Assume that for each \( \theta \neq \theta_0, \) there is an open set \( N_\theta \) such that \( \theta \in N_\theta \) and that \( E_{\theta_0}Z(N_\theta, X_i) \geq -\infty. \) Also assume that \( f_1(x|\cdot) \) is continuous at \( \theta \) for every \( \theta, \) a.s. \( [P_{\theta_0}]. \) Then, if \( \hat{\theta}_n \) is the MLE of \( \theta \) corresponding to \( n \) observations, it holds that \( \lim_{n \to \infty} \hat{\theta}_n = \theta_0, \) a.s. \( [P_{\theta_0}]. \)

2.1.1 Verification of strong consistency of MLE in our SDE set-up

Following Delattre et al. (2012) we assume that the parameter space \( \Omega \) is compact. To verify the conditions of Theorem 1 in our case, we note that for any \( x, f_1(x|\theta) = \lambda_1(x, \theta) = \lambda(x, \theta) \) given by (1.6).
which is clearly continuous in \( \theta \). Also, it follows from the proof of Proposition 7 of \cite{Delattre et al. 2012} that for every \( \theta \neq \theta_0 \),

\[
\log \frac{f_1(x|\theta_0)}{f_1(x|\theta)} = \frac{1}{2} \log \left( \frac{1 + \omega^2 V_1}{1 + \omega_0^2 V_1} \right) + \frac{1}{2} \frac{(\omega_0^2 - \omega^2)U_1^2}{(1 + \omega^2 V_1)(1 + \omega_0^2 V_1)} + \frac{\mu^2 V_1}{2(1 + \omega^2 V_1)} - \frac{\mu U_1}{1 + \omega^2 V_1} - \frac{\mu_0 U_1}{1 + \omega_0^2 V_1}
\]

\[
\geq - \frac{1}{2} \left\{ \log \left( \frac{1 + \omega^2}{\omega_0^2} \right) + \frac{|\omega_0^2 - \omega^2|}{\omega^2} \right\} - \frac{1}{2} \frac{|\omega_0^2 - \omega^2|}{\omega^2} \left( \frac{U_1}{1 + \omega_0^2 V_1} \right)^2 \left( 1 + \frac{\omega^2}{\omega_0^2} \right)
\]

\[
- |\mu| \left| \frac{U_1}{1 + \omega_0^2 V_1} \right| \left( \frac{1 + \omega_0^2 - \omega^2}{\omega^2} \right) - \left| \frac{\mu^2 V_1}{2(1 + \omega_0^2 V_1)} - \frac{\mu_0 U_1}{1 + \omega_0^2 V_1} \right|. \tag{2.1}
\]

Taking \( N_0 = (\omega, \omega_0^2) \times (\omega, \omega_2^2) \), and noting that \( E_{\theta_0} \left( \frac{U_1}{1 + \omega_0^2 V_1} \right)^2 \), \( E_{\theta_0} \left| \frac{U_1}{1 + \omega_0^2 V_1} \right| \) and \( E_{\theta_0} \left( \frac{\mu^2 V_1}{2(1 + \omega_0^2 V_1)} \right) \) are finite due to Lemma 1 of \cite{Delattre et al. 2012}, it follows that \( E_{\theta_0} Z(\theta, X_1) > -\infty \). Hence, \( \theta_n \xrightarrow{a.s.} \theta_0 \) \( [P_{\theta_0}] \).

### 2.2 Asymptotic normality of MLE

To verify asymptotic normality of MLE we invoke the following theorem provided in \cite{Schervish 1995} (Theorem 7.63):

**Theorem 2 \cite{Schervish 1995}** Let \( \Omega \) be a subset of \( \mathbb{R}^d \), and let \( \{X_n\}_{n=1}^\infty \) be conditionally iid given \( \theta \) each with density \( f_1(\cdot|\theta) \). Let \( \hat{\theta}_n \) be an MLE. Assume that \( \hat{\theta}_n \xrightarrow{p} \theta \) under \( P_\theta \) for all \( \theta \). Assume that \( f_1(x|\theta) \) has continuous second partial derivatives with respect to \( \theta \) and that differentiation can be passed under the integral sign. Assume that there exists \( H_r(x, \theta) \) such that, for each \( \theta_0 \in \text{int}(\Omega) \) and each \( k, j \),

\[
\sup_{||\theta - \theta_0|| \leq r} \left| \frac{\partial^2}{\partial \theta_k \partial \theta_j} \log f_{X_1|\theta}(x|\theta_0) - \frac{\partial^2}{\partial \theta_k \partial \theta_j} \log f_{X_1|\theta}(x|\theta) \right| \leq H_r(x, \theta_0), \tag{2.2}
\]

with

\[
\lim_{r \to 0} E_{\theta_0} H_r (X, \theta_0) = 0. \tag{2.3}
\]

Assume that the Fisher information matrix \( I(\theta) \) is finite and non-singular. Then, under \( P_{\theta_0} \),

\[
\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{d} N \left( 0, I^{-1}(\theta_0) \right). \tag{2.4}
\]

#### 2.2.1 Verification of the above regularity conditions for asymptotic normality in our SDE set-up

In Section 2.1 we proved almost sure consistency of the MLE \( \hat{\theta}_n \) in the SDE set-up. Hence, \( \hat{\theta}_n \xrightarrow{p} \theta \) under \( P_\theta \) for all \( \theta \). In the proof of Proposition 5, \cite{Delattre et al. 2012} show that differentiation can be passed under the integral sign. Letting \( \gamma_1(\theta) = \frac{U_1}{1 + \omega_0^2 V_1} \) and \( I_1 = \frac{V_1}{1 + \omega_0^2 V_1} \), note that (see the proof of Proposition 6 of \cite{Delattre et al. 2012})

\[
\frac{\partial^2}{\partial \mu^2} \log f_1(x|\theta) = -I_1(\omega^2), \quad \frac{\partial^2}{\partial \mu \partial \omega^2} \log f_1(x|\theta) = -\gamma_1(\theta)I_1(\omega^2); \tag{2.5}
\]

\[
\frac{\partial^2}{\partial \omega^2 \partial \omega_0^2} \log f_1(x|\theta) = -\frac{1}{2} \left( 2\gamma_1(\theta)I_1(\omega^2) - I_1^2(\omega^2) \right). \tag{2.6}
\]

It follows from (2.5) and (2.6) that in our case \( \frac{\partial^2}{\partial \theta_k \partial \theta_j} \log f_1(x|\theta) \) is differentiable in \( \theta = (\mu, \omega^2) \), and the derivative has finite expectation; see the proof of Proposition 8 of \cite{Delattre et al. 2012}. Hence,
\( (2.2) \) and \( (2.3) \) clearly hold. That the information matrix \( \mathcal{I}(\theta) \) is finite and positive-definite, are shown in [Delattre et al. (2012)]. Hence, asymptotic normality of the MLE, of the form \( (2.4) \), holds in our case.

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

We now consider the case where the processes \( X_i(\cdot); i = 1, \ldots, n \), are independently, but not identically distributed. This happens when we no longer enforce the restrictions \( T_i = T \) and \( x^i = x \) for \( i = 1, \ldots, n \). However, we do assume that the sequences \( \{T_1, T_2, \ldots\} \) and \( \{x^1, x^2, \ldots, \} \) are sequences entirely contained in compact sets \( \mathfrak{T} \) and \( \mathfrak{X} \), respectively. Due to compactness, there exist convergent subsequences with limits in \( \mathfrak{T} \) and \( \mathfrak{X} \). Abusing notation, we continue to denote the convergent subsequences as \( \{T_1, T_2, \ldots\} \) and \( \{x^1, x^2, \ldots, \} \). Let the limits be \( T^\infty \in \mathfrak{T} \) and \( x^\infty \in \mathfrak{X} \).

Now, since the distributions of the processes \( X_i(t) \) are uniquely defined on the space of real, continuous functions \( \mathcal{C}([0, T] \mapsto \mathbb{R}) = \{f : [0, T] \mapsto \mathbb{R} \text{ such that } f \text{ is continuous}\} \), given any \( t \in [0, T_i] \), \( f(t) \) is clearly a continuous function of the initial value \( f(0) = x \). To emphasize dependence on \( x \), we denote the function as \( f(t, x) \). In fact, for any \( \epsilon > 0 \), there exists \( \delta_\epsilon > 0 \) such that whenever \( |x_1 - x_2| < \delta_\epsilon \), \( |f(t, x_1) - f(t, x_2)| < \epsilon \) for all \( t \in [0, T_i] \).

Henceforth, we denote the process associated with the initial value \( x \) and time point \( t \) as \( X(t, x) \), and by \( \phi(x) \) the random effect parameter associated with the initial value \( x \) such that \( \phi(x^1) = \phi_i \). We assume that \( \phi(x) \) is a real-valued, continuous function of \( x \), and that for \( k \geq 1 \),

\[
\sup_{x \in \mathfrak{X}} E |\phi(x)|^{2k} < \infty. \tag{3.1}
\]

For \( x \in \mathfrak{X} \) and \( T \in \mathfrak{T} \), let

\[
U(x, T) = \int_0^T \frac{b(X(s, x))}{\sigma^2(X(s, x))} dX(s, x); \tag{3.2}
\]
\[
V(x, T) = \int_0^T \frac{b^2(X(s, x))}{\sigma^2(X(s, x))} ds. \tag{3.3}
\]

Clearly, \( U(x^i, T_i) = U_i \) and \( V(x^i, T_i) = V_i \), where \( U_i \) and \( V_i \) are given by \( (1.2) \). Analogous to assumption \( (1.3) \), here we assume that

\[
V(x, T) < \infty \tag{3.4}
\]

almost surely for every \( x \in \mathfrak{X} \) and \( T \in \mathfrak{T} \). Then the moments of uniformly integrable continuous functions of \( U(x, T) \), \( V(x, T) \) and \( \theta \) are continuous in \( x \), \( T \) and \( \theta \). The result is formalized as Theorem 3 the proof of which is presented in the Appendix.

**Theorem 3** Let \( h(u, v, \theta) \) be any continuous function of \( u, v \) and \( \theta \), such that for any sequences \( \{x_m\}_{m=1}^\infty \) and \( \{T_m\}_{m=1}^\infty \), converging to \( \tilde{x} \), \( \tilde{T} \) and \( \tilde{\theta} \), respectively, for any \( \tilde{x} \in \mathfrak{X} \), \( \tilde{T} \in \mathfrak{T} \) and \( \tilde{\theta} \in \Omega \), the sequence \( \{h(U(x_m, T_m), V(x_m, T_m), \theta_m)\}_{m=1}^\infty \) is uniformly integrable. Then, as \( m \to \infty \),

\[
E[h(U(x_m, T_m), V(x_m, T_m), \theta_m)] \to E \left[ h \left( U(\tilde{x}, \tilde{T}), V(\tilde{x}, \tilde{T}), \tilde{\theta} \right) \right]. \tag{3.5}
\]

**Corollary 4** As in [Delattre et al. (2012)], consider the function

\[
h(u, v) = \exp \left( \psi \frac{u}{1 + \xi v} \right), \tag{3.6}
\]

where \( \psi \in \mathbb{R} \) and \( \xi \in \mathbb{R}^+ \). Then, for any sequences \( \{x_m\}_{m=1}^\infty \) and \( \{T_m\}_{m=1}^\infty \) converging to \( \tilde{x} \) and \( \tilde{T} \), for any \( \tilde{x} \in \mathfrak{X} \) and \( T \in \mathfrak{T} \), and for \( k \geq 1 \),

\[
E[h(U(x_m, T_m), V(x_m, T_m))]^k \to E \left[ h \left( U(\tilde{x}, \tilde{T}), V(\tilde{x}, \tilde{T}) \right) \right]^k, \tag{3.7}
\]
as $m \to \infty$.

The proof of the above corollary only entails proving uniform integrability of \{h(U(x_m, T_m), V(x_m, T_m))\}_{m=1}^{\infty}, which simply follows from the proof of Lemma 1 of Delattre et al. (2012).

Note that in our case, the Kullback-Leibler distance and Fisher’s information are expectations of functions of the form $h(u, v, \theta)$, continuous in $u$, $v$ and $\theta$. The upper bounds provided in Delattre et al. (2012). Corollary 4 and compactness of $\Omega$, can be used to easily verify uniform integrability of the relevant sequences. It follows that in our situation the Kullback-Leibler distance, which we now denote by $K_{x,T}(\theta_0, \theta) \ (\text{or } K_{x,T}(\theta, \theta_0))$ to emphasize dependence on $x$ and $T$, apart from $\theta$, are continuous in $\theta$, $x$ and $T$. Similarly, the elements of the Fisher’s information matrix $\mathcal{I}_{x,T}(\theta)$ are continuous in $\theta$, $x$ and $T$.

For $x = x^k$ and $T = T_k$, we denote the Kullback-Leibler distance and the Fisher’s information as $K_k(\theta_0, \theta)$ ($K_k(\theta, \theta_0)$) and $\mathcal{I}_k(\theta)$, respectively.

Continuity of $K_{x,T}(\theta_0, \theta)$ (or $K_{x,T}(\theta, \theta_0)$) and $\mathcal{I}_{x,T}(\theta_0)$ with respect to $x$ and $T$ ensures that as $x^k \to x^\infty$ and $T_k \to T^\infty$, $K_{x^k,T^k}(\theta, \theta_0) \to K_{x^\infty,T^\infty}(\theta, \theta_0) = K(\theta, \theta_0)$, say. Similarly, $K_{x^k,T^k}(\theta, \theta_0) \to K(\theta, \theta_0)$ and $\mathcal{I}_{x^k,T^k}(\theta) \to \mathcal{I}_{x^\infty,T^\infty}(\theta) = \mathcal{I}(\theta)$, say. Since $X^\infty$ and $T^\infty$ are contained in the respective compact sets, the limits $K(\theta, \theta), K(\theta, \theta_0)$ and $\mathcal{I}(\theta)$ are well-defined Kullback-Leibler divergences and Fisher’s information, respectively.

From the above limits, it follows that for any $\theta \in \Omega$,

\[
\lim_{n \to \infty} \frac{\sum_{k=1}^{n} K_k(\theta_0, \theta)}{n} = K(\theta_0, \theta); \quad (3.8)
\]

\[
\lim_{n \to \infty} \frac{\sum_{k=1}^{n} K_k(\theta, \theta_0)}{n} = K(\theta_0, \theta); \quad (3.9)
\]

\[
\lim_{n \to \infty} \frac{\sum_{k=1}^{n} \mathcal{I}_k(\theta)}{n} = \mathcal{I}(\theta). \quad (3.10)
\]

We investigate consistency and asymptotic normality of MLE in our case using the results of Hoadley (1971). The limit results (3.8), (3.9) and (3.10) will play important roles in our proceedings.

### 3.1 Consistency and asymptotic normality of MLE in the non-iid set-up

Following Hoadley (1971) we define the following:

\[
R_i(\theta) = \log \frac{f_i(X_i|\theta)}{f_i(X_i|\theta_0)} \quad \text{if } f_i(X_i|\theta_0) > 0
\]

\[
= 0 \quad \text{otherwise.} \quad (3.11)
\]

\[
R_i(\theta, \rho) = \sup \{ R_i(\xi) : \| \xi - \theta \| \leq \rho \}
\]

\[
V_i(r) = \sup \{ R_i(\theta) : \| \theta \| > r \}. \quad (3.12)
\]

Following Hoadley (1971) we denote by $r_i(\theta), r_i(\theta, \rho)$ and $v_i(r)$ to be expectations of $R_i(\theta), R_i(\theta, \rho)$ and $V_i(r)$ under $\theta_0$; for any sequence \{a_i; i = 1, 2, \ldots \} we denote $\sum_{i=1}^{n} a_i/n$ by \(\bar{a}_n\).

Hoadley (1971) proved that if the following regularity conditions are satisfied, then the MLE $\hat{\theta}_n \overset{P}{\to} \theta_0$:

1. $\Omega$ is a closed subset of $\mathbb{R}^d$.
2. $f_i(X_i|\theta)$ is an upper semicontinuous function of $\theta$, uniformly in $i$, a.s. $[P_{\theta_0}]$.
3. There exist $\rho^* = \rho^*(\theta) > 0$, $r > 0$ and $0 < K^* < \infty$ for which
   (i) $E_{\theta_0}[R_i(\theta, \rho)]^2 \leq K^*, \quad 0 \leq \rho \leq \rho^*$;
(ii) $E_{\theta_0} \left[ V_1(r) \right]^2 \leq K^*.$

(4) (i) $\lim_{n \to \infty} \bar{r}_n(\theta) < 0, \quad \theta \neq \theta_0;$

(ii) $\lim_{n \to \infty} \bar{v}_n(r) < 0.$

(5) $R_i(\theta, \rho)$ and $V_i(r)$ are measurable functions of $X_i.$

Actually, conditions (3) and (4) can be weakened but these are more easily applicable (see [Hoadley 1971]) for details.

### 3.1.1 Verification of the regularity conditions

Since $\Omega$ is compact in our case, the first regularity condition clearly holds.

For the second regularity condition, note that given $X_i$, $f_i(X_i|\theta)$ is continuous, in fact, uniformly continuous in $\theta$ in our case, since $\Omega$ is compact. Hence, for any given $\epsilon > 0$, there exists $\delta_i(\epsilon) > 0$, independent of $\theta$, such that $||\theta_1 - \theta_2|| < \delta_i(\epsilon)$ implies $|f(X_i|\theta_1) - f(X_i|\theta_2)| < \epsilon$. Now consider a strictly positive function $\delta_{x,T}(\epsilon)$, continuous in $x \in \mathcal{X}$ and $T \in \mathcal{T}$, such that $\delta_{x,T}(\epsilon) = \delta_i(\epsilon)$. Let $\delta(\epsilon) = \inf_{x \in \mathcal{X}, T \in \mathcal{T}} \delta_{x,T}(\epsilon)$. Since $\mathcal{X}$ and $\mathcal{T}$ are compact, it follows that $\delta(\epsilon) > 0$. Now it holds that $||\theta_1 - \theta_2|| < \delta(\epsilon)$ implies $|f(X_i|\theta_1) - f(X_i|\theta_2)| < \epsilon$, for all $i$. Hence, the second regularity condition is satisfied.

Let us now focus attention on condition (3)(i). It follows from (2.1) that

$$R_i(\theta) \leq \frac{1}{2} \left\{ \log \left( 1 + \frac{\omega^2}{\omega_0^2} \right) + \left| \frac{\omega^2 - \omega_0^2}{\omega^2} \right| \right\} + \frac{1}{2} \left| \frac{\omega^2 - \omega_0^2}{\omega^2} \right| \left( 1 + \frac{\omega_0^2}{\omega^2} \right)^2 \left( 1 + \frac{\omega^2}{\omega_0^2} \right)$$

$$+ |\mu| \left| \frac{U_i}{1 + \omega_0^2 V_i} \right| \left( 1 + \frac{\omega_0^2}{\omega^2} \right)^2 \left( 1 + \frac{\omega_0^2}{\omega^2} \right) + \left( \frac{\mu_0 V_i}{2(1 + \omega_0^2 V_i)} \right).$$

(3.14)

Let us denote $\{ \xi : ||\xi - \theta|| \leq \rho \}$ by $S(\rho, \theta)$. Here $0 < \rho < \rho^*(\theta)$, and $\rho^*(\theta)$ is so small that $S(\rho, \theta) \subset \Omega$ for all $\rho \in (0, \rho^*(\theta))$. It then follows from (3.14) that

$$\sup_{\xi \in S(\rho, \theta)} R_i(\xi) \leq \sup_{(\mu, \omega^2) \in S(\rho, \theta)} \frac{1}{2} \left\{ \log \left( 1 + \frac{\omega^2}{\omega_0^2} \right) + \left| \frac{\omega^2 - \omega_0^2}{\omega^2} \right| \right\}$$

$$+ \left( \frac{U_i}{1 + \omega_0^2 V_i} \right)^2 \times \sup_{(\mu, \omega^2) \in S(\rho, \theta)} \left| \frac{\omega^2 - \omega_0^2}{\omega^2} \right| \left( 1 + \frac{\omega_0^2}{\omega^2} \right)$$

$$+ \left( \frac{U_i}{1 + \omega_0^2 V_i} \right) \times \left| \mu \right| \left( 1 + \frac{\omega_0^2}{\omega^2} \right)|$$

$$+ \left( \frac{\mu_0 V_i}{2(1 + \omega_0^2 V_i)} \right) + \left( \frac{\mu_0 U_i}{1 + \omega_0^2 V_i} \right).$$

(3.15)

The supremums in (3.15) are finite due to compactness of $S(\rho, \theta)$. Since under $P_{\theta_0}$, $U_i/(1 + \omega_0^2 V_i)$ admits moments of all orders and $0 < L_i(\omega_0^2) = \frac{V_i}{1 + \omega_0^2 V_i} < \frac{1}{\omega_0^2}$ (see [Delattre et al. 2012]), it follows from (3.15) that

$$E_{\theta_0} [R_i(\theta, \rho)]^2 \leq K_i(\theta),$$

(3.16)

where $K_i(\theta) = K(x^i, T_i, \theta)$, with $K(x, T, \theta)$ being a continuous function of $(x, T, \theta)$, continuity being a consequence of Theorem 3. Since because of compactness of $\mathcal{X}$, $\mathcal{T}$ and $\Omega$,

$$K_i(\theta) \leq \sup_{x \in \mathcal{X}, T \in \mathcal{T}, \theta \in \Omega} K(x, T, \theta) < \infty,$$

regularity condition (3)(i) follows.
To verify condition (3)(ii), first note that we can choose \( r > 0 \) such that \( \| \theta_0 \| < r \) and \( \{ \theta \in \Omega : \| \theta \| > r \} \neq \emptyset \). It then follows that \( \sup_{\theta \in \Omega : \| \theta \| > r} R_i(\theta) \leq \sup_{\theta \in \Omega} R_i(\theta) \) for every \( i \geq 1 \). The right hand side is bounded by the same expression as the right hand side of \((3.15)\), with only \( S(\theta, \rho) \) replaced with \( \Omega \). The rest of the verification follows in the same way as verification of (3)(i).

To verify condition (4)(i) note that by \((3.8)\)

\[
\lim_{n \to \infty} \bar{r}_n = - \lim_{n \to \infty} \frac{\sum_{i=1}^{n} K_i(\theta_0, \theta)}{n} = -K(\theta_0, \theta) < 0 \quad \text{for} \quad \theta \neq \theta_0. \tag{3.17}
\]

In other words, (4)(i) is satisfied.

To verify (4)(ii) we first show that \( \lim \bar{v}_n \) exists for suitably chosen \( r > 0 \). Then we prove that the limit is negative. To see that the limit exists, we first write \( R_{x,T}(\theta) = -K_{x,T}(\theta_0, \theta) \). Clearly, \( R_i(\theta) = R_{x_i,T_i}(\theta) \). Using the arguments provided in the course of verification of (3)(ii), and the moment existence result of [Delattre et al. 2012], yield

\[
\sup_{i \geq 1} E_{\theta_0} \left[ \sup_{\theta \in \Omega : \| \theta \| > r} R_i(\theta) \right]^2 \leq \sup_{i \geq 1} E_{\theta_0} \left[ \sup_{\theta \in \Omega} R_i(\theta) \right]^2 \leq \sup_{x \in \mathcal{X}, T \in \mathcal{T}} K_1(x, T), \tag{3.18}
\]

where \( K_1(x, T) \) is a continuous function of \( x \) and \( T \). That \( K_1(x, T) \) is continuous in \( x \) and \( T \) follows from Theorem [3] the required uniform integrability follows due to finiteness of the moments for every \( x \in \mathcal{X} \) and \( T \in \mathcal{T} \), and compactness of \( \mathcal{X} \) and \( \mathcal{T} \). Now, because of compactness of \( \mathcal{X} \) and \( \mathcal{T} \) it also follows that the right hand side of \((3.18)\) is finite, proving uniform integrability of \( \left\{ \sup_{\theta \in \Omega : \| \theta \| > r} R_i(\theta) \right\}_{i=1}^{\infty} \).

Hence, it follows from Theorem [3] that \( v_{x,T} = E_{\theta_0} \left[ \sup_{\theta \in \Omega : \| \theta \| > r} R_{x,T}(\theta) \right] \) is continuous in \( x \) and \( T \).

Since \( x^i \to x^\infty \) and \( T_i \to T^\infty \),

\[
\bar{v}_i = v_{x^i, T_i} \to \bar{v}_{x^\infty, T^\infty}.
\]

Since \( \bar{v}_{x,T} \) is well-defined for every \( x \in \mathcal{X}, T \in \mathcal{T} \), and since \( x^\infty \in \mathcal{X}, T^\infty \in \mathcal{T} \), \( \bar{v}_{x^\infty, T^\infty} \) is also well-defined. It follows that

\[
\lim_{n \to \infty} \bar{v}_n = \bar{v}_{x^\infty, T^\infty}
\]

exists.

To show that the limit \( \lim_{n \to \infty} \bar{v}_n \) is negative, let us first re-write \( \mathcal{V}_i(r) \) as

\[
\mathcal{V}_i(r) = \inf_{\{ \theta \in \Omega : \| \theta \| > r \}} \left[ \log \frac{f_i(X_1|\theta)}{f_i(\theta)} \right] \leq \inf_{\{ \theta \in \Omega : \| \theta \| \geq r \}} \left[ \log \frac{f_i(X_1|\theta)}{f_i(\theta)} \right] = -\log \frac{f_i(X_1|\theta)}{f_i(X_1|\theta^*(X_1))}, \tag{3.19}
\]

for some \( \theta^*(X_1) \), depending upon \( X_1 \), contained in \( \Omega_r = \Omega \cap \{ \theta : \| \theta \| \geq r \} \). Recall that we chose \( r > 0 \) such that \( \| \theta_0 \| < r \) and \( \Omega \cap \{ \theta : \| \theta \| > r \} \neq \emptyset \), so that \( \theta^*_i(X_1) \neq \theta_0 \) as \( \| \theta^*_i(X_1) \| \geq r > \| \theta_0 \| \) for all \( X_1 \). It is important to observe that \( \theta^*_i(X_1) \) can not be a one-to-one function of \( X_1 \equiv (U_i, V_i) \). To see this, first observe that for any given constant \( c \), the equation \( f_i(X_1|\theta_0) - f_i(X_1|\theta) = c \), equivalently, the equation \( \log f_i(U_i, V_i|\theta_0) - \log f_i(U_i, V_i|\theta) = c \), admits infinite number of solutions in \( (U_i, V_i) \), for any given \( \theta = (\mu, \sigma^2) \). Hence, for \( \theta^*_i(X_i) = \varphi \) such that \( \inf_{\{ \theta : \| \theta \| \geq r \}} \left[ \log \frac{f_i(X_1|\theta)}{f_i(\theta)} \right] = \log \frac{f_i(X_1|\theta)}{f_i(X_1|\varphi)} = c \), there exist infinitely many values of \( (U_i, V_i) \) with the same infimum \( c \) for the same value \( \varphi \), thereby proving that \( \theta^*(X_i) \) is a many-to-one function of \( X_i \). A consequence of this is non-degeneracy of the conditional...
distribution of $X_i$, given $\theta^*(X_i)$, which ensures that $E_{X_i|\theta^*_i(X_i),\theta_0} \left[ \log \frac{f_i(X_i|\theta_0)}{f_i(X_i|\theta^*_i(X_i))} \right] = K_i(\theta_0, \theta^*_i(X_i))$ is well-defined and strictly positive, since $\theta^*_i(X_i) \neq \theta_0$.

Given the above arguments, now note that,

$$E_{\theta_0} \left[ \log \frac{f_i(X_i|\theta_0)}{f_i(X_i|\theta^*_i(X_i))} \right] = E_{\theta_0^*|\theta_0} E_{X_i|\theta_0^*|X_i} \left[ \log \frac{f_i(X_i|\theta_0)}{f_i(X_i|\theta^*_i(X_i))} \right]$$

$$= E_{\theta_0^*|\theta_0} \left[ K_i(\theta_0, \varphi_i) \right]$$

$$\geq E_{\theta_0^*|\theta_0} \left[ \inf_{\varphi_i \in \Omega_r} K_i(\theta_0, \varphi_i) \right]$$

$$= K_i(\theta_0, \varphi^*_i),$$

(3.20)

where $\varphi^*_i \in \Omega_r$ is where the infimum of $K_i(\theta_0, \varphi_i)$ is achieved. Since $\varphi^*_i$ is independent of $X_i$, the last step (3.20) follows. Hence,

$$E_{\theta_0} V_i(r) \leq -K_i(\theta_0, \varphi^*_i) \leq -\inf_{x \in \bar{X}, T \in \Omega, \theta \in \Omega_r} K_{x,T}(\theta, \varphi) = -K_{x^*,T^*}(\theta_0, \varphi^*),$$

(3.21)

for some $x^* \in \bar{X}$, $T^* \in \Omega$ and $\varphi^* \in \Omega_r$. Since $K_{x^*,T^*}(\theta_0, \varphi^*)$ is a well-defined Kullback-Leibler distance, it is strictly positive since $\varphi^* \neq \theta_0$. Hence, it follows from (3.21) and the fact that $K_{x^*,T^*}(\theta_0, \varphi^*) > 0$, that

$$\lim_{n \to \infty} \bar{v}_n = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} E_{\theta_0} V_i(r)}{n}$$

$$\leq -\lim_{n \to \infty} \frac{\sum_{i=1}^{n} K_i(\theta_0, \varphi^*_i)}{n}$$

$$\leq -K_{x^*,T^*}(\theta_0, \varphi^*)$$

$$< 0.$$

Thus, condition (4)(ii) holds.

Regularity condition (5) holds because for any $\theta \in \Omega$, $R_i(\theta)$ is an almost surely continuous function of $X_i$ rendering it measurable for all $\theta \in \Omega$, and due to the fact that suprema of measurable functions are measurable.

In other words, in the non-\textit{iid} set-up in the SDE framework it holds that $\hat{\theta}_n \overset{p}{\to} \theta_0$.

3.2 Asymptotic normality of \textit{MLE} in the non-\textit{iid} set-up

Let $\zeta_i(x, \theta) = \log f_i(x|\theta)$; also, let $\zeta'_i(x, \theta)$ be the $d \times 1$ vector with $j$-th component $\zeta'_{i,j}(x, \theta) = \frac{\partial}{\partial \theta_j} \zeta_i(x, \theta)$, and let $\zeta''_i(x, \theta)$ be the $d \times d$ matrix with $(j,k)$-th element $\zeta''_{i,j,k}(x, \theta) = \frac{\partial^2}{\partial \theta_j \partial \theta_k} \zeta_i(x, \theta)$.

For proving asymptotic normality in the non-\textit{iid} framework, \textit{Hoadley} (1971) assumed the following regularity conditions:

1. $\Omega$ is an open subset of $\mathbb{R}^d$.
2. $\hat{\theta}_n \overset{p}{\to} \theta_0$.
3. $\zeta'_i(X_i, \theta)$ and $\zeta''_i(X_i, \theta)$ exist a.s. $[P_{\theta_0}]$.
4. $\zeta''_i(X_i, \theta)$ is a continuous function of $\theta$, uniformly in $i$, a.s. $[P_{\theta_0}]$, and is a measurable function of $X_i$.
5. $E[\zeta'_i(X_i, \theta)] = 0$ for $i = 1, 2, \ldots$.
(6) $\mathcal{I}_x(\theta) = E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \ln g(X_1, \theta) \right] = -E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \ln g(X_1, \theta) \right]$, where for any vector $y$, $y^T$ denotes the transpose of $y$.

(7) $\tilde{I}_n(\theta) \to \tilde{I}(\theta)$ as $n \to \infty$ and $\tilde{I}(\theta)$ is positive definite.

(8) $E_{\theta_0} \left| \zeta_{i,j}(X_i, \theta_0) \right|^3 \leq K_2$, for some $0 < K_2 < \infty$.

(9) There exist $\epsilon > 0$ and random variables $B_{i,j,k}(X_i)$ such that

(i) $\sup \left\{ \left| \zeta_{i,j}(X_i, \xi) \right| : \left| \xi - \theta_0 \right| \leq \epsilon \right\} \leq B_{i,j,k}(X_i)$.

(ii) $E_{\theta_0} \left| B_{i,j,k}(X_i) \right|^{1+\delta} \leq K_2$, for some $\delta > 0$.

Condition (8) can be weakened but is relatively easy to handle. Under the above regularity conditions, Hoadley (1971) prove that

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{d} N \left( 0, \tilde{I}^{-1}(\theta_0) \right).$$

(3.22)

### 3.2.1 Validation of asymptotic normality of MLE in the non-iid SDE set-up

Note that condition (1) requires the parameter space $\Omega$ to be an open subset. However, the proof of asymptotic normality presented in Hoadley (1971) continues to hold for compact $\Omega$, since for any open cover of $\Omega$ we can extract a finite subcover, consisting of open sets.

Conditions (2), (3), (5), (6) are clearly valid in our case. Condition (4) can be verified in exactly the same way as condition (2) of Section 3.1 is verified; measurability of $\zeta''(X_i, \theta)$ follows due its continuity with respect to $X_i$. Condition (7) simply follows from (3.10) and condition (8) holds due to finiteness of the moments for every $x \in \mathcal{X}$, $T \in \mathcal{T}$, and compactness of $\mathcal{X}$ and $\mathcal{T}$.

For conditions (9)(i) and (9)(ii) note that $\zeta''_{i,j}(X_i, \theta)$ for $j, k = 1, 2$ are given by

$$\frac{\partial^2}{\partial \omega^2} \ln f(X_i|\theta) = -I_i(\omega^2),$$

$$\frac{\partial^2}{\partial \omega^2} \ln f(X_i|\theta) = -\gamma_i(\theta)I_i(\omega^2),$$

and

$$\frac{\partial^2}{\partial \omega^2} \ln f(X_i|\theta) = -\frac{1}{2} \left( 2\gamma_i^2(\theta)I_i^2(\omega^2) - I_i^2(\omega^2) \right).$$

Also since Delattre et al. (2012) establish

$$\sup_{\theta \in \Omega} \left| \gamma_i(\theta) \right| \leq \left| \frac{U_i}{1 + \omega_0^2 V_i} \right| \left( 2 + \frac{\omega_0^2}{\omega^2} \right) + \frac{\bar{\mu}}{\omega^2},$$

(3.23)

it follows from (4.23), the fact that $0 < I_i(\omega^2) < 1/\omega^2$, finiteness of moments of all orders for every $x \in \mathcal{X}$, $T \in \mathcal{T}$, and compactness of $\mathcal{X}$ and $\mathcal{T}$, that conditions (9)(i) and (9)(ii) hold.

In other words, in our non-iid SDE case we have asymptotic normality of the form (3.22).

### Appendix

**Proof of Theorem 3.** We can decompose (3.2) as

$$U(x, T) = \int_0^T \frac{b(X(s, x))}{\sigma^2(X(s, x))} \phi(x) b(X(s, x)) ds$$

$$+ \int_0^T \frac{b(X(s, x))}{\sigma^2(X(s, x))} dX(s, x) - \phi(x) b(X(s, x)) ds$$

$$= \phi(x) \int_0^T \frac{b^2(X(s, x))}{\sigma^2(X(s, x))} ds$$

$$+ \int_0^T \frac{b(X(s, x))}{\sigma^2(X(s, x))} dW(s),$$

(3.24)

(3.25)

$$= \phi(x) U^{(1)}(x, T) + U^{(2)}(x, T), \quad \text{(say)},$$

(3.26)

where $W(s)$ is the standard Weiner process defined on $[0, T]$. 

Given the process $X(\cdot, \cdot)$, continuity of (3.24) with respect to $x$ and $T$ can be seen as follows. Let $T_1, T_2 \in \mathcal{T}$; without loss of generality, let $T_2 > T_1$. Also, let $x_1, x_2 \in \mathcal{X}$. Then,

$$
|U^{(1)}(x_1, T_1) - U^{(1)}(x_2, T_2)|
= \left| \int_0^{T_1} \frac{b^2(X(s, x_1))}{\sigma^2(X(s, x_1))} ds - \int_0^{T_2} \frac{b^2(X(s, x_2))}{\sigma^2(X(s, x_2))} ds \right|
= \left| \int_0^{T_1} \frac{b^2(X(s, x_1))}{\sigma^2(X(s, x_1))} ds - \int_0^{T_2} \frac{b^2(X(s, x_2))}{\sigma^2(X(s, x_2))} ds \right|
\leq \int_0^{T_1} \left| \frac{b^2(X(s, x_1))}{\sigma^2(X(s, x_1))} - \frac{b^2(X(s, x_2))}{\sigma^2(X(s, x_2))} \right| ds
+ \int_0^{T_2} \left| \frac{b^2(X(s, x_1))}{\sigma^2(X(s, x_1))} - \frac{b^2(X(s, x_2))}{\sigma^2(X(s, x_2))} \right| ds
\leq T_{\max} \sup_{s \in [0, T_1]} \left| \frac{b^2(X(s, x_1))}{\sigma^2(X(s, x_1))} - \frac{b^2(X(s, x_2))}{\sigma^2(X(s, x_2))} \right|
+ |T_2 - T_1| \sup_{s \in [T_1, T_2], x \in \mathcal{X}} \left| \frac{b^2(X(s, x))}{\sigma^2(X(s, x))} \right|
\leq T_{\max} \sup_{s \in [0, T_1]} \left| \frac{b^2(X(s, x_1))}{\sigma^2(X(s, x_1))} - \frac{b^2(X(s, x_2))}{\sigma^2(X(s, x_2))} \right| + C_2|T_2 - T_1|,
$$

(3.27)

(3.28)

(3.29)

where $T_{\max} = \sup \mathcal{T}$; $s^* \in [0, T_1]$ is such that the supremum in (3.27) is attained. That there exists such $s^*$ is clear due to continuity of the functions in $s$ and compactness of the interval. In (3.29), $C_2$ is the upper bound for the function $\left| \frac{b^2(X(s, x))}{\sigma^2(X(s, x))} \right|$.

Since $X(\cdot, x)$ is continuous in $x$, due to continuity of $b(\cdot)$ and $\sigma^2(\cdot)$, for any $\epsilon > 0$, one can choose $\delta_1(\epsilon) > 0$ such that $|x_1 - x_2| < \delta_1(\epsilon)$ implies

$$
\left| \frac{b^2(X(s^*, x_1))}{\sigma^2(X(s^*, x_1))} - \frac{b^2(X(s^*, x_2))}{\sigma^2(X(s^*, x_2))} \right| < \frac{\epsilon}{2T_{\max}},
$$

so that the first term in (3.29) is less than $\epsilon/2$. Choosing $\delta_2(\epsilon) = \frac{\epsilon}{2T_{\max}}$ yields that if $|T_2 - T_1| < \delta_2(\epsilon)$, then the second term in (3.29) is less than $\epsilon/2$. This shows continuity of $U^{(1)}(x, T)$ with respect to $x$ and $T$ for given $X(\cdot, \cdot)$. It follows that, for sequences $\{x_m\}_{m=1}^\infty$, $\{T_m\}_{m=1}^\infty$ such that $x_m \to \tilde{x}$ and $T_m \to \tilde{T}$ as $m \to \infty$,

$$
U^{(1)}(x_m, T_m) \xrightarrow{\mathcal{L}} U^{(1)}(\tilde{x}, \tilde{T}).
$$

(3.30)

It is also clear that

$$
\phi(x_m) \xrightarrow{\mathcal{L}} \phi(\tilde{x}).
$$

(3.31)

Now note that due to assumptions (3.1), (3.4) and compactness of $\mathcal{X}$ and $\mathcal{T}$, we have, for any $k \geq 1$,

$$
\sup_{m \geq 1} E \left[ \phi(x_m)U^{(1)}(x_m, T_m) \right]^{2k} < \infty,
$$

(3.32)

for all $m \geq 1$, ensuring requisite uniform integrability. Hence, it follows that

$$
E \left[ \phi(x_m)U^{(1)}(x_m, T_m) - \phi(\tilde{x})U^{(1)}(\tilde{x}, \tilde{T}) \right]^2 \to 0.
$$

(3.33)

Let us now deal with $U^{(2)}(x, T)$ given by (3.25). Letting for any set $A$, $\delta_A(s) = 1$ if $s \in A$ and 0
otherwise, be the indicator function, we define

\[
Q(x_m, T_m) = \int_0^{T_{\text{max}}} \left[ \frac{b(X(s, x_m))}{\sigma(X(s, x_m))} \delta_{[0,T_m]}(s) - \frac{b(X(s, \tilde{x}))}{\sigma(X(s, \tilde{x}))} \delta_{[0,T]}(s) \right]^2 ds
\]

\[
= \int_0^{T_{\text{max}}} \frac{b^2(X(s, x_m))}{\sigma^2(X(s, x_m))} \delta_{[0,T_m]}(s) ds + \int_0^{T_{\text{max}}} \frac{b^2(X(s, \tilde{x}))}{\sigma^2(X(s, \tilde{x}))} \delta_{[0,T]}(s) ds
\]

\[
- 2 \int_0^{T_{\text{max}}} \frac{b(X(s, x_m))}{\sigma(X(s, x_m))} \frac{b(X(s, \tilde{x}))}{\sigma(X(s, \tilde{x}))} \delta_{[0,T_m]}(s) ds
\]

\[
= \int_0^{T_m} \frac{b^2(X(s, x_m))}{\sigma^2(X(s, x_m))} ds + \int_0^T \frac{b^2(X(s, \tilde{x}))}{\sigma^2(X(s, \tilde{x}))} ds
\]

\[
- 2 \int_{\min(T_m, T)}^{\min(T_m, T)} \frac{b(X(s, x_m))}{\sigma(X(s, x_m))} \frac{b(X(s, \tilde{x}))}{\sigma(X(s, \tilde{x}))} ds.
\]

(3.34)

It follows in the same way as in the proof of continuity of \(U^{(1)}(\cdot, \cdot)\) that the first and the third integrals in (3.34) associated with the function \(Q(\cdot, \cdot)\), are continuous at \((\tilde{x}, \tilde{T})\). As a result, for given \(X(\cdot, \cdot), Q(x_m, T_m) \to 0\) as \(m \to \infty\). It follows that \(Q(x_m, T_m) \xrightarrow{L^2} 0\).

Now note that

\[
Q(x_m, T_m) \leq 2 \left[ \left( \int_0^{T_{\text{max}}} \frac{b^2(X(s, x_m))}{\sigma^2(X(s, x_m))} ds \right)^2 + \left( \int_0^{T_{\text{max}}} \frac{b^2(X(s, \tilde{x}))}{\sigma^2(X(s, \tilde{x}))} ds \right)^2 \right],
\]

so that, for any \(k \geq 2\),

\[
E[Q(x_m, T_m)]^k \leq 2^k E \left[ \left( \int_0^{T_{\text{max}}} \frac{b^2(X(s, x_m))}{\sigma^2(X(s, x_m))} ds \right)^{2k} + \left( \int_0^{T_{\text{max}}} \frac{b^2(X(s, \tilde{x}))}{\sigma^2(X(s, \tilde{x}))} ds \right)^{2k} \right].
\]

(3.35)

Since, by assumption (3.4), \(\int_0^T \frac{b^2(X(s, x))}{\sigma^2(X(s, x))} ds < \infty\) for any \(x \in \mathcal{X}\) and \(T \in \mathcal{T}\), and since \(\mathcal{X}\) and \(\mathcal{T}\) are compact, it follows that

\[
\sup_{m \geq 1} E[Q(x_m, T_m)]^k < \infty,
\]

guaranteeing uniform integrability. Hence,

\[
E[Q(x_m, T_m)] \to 0, \quad \text{as } m \to \infty.
\]

(3.36)

By Itô isometry (see, for example, Øksendal (2003)) it then follows that

\[
E \left[ \int_0^{T_{\text{max}}} \frac{b(X(s, x_m))}{\sigma(X(s, x_m))} \delta_{[0,T_m]}(s) dW(s) - \int_0^{T_{\text{max}}} \frac{b(X(s, \tilde{x}))}{\sigma(X(s, \tilde{x}))} \delta_{[0,T]}(s) dW(s) \right]^2 \to 0.
\]

(3.37)

That is,

\[
E \left[ \int_0^{T_m} \frac{b(X(s, x_m))}{\sigma(X(s, x_m))} dW(s) - \int_0^T \frac{b(X(s, \tilde{x}))}{\sigma(X(s, \tilde{x}))} dW(s) \right]^2 \to 0.
\]

(3.38)

It follows that

\[
U^{(2)}(x_m, T_m) \xrightarrow{L^2} U^{(2)}(\tilde{x}, \tilde{T}).
\]

(3.39)

Using the Burkholder-Davis-Gundy inequality (see, for example, Delattre et al. (2012)) we obtain

\[
E \left[ U^{(2)}(x_m, T_m) \right]^{2k} \leq C_k E \left[ \int_0^{T_m} \frac{b^2(X(s, x_m))}{\sigma^2(X(s, x_m))} ds \right]^k.
\]

(3.40)
Again, due to assumption (3.4) and compactness of $X$ and $T$ it follows that $\sup_{m \geq 1} E \left[ U^{(2)}(x_m, T_m) \right]^{2k} < \infty$, so that uniform integrability is assured. It follows that

$$E \left[ U^{(2)}(x_m, T_m) - U^{(2)}(\tilde{x}, \tilde{T}) \right]^2 \to 0.$$  \hfill (3.41)

From (3.33) and (3.41) it follows, using the Cauchy-Schwartz inequality, that

$$E \left[ U(x_m, T_m) - U(\tilde{x}, \tilde{T}) \right]^2 \leq E \left[ \phi(x_m)U^{(1)}(x_m, T_m) - \phi(\tilde{x})U^{(1)}(\tilde{x}, \tilde{T}) \right]^2 + 2 \sqrt{E \left[ \phi(x_m)U^{(1)}(x_m, T_m) - \phi(\tilde{x})U^{(1)}(\tilde{x}, \tilde{T}) \right]^2 E \left[ U^{(2)}(x_m, T_m) - U^{(2)}(\tilde{x}, \tilde{T}) \right]^2} \to 0.$$  \hfill (3.42)

Since $V(x, T) = U^{(1)}(x, T)$, due to (3.30) and assumption (3.4) (the latter ensuring uniform integrability), it easily follows that

$$E \left[ V(x_m, T_m) - V(\tilde{x}, \tilde{T}) \right]^2 \to 0.$$  \hfill (3.43)

Let $G(x, T) = (U(x, T), V(x, T))$. Then it follows from (3.42) and (3.43), that

$$G(x_m, T_m) \xrightarrow{\mathcal{L}} G(\tilde{x}, \tilde{T}).$$  \hfill (3.44)

That is,

$$\left( U(x_m, T_m), V(x_m, T_m) \right) \xrightarrow{\mathcal{L}} \left( U(\tilde{x}, \tilde{T}), V(\tilde{x}, \tilde{T}) \right).$$  \hfill (3.45)

Consider also a sequence $\{\theta_m\}_{m=1}^{\infty}$ in $\Omega$, converging to $\tilde{\theta} \in \Omega$. Then, for any function $h(u, v, \theta)$, which is continuous in $u$, $v$ and $\theta$, and such that the sequence $\{h(U(x_m, T_m), V(x_m, T_m), \theta_m)\}_{m=1}^{\infty}$ is uniformly integrable, we must have

$$E [h(U(x_m, T_m), V(x_m, T_m), \theta_m)] \to E \left[ h(U(\tilde{x}, \tilde{T}), V(\tilde{x}, \tilde{T}), \tilde{\theta}) \right],$$  \hfill (3.46)

ensuring continuity of $E [h(U(x, T), V(x, T), \theta)]$ with respect to $x$, $T$ and $\theta$.

**References**

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.

Hoadley, B. (1971). Asymptotic Properties of Maximum Likelihood Estimators for the Independent not Identically Distributed Case. *The Annals of Mathematical Statistics*, 42, 1977–1991.

Øksendal, B. (2003). *Stochastic Differential Equations*. Springer-Verlag, New York.

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