Asymptotic Theory of Bayes Factor in Stochastic Differential Equations with Random Effects

Trisha Maitra and Sourabh Bhattacharya∗

Abstract

Research on model selection in the context of stochastic differential equations (SDE’s) is almost non-existent in the literature. In particular, when a system of SDE’s is considered, as in random effects models, the problem of model selection has not been hitherto investigated.

In this article, we consider a system of SDE’s for modeling random effects, and address the question of model selection using Bayes factors. Specifically, we develop the asymptotic theory of Bayes factors when the observed processes associated with the systems of SDE’s are independently and identically distributed, as well as when they are independently but not identically distributed.

Keywords: Bayes factor; Consistency; Kullback-Leibler divergence; Martingale; Random effects; Stochastic differential equations.

1 Introduction

Random effects models have a rich tradition in statistical applications where it is important to account for variabilities between and within subjects. When “within” variability is caused by some random component varying continuously with time, stochastic differential equations (SDE’s) have important roles to play for modeling the temporal component. Unfortunately, SDE based random effects models are somewhat rare in the statistical literature, in spite of their importance. We refer to Delattre et al. (2013) for a brief review, who also undertake theoretical and classical asymptotic investigation of a class of random effects models based on SDE’s. They specifically deal with the following form: for $i = 1, \ldots, n$,

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

(1.1)

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. The function $b(x, \varphi)$, which is the drift function, is a known, real-valued function on $\mathbb{R} \times \mathbb{R}^d$ ($\mathbb{R}$ is the real line and $d$ is the dimension), and the function $\sigma: \mathbb{R} \mapsto \mathbb{R}$ is the known diffusion coefficient. In statistical applications related to random effects, $X_i(\cdot)$ models the $i$-th individual. The SDE’s given by (1.1) are driven by independent standard Wiener 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, which are assumed by Delattre et al. (2013) to be independent of the Brownian motions and independently and identically distributed (iid) random variables with some common distribution. Delattre et al. (2013) impose the following assumptions for ensuring existence of strong solution of (1.1).

(H1) (i) The function $(x, \varphi) \mapsto b(x, \varphi)$ is $C^1$ (differentiable with continuous first derivative) on $\mathbb{R} \times \mathbb{R}^d$, and such that there exists $K > 0$ so that

$$b^2(x, \varphi) \leq K(1 + x^2 + |\varphi|^2),$$

for all $(x, \varphi) \in \mathbb{R} \times \mathbb{R}^d$.

(ii) The function $\sigma(\cdot)$ is $C^1$ on $\mathbb{R}$ and

$$\sigma^2(x) \leq K(1 + x^2),$$

for all $x \in \mathbb{R}$.

∗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.
Let $X_i^\varphi$ be associated with the SDE of the form (1.1) with drift function $b(x, \varphi)$. Also letting $Q^\varphi_{x,T_i}$ denote the joint distribution of $\{X_i^\varphi(t); t \in [0, T_i]\}$, it is assumed that for $i = 1, \ldots, n$, and for all $\varphi, \varphi'$, the following holds:

$$Q^\varphi_{x,T_i} \left( \int_0^{T_i} \frac{b^2(X_i^\varphi(t), \varphi')}{\sigma^2(X_i^\varphi(t))} dt < \infty \right) = 1.$$ 

For $g = \frac{\partial h}{\partial \varphi^j}$, $j = 1, \ldots, d$, there exist $c > 0$ and some $\gamma \geq 0$ such that

$$\sup_{\varphi \in \mathbb{R}^d} \frac{|g(x, \varphi)|}{\sigma^2(x)} \leq c (1 + |x|^\gamma).$$

For the sake of convenience, Delattre et al. (2013), Maitra and Bhattacharya (2014a) and Maitra and Bhattacharya (2014b) assume $b(x, \phi_i) = \phi_i b(x)$. With this assumption (H3) is no longer required.

Exploiting the linearity assumption $b(x, \phi_i) = \phi_i b(x)$, and assuming $\phi_i$ to be Gaussian random variables with mean $\mu$ and variance $\omega^2$, Delattre et al. (2013) obtained a closed form of the likelihood of $(\mu, \omega^2)$ and proved convergence in probability and asymptotic normality of the maximum likelihood estimator of $(\mu, \omega^2)$ under the iid set-up, that is, when $x_i = x$ and $T_i = T$, for $i = 1, \ldots, n$. Maitra and Bhattacharya (2014a) proved strong consistency and asymptotic normality of the maximum likelihood estimator under weaker assumptions in the iid set-up. They also proved weak convergence and asymptotic normality under the independent but non-identical set-up, which they refer to as the non-iid setup. Maitra and Bhattacharya (2014b) investigated Bayesian asymptotics in the iid and non-iid set-ups, as a follow-up of Maitra and Bhattacharya (2014a).

In this article, we explore the model selection problem in these SDE based random effects models, in both iid and non-iid contexts. In our knowledge, the problem of model selection has not been addressed in the random effects based SDE models, although model selection using intrinsic and fractional Bayes factors has been considered by Sivaganesan and Lingham (2002) with three particular diffusion models in single equation set-ups. In this paper, we prove consistency of the relevant Bayes factors in our SDE-based random effects models. In the iid case we develop our asymptotic theory based on a general result already existing in the literature. However, for the non-iid situation we develop a general theorem, and prove our result on SDE as a special case of our theorem.

The rest of our article is structured as follows. In Section 2 we illustrate the problem of model selection in systems of SDE’s. In Section 3 we investigate consistency of the Bayes factor when the SDE models being compared form an iid system of equations. In Section 4 we develop a general asymptotic theory of Bayes factors in the non-iid situation, and then in Section 5 we investigate consistency of the Bayes factor when the system of SDE’s are non-iid. We provide concluding remarks in Section 6.

## 2 Illustration of the model selection problem in the SDE set-up

First, we consider the following two systems of SDE models for $i = 1, 2, \ldots, n$:

$$dX_i(t) = \phi_i b_0(X_i(t))dt + \sigma_0(X_i(t))dW_i(t) \quad (2.1)$$

and

$$dX_i(t) = \phi_i b_1(X_i(t))dt + \sigma_1(X_i(t))dW_i(t) \quad (2.2)$$

where, $X_i(0) = x_i$ is the initial value of the stochastic process $X_i(t)$, which is assumed to be continuously observed on the time interval $[0, T_i]; T_i > 0$ for all $i$ and assumed to be known.

We assume that (2.1) represents the true model and (2.2) is any other model. To fix ideas, let us first
define the following quantities:

\[ U_{i,j} = \int_0^{T_i} \frac{b_j(X_i(s))}{\sigma_j^2(X_i(s))} \, dX_i(s), \quad V_{i,j} = \int_0^{T_i} \frac{b_j^2(X_i(s))}{\sigma_j^2(X_i(s))} \, ds \]  

(2.3)

for \( j = 0, 1 \) and \( i = 1, \ldots, n \). Also, for \( j = 0, 1 \), let

\[ f_j(X_i) = \exp \left( \phi_{i,j} U_{i,j} - \frac{\phi_{i,j}^2}{2} V_{i,j} \right) \]  

(2.4)

so that \( f_0(X_i) \) denotes the true density and \( f_1(X_i) \) stands for any other density.

In reality, \( b_0 \) and \( \sigma_0 \) may be piecewise linear or convex combinations of linear functions, where the number of linear functions involved (and hence, the number of associated intercept and slope parameters) may be unknown. That is, not only the values of the parameters, but also the number of the parameters may be unknown in reality. In general, \( b_0 \) and \( \sigma_0 \) may be any functions, linear or non-linear, satisfying some desirable conditions. Linearity assumptions may be convenient, but need not necessarily be unquestionable. In other words, modeling these functions in the SDE context is a challenging exercise, and hence the issue of model selection must play an important role even in the SDE set-up.

For the sake of generality, we denote \( b_j(\cdot) \) and \( \sigma_j(\cdot) \) as \( b_{\beta_j}(\cdot) \) and \( \sigma_{\gamma_j}(\cdot) \), where \( \beta_j \) and \( \gamma_j \) are vectors of parameters associated with the \( j \)-th model. We accommodate the possibility that the dimensions of \( \beta_0, \beta_1, \gamma_0, \gamma_1 \) may be different. We also set \( \phi_{i,j} = h(x^i, \xi_j) \), where, \( h : \mathbb{R} \times \mathcal{C} \rightarrow \mathbb{R} \) is a real-valued function continuous in both arguments and defined on some compact domain \( \mathcal{C}, \xi_j \) are finite-dimensional parameters having some (perhaps hierarchical) distribution on \( \mathcal{C} \). Thus, the distribution of the random effects \( \phi_{i,j} \) is induced by the distribution of \( \xi_j \).

For \( j = 0, 1 \), let \( \theta_j = (\xi_j, \beta_j, \gamma_j) \), where \( \theta_0 \) is the set of true parameters, and \( \theta_1 \) are unknown parameters on which a prior \( \pi \) is assumed. Let \( \Theta \) be the parameter space on which a prior probability measure is proposed. In view of the aforementioned discussion on parameterization, let us denote \( f_j(X_i), U_{i,j}, V_{i,j}, \) and \( \phi_{i,j} \) by \( f_{\theta_j}(X_i), U_{i,\theta_j}, V_{i,\theta_j}, \) and \( h(x^i, \xi_j) \), respectively. An useful relation between \( U_{i,\theta_j} \) and \( V_{i,\theta_j} \), which we will often make use of in this paper is as follows.

\[ U_{i,\theta_j} = h(x^i, \xi_j)V_{i,\theta_j} + \int_0^{T_i} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} \, dW_i(s). \]  

(2.5)

Also note that

\[ E_{\theta_0} \left[ \int_0^{T_i} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} \, dW_i(s) \right] = 0. \]  

(2.6)

Now, let

\[ R_n(\theta_1) = \prod_{i=1}^n \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}. \]

We are interested in asymptotic properties of the Bayes factor, given by, \( I_0 \equiv 1 \) and for \( n \geq 1 \),

\[ I_n = \int_{\Theta} R_n(\theta_1) \pi(d\theta_1), \]  

(2.7)

as \( n \rightarrow \infty \).

Next, we state some conditions and prove some results required for the development of our model selection theory.

### 2.1 Requisite assumptions and results

We assume the following conditions:

- ...
(H1') The parameter space $\Theta$ is compact.

(H2') For $j = 0, 1$, $b_{\beta_j}(\cdot)$ and $\sigma_{\gamma_j}(\cdot)$ are $C^1$ on $\mathbb{R}$ and satisfy $b_{\beta_j}^2(x) \leq K_1(1 + x^2 + \|\beta_j\|^2)$ and $\sigma_{\gamma_j}^2(x) \leq K_2(1 + x^2 + \|\gamma_j\|^2)$ for all $x \in \mathbb{R}$, for some $K_1, K_2 > 0$. Now, due to (H1') the latter boils down to assuming $b_{\beta_j}^2(x) \leq K_1(1 + x^2)$ and $\sigma_{\gamma_j}^2(x) \leq K_2(1 + x^2)$ for all $x \in \mathbb{R}$, for some $K > 0$.

Because of (H2') it follows from Theorem 4.4 of [Mao (2011)], page 61, that for all $T_i > 0$, and any $k \geq 2$,

$$E \left( \sup_{s \in [0,T_i]} |X_i(s)|^k \right) \leq \left( 1 + 3k^{-1}E|X_i(0)|^k \right) \exp (\vartheta T_i),$$

where

$$\vartheta = \frac{1}{6} (18K)^{\frac{1}{2}} T_i \left[ T_i \left( 3 \right) + \left( \frac{k^3}{2(k-1)} \right)^{\frac{1}{2}} \right].$$

We further assume:

(H3') For every $x$, $b_{\beta_j}(x)$ and $\sigma_{\gamma_j}(x)$ are continuous in $\beta_j$ and $\gamma_j$, respectively, for $j = 0, 1$.

(H4') For $j = 0, 1$,

$$\frac{b_{\beta_j}^2(x)}{\sigma_{\gamma_j}^2(x)} \leq K_{\beta_j,\gamma_j} \left( 1 + x^2 + \|\beta_j\|^2 + \|\gamma_j\|^2 \right),$$

where $K_{\beta_j,\gamma_j}$ is continuous in $(\beta_j, \gamma_j)$.

We then have the following lemma, which will be useful for proving our main results.

**Lemma 1** Assume (H1') – (H4'). Then for $j = 0, 1$, for all $\theta_i \in \Theta$, for $k \geq 1$,

$$E_{\theta_0} [U_{i,\theta_i}]^k < \infty, \quad (2.9)$$

$$E_{\theta_0} [V_{i,\theta_i}]^k < \infty. \quad (2.10)$$

Moreover, for $j = 1$, the above expectations are continuous in $\theta_1$.

**Proof.** We first consider $k = 1$. Observe that (2.10) follows from (H4') and (H1'). Also note that since $E_{\theta_0} (U_{i,\theta_i}) = E_{\theta_0} (h(x^i, s_j)U_{i,\theta_i})$ by (2.5) and (2.6), (2.9) is implied by (2.10). To see that the moments are continuous in $\theta_1$, let $\{\theta_1^{(m)}\}_{m=1}^{\infty}$ be a sequence converging to $\theta_1$ as $m \to \infty$. Due to (H3'),

$$\frac{b_{\theta_1^{(m)}}^2(X_i(s))}{\sigma_{\theta_1^{(m)}}^2(X_i(s))} \to \frac{b_{\theta_1}^2(X_i(s))}{\sigma_{\theta_1}^2(X_i(s))},$$

as $m \to \infty$, for any given sample path $\{X_i(s) : s \in [0, T_i]\}$. Assumption (H4') implies that $\frac{b_{\theta_1^{(m)}}^2(X_i(s))}{\sigma_{\theta_1^{(m)}}^2(X_i(s))}$ is dominated by $\sup_{\theta_1 \in \Theta} K_{\beta_1,\gamma_1} \left( 1 + |X_i(s)|^2 + \sup_{\theta_1 \in \Theta} \|\beta_1\|^2 + \sup_{\theta_1 \in \Theta} \|\gamma_1\|^2 \right)$. Since $X_i(s)$ is continuous on $[0, T_i]$, (guaranteed by (H2'); see [Delattre et al. (2013)]), it follows that $\int_0^{T_i} [X_i(s)]^2 \, ds < \infty$, which, in turn guarantees, in conjunction with compactness of $\Theta$, that the upper bound is integrable. Hence, $V_{i,\theta_1^{(m)}} \to V_{i,\theta_1}$, almost surely. Now, for all $m \geq 1$,

$$V_{i,\theta_1^{(m)}} < T_i \sup_{\theta_1 \in \Theta} K_{\beta_1,\gamma_1} \left( 1 + \sup_{s \in [0,T_i]} |X_i(s)|^2 + \sup_{\theta_1 \in \Theta} \|\beta_1\|^2 + \sup_{\theta_1 \in \Theta} \|\gamma_1\|^2 \right).$$
Since \( E \left( \sup_{s \in [0,T_1]} |X_i(s)|^2 \right) < \infty \) by (2.8), it follows that \( E_{\theta_0} \left( V_i, \theta_1^{(m)} \right) \rightarrow E_{\theta_0} \left( V_i, \theta_1^* \right) \), as \( \theta_1^{(m)} \rightarrow \theta_1^* \).

Hence, \( E_{\theta_0} \left( V_i, \theta_1 \right) \) is continuous in \( \theta_1 \). Since \( E_{\theta_0} \left( U_{i,\theta_1} \right) = h(x^i, \xi_1) E_{\theta_0} \left( V_i, \theta_1 \right) \), and since \( h(x^i, \xi_1) \) is continuous in \( \xi_1 \), it follows that \( E_{\theta_0} \left( U_{i,\theta_1} \right) \) is continuous in \( \theta_1 \).

We now consider \( k \geq 2 \). Note that, due to (H4'), and the inequality \((a + b)^k \leq 2^{k-1}(|a|^k + |b|^k)\) for \( k \geq 2 \) and any \( a, b \),

\[
E_{\theta_0} \left( V_i, \theta_1 \right)^k \leq 2^{k-1} T_i^{k} K_{\beta_j, \gamma_j} \left( 1 + \|\beta_j\|^2 + \|\gamma_j\|^2 \right)^k + 2^{k-1} T_i^{k} K_{\beta_j, \gamma_j} E \left( \sup_{s \in [0,T_1]} |X_i(s)|^{2k} \right).
\]

Since \( E \left( \sup_{s \in [0,T_1]} |X_i(s)|^{2k} \right) < \infty \) due to (2.8), and because \( K_{\beta_j, \gamma_j} \), \( \|\beta_j\| \) and \( \|\gamma_j\| \) are continuous in compact \( \Theta \), (2.10) follows. To see that (2.9) holds, note that, due to (2.5) and \((a + b)^k \leq 2^{k-1}(|a|^k + |b|^k)\),

\[
E_{\theta_0} \left( U_{i,\theta_1} \right)^k \leq 2^{k-1} h^k(x^i, \xi_j) E_{\theta_0} \left( V_i, \theta_1 \right)^k + 2^{k-1} E_{\theta_0} \left( \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} dW_i(s) \right)^k.
\]  

(2.11)

Since, due to (H4') and (2.8),

\[
E_{\theta_0} \left( \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} ds \right)^k \leq \infty,
\]

Theorem 7.1 of [Maio 2011] shows that

\[
E_{\theta_0} \left( \left( \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} dW_i(s) \right)^k \right) \leq \left( \frac{k(k - 1)}{2} \right)^{\frac{k}{2} - 1} T_i^{k - 2} E_{\theta_0} \left( \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} ds \right)^k. \quad \text{(2.12)}
\]

Combining (2.11) with (2.12), it follows that \( E_{\theta_0} \left( U_{i,\theta_1} \right)^k \leq \infty \).

As regards continuity of the moments for \( k \geq 2 \), first note that in the context of \( k = 1 \), we have shown almost sure continuity of \( V_i, \theta_1 \) with respect to \( \theta_1 \). Hence, \( V_{i,\theta_1}^k \) is almost surely continuously with respect to \( \theta_1 \). That is, \( \theta_1^{(m)} \rightarrow \theta_1^* \) implies \( V_{i,\theta_1^{(m)}}^k \rightarrow V_{i,\theta_1^*}^k \), almost surely. Once again, dominated convergence theorem allows us to conclude that \( E_{\theta_0} \left( V_{i,\theta_1^{(m)}} \right)^k \rightarrow E_{\theta_0} \left( V_{i,\theta_1^*} \right)^k \), implying continuity of \( \theta_1^{(m)} \) with respect to \( \theta_1 \). To see continuity of \( E_{\theta_0} \left( U_{i,\theta_1} \right)^k \), first note that

\[
E_{\theta_0} \left[ \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} ds - \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} \right]^2 \rightarrow 0,
\]

as \( m \rightarrow \infty \). The result follows as before by first noting pointwise convergence, and then using (H4') and then (2.8), along with (H1'). By Itô isometry it holds that

\[
E_{\theta_0} \left[ \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} dW_i(s) - \int_0^{T_1} \frac{b_{\beta_j}(X_i(s))}{\sigma_{\gamma_j}(X_i(s))} dW_i(s) \right]^2 \rightarrow 0.
\]
Hence,
\[
\int_0^T \frac{b_{\beta_1}(m) (X_i(s))}{\sigma_{\gamma_1}(m)(X_i(s))} dW_i(s) \to \int_0^T \frac{b_{\beta_1}(X_i(s))}{\sigma_{\gamma_1}(X_i(s))} dW_i(s)
\]
in probability, as \( m \to \infty \). Since, \( V_{i,\theta_1}^{(m)} \to V_{i,\theta_1} \) and \( h \left( x^i, \xi_1^{(m)}(s) \right) \to h \left( x^i, \xi_1^* \right) \) almost surely as \( m \to \infty \), it follows from (2.5) that \( U_{i,\theta_1}^{(m)} \to U_{i,\theta_1} \) in probability, so that \( U_{i,\theta_1}^{(m)} \to U_{i,\theta_1}^k \) in probability. Using (H4’), (2.8) and (H1’), it is easily seen, using the same methods associated with (2.11) and (2.12), that \( \sup_m E_{\theta_0} \left( U_{i,\theta_1}^{(m)} \right) < \infty \), proving that \( \left\{ U_{i,\theta_1}^{(m)} \right\}_{m=1}^\infty \) is uniformly integrable. Hence, \( E_{\theta_0} \left( U_{i,\theta_1}^{(m)} \right)^k \to E_{\theta_0} \left( U_{i,\theta_1} \right)^k \). In other words, \( E_{\theta_0} (U_{i,\theta_1})^k \) is continuous in \( \theta_1 \).

3 Consistency of Bayes factor in the SDE based iid set-up

We first consider the iid set-up; in other words, we assume that \( x^i = x \), \( T_i = T \) for \( i = 1, \ldots, n \), and \( j = 0, 1 \). We shall relax these assumptions subsequently when we take up the non-iid (that is, independent, but non-identical) case.

3.1 A general result on consistency of Bayes factor in the iid set-up

To investigate consistency of the Bayes factor, we resort to a general result in the iid set-up developed by Walker (2004) (see also Walker et al. (2004)). To state the result we first define some relevant notation which apply to both parametric and nonparametric problems. For any \( x \) in the appropriate domain, let
\[
\hat{f}_n(x) = \int f(x) \pi_n(df)
\]
be the posterior predictive density, where \( \pi_n \) stands for the posterior of \( f \), given by
\[
\pi_n(A) = \frac{\int_A \prod_{i=1}^n f(X_i) \pi(df)}{\int \prod_{i=1}^n f(X_i) \pi(df)}
\]
and let
\[
\hat{f}_{n,A}(x) = \int f(x) \pi_{n,A}(df)
\]
be the posterior predictive density restricted to the set \( A \), that is, for the prior probability \( \pi(A) > 0 \),
\[
\pi_{n,A}(df) = \frac{I_A(f) \pi_n(df)}{\int_A \pi_n(df)}
\]
where \( I_A \) denotes the indicator function of the set \( A \).

Clearly, the above set-up is in accordance with the iid situation. The following theorem of Walker (2004) is appropriate for our iid set-up.

Theorem 2 (Walker (2004)) Assume that
\[
\pi \left( f : K(f, f_0) < \epsilon \right) > 0, \quad (3.1)
\]
only for, and for all \( \epsilon > \delta \), for some \( \delta \geq 0 \), and that for all \( \epsilon > 0 \),
\[
\liminf_n K \left( f_0, \hat{f}_{n,A} \right) \geq \epsilon, \quad (3.2)
\]
when \( A(\epsilon) = \{ f : K(f_0, f) > \epsilon \} \). Property (3.1) is the Kullback-Leibler property and (3.2) has been
referred to as the $Q^*$ property by Walker (2004). Assume further that

$$\sup_n \text{Var} \left( \frac{I_{n+1}}{I_n} \right) < \infty.$$  \hfill (3.3)

Then,

$$n^{-1} \log (I_n) \to -\delta,$$  \hfill (3.4)

almost surely.

The following corollary provides the result on asymptotic comparison between two models using Bayes factors, in the iid case.

**Corollary 3** (Walker (2004)) Let $R_n(f) = \prod_{i=1}^n f(X_i)_{f_0(X_i)}$. For $j = 1, 2$, let

$$I_{jn} = \int R_n(f) \pi_j(df),$$

where $\pi_1$ and $\pi_2$ are two different priors on $f$. Let $B_n = I_{1n}/I_{2n}$ denote the Bayes factor for comparing the two models associated with $\pi_1$ and $\pi_2$. If $\pi_1$ and $\pi_2$ have the Kullback-Leibler property (3.1) with $\delta = \delta_1$ and $\delta = \delta_2$ respectively, satisfy the $Q^*$ property (3.2), and (3.3) with $I_n = I_{jn}$, for $j = 1, 2$, then

$$n^{-1} \log B_n \to \delta_2 - \delta_1,$$

almost surely.

### 3.2 Verification of Theorem 2 in iid SDE set-up

In our parametric case, $f_0 \equiv f_{\theta_0}$ and $f \equiv f_{\theta_1}$. In this iid set-up, the Kullback-Leibler (KL) divergence measure between $f_0$ and $f_1$ is given by

$$K(f_{\theta_0}, f_{\theta_1}) = \frac{h(x, \xi_0)^2}{2} E_{\theta_0}(V_1, \theta_0) - \frac{h(x, \xi_1)^2}{2} E_{\theta_0}(V_1, \theta_1),$$  \hfill (3.5)

where $E_{\theta_0} \equiv E_{f_{\theta_0}}$. The result easily follows from (2.5) and (2.6). Now let

$$\delta = \min_{\Theta} K(f_{\theta_0}, f_{\theta_1})$$

$$\delta = \min_{\Theta} \left\{ \frac{h^2(x, \xi_0)}{2} E_{\theta_0}(V_1, \theta_0) - \frac{h^2(x, \xi_1)}{2} E_{\theta_0}(V_1, \theta_1) \right\}. \hfill (3.6)

Since $E_{\theta_0}(V_1, \theta_1)$ is continuous in $\theta_1$, compactness of $\Theta$ guarantees that $0 \leq \delta < \infty$.

#### 3.2.1 Verification of (3.1)

To see that (3.1) holds in our case for any prior dominated by Lebesgue measure, first let us define

$$K^*(f_{\theta^*}, f_{\theta_1}) = K(f_{\theta_0}, f_{\theta_1}) - K(f_{\theta_0}, f_{\theta^*}),$$  \hfill (3.7)

where $f_{\theta^*} = \arg\min_{\Theta} K(f_{\theta_0}, f_{\theta_1})$. Now, let us choose any prior $\pi$ such that $\frac{d\pi}{d\theta} = g$ where $g$ is a continuous positive density with respect to Lebesgue measure, where, by “positive” density, we mean a density excluding any interval of null measure. For any $c > 0$, we then need to show that

$$\pi(\theta_1 \in \Theta : \delta \leq K(f_{\theta_0}, f_{\theta_1}) < \delta + c) > 0,$$
for any prior $\pi$ dominated by Lebesgue measure. This is equivalent to showing

$$\pi(\theta_1 \in \Theta : 0 \leq K^*(f_{\theta^*}, f_{\theta_1}) < c) > 0,$$

for any prior $\pi$ dominated by Lebesgue measure.

Since $K(f_{\theta_0}, f_{\theta_1})$ is continuous in $\theta_1$, so is $K^*(f_{\theta^*}, f_{\theta_1})$. Compactness of $\Theta$ ensures uniform continuity of $K^*(f_{\theta_0}, f_{\theta_1})$. Hence, for any $c > 0$, there exists $\epsilon_c$ independent of $\theta_1$, such that $\|\theta_1 - \theta^*\| < \epsilon_c$ implies $K^*(f_{\theta^*}, f_{\theta_1}) < c$. Then,

$$\pi(\theta_1 \in \Theta : 0 \leq K^*(f_{\theta^*}, f_{\theta_1}) < c) \geq \pi(\theta_1 \in \Theta : \|\theta_1 - \theta^*\| < \epsilon_c) \geq \nu(\{\theta_1 \in \Theta : \|\theta_1 - \theta^*\| < \epsilon_c\}) > 0,$$

(3.8)

where $\nu$ stands for Lebesgue measure. In other words, (3.1) holds in our case.

### 3.2.2 Verification of (3.2)

To see that (3.2) also holds in our SDE set-up, first note that in our case

$$\hat{f}_{nA(\epsilon)}(x) = \frac{\int_{A(\epsilon)} f_{\theta_1}(x) \pi_n(d\theta_1)}{\int_{A(\epsilon)} \pi_n(d\theta_1)},$$

(3.9)

with

$$A(\epsilon) = \{\theta_1 \in \Theta : K(f_{\theta_0}, f_{\theta_1}) \geq \epsilon\} = \left\{\theta_1 : h^2(x, \xi_0) - \frac{h^2(x, \xi_1)}{2} E_{\theta_0}(V_{1, \theta_0}) > \epsilon\right\} \quad (3.10)$$

for any $\epsilon > 0$. Note that, here we have replaced $K(f_{\theta_0}, f_{\theta_1}) > \epsilon$ with $K(f_{\theta_0}, f_{\theta_1}) \geq \epsilon$ in the definition of $A(\epsilon)$ because of continuity of the posterior of $\theta_1$. Note that

$$\hat{f}_{nA(\epsilon)}(X) \leq \sup_{\theta_1 \in A(\epsilon)} f_{\theta_1}(X) = f_{\theta_1^*(X)}(X),$$

(3.11)

where $\theta_1^*(X)$, which depends upon $X$, is the maximizer lying in the compact set $A(\epsilon)$. Now note that

$$K(f_{\theta_0}, \hat{f}_{nA(\epsilon)}) = E_{\theta_0} \left[ \log f_{\theta_0}(X) - \log \hat{f}_{nA(\epsilon)}(X) \right]$$

$$\geq E_{\theta_0} \left[ \log f_{\theta_0}(X) - \log f_{\theta_1^*(X)}(X) \right]$$

$$= E_{\theta_0} \left[ \log \frac{f_{\theta_0}(X)}{f_{\theta_1^*(X)}(X)} \right]$$

$$= E_{\theta_1^*(X)} E_{\theta_0} \left[ \log f_{\theta_0}(X) \right] \frac{1}{f_{\theta_1^*(X)}(X)}$$

$$= E_{\theta_1^*(X)} E_{\theta_0} \left[ \log \left( \frac{f_{\theta_0}(X)}{f_{\theta_1^*(X)}(X)} \right) \right]$$

$$= E_{\theta_1^*(X)} E_{\theta_0} K(f_{\theta_0}, f_{\theta_1})$$

$$\geq E_{\theta_1^*(X)} \inf_{\theta_1 \in A(\epsilon)} K(f_{\theta_0}, f_{\theta_1})$$

$$= E_{\theta_1^*(X)} K(f_{\theta_0}, f_{\theta^*})$$

$$\geq \epsilon,$$

(3.12)

where $\theta^* = \arg \min_{\theta \in A(\epsilon)} K(f_{\theta_0}, f_{\theta})$. Hence, (3.2) is satisfied in our SDE set-up.
3.2.3 Verification of (3.3)

We now prove that (3.3) also holds. It is straightforward to verify that

\[
\frac{I_{n+1}}{I_n} = \frac{\hat{f}_{n+1}(X_{n+1})}{f_{\theta_1}(X_{n+1})},
\]

where

\[
\hat{f}_{n+1}(\cdot) = E_{\theta_1|X_1,\ldots,X_n}[f_{\theta_1}(\cdot)]
\]

is the posterior predictive distribution of \( f_{\theta_1}(\cdot) \), with respect to the posterior of \( \theta_1 \), given \( X_1,\ldots,X_n \).

In (3.14), \( E_{\theta_1|X_1,\ldots,X_n} \) denotes expectation with respect to the posterior of \( \theta_1 \) given \( X_1,\ldots,X_n \).

First note that, since

\[
\log [f_{\theta_0}(X_{n+1})] = h(x,\xi_0)U_{n+1,\theta_0} - \frac{h^2(x,\xi_0)}{2}V_{n+1,\theta_0},
\]

it follows from Lemma 1 that the moments of all orders of \( \log [f_{\theta_0}(X_{n+1})] \) exist and are finite. Also, since \( X_i \) are iid, the moments are the same for every \( n = 1,2,\ldots \). In other words,

\[
\sup_n \text{Var} (\log f_{\theta_0}(X_{n+1})) < \infty.
\]

Then observe that for any given \( X_{n+1} \), using compactness of \( \Theta \) and continuity of \( f_{\theta_1}(X_{n+1}) \) with respect to \( \theta_1 \),

\[
f_{\theta_1^*}(X_{n+1})(X_{n+1}) = \inf_{\theta_1 \in \Theta} f_{\theta_1}(X_{n+1}) \leq \hat{f}_{n+1}(X_{n+1}) \leq \sup_{\theta_1 \in \Theta} f_{\theta_1}(X_{n+1}) = f_{\theta_1^{**}}(X_{n+1})(X_{n+1}),
\]

where \( \theta_1^*(X_{n+1}) = \arg \min_{\theta_1 \in \Theta} f_{\theta_1}(X_{n+1}) \) and \( \theta_1^{**}(X_{n+1}) = \arg \max_{\theta_1 \in \Theta} f_{\theta_1}(X_{n+1}) \). Clearly, \( \theta_1^*(X_{n+1}), \theta_1^{**}(X_{n+1}) \in \Theta \), for any given \( X_{n+1} \). Moreover,

\[
\theta_1^*(X_{n+1}) = (\xi_1^*,\beta_1^*,\gamma_1^*), \quad \theta_1^{**}(X_{n+1}) = (\xi_1^{**},\beta_1^{**},\gamma_1^{**})
\]

where each component of \( \theta_1^*(X_{n+1}) \) and \( \theta_1^{**}(X_{n+1}) \) depends on \( X_{n+1} \). It follows from the above inequality that

\[
-|h(\xi_1^*(X_{n+1}))| \left| U_{n+1,\theta_1^*(X_{n+1})} \right| - \frac{h^2(\xi_1^*(X_{n+1}))}{2}V_{n+1,\theta_1^*(X_{n+1})} \\
\leq h(\xi_1^*(X_{n+1}))U_{n+1,\theta_1^*(X_{n+1})} - \frac{h^2(\xi_1^*(X_{n+1}))}{2}V_{n+1,\theta_1^*(X_{n+1})} \\
\leq \log \hat{f}_{n+1}(X_{n+1}) \\
\leq h(\xi_1^{**}(X_{n+1}))U_{n+1,\theta_1^{**}(X_{n+1})} - \frac{h^2(\xi_1^{**}(X_{n+1}))}{2}V_{n+1,\theta_1^{**}(X_{n+1})} \\
\leq -|h(\xi_1^{**}(X_{n+1}))| \left| U_{n+1,\theta_1^{**}(X_{n+1})} \right| + \frac{h^2(\xi_1^{**}(X_{n+1}))}{2}V_{n+1,\theta_1^{**}(X_{n+1})}
\]

Hence, \( E_{\theta_0} \left( \log \hat{f}_{n+1}(X_{n+1}) \right)^2 \) lies between \( E_{\theta_0} \left( |h(\xi_1^*(X_{n+1}))| \left| U_{n+1,\theta_1^*(X_{n+1})} \right| + \frac{h^2(\xi_1^*(X_{n+1}))}{2}V_{n+1,\theta_1^*(X_{n+1})} \right)^2 \)

and \( E_{\theta_0} \left( |h(\xi_1^{**}(X_{n+1}))| \left| U_{n+1,\theta_1^{**}(X_{n+1})} \right| + \frac{h^2(\xi_1^{**}(X_{n+1}))}{2}V_{n+1,\theta_1^{**}(X_{n+1})} \right)^2 \).

We obtain uniform lower and upper bounds of the above two expressions in the following manner.

For the upper bound of the latter we first take supremum of the expectation with respect to \( X_{n+1} \), conditional on \( \theta_1^*(X_{n+1}) = \psi \), over \( \psi \in \Theta \), and then take expectation with respect to \( X_{n+1} \). Since \( \psi \in \Theta \), compactness of \( \Theta \) and Lemma 1 ensure that the moments of any given order of the above expression is uniformly bounded above. Analogously, we obtain a uniform lower bound replacing the
supremum with infimum. In the same way we obtain uniform lower and upper bounds of the other expression. The uniform bounds on the second order moments, in turn, guarantee that
\[
\sup_n \Var \left( \log \hat{f}_{n+1}(X_{n+1}) \right) < \infty.
\] (3.17)
Combining (3.16) and (3.17) and using the Cauchy-Schwartz inequality for the covariance term associated with \( \Var \left( \log \hat{f}_{n+1}(X_{n+1}) - \log f_{\theta_0}(X_{n+1}) \right) \) shows that (3.3) holds in our set-up.

**Theorem 4** Assume the iid case of the SDE based random effects set-up and conditions \((H^I') - (H^I')\). Then (3.4) holds.

We also have the following corollary motivated by (3) in our iid SDE context.

**Corollary 5** For \( j = 1, 2 \), let \( R_{jn}(\theta_j) = \prod_{i=1}^{n} \frac{f_{\theta_j}(X_i)}{f_{\theta_0}(X_i)} \), where \( \theta_1 \) and \( \theta_2 \) are two different finite sets of parameters, perhaps with different dimensionalities, associated with the two models to be compared. For \( j = 1, 2 \), let
\[
I_{jn} = \int R_{jn}(\theta_j) \pi_j(d\theta_j),
\]
where \( \pi_j \) is the prior on \( \theta_j \). Let \( B_n = I_{1n}/I_{2n} \) as before. Assume the iid case of the SDE based random effects set-up and suppose that both the models satisfy conditions \((H^I') - (H^I')\) and have the Kullback-Leibler property with \( \delta = \delta_1 \) and \( \delta = \delta_2 \) respectively. Then
\[
n^{-1} \log B_n \to \delta_2 - \delta_1,
\]
almost surely.

# 4 General asymptotic theory of Bayes factor in the non-iid set-up

In this section, we first develop a general asymptotic theory of Bayes factors in the non-iid set-up, and then obtain the result for the non-iid SDE set-up as a special case of our general theory.

## 4.1 The basic set-up

We assume that \( X_i \sim f_{0i} \) independently for \( i = 1, \ldots, n \), where \( f_{0i} \) is the true density (with respect to some \( \sigma \)-finite measure) of the \( i \)-th observation. Under the proposed model, we assume that the observed data has the following distribution: \( X_i \sim f_i \) independently for \( i = 1, \ldots, n \), where \( f_i \) is some other density (with respect to some \( \sigma \)-finite measure).

We assume that for \( i = 1, \ldots, n \), \( X_i \sim f_{0i} \), i.e. the true density function corresponding to ith individual is \( f_{0i} \). Considering another arbitrary density \( f_i \) for individual \( X_i \) we investigate consistency of the Bayes factor in this general non-iid set-up. For our purpose we introduce the following two properties:

1. **Kullback-Leibler \((\delta)\) property in the non-iid set-up:**

Let us denote Kullback-Leibler divergence measure between \( f_{0i} \) and \( f_i \) by \( K(f_{0i}, f_i) \). Also, assume that the limit
\[
K^\infty(f_0, f) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E \left[ \log \frac{f_{0i}(X_i)}{f_i(X_i)} \right] = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} K(f_{0i}, f_i)
\] (4.1)
exists almost surely with respect to the prior \( \pi \) on \( f \). Let the prior distribution \( \pi \) satisfy
\[
\pi \left( f : \inf_i K(f_{0i}, f_i) \geq \delta \right) = 1,
\] (4.2)
for some \( \delta \geq 0 \). Then we say that \( \pi \) has the Kullback-Leibler \( (\delta) \) property if, for any \( c > 0 \),
\[
\pi(f : \delta \leq K^n(f_0, f) \leq \delta + c) > 0.
\] (4.3)

2. \( Q^* \) property in the non-iid set-up:

Let us denote the posterior distribution corresponding to \( n \) observations by \( \pi_n \). We denote \( \pi(df_1, df_2, \ldots, df_n) \) by \( \pi(df) \). For any set \( A \),
\[
\pi_n(A) = \frac{\int_A \prod_{i=1}^n f_i(X_i) \pi(df)}{\int \prod_{i=1}^n f_i(X_i) \pi(df)}
\]
denotes the posterior probability of \( A \). Let
\[
R_n(f_1, f_2, \ldots, f_n) = \prod_{i=1}^n \frac{f_i(X_i)}{f_{0i}(X_i)}.
\]

Let us define the posterior predictive density by
\[
\hat{f}_n(X_n) = \int f_n(X_n) \pi_n(df_n),
\]
and
\[
\hat{f}_{nA}(X_n) = \int f_n(X_n) \pi_{nA}(df_n)
\]
to be the posterior predictive density with posterior restricted to the set \( A \), that is, for \( \pi(A) > 0 \),
\[
\pi_{nA}(df_n) = \frac{I_A(f_n) \pi_n(df_n)}{\int_A \pi_n(df_n)}.
\]

Then we say that the prior has the property \( Q^* \) in the non-iid set-up if the following holds for any \( \epsilon > 0 \):
\[
\liminf_{n} K(f_0, \hat{f}_{n, A_n(\epsilon)}) \geq \epsilon,
\] (4.4)
when
\[
A_n(\epsilon) = \{ f_n : K(f_0, f_n) \geq \epsilon \}.
\] (4.5)

Let \( I_0 \equiv 1 \) and for \( n \geq 1 \), let us define
\[
I_n = \int R_n(f_1, f_2, \ldots, f_n) \pi(df),
\] (4.6)
which is relevant for the study of the Bayes factors. Regarding convergence of \( I_n \), we formulate the following theorem.

**Theorem 6** Assume the non-iid set-up and that the limit (4.1) exists almost surely with respect to the prior \( \pi \). Also assume that the prior \( \pi \) satisfies (4.2), has the Kullback-Leibler \( (\delta) \) and \( Q^* \) properties given by (4.3) and (4.4), respectively. Assume further that
\[
\sup_i E \left[ \log \frac{f_{0i}(X_i)}{f_i(X_i)} \right]^2 < \infty
\] (4.7)
and
\[
\sup_n E \left[ \log \frac{I_n}{I_{n-1}} \right]^2 < \infty.
\] (4.8)

Then
\[
n^{-1} \log I_n \to -\delta,
\] (4.9)
almost surely as \( n \to \infty \).

**Proof.** Let us consider the martingale sequence

\[
S_N = \sum_{n=1}^{N} \left[ \log \left( \frac{I_n}{I_{n-1}} \right) + \mathcal{K}(f_{0n}, \hat{f}_n) \right],
\]

which is a martingale because \( E[\log (I_n/I_{n-1})|X_1, X_2, \ldots, X_{n-1}] = -\mathcal{K}(f_{0n}, \hat{f}_n) \). Using the above it can be verified that if (4.8) holds, implying

\[
\sum_{n=1}^{\infty} n^{-2} \text{Var} \left[ \log \left( \frac{I_n}{I_{n-1}} \right) \right] < \infty,
\]

then \( S_N/N \to 0 \) almost surely. Therefore

\[
N^{-1} \log I_N + N^{-1} \sum_{n=1}^{N} \mathcal{K}(f_{0n}, \hat{f}_n) \to 0,
\]

(4.10)
aalmost surely, as \( N \to \infty \).

Now consider \( N^{-1} \sum_{i=1}^{N} \log \frac{f_{0i}(X_i)}{f_i(X_i)} \). If (4.7) holds, implying

\[
\sum_{i=1}^{\infty} i^{-2} \text{Var} \left[ \log \left( \frac{f_{0i}(X_i)}{f_i(X_i)} \right) \right] < \infty,
\]

then by Kolmogorov’s strong law of large numbers in the independent but non-identical case,

\[
\frac{1}{N} \sum_{i=1}^{N} \log \frac{f_{0i}(X_i)}{f_i(X_i)} \to \mathcal{K}^\infty(f_0, f)
\]

almost surely, as \( N \to \infty \). Let \( \mathcal{N}_0(c) = \{ f : \delta \leq \mathcal{K}^\infty(f_0, f) \leq \delta + c \} \), where \( c > 0 \). Now, note that,

\[
I_N = \int \prod_{i=1}^{N} \frac{f_i(X_i)}{f_0(X_i)} \pi(d\bar{f}) \\
\geq \int_{\mathcal{N}_0(c)} \exp \left( \sum_{i=1}^{N} \log \frac{f_i(X_i)}{f_0(X_i)} \right) \pi(d\bar{f}) \\
= \int_{\mathcal{N}_0(c)} \exp \left( -\sum_{i=1}^{N} \log \frac{f_{0i}(X_i)}{f_i(X_i)} \right) \pi(d\bar{f}).
\]

By Jensen’s inequality,

\[
\frac{1}{N} \log I_N \geq -\int_{\mathcal{N}_0(c)} \frac{1}{N} \left( \sum_{i=1}^{N} \log \frac{f_{0i}(X_i)}{f_i(X_i)} \right) \pi(d\bar{f}) \quad (4.11)
\]

The integrand on the right hand side converges to \( \mathcal{K}^\infty(f_0, f) \), pointwise for every \( f \), given any sequence \( \{ X_i \}_{i=1}^{\infty} \) associated with the complement of some null set. Since, for all such sequences, uniform integrability of the integrand is guaranteed by (4.7), it follows that the right hand side of (4.11) converges
to $-\int_{\mathcal{X}_0(c)} \mathcal{K}^\infty(f_0, f) \pi(\tilde{df})$ almost surely. Hence, almost surely,

$$
\liminf_N N^{-1} \log I_N \geq -\int_{\mathcal{X}_0(c)} \mathcal{K}^\infty(f_0, f) \pi(\tilde{df}) \\
\geq -(\delta + c) \pi(\mathcal{N}_0(c)) \\
\geq -(\delta + c).
$$

Since $c > 0$ is arbitrary, it follows that

$$
\liminf_N N^{-1} \log I_N \geq -\delta,
$$

almost surely. Now, due to (4.2) it follows that $\mathcal{K}(f_{0n}, f_n) \geq \delta$ for all $n$ with probability 1, so that $\mathcal{K}(f_{0n}, \hat{f}_n) = \mathcal{K}(f_{0n}, \hat{f}_n, A_n(\delta))$, where $A_n(\delta)$ is given by (4.5). By the $Q^*$ property (4.4) it implies that

$$
\liminf_N N^{-1} \sum_{n=1}^N \mathcal{K}(f_{0n}, \hat{f}_n) \geq \delta.
$$

Hence, it follows from (4.10) that

$$
\limsup_N N^{-1} \log I_N \leq -\delta.
$$

Combining (4.12) and (4.13) it follows that

$$
\lim_{N \to \infty} N^{-1} \log I_N = -\delta,
$$

almost surely. ■

**Corollary 7** For $j = 1, 2$, let

$$
I_{jn} = \int R_n(f_1, \ldots, f_n) \pi_j(\tilde{df}),
$$

where $\pi_1$ and $\pi_2$ are two different priors on $f$. Let $B_n = I_{1n}/I_{2n}$ denote the Bayes factor for comparing the two models associated with $\pi_1$ and $\pi_2$. If both the models satisfy the conditions of Theorem 6 and satisfy the Kullback-Leibler property with $\delta = \delta_1$ and $\delta = \delta_2$ respectively, then

$$
n^{-1} \log B_n \to \delta_2 - \delta_1,
$$

almost surely.

5 **Specialization of non-ide asymptotic theory of Bayes factors to non-ide SDE set-up**

In this section we relax the restrictions $T_i = T$ and $x^i = x$ for $i = 1, \ldots, n$. In other words, here we deal with the set-up where the processes $X_i(.)$: $i = 1, \ldots, n$, are independently, but not identically distributed. Following Maitra and Bhattacharya (2014a), Maitra and Bhattacharya (2014b) we assume the following:

(H5') The sequences $\{T_1, T_2, \ldots\}$ and $\{x^1, x^2, \ldots\}$ are sequences in compact sets $\mathfrak{S}$ and $\mathfrak{X}$, respectively, so that there exist convergent subsequences with limits in $\mathfrak{S}$ and $\mathfrak{X}$. 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 $T^\infty \in \mathfrak{S}$ and $x^\infty \in \mathfrak{X}$.
Following [Maitra and Bhattacharya (2014a)], we denote the process associated with the initial value $x$ and time point $t$ as $X(t, x)$, so that $X(t, x') = X_i(t)$, and $X_i = \{X_i(t); t \in [0, T_i]\}$. We also denote by $\phi(x)$ the random effect parameter associated with the initial value $x$ such that $\phi(x') = \phi_i$. As in the iid set-up, here also we assume that $\phi(x)$ is parameterized by $\xi$; in other words, we assume that

$$
\phi(x) = h(x, \xi),
$$

where $\xi \in \mathcal{C}$ is a parameter belonging to some compact set $\mathcal{C}$ and $h : \mathbb{R} \times \mathcal{C} \mapsto \mathbb{R}$ is a function continuous in both the arguments.

For $x \in \mathcal{X}$, $T \in \mathcal{T}$, and $\theta \in \Theta$, let

$$
U(x, T, \theta) = \int_0^T \frac{b\varphi(X(s, x))}{\sigma^2(X(s, x))} dX(s, x); \quad (5.2)
$$

$$
V(x, T, \theta) = \int_0^T \frac{\varphi^2(X(s, x))}{\sigma^2(X(s, x))} ds. \quad (5.3)
$$

Let $\theta = (\xi, \beta, \gamma)$ denote the set of finite number of parameters, where $\beta$ and $\gamma$ have the same interpretation as in the iid set-up.

In this non-iid set-up $f_{0i} = f_{0i, x^i, T_i}$ and $f_i = f_{1i, x^i, T_i}$. An extension of Lemma 1 incorporating $x$ and $T$ shows that moments of $U(x, T, \theta)$, $V(x, T, \theta)$, of all orders exist, and are continuous in $x$, $T$ and $\theta$. Formally, we have the following lemma.

**Lemma 8** Assume (H1$'$) – (H5$'$). Then for all $x \in \mathcal{X}$, $T \in \mathcal{T}$ and $\theta_1 \in \Theta$, for $j = 0, 1$,

$$
\mathbf{E}_{\theta_0} [U(x, T, \theta_j)]^k < \infty, \quad (5.4)
$$

$$
\mathbf{E}_{\theta_0} [V(x, T, \theta_j)]^k < \infty. \quad (5.5)
$$

Moreover, the above expectations are continuous in $(x, T, \theta_1)$.

**Proof.** The proofs of (5.4) and (5.5) follow in the same way as the proofs of (2.9) and (2.10), using compactness of $\mathcal{X}$ and $\mathcal{T}$ in addition to that of $\Theta$.

For the proofs of continuity of the moments, note that as in the iid case, uniform integrability is ensured by (H4$'$), (2.8) and compactness of the sets $\Theta$, $\mathcal{X}$ and $\mathcal{T}$. The rest of the proof is almost the same as the proof of Theorem 5 of [Maitra and Bhattacharya (2014a)].

In particular, the Kullback-Leibler distance is continuous in $x$, $T$ and $\theta_1$. For $x = x^k$ and $T = T_k$, continuity of $\mathcal{K}(f_{\theta_0, x^k, T_k}, f_{\theta_1, x^k, T_k})$ with respect to $x$ and $T$ ensures that as $x^k \rightarrow x^\infty$ and $T_k \rightarrow T^\infty$, $\mathcal{K}(f_{\theta_0, x^k, T_k}, f_{\theta_1, x^k, T_k})$ is a well-defined Kullback-Leibler divergence. Consequently (see [Maitra and Bhattacharya (2014a)], Maitra and Bhattacharya (2014b)), the following holds for any $\theta_1 \in \Theta$,

$$
\lim_{n \rightarrow \infty} \sum_{k=1}^{n} \frac{\mathcal{K}(f_{\theta_0, x^k, T_k}, f_{\theta_1, x^k, T_k})}{n} = \mathcal{K}(f_{\theta_0, x^\infty, T^\infty}, f_{\theta_1, x^\infty, T^\infty}). \quad (5.6)
$$

Even in this non-iid context, the Bayes factor is of the same form as (2.7); however, for $j = 0, 1$, $U_{x^i, T_i, \theta}$ and $V_{x^i, T_i, \theta}$ are not identically distributed for $i = 1, \ldots, n$. Next, we establish consistency of Bayes factor in the non-iid SDE set-up by verifying the sufficient conditions of Theorem 6.

### 5.1 Verification of (4.2) and the Kullback-Leibler property in the non-iid set-up

Firstly, note that in our case,

$$
\mathcal{K}^\infty (f_0, f) = \mathcal{K}^\infty (f_{\theta_0}, f_{\theta_1}) = \mathcal{K}(f_{\theta_0, x^\infty, T^\infty}, f_{\theta_1, x^\infty, T^\infty}), \quad (5.7)
$$
where the rightmost side, given by (5.6), clearly exists almost surely with respect to $\theta_1$ and is also continuous in $\theta_1$.

Now note that compactness of $X$ and $Î"$ along with continuity of the function $h$ and $K(f_{0,x,T},f_{1,x,T})$ with respect to $x$ and $T$ implies

$$
\psi(\theta_1) = \inf_{x \in X, \, T \in Î"} K(f_{0,x,T},f_{1,x,T}) = \inf_{x \in X, \, T \in Î"} K(f_{0,x,T},f_{1,x,T})
$$

$$
= \inf_{x \in X, \, T \in Î"} \left\{ \frac{h^2(x,\xi_0)}{2} E_{\theta_0}(V_{x,T,\theta_0}) - \frac{h^2(x,\xi_1)}{2} E_{\theta_0}(V_{x,T,\theta_1}) \right\}
$$

$$
= \left\{ \frac{h^2(x^*(\theta_1),\xi_0)}{2} E_{\theta_0}(V_{x^*(\theta_1),T^*(\theta_1),\theta_0}) - \frac{h^2(x^*(\theta_1),\xi_1)}{2} E_{\theta_0}(V_{x^*(\theta_1),T^*(\theta_1),\theta_1}) \right\},
$$

(5.8)

where $x^*(\theta_1) \in X$, $T^*(\theta_1) \in Î"$ depend upon $\theta_1$. Then, considering the constant correspondence function $\Gamma(\theta_1) = X \otimes Î", \forall \theta_1 \in \Theta$, where “$\otimes$” indicates Cartesian product, we note that $\Gamma$ is both upper and lower hemicontinuous (hence continuous), and also compact-valued. Hence, Berge’s maximum theorem (Berge (1963)) guarantees that (5.8) is a continuous function of $\theta_1$.

Because of continuity of $\psi(\theta_1)$ in $\theta_1$, the set $\{ \theta_1 : \psi(\theta_1) \geq \delta \}$ is open and can be assigned any desired probability by choosing appropriate priors dominated by the Lebesgue measure. That is, we can assign prior probability one to this set by choosing appropriate priors dominated by the Lebesgue measure. Now, because of the inequality

$$
\pi \left( \theta_1 : \inf_{x^1 \in X, \, T^1 \in Î"} K(f_{0,x^1,T^1},f_{1,x^1,T^1}) \geq \delta \right) \geq \pi (\theta_1 : \psi(\theta_1) \geq \delta),
$$

and since we choose $\pi$ such that $\pi (\theta_1 : \psi(\theta_1) \geq \delta) = 1$, it follows that

$$
\pi \left( \theta_1 : \inf_{x^1 \in X, \, T^1 \in Î"} K(f_{0,x^1,T^1},f_{1,x^1,T^1}) \geq \delta \right) = 1,
$$

satisfying (4.2).

The Kullback-Leibler property of the Lebesgue measure dominated $\pi$ easily follows from continuity of $K^\infty(f_{0\theta},f_{1\theta})$ in $\theta_1$.

5.2 Verification of the $Q^*$ property in the non-iid set-up

Observe that in this situation, for any $\epsilon > 0$,

$$
A_n(\epsilon) = \{ f_n : K(f_{0n},f_n) \geq \epsilon \}
$$

$$
= \{ \theta_1 : K(f_{0n,x^n,T_n},f_{1n,x^n,T_n}) \geq \epsilon \}
$$

Then note that

$$
\hat{f}_{nA_n(\epsilon)}(Z) \leq \sup_{\theta_1 \in A_n(\epsilon)} f_{\theta_1,x^n,T_n}(Z) = f_{\theta_1^*(Z,x^n,T_n)}(Z),
$$

(5.9)
where $\theta^*_1(Z, x^n, T_n)$, which depends upon $Z, x^n, T_n$, is the maximizer lying in the compact set $A_n(\epsilon)$. Now

$$
\mathcal{K}(f_{\theta_0, x^n, T_n}, \hat{f}_{nA_n(\epsilon)}) = E_{\theta_0} \left[ \log f_{\theta_0, x^n, T_n}(Z) \right] - E_{\theta_0} \left[ \log \hat{f}_{nA_n(\epsilon)}(Z) \right]
\geq E_{\theta_0} \left[ \log f_{\theta_0, x^n, T_n}(Z) \right] - E_{\theta_0} \left[ \log f_{\theta^*_1(Z, x^n, T_n)}(Z) \right]
= E_{\theta_0} \left( \log \frac{f_{\theta_0, x^n, T_n}(Z)}{f_{\theta^*_1(Z, x^n, T_n)}(Z)} \right)
= E_{\theta^*_1(Z, x^n, T_n)}(\theta) E_{Z} |\theta^*_1(Z, x^n, T_n) = \xi, \theta_0(\log \frac{f_{\theta_0, x^n, T_n}(Z)}{f_{\theta^*_1(Z, x^n, T_n)}(Z)})
\geq E_{\theta^*_1(Z, x^n, T_n)}(\theta_0) \inf_{\eta \in A_n(\epsilon)} \mathcal{K}(f_{\theta_0, x^n, T_n}, f_{\eta, x^n, T_n})
\geq E_{\theta^*_1(Z, x^n, T_n)}(\theta_0) \mathcal{K}(f_{\theta_0, x^n, T_n}, f_{\eta^*_n, x^n, T_n})
\geq \epsilon,
$$

where $\eta^*_n = \arg \min \mathcal{K}(f_{\theta_0, x^n, T_n}, f_{\eta, x^n, T_n}) \in A_n(\epsilon)$, due to compactness of $A_n(\epsilon)$. Hence, (4.4) is satisfied in our non-iid SDE set-up.

### 5.3 Verification of (4.7)

From Lemma 8 it follows that $E \left\{ \log \frac{f_{\theta_1, x, T}(Z)}{f_{\theta_1, x, T}(Z)} \right\}^2$ exists and is continuous in $\theta_1, x$ and $T$. Then compactness of $\Theta, \mathcal{X}$ and $\mathcal{Z}$ ensures (4.7).

### 5.4 Verification of (4.8)

For the non-iid case, the following identity holds:

$$
\frac{I_{n+1}}{T_n} = \frac{\hat{f}_{n+1}(X_{n+1})}{f_{0,n+1}(X_{n+1})} = \frac{\hat{f}_{n+1, T_{n+1}}(X_{n+1})}{f_{\theta_1, x^{n+1}, T_{n+1}}(X_{n+1})},
$$

where

$$
\hat{f}_{n+1, T_{n+1}}(\cdot) = E_{\theta_1|X_1, \ldots, X_n} \left[ f_{\theta_1, x^{n+1}, T_{n+1}}(\cdot) \right]
$$

is the posterior predictive distribution of $f_{\theta_1, x^{n+1}, T_{n+1}}(\cdot)$, with respect to the posterior of $\theta_1$, given $X_1, \ldots, X_n$.

Now since $\log f_{\theta_1, x^{n+1}, T_{n+1}}(X_{n+1}) = h(x^{n+1}, \xi_0) U_{x^{n+1}, T_{n+1}} \theta_0 - \frac{h^2(x^{n+1}, \xi_0)}{2} V_{x^{n+1}, T_{n+1}} \theta_0$, using Lemma 8 and compactness of $\Theta, \mathcal{X}$ and $\mathcal{Z}$, it is easy to see that the moments of $\log f_{\theta_1, x^{n+1}, T_{n+1}}(X_{n+1})$ are uniformly bounded above. So, we have

$$
\sup_n E \left( \log f_{\theta_1, x^{n+1}, T_{n+1}}(X_{n+1}) \right)^2 < \infty.
$$

(5.13)
As in the iid case, also we have
\[
 f_{\theta_1}(X_{n+1}, x^{n+1}, T_{n+1}) (X_{n+1}) = \inf_{\theta_1 \in \Theta} f_{\theta_1, x^{n+1}, T_{n+1}} (X_{n+1})
 \leq \hat{f}_{x^{n+1}, T_{n+1}}(X_{n+1})
 \leq \sup_{\theta_1 \in \Theta} f_{\theta_1, x^{n+1}, T_{n+1}} (X_{n+1}) = f_{\theta_1^*(X_{n+1}), x^{n+1}, T_{n+1}} (X_{n+1}),
\]
where \( \theta_1^*(X_{n+1}, x^{n+1}, T_{n+1}) = \arg\max_{\theta_1 \in \Theta} f_{\theta_1, x^{n+1}, T_{n+1}} (X_{n+1}) \) and \( \theta_1^{**}(X_{n+1}, x^{n+1}, T_{n+1}) = \arg\max_{\theta_1 \in \Theta} f_{\theta_1, x^{n+1}, T_{n+1}} (X_{n+1}) \). Moreover,
\[
 \theta_1^*(X_{n+1}, x^{n+1}, T_{n+1}) = (\xi_1^*, \beta_1^*, \gamma_1^*), \quad \theta_1^{**}(X_{n+1}, x^{n+1}, T_{n+1}) = (\xi_1^{**}, \beta_1^{**}, \gamma_1^{**})
\]
where each component of \( \theta_1^*(X_{n+1}, x^{n+1}, T_{n+1}) \) and \( \theta_1^{**}(X_{n+1}, x^{n+1}, T_{n+1}) \) depends on \( X_{n+1}, x^{n+1}, T_{n+1} \).

It follows, as in the iid case, that
\[
 - \left| h(x^{n+1}, \xi_1^*(X_{n+1}, x^{n+1}, T_{n+1})) \right| U_{\theta_1^*}(X_{n+1}, x^{n+1}, T_{n+1}) \leq \frac{h^2(x^{n+1}, \xi_1^*(X_{n+1}, x^{n+1}, T_{n+1}))}{2} V_{\theta_1^*}(X_{n+1}, x^{n+1}, T_{n+1})
\]
\[
 \leq h(x^{n+1}, \xi_1^{**}(X_{n+1}, x^{n+1}, T_{n+1})) U_{\theta_1^{**}}(X_{n+1}, x^{n+1}, T_{n+1}) \leq \frac{h^2(x^{n+1}, \xi_1^{**}(X_{n+1}, x^{n+1}, T_{n+1}))}{2} V_{\theta_1^{**}}(X_{n+1}, x^{n+1}, T_{n+1})
\]
\[
 \leq \left| h(x^{n+1}, \xi_1^{**}(X_{n+1}, x^{n+1}, T_{n+1})) \right| U_{\theta_1^{**}}(X_{n+1}, x^{n+1}, T_{n+1}) \leq \frac{h^2(x^{n+1}, \xi_1^{**}(X_{n+1}, x^{n+1}, T_{n+1}))}{2} V_{\theta_1^{**}}(X_{n+1}, x^{n+1}, T_{n+1})
\]

Proceeding in the same way as in the iid case, and exploiting Lemma\(^8\) we obtain
\[
 \sup_n E \left( \log \hat{f}_{x^{n+1}, T_{n+1}}(X_{n+1}) \right)^2 < \infty. \quad (5.14)
\]

Thus, as in the iid set-up, (4.8) follows from (5.13) and (5.14).

We formalize the above arguments in the form of a theorem in our non-iid SDE set-up.

**Theorem 9** Assume the non-iid case of the SDE based random effects set-up and conditions (H1') – (H5'). Then (4.9) holds.

As in the previous cases, the following corollary provides asymptotic comparison between two models using Bayes factor in the non-iid SDE set-up.

**Corollary 10** For \( j = 1, 2 \), let \( R_{jn}(\theta_j) = \prod_{i=1}^n f_{\theta_j, x^i, T_i}(X_i) \), where \( \theta_1 \) and \( \theta_2 \) are two different finite sets of parameters, perhaps with different dimensionalities, associated with the two models to be compared. For \( j = 1, 2 \), let
\[
 I_{jn} = \int R_{jn}(\theta_j) \pi_j(d\theta_j),
\]
where \( \pi_j \) is the prior on \( \theta_j \). Let \( B_n = I_{1n}/I_{2n} \) as before. Assume the non-iid case of the SDE based random effects set-up and suppose that both the models satisfy (H1') – (H5'), and have the Kullback-
Leibler property with $\delta = \delta_1$ and $\delta = \delta_2$ respectively. Then

$$n^{-1} \log B_n \to \delta_2 - \delta_1,$$

almost surely.

6 Conclusion

In this article we have investigated the asymptotic theory of Bayes factors when the models are associated with systems of SDE’s. Such an undertaking, according to our knowledge, is a first-time effort which did not hitherto take place in the literature. Here we have addressed the Bayes factor asymptotics when the systems of SDE’s being compared are either iid or non-iid. In developing the asymptotic theory of Bayes factors in the non-iid situation, we proposed and proved general results on Bayes factor asymptotics in non-iid situations, which should be of independent interest.

Note that, our asymptotic theory for non-iid situations also readily extends to model comparison problems when one of the models is associated with an iid system of SDE’s and another with a non-iid system of SDE’s. For instance, if the true model is associated with an iid system, then $f_{0i} \equiv f_0 \equiv f_{\theta_0}$, and the rest of the theory remains the same as our non-iid theory of Bayes factors. The case when the other model is associated with an iid system is analogous.

Acknowledgments

The first author gratefully acknowledges her CSIR Fellowship, Govt. of India.

References

Berge, C. (1963). *Espaces Topologiques*. Dunod, Paris. Translated version: Topological Spaces. Translated by E. M. Patterson, Oliver and Boyd, Edinburgh.

Delattre, M., Genon-Catalot, V., and Samson, A. (2013). Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects. *Scandinavian Journal of Statistics*, 40, 322–343.

Maitra, T. and Bhattacharya, S. (2014a). On Asymptotics Related to Classical Inference in Stochastic Differential Equations with Random Effects. Submitted. Available at “http://arxiv.org/abs/1407.3968”.

Maitra, T. and Bhattacharya, S. (2014b). On Bayesian Asymptotics in Stochastic Differential Equations with Random Effects. Submitted. Available at “http://arxiv.org/abs/1407.3971”.

Mao, X. (2011). *Stochastic Differential Equations and Applications*. Woodhead Publishing India Private Limited, New Delhi, India.

Sivaganesan, S. and Lingham, R. T. (2002). On the Asymptotic of the Intrinsic and Fractional Bayes Factors for Testing Some Diffusion Models. *Annals of the Institute of Statistical Mathematics*, 54, 500–516.

Walker, S. G. (2004). Modern Bayesian Asymptotics. *Statistical Science*, 19, 111–117.

Walker, S. G., Damien, P., and Lenk, P. (2004). On Priors With a Kullback-Leibler Property. *Journal of the American Statistical Association*, 99, 404–408.