Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 3, 629–657
https://doi.org/10.1214/19-BJPS439

On classical and Bayesian asymptotics in state space stochastic differential equations

Trisha Maitra and Sourabh Bhattacharya

Indian Statistical Institute

Abstract. In this article, we investigate consistency and asymptotic normality of the maximum likelihood and the posterior distribution of the parameters in the context of state space stochastic differential equations (*SDE*s). We then extend our asymptotic theory to random effects models based on systems of state space *SDE*s, covering both independent and identical and independent but non-identical collections of state space *SDE*s. We also address asymptotic inference in the case of multidimensional linear random effects, and in situations where the data are available in discretized forms. It is important to note that asymptotic inference, either in the classical or in the Bayesian paradigm, has not been hitherto investigated in state space *SDE*s.

1 Introduction

State-space models are well known for their versatility in modeling complex dynamic systems in the context of discrete time, and have important applications in various disciplines like engineering, medicine, finance and statistics. As is also well known, most time series models of interest can be expressed in the form of state space models; see, for example, Durbin and Koopman (2001) and Shumway and Stoffer (2011). Discrete time state space models are characterized by a latent, unobserved stochastic process, $X = \{X(t); t = 0, 1, 2, \dots\}$ and another stochastic process $Y = \{Y(t); t = 0, 1, 2, \dots\}$, the distribution of which depends upon X. The observed time series data are modeled by the conditional distribution of Y given X, where X is assumed to have some specified distribution. An important special case of such discrete state space models is the hidden Markov model. Here X is assumed to be a Markov chain, the distribution of $Y(t)$ depends upon $X(t)$, and conditionally on $X(t)$'s, $Y(t)$'s are independent. Such models have important applications in engineering, finance, biology, statistics; see, for example, Elliott, Aggoun and Moore (1995) and Cappé, Moulines and Rydén (2005).

However, when the time is continuous, research on state space or hidden Markov models seem to be much scarce. Ideally, one should consider a pair of stochastic differential equations (*SDE*s) whose solutions would be the continuous time processes $Y = \{Y(t) : t \in [0, T]\}$ and $X = \{X(t) : t \in [0, T]\}$. In fact, the *SDE* with solution Y should depend upon X. Since the solutions of *SDE*s under general regularity conditions are Markov processes (see, for example, Mao (2011)), X would turn out to be a Markov process, and conditionally on X, Y would also be a Markov process. Thus, such an approach could be interpreted as continuous time versions of the traditional discrete time hidden Markov model based approach. Continuous time models closely resembling the above-mentioned type exists in the literature, but rather than estimating relevant parameters, filtering theory has been considered. For instance, Stratonovich (1968), Jazwinski (1970), Maybeck (1979, 1982), Särkkä (2006), Crisan and Rozovskii (2011) consider the filtering problem in state space *SDE*s of the following type:

$$dY(t) = b_Y\big(X(t), t\big)\,dt + dW_Y(t); \tag{1.1}$$

Key words and phrases. Asymptotic normality, Kullback–Leibler divergence, posterior consistency, random effects, state space stochastic differential equations, stochastic stability.

Received December 2017; accepted March 2019.

$$dX(t) = b_X\big(X(t), t\big)\,dt + \sigma_X\big(X(t), t\big)\,dW_X(t), \tag{1.2}$$

where W_Y and W_X are independent standard Wiener processes, b_Y, b_X are real-valued drift functions, and σ_X is the real-valued diffusion coefficient. The *SDE*s are assumed to satisfy the usual regularity conditions that guarantee existence of strong solutions; see, for example, Arnold (1974), Øksendal (2003), Mao (2011). The purpose of filtering theory is to compute the posterior distribution of the latent process conditional on the observed process. This can be obtained from the the continuous-time optimal filtering equation, which is, in fact, the Kushner–Stratonovich equation (Kushner (1964), Bucy (1965)). Note that (see Särkkä (2012), for example) it is possible to obtain the latter as continuous-time limits of the Bayesian filtering equations. The so-called Zakai equation (Zakai (1969)) provides a simplified form by removing the non-linearity in the Kushner–Stratonovich equation. In the special case of (1.1) and (1.2), with $b_Y(X(t), t) = L(t)X(t)$, $b_X(X(t), t) = H(t)X(t)$ and $\sigma_X(X(t), t) = \sigma_X(t)$, exact solution of the filtering problem, known as the Kalman–Bucy filter (Kalman and Bucy (1961)), can be obtained. In the non-linear cases various approximations are employed; see Crisan and Rozovskii (2011), Särkkä (2007), Särkkä and Sarmavuori (2013), among others.

In pharmocokinetic/pharmacodynamic contexts, the following type of model is regarded as the state space model, assuming $\{Y_1, \ldots, Y_n\}$ are observed at discrete times $\{t_1, \ldots, t_n\}$:

$$Y_j = b_Y(X_{t_j}, \theta) + \sigma_Y(X_{t_j}, \theta)\varepsilon_j; \quad \varepsilon_j \overset{\text{i.i.d.}}{\sim} N(0, 1); \tag{1.3}$$

$$dX(t) = b_X\big(X(t), t, \theta\big)\,dt + \sigma_X\big(X(t), \theta\big)\,dW_X(t), \tag{1.4}$$

where b_Y and σ_Y are appropriate real-valued functions, and θ denotes the set of relevant parameters. The standard choices of σ_Y are $\sigma_Y(x, \theta) = \sigma$ (homoscedastic model) and $\sigma_Y(x, \theta) = a + \sigma b_Y(x, \theta)$ (heterogeneous model), and b_Y is usually chosen to be a linear function. Thus, even though the latent process X is described as the solution of the *SDE* (1.4), the model for the (discretely) observed data is postulated to be arising from independent normal distributions, conditional on the discretized version of the diffusion process X. This simplifies inference proceedings to a large extent, particularly when the Markov transition model associated with (1.4) is available explicitly. Here we recall that under suitable regularity conditions, the solution of (1.4) is a continuous time Markov process (see, for example, Arnold (1974), Øksendal (2003), Mao (2011)). If the Markov transition model is not available in closed form, then various approximations are proposed in the literature to approximate the likelihood of θ, using which the *MLE* of θ or the posterior distribution of θ is obtained. Under special cases, for instance, when $\sigma_Y(x, \theta) = \sigma$, $b_Y(x, \theta) = b_\theta x$, $\sigma_X(x, \theta) = \sigma_\theta$, $b_X(x_t, t, \theta) = a_\theta x_t + c_\theta(t)$, an explicit form of the likelihood (based on discretization) is available, and the resulting *MLE* has been shown to be consistent and asymptotically normal by Favetto and Samson (2010), but in more general, non-linear situations, theoretical results do not seem to be available. A comprehensive account of the methods of approximating the *MLE* and posterior distribution of θ, with discussion of related computational issues and theoretical results, have been provided in Donnet and Samson (2013).

Our interest in this article is primarily the investigation of asymptotic parametric inference, as $T \to \infty$, from both classical and Bayesian perspectives, in the context of state space models where the models for the observed data as well as the latent process, are both described by *SDE*s. In Section 2.2, we show that such asymptotic parametric inference also addresses consistency of the so-called particle filtering problem associated with the joint posterior distribution of the parameters and the latent states $X(t)$ given the data $\{Y(s) : 0 \leq s \leq t\}$. For relatively recent research, works on the particle filtering problem in non-*SDE* setups, see, for example, Chopin, Jacob and Papaspiliopoulos (2013), Crisan and Miguez (2013), Urteaga, Bugallo and Djuric (2016), Martino et al. (2017).

In our knowledge, asymptotic inference in such models has not been hitherto investigated. In our proceedings we assume a somewhat generalized version of the state space *SDE*s described by (1.1) and (1.2) in that the drift function b_Y depends upon $Y(t)$, in addition to $X(t)$ and t; moreover, we assume that there is a diffusion coefficient $\sigma_Y(Y(t), X(t), t)$ associated with the Wiener process $W_Y(t)$ that drives the observational *SDE* (1.1); a practical instance of such a state space model in the case of bacterial growth can be found in Møller et al. (2012). We further assume that there is a common set of parameters θ associated with both the *SDE*s, which are of interest. In particular, we assume that there exist appropriate real-valued, known, functions for θ, $\psi_Y(\theta)$ and $\psi_X(\theta)$, such that the drift functions are $\psi_Y(\theta)b_Y(Y(t), X(t), t)$ and $\psi_X(\theta)b_X(X(t), t)$, respectively. In Section 4, we clarify that $\psi_Y(\theta)$ and $\psi_X(\theta)$ offers very general scope of parameterizations by mapping the perhaps high-dimensional (although finite-dimensional) quantity θ to appropriate real-valued functional forms composed of the elements of θ. We also assume that the diffusion coefficients of the respective *SDE*s are independent of θ. A key assumption in our approach to asymptotic investigation is that X is stochastically stable. In a nutshell, in this article, by stochastic stability of X we mean that

$$\left|x(t)\right| \le \xi\lambda(t) \quad \text{for all } t \ge 0, \tag{1.5}$$

almost surely, for all initial values $x(0) \in \mathbb{R}$, where $\lambda(t) \to 0$ as $t \to \infty$, and ξ is a non-negative, finite random variable depending upon $x(0)$. For comprehensive details regarding various versions of stochastic stability of solutions of *SDE*s, see Mao (2011).

It is to be noted that our model clearly corresponds to a dependent setup, and establishment of asymptotic results are therefore can not be achieved by the state-of-the-art methods that typically deal with at least independent situations. For Bayesian asymptotics, we find the consistency results of Shalizi (2009) and the general result on posterior asymptotic normality of Schervish (1995) useful for our purpose, while for classical asymptotics we obtain a suitable asymptotic approximation to the target log-likelihood, which helped us establish strong consistency, as well as asymptotic normality of the *MLE*.

Once we establish classical and Bayesian asymptotic results associated with our state space *SDE* model, we then extend our model to random effects state space model (see Delattre, Genon-Catalot and Samson (2013), for instance, for *SDE* based random effects model), where we model each time series data available on n individuals using our state space model, assuming that the effects $\psi_{Y_i}(\theta)$ and $\psi_{X_i}(\theta)$ for individual i are parameterized by θ, which is the parameter of interest. From the classical point of view, this is not a random effects model technically since θ is treated as a fixed quantity, but from the Bayesian viewpoint, a prior on θ renders the effects random. Slightly abusing terminology for the sake of convenience, we continue to call the model random effects stochastic *SDE*, from both classical and Bayesian perspectives. Under such random effects *SDE* model we seek asymptotic classical and Bayesian inference on θ as both number of individuals, n, and the domain of observations $[0, T_i]$; $i = 1, \dots, n$ increase indefinitely. For our purpose we assume $T_i = T$ for each i. Here we remark that Donnet and Samson (2013) discuss population *SDE* models with measurement errors; see also Overgaard et al. (2005), Donnet and Samson (2008), Yan et al. (2014), Leander et al. (2015); for the ith individual such models are of the same form as (1.3) and (1.4), but specifics depending upon i, and with θ replaced with ϕ_i, where $\{\phi_1, \dots, \phi_n\}$ are independently and identically distributed with some distribution with parameter θ, say, which is one of the parameters of interest. This is a genuine random effects model unlike ours, but here only the latent process X is based upon *SDE*. Theoretical results do not exist for this setup; see Donnet and Samson (2013). On the other hand, even though our random effects state space *SDE* model is completely based upon *SDE*s, the simplified form of the effects, parameterized by a common θ, enables us to obtain desired asymptotic results for both classical and Bayesian paradigms. Indeed, in our case it is certainly possible to postulate a genuine random effects state space *SDE* model by replacing $\psi_{Y_i}(\theta)$ and $\psi_{X_i}(\theta)$ with i.i.d. random effects ϕ_{Y_i} and ϕ_{X_i}, having distributions parameterized by quantities of inferential interest θ_Y and θ_X, say, but in this setup complications arise regarding handling the observed integrated likelihood and its associated bounds, which does not assist in our asymptotic investigations.

Discretization of our state space *SDE* models is essential for practical applications such as in fields of pharmacokinetics/pharmacodynamics, where continuous time data are usually unavailable. We show in the supplement (Maitra and Bhattacharya (2019)) that the same asymptotic results go through in discretized situations.

In our proceedings with each setup, we first investigate Bayesian consistency, then consistency and asymptotic normality of the *MLE*, and finally asymptotic posterior normality. One reason behind this sequence is that the proofs of the results on posterior normality depend upon the proofs of the results of consistency and asymptotic normality of *MLE*, which, in turn, depend upon the proofs associated with Bayesian posterior consistency. Moreover, adhering to this sequence allows us to introduce the assumptions in a sequential manner, so that an overall logical order could be maintained throughout the paper.

The rest of our article is organized as follows. In Section 2, we introduce our state space *SDE* model and provide an overview of the asymptotic results in Section 3. We list the various sets of assumptions including stochastic stability of the solution of the latent *SDE*, in Section 4. Development of the asymptotic theory requires asymptotic approximation of the true and observed likelihoods. Such asymptotic approximations are developed in Section 5, under suitable regularity conditions. Next, in Section 6, with further regularity conditions, we prove posterior convergence of θ by proving validity of the conditions of Shalizi stated formally for our state space *SDE* setup in Section S-1 of the supplement (Maitra and Bhattacharya (2019)). We prove strong consistency and asymptotic normality of the *MLE* in Section 7, under further extra assumptions. With a few more regularity conditions, In Section 8 we establish asymptotic posterior normality of θ. We introduce random effects state space *SDE* models in Section 9 and provide a briefing of the asymptotic results, with the details in Section S-5 of the supplement (Maitra and Bhattacharya (2019)). Finally, in Section 10 we provide a brief summary of our work, discuss some key issues, and identify future research agenda. The extension of our theory for state space *SDE* models with multidimensional linear random effects and in the case of discretized data are discussed, respectively, in Sections S-6 and S-7 of the supplement (Maitra and Bhattacharya (2019)).

2 State space *SDE*

2.1 True and postulated state space *SDE* models

First, consider the following "true" state space *SDE*:

$$dY(t) = \phi_{Y,0} b_Y\big(Y(t), X(t), t\big)\,dt + \sigma_Y\big(Y(t), X(t), t\big)\,dW_Y(t); \tag{2.1}$$

$$dX(t) = \phi_{X,0} b_X\big(X(t), t\big)\,dt + \sigma_X\big(X(t), t\big)\,dW_X(t), \tag{2.2}$$

for $t \in [0, b_T]$, where $b_T \to \infty$, as $T \to \infty$. The first *SDE*, namely, (2.1) is the true observational *SDE* and is associated with the observed data. The second *SDE* (2.2) is the true evolutionary, unobservable *SDE*. In the above two equations, we assume that $\phi_{Y,0}$ and $\phi_{X,0}$ are both explained by a "true" set of parameters θ_0, through known but perhaps different functions of θ_0. In other words, we assume that $\phi_{Y,0} = \psi_Y(\theta_0)$ and $\phi_{X,0} = \psi_X(\theta_0)$, where ψ_Y and ψ_X are known functions. Note that this is a general formulation, where we allow the possibility $\theta_0 = (\theta_{Y,0}, \theta_{X,0})$ and choice of ψ_Y and ψ_X such that $\psi_Y(\theta_0) = \theta_{Y,0}$ and $\psi_X(\theta_0) = \theta_{X,0}$, for scalars $\theta_{Y,0}$ and $\theta_{X,0}$. In this instance, the observational and evolutionary *SDE*s have their own sets of parameters. We also allow common subsets of the parameter vector θ_0 to feature in the two *SDE*s. For instance, $\psi_Y(\theta_0) = \theta_{Y,0} + \theta_{X,0}$ and $\psi_X(\theta_0) = \theta_{X,0}$. Indeed, θ_0 can be any finite-dimensional vector, appropriately mapped to the real line by ψ_Y and ψ_X. We wish to learn about the set of parameters θ_0, which would enable learning about $\phi_{Y,0}$ and $\phi_{X,0}$ simultaneously. For our purpose, we assume that $(\psi_Y(\theta), \psi_X(\theta))$ is identifiable in θ, that is, $(\psi_Y(\theta_1), \psi_X(\theta_1)) = (\psi_Y(\theta_2), \psi_X(\theta_2))$ implies $\theta_1 = \theta_2$.

Our modeled state space *SDE* is analogously given, for $t \in [0, b_T]$ by:

$$dY(t) = \phi_Y b_Y\big(Y(t), X(t), t\big)\,dt + \sigma_Y\big(Y(t), X(t), t\big)\,dW_Y(t); \tag{2.3}$$

$$dX(t) = \phi_X b_X\big(X(t), t\big)\,dt + \sigma_X\big(X(t), t\big)\,dW_X(t), \tag{2.4}$$

where $\phi_Y = \psi_Y(\theta)$ and $\phi_X = \psi_X(\theta)$.

Throughout, we assume that the initial values associated with the *SDE*s (2.1), (2.2), (2.3) and (2.4), are non-random. It is worth mentioning in this context that for stochastic stability it is enough to assume non-randomness of the initial value; see Mao ((2011), page 111), for a proof of this.

We wish to establish consistency and asymptotic normality of the maximum likelihood estimator (*MLE*) and the posterior distribution of θ, as $T \to \infty$. For technical reasons, we shall consider the likelihood for $t \in [a_T, b_T]$, where $a_T \to \infty$ and $(b_T - a_T) \to \infty$, as $T \to \infty$. In particular, we assume that $(b_T - a_T) \geq T$.

2.2 Connection of parametric asymptotic inference with the asymptotics of the particle filtering problem

As already mentioned, in this article we focus on classical and Bayesian asymptotic inference on the parameter θ. However, such asymptotic parametric inference automatically leads to asymptotic inference regarding the particle filtering problem. To clarify, first let $\mathcal{Y}_t = \{Y(s) : 0 \leq s \leq t\}$, for $t \in [0, b_T]$, and let $\hat{\theta}_T$ denote the *MLE* of θ or the posterior expectation of θ, given the data $\mathcal{Y}_T$. Then provided that $\hat{\theta}_T \to \theta_0$ almost surely (or in probability), for each $t \in [0, b_T]$, the posterior distribution $\pi(X(t)|\hat{\theta}_T, \mathcal{Y}_t) \to \pi(X(t)|\theta_0, \mathcal{Y}_t)$, as $T \to \infty$, almost surely (or in probability), if $\pi(X(t)|\theta, \mathcal{Y}_t)$ is continuous in θ. As a simple example, let us assume that $b_Y(Y(t), X(t), t) = L(t)X(t)$, $b_X(X(t), t) = H(t)X(t)$, $\sigma_Y(Y(t), X(t), t) \equiv 1$ and $\sigma_X(X(t), t) = \sigma_X(t)$. Also, let us assume that $\psi_Y(\theta)$ and $\psi_X(\theta)$ are continuous in θ. Then the Kalman–Bucy filter ensures that $\pi(X(t)|\theta, \mathcal{Y}_t)$ is a Gaussian density with mean and variances depending upon t, and the density is continuous in θ. Letting $\mathcal{X}_t = \{X(s) : 0 \leq s \leq b_T\}$, we similarly have $\pi(\mathcal{X}_t|\hat{\theta}_T, \mathcal{Y}_t) \to \pi(\mathcal{X}_t|\theta_0, \mathcal{Y}_t)$, as $T \to \infty$, almost surely (or in probability).

3 A brief overview of the main asymptotic results

3.1 Posterior convergence of $\boldsymbol{\theta}$

Our main result on posterior convergence of θ is based on verification of a general posterior convergence result of Shalizi (2009), which amounts to validating seven regularity conditions required by Shalizi's result, which we denote by (A1)–(A7). We present the assumptions and the result of Shalizi in Section S-1 of the supplement (Maitra and Bhattacharya (2019)). The most essential notions, the key assumption, and our main result on posterior convergence with a brief sketch of the proof utilizing the key assumption of Shalizi, are presented below.

Let $\mathcal{F}_T = \sigma(\{Y(s) : s \in [a_T, b_T]\})$ denote the σ-algebra generated by $\{Y(s) : s \in [a_T, b_T]\}$. Let $\mathcal{T}$ denote the σ-algebra associated with the d (≥ 1)-dimensional parameter space Θ.

Let $p_T(\theta_0)$ denote the marginal likelihood of $\{Y(t) : t \in [a_T, b_T]\}$ of the true model (2.1) and (2.2). Also, let $L_T(\theta)$ be the modeled likelihood of $\{Y(t) : t \in [a_T, b_T]\}$ of the postulated model (2.3) and (2.4). We denote $\frac{L_T(\theta)}{p_T(\theta_0)}$ by $R_T(\theta)$. For every $\theta \in \Theta$, the Kullback–Leibler divergence rate is given by

$$h(\theta) = \lim_{T\to\infty} \frac{1}{b_T - a_T} E_{\theta_0}\bigl(-\log R_T(\theta)\bigr),$$

where E_{θ_0} denotes the expectation is with respect to the true model.

For $A \subseteq \Theta$, let

$$h(A) = \operatorname*{ess\,inf}_{\theta\in A} h(\theta);$$

$$J(\theta) = h(\theta) - h(\Theta);$$

$$J(A) = \operatorname*{ess\,inf}_{\theta\in A} J(\theta).$$

The above essential infimums are with respect to the prior π assigned for θ.

With the above notions, our posterior convergence results are summarized by Theorem 1.

Theorem 1. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *For the prior π on θ, consider any set $A \in \mathcal{T}$ with $\pi(A) > 0$ and $h(A) > h(\Theta)$. Then, under suitable assumptions, almost surely,*

$$\lim_{T\to\infty} \pi(A|\mathcal{F}_T) = 0.$$

Moreover, if the set A satisfies another technical condition, then almost surely,

$$\lim_{T\to\infty} \frac{1}{b_T - a_T} \log \pi(A|\mathcal{F}_T) = -J(A).$$

Sketch of the proof. The proof follows by verifying the seven assumptions of Shalizi, which are shown to hold under appropriate conditions. The most important result guiding posterior convergence is the asymptotic equipartition property, which is given in this case by

$$\begin{aligned}
&\frac{1}{b_T - a_T} \log R_T(\theta) \\
&\quad \to -\frac{1}{2}\bigl[K_Y(\phi_Y - \phi_{Y,0})^2 + K_X(\phi_X - \phi_{X,0})^2 + K_X\bigl(\phi_{X,0}^2 - \phi_X^2\bigr)\bigr] \\
&\quad = -h(\theta),
\end{aligned}$$

where

$$\begin{aligned}
h(\theta) &= \frac{1}{2}\bigl[K_Y(\phi_Y - \phi_{Y,0})^2 + K_X(\phi_X - \phi_{X,0})^2 + K_X\bigl(\phi_{X,0}^2 - \phi_X^2\bigr)\bigr] \\
&= \frac{1}{2}\bigl[K_Y\bigl(\psi_Y(\theta) - \psi_Y(\theta_0)\bigr)^2 + K_X\bigl(\psi_X(\theta) - \psi_X(\theta_0)\bigr)^2 \\
&\quad + K_X\bigl(\psi_X^2(\theta_0) - \psi_X^2(\theta)\bigr)\bigr].
\end{aligned}$$

In the above, K_X (> 0) and K_Y (> 0) are the limits of the bounds of $b_Y^2(y,x,t)/\sigma_Y^2(y,x,t)$ and $b_X^2(x,t)/\sigma_X^2(x,t)$, respectively, as $T \to \infty$.

This result is achieved using the following approximations proved subsequently: $p_T(\theta_0) \overset{\text{a.s.}}{\sim} \hat{p}_T(\theta_0)$ and $L_T(\theta) \overset{\text{a.s.}}{\sim} \hat{L}_T(\theta)$, where

$$\begin{aligned}
\hat{p}_T(\theta_0) = \exp\biggl(&\frac{(b_T - a_T)K_Y\phi_{Y,0}^2}{2} + \phi_{Y,0}\sqrt{K_Y}\bigl(W_Y(b_T) - W_Y(a_T)\bigr) \\
&+ (b_T - a_T)K_X\phi_{X,0}^2\biggr),
\end{aligned}$$

$$\hat{L}_T(\theta) = \exp\Big((b_T - a_T)K_Y\phi_Y\phi_{Y,0} + \phi_Y\sqrt{K_Y}\big(W_Y(b_T) - W_Y(a_T)\big) - \frac{(b_T - a_T)K_Y\phi_Y^2}{2} + (b_T - a_T)K_X\phi_X\phi_{X,0}\Big),$$

and then noting that, as $T \to \infty$,

$$\begin{aligned}
&\frac{1}{b_T - a_T}\log R_T(\theta) \\
&\quad = \frac{1}{b_T - a_T}\log\left(\frac{L_T(\theta)}{p_T(\theta_0)}\right) \\
&\quad \overset{\text{a.s.}}{\sim} -\frac{K_Y}{2}(\phi_Y - \phi_{Y,0})^2 + \sqrt{K_Y}(\phi_Y - \phi_{Y,0})\frac{(W_Y(b_T) - W_Y(a_T))}{b_T - a_T} \\
&\qquad - \frac{K_X}{2}(\phi_X - \phi_{X,0})^2 + \frac{K_X}{2}(\phi_X^2 - \phi_{X,0}^2) \\
&\quad \xrightarrow{\text{a.s.}} -\frac{1}{2}\big[K_Y(\phi_Y - \phi_{Y,0})^2 + K_X(\phi_X - \phi_{X,0})^2 + K_X(\phi_{X,0}^2 - \phi_X^2)\big].
\end{aligned}$$

It is important to note that compactness of the parameter space Θ is not necessary for Theorem 1 to hold. Instead, we constructed appropriate "sieves" of the form $\mathcal{G}_T = \{\theta : |\psi_Y(\theta)| \leq \exp(\beta(b_T - a_T))\}$ with $\beta > 2h(\Theta)$ that are compact for each T and increasing in T and such the prior probability of the complement $\mathcal{G}_T^c$ is exponentially small, and satisfies some other technical conditions that essentially guarantee posterior convergence, along with the asymptotic equipartition property. □

Remark 2. In particular, let $A_\varepsilon = \{\theta \in \Theta : h(\theta) > h(\Theta) + \varepsilon\}$, for $\varepsilon > 0$. Then note that $h(A_\varepsilon) > h(\Theta)$, for any $\varepsilon > 0$. Let $\pi(A_\varepsilon) > 0$. Then by the first part of Theorem 1, $\pi(A_\varepsilon|\mathcal{F}_T) \to 0$, almost surely, as $T \to \infty$, for any $\varepsilon > 0$. It is also important to note that if θ_0 belongs to the support of the prior on Θ, then $h(\Theta) = 0$. In this case, the posterior probability of $A_\varepsilon = \{\theta \in \Theta : h(\theta) > \varepsilon\}$ tends to zero almost surely, for any $\varepsilon > 0$.

3.2 Consistency of the *MLE* of θ

Let $\theta \in \Theta \subseteq \mathbb{R}^d$, where Θ is the $d(\geq 1)$-dimensional, compact parameter space. Our main result on consistency of the *MLE* of θ can be formalized as the following theorem.

Theorem 3. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Then under appropriate regularity conditions the MLE $\hat{\theta}_T$ of θ is strongly consistent in the sense that $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta_0$.*

Sketch of the proof. Identifiability of the model and uniqueness of the *MLE* follow from our assumptions. To prove strong consistency of the *MLE*, we first note that the *MLE* can be approximated by maximizing the function

$$\tilde{g}_T(\theta) = g_{Y,T}(\theta) + g_{X,T}(\theta)$$

with respect to θ, where

$$g_{Y,T}(\theta) = -\frac{K_Y}{2}\big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)^2 + \sqrt{K_Y}\big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T};$$

$$g_{X,T}(\theta) = -\frac{K_X}{2}\left(\psi_X(\theta) - \psi_X(\theta_0)\right)^2 + \frac{K_X}{2}\left(\psi_X^2(\theta) - \psi_X^2(\theta_0)\right).$$

Letting $\hat{\theta}_T$ denote the *MLE*, note that

$$0 = \tilde{g}'_T(\hat{\theta}_T) = \tilde{g}'_T(\theta_0) + \tilde{g}''_T(\theta^*_T)(\hat{\theta}_T - \theta_0),$$

where θ^*_T lies between θ_0 and $\hat{\theta}_T$. Since $\tilde{g}'_T(\theta_0) \xrightarrow{\text{a.s.}} 0$ as $T \to \infty$ and since $\tilde{g}''_T(\theta^*_T)$ is positive definite for $T \geq 1$ under appropriate assumptions, it holds that $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta_0$, as $T \to \infty$. □

Remark 4. Note that compactness of Θ is not necessary for Bayesian consistency, in contrast with consistency of the *MLE*.

3.3 Asymptotic normality of the *MLE* of $\boldsymbol{\theta}$

For asymptotic normality of the *MLE* of θ, the result is summarized by the following theorem.

Theorem 5. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Then under suitable assumptions the MLE of θ is asymptotically normal in the sense that $\sqrt{b_T - a_T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} N_d(0, \mathcal{I}^{-1}(\theta_0))$. Here $\mathcal{I}(\theta)$ is the matrix with (j,k)th element given by*

$$\{\mathcal{I}(\theta)\}_{jk} = K_Y\left[\frac{\partial\psi_Y(\theta)}{\partial\theta_j}\frac{\partial\psi_Y(\theta)}{\partial\theta_k}\right].$$

Sketch of the proof. Asymptotic normality follows easily from the above developments on consistency of *MLE*, and the fact that $\theta^*_T \xrightarrow{\text{a.s.}} \theta_0$, and $\frac{W_Y(b_T)-W_Y(a_T)}{b_T-a_T} \xrightarrow{\text{a.s.}} 0$, as $T \to \infty$.

Observe that $\{\mathcal{I}(\theta_0)\}_{jk}$ is the covariance between the jth and the kth components of $\sqrt{b_T - a_T}\,\tilde{g}'_T(\theta_0)$, and so $\mathcal{I}(\theta_0)$ is non-negative definite. □

3.4 Asymptotic posterior normality of $\boldsymbol{\theta}$

We prove posterior normality of θ by verifying the seven regularity conditions of Theorem 7.102 of Schervish (1995).

Theorem 6. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Then denoting $\Psi_T = (b_T - a_T)^{\frac{1}{2}}\mathcal{I}^{\frac{1}{2}}(\theta_0)(\theta - \hat{\theta}_T)$, for each compact subset B of $\mathbb{R}^d$ and each $\varepsilon > 0$, the following holds under appropriate assumptions*:

$$\lim_{T\to\infty} P_{\theta_0}\left(\sup_{\Psi_T \in B}\left|\pi(\Psi_T|\mathcal{F}_T) - \varrho(\Psi_T)\right| > \varepsilon\right) = 0,$$

where $\varrho(\cdot)$ denotes the density of the standard normal distribution.

Sketch of the proof. Here we assume that Θ is compact which enables us to uniformly approximate $\frac{1}{b_T-a_T}\log R_T(\theta)$ by $g_{Y,T}(\theta) + g_{X,T}(\theta)$ for $\theta \in \Theta$. Hence, $\frac{1}{b_T-a_T}\log L_T(\theta)$ can be uniformly approximated by $\frac{1}{b_T-a_T}\tilde{\ell}_T(\theta) = g_{Y,T}(\theta) + g_{X,T}(\theta) + \frac{1}{b_T-a_T}\log p_T(\theta_0)$, for $\theta \in \Theta$. This is the key idea, and by working with the first three differentials of $\tilde{\ell}_T(\theta)$, in conjunction with Taylor's series expansion and our proven result that $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta_0$, the seven regularity conditions of Theorem 7.102 of Schervish (1995) are relatively straightforward to verify. □

4 Regularity conditions

4.1 Assumptions regarding b_Y and σ_Y

(H1) For every $T > 0$, and integer $\eta \geq 1$, given any x, there exists a positive constant $K_{Y,x,T,\eta}$ such that for all $t \in [0, b_T]$ and all (y_1, y_2) with $\max\{y_1, y_2\} \leq \eta$,

$$\max\{[b_Y(y_1, x, t) - b_Y(y_2, x, t)]^2, [\sigma_Y(y_1, x, t) - \sigma_Y(y_2, x, t)]^2\} \\ \leq K_{Y,x,T,\eta}|y_1 - y_2|^2.$$

(H2) For every $T > 0$, given any x, there exists a positive constant $K_{x,T}$ such that for all $(y, t) \in \mathbb{R} \times [0, T]$,

$$\max\{b_Y^2(y, x, t), \sigma_Y^2(y, x, t)\} \leq K_{x,T}(1 + y^2).$$

(H3) For every $T > 0$, there exist positive constants $K_{Y,1,T}$, $K_{Y,2,T}$, $\alpha_{Y,1}$, $\alpha_{Y,2}$ such that for all $(x, t) \in \mathbb{R} \times [0, b_T]$,

$$K_{Y,1,T}(1 - \alpha_{Y,1}x^2) \leq \frac{b_Y^2(y, x, t)}{\sigma_Y^2(y, x, t)} \leq K_{Y,2,T}(1 + \alpha_{Y,2}x^2),$$

where $K_{Y,1,T} \to K_Y$ and $K_{Y,2,T} \to K_Y$ as $T \to \infty$; K_Y being a positive constant. We further assume that for $j = 1, 2$, $(b_T - a_T)|K_{Y,j,T} - K_Y| \to 0$, as $T \to \infty$.

In (H3), we have assumed that the bounds of $\frac{b_Y^2(y,x,t)}{\sigma_Y^2(y,x,t)}$ do not depend upon y, which is somewhat restrictive. Dependence of the bounds on y can be insisted upon, but at the cost of the assumption of stochastic stability of Y in addition to that of X. See Section 10 for details regarding the modified assumption. All our results remain intact under the modified assumption. It is also important to clarify that the lower bound in (H3), when utilized in our *SDE* context, becomes non-negative after possibly a few time steps, thanks to the stochastic stability assumption which ensures (1.5).

4.2 Assumptions regarding b_X and σ_X

(H4) $b_X(0, t) = 0 = \sigma_X(0, t)$ for all $t \geq 0$.

(H5) For every $T > 0$, and integer $\eta \geq 1$, there exists a positive constant $K_{T,\eta}$ such that for all $t \in [0, b_T]$ and all (x_1, x_2) with $\max\{x_1, x_2\} \leq \eta$,

$$\max\{[b_X(x_1, t) - b_X(x_2, t)]^2, [\sigma_X(x_1, t) - \sigma_X(x_2, t)]^2\} \leq K_{T,\eta}|x_1 - x_2|^2.$$

(H6) For every $T > 0$, there exists a positive constant K_T such that for all $(x, t) \in \mathbb{R} \times [0, b_T]$,

$$\max\{b_X^2(x, t), \sigma_X^2(x, t)\} \leq K_T(1 + x^2).$$

(H7) For every $T > 0$, there exist positive constants $K_{X,1,T}$, $K_{X,2,T}$, $\alpha_{X,1}$, $\alpha_{X,2}$ such that for all $(x, t) \in \mathbb{R} \times [0, b_T]$,

$$K_{X,1,T}(1 - \alpha_{X,1}x^2) \leq \frac{b_X^2(x, t)}{\sigma_X^2(x, t)} \leq K_{X,2,T}(1 + \alpha_{X,2}x^2),$$

where $K_{X,1,T} \to K_X$ and $K_{X,2,T} \to K_X$, as $T \to \infty$; K_X being a positive constant. We also assume that for $j = 1, 2$, $(b_T - a_T)|K_{X,j,T} - K_X| \to 0$, as $T \to \infty$.

4.3 Further assumptions ensuring almost sure stochastic stability of $X(t)$

Let $\mathcal{C}$ denote the family of all continuous non-decreasing functions $f : \mathbb{R}^+ \mapsto \mathbb{R}^+$ such that $f(0) = 0$ and $f(r) > 0$ when $r > 0$.

Let $S_h = \{x \in \mathbb{R} : |x| < h\}$ and $\mathbb{C}(S_h \times [0, \infty); \mathbb{R}^+)$ denote the family of all continuous functions $V(x,t)$ from $S_h \times [0, \infty)$ to $\mathbb{R}^+$ with continuous first partial derivatives with respect to x and t. Also, let $\mathfrak{C}(S_h \times [0, \infty); \mathbb{R}^+)$, where $0 < h \leq \infty$, denote the family of non-negative functions $V(x,t)$ defined on $S_h \times \mathbb{R}^+$ such that they are continuously twice differentiable in x and once in t. Let

$$LV(x,t) = V_t(x,t) + V_x(x,t) b_X(x,t) + \frac{1}{2}\sigma_X^2(x,t) V_{xx}(x,t),$$

where $V_t = \frac{\partial V}{\partial t}$, $V_x = \frac{\partial V}{\partial x}$, and $V_{xx} = \frac{\partial^2 V}{\partial x^2}$.

With these definitions and notations, we now make the following assumption:

(H8) Let $p > 0$ and let there exist a function $V \in \mathfrak{C}(S_h \times [0, \infty); \mathbb{R}^+)$, a continuous non-decreasing function $\gamma : \mathbb{R}^+ \mapsto \mathbb{R}^+$ such that $\gamma(t) \to \infty$ as $t \to \infty$, and a continuous function $\breve{\eta} : \mathbb{R}^+ \mapsto \mathbb{R}^+$ such that $\int_0^\infty \breve{\eta}(t) < \infty$. Assume that for $x \neq 0$, $t \geq 0$,

$$\gamma(t)|x|^p \leq V(x,t) \quad \text{and} \quad LV(x,t) \leq \breve{\eta}(t).$$

Thanks to Theorem 6.2 of Mao ((2011), page 145), assumption (H8) ensures that stochastic stability of X of the form $|x(t)| \leq \xi\lambda(t)$for all$t \geq 0$ holds almost surely, for all initial values $x(0) \in \mathbb{R}$ with

$$\lambda(t) = [\gamma(t)]^{-\frac{1}{p}},$$

where ξ is a non-negative, finite random variable depending upon $x(0)$.

5 Asymptotic approximations of the true and modeled likelihoods of the state space *SDE*s

Let us define

$$v_{Y|X,T} = \int_{a_T}^{b_T} \frac{b_Y^2(Y(s), X(s), s)}{\sigma_Y^2(Y(s), X(s), s)} ds; \tag{5.1}$$

$$u_{Y|X,T} = \int_{a_T}^{b_T} \frac{b_Y(Y(s), X(s), s)}{\sigma_Y^2(Y(s), X(s), s)} dY(s); \tag{5.2}$$

$$v_{X,T} = \int_{a_T}^{b_T} \frac{b_X^2(X(s), s)}{\sigma_X^2(X(s), s)} ds; \tag{5.3}$$

$$u_{X,T} = \int_{a_T}^{b_T} \frac{b_X(X(s), s)}{\sigma_X^2(X(s), s)} dX(s). \tag{5.4}$$

Due to (H3) and (H7), the following hold:

$$K_{Y,\xi,T,1} \leq v_{Y|X,T} \leq K_{Y,\xi,T,2}; \tag{5.5}$$

$$K_{X,\xi,T,1} \leq v_{X,T} \leq K_{X,\xi,T,2}, \tag{5.6}$$

where

$$K_{Y,\xi,T,1} = K_{Y,1,T}\left((b_T - a_T) - \alpha_{Y,1}\xi^2 \int_{a_T}^{b_T} \lambda^2(s)\, ds\right); \tag{5.7}$$

$$K_{Y,\xi,T,2} = K_{Y,2,T}\left((b_T - a_T) + \alpha_{Y,2}\xi^2 \int_{a_T}^{b_T} \lambda^2(s)\,ds\right); \tag{5.8}$$

$$K_{X,\xi,T,1} = K_{X,1,T}\left((b_T - a_T) - \alpha_{X,1}\xi^2 \int_{a_T}^{b_T} \lambda^2(s)\,ds\right); \tag{5.9}$$

$$K_{X,\xi,T,2} = K_{X,2,T}\left((b_T - a_T) + \alpha_{X,2}\xi^2 \int_{a_T}^{b_T} \lambda^2(s)\,ds\right). \tag{5.10}$$

To proceed, we shall make use of the following relationships between $u_{Y|X,T}$, $v_{Y|X,T}$ and $u_{X,T}$, $v_{X,T}$ under the true state space *SDE* model described by (2.1) and (2.2):

$$u_{Y|X,T} = \phi_{Y,0} v_{Y|X,T} + \int_{a_T}^{b_T} \frac{b_Y(Y(s), X(s), s)}{\sigma_Y(Y(s), X(s), s)}\,dW_Y(s); \tag{5.11}$$

$$u_{X,T} = \phi_{X,0} v_{X,T} + \int_{a_T}^{b_T} \frac{b_X(X(s), s)}{\sigma_X(X(s), s)}\,dW_X(s). \tag{5.12}$$

Let

$$I_{Y,X,T} = \int_{a_T}^{b_T} \frac{b_Y(Y(s), X(s), s)}{\sigma_Y(Y(s), X(s), s)}\,dW_Y(s);$$

$$I_{X,T} = \int_{a_T}^{b_T} \frac{b_X(X(s), s)}{\sigma_X(X(s), s)}\,dW_X(s).$$

Because of (5.5), (5.6), (5.11) and (5.12) the following hold:

$$\begin{aligned} \phi_{Y,0} K_{Y,\xi,T,1} + I_{Y,X,T} &\le u_{Y|X,T} \le \phi_{Y,0} K_{Y,\xi,T,2} + I_{Y,X,T}; \\ \phi_{X,0} K_{X,\xi,T,1} + I_{X,T} &\le u_{X,T} \le \phi_{X,0} K_{X,\xi,T,2} + I_{X,T}. \end{aligned} \tag{5.13}$$

5.1 True likelihood and its asymptotic approximation

First note that $\exp(\phi_{Y,0} u_{Y|X,T} - \frac{\phi_{Y,0}^2}{2} v_{Y|X,T})$ is the conditional density of Y given X, with respect to $Q_{T,Y|X}$, the probability measure associated with (2.1) on $[a_T, b_T]$, assuming null drift. Also, $\exp(\phi_{X,0} u_{X,T} - \frac{\phi_{X,0}^2}{2} v_{X,T})$ is the marginal density of X with respect to $Q_{T,X}$, the probability measure associated with the latent state *SDE* (2.2) on $[a_T, b_T]$, but assuming null drift. These are standard results; see for example, Lipster and Shiryaev (2001), Øksendal (2003), Delattre, Genon-Catalot and Samson (2013).

It then follows that the marginal likelihood under the true model (2.1) and (2.2) is the marginal density of $\{Y(t) : t \in [a_T, b_T]\}$, given by

$$\begin{aligned} p_T(\theta_0) &= \int \exp\left(\phi_{Y,0} u_{Y|X,T} - \frac{\phi_{Y,0}^2}{2} v_{Y|X,T}\right) \\ &\qquad \times \exp\left(\phi_{X,0} u_{X,T} - \frac{\phi_{X,0}^2}{2} v_{X,T}\right) dQ_{T,X} \\ &= E_{T,X}\left[\exp\left(\phi_{Y,0} u_{Y|X,T} - \frac{\phi_{Y,0}^2}{2} v_{Y|X,T}\right)\right. \\ &\qquad \left.\times \exp\left(\phi_{X,0} u_{X,T} - \frac{\phi_{X,0}^2}{2} v_{X,T}\right)\right], \end{aligned} \tag{5.14}$$

where $E_{T,X}$ denotes expectation with respect to $Q_{T,X}$. The following lemma proved in supplement (Maitra and Bhattacharya (2019)) formalizes the dominating measure with respect to which $p_T(\theta_0)$ is the Radon–Nikodym derivative.

Lemma 7. *The likelihood given by (5.14) is the density of* $\{Y(t) : t \in [a_T, b_T]\}$ *with respect to* $Q_{T,Y}$*, where for any relevant measurable set* A,

$$Q_{T,Y}(A) = \int_{\mathfrak{X}_T} dQ_{T,Y|X}(A)\, dQ_{T,X} = \int_A \int_{\mathfrak{X}_T} dQ_{T,Y|X}\, dQ_{T,X}.$$

In the above, $\mathfrak{X}_T$ *stands for the sample space of* $\{X(t) : t \in [a_T, b_T]\}$.

It is important to remark that our likelihood (5.14) is of a very general form and does not usually admit a closed form expression, but this is not at all a requirement for our asymptotic purpose. Closed form expressions may be necessary when it is of interest to directly maximize the likelihood with respect to the parameters, and in such cases, more stringent assumptions regarding the *SDE*s are necessary. See, for example, Frydman and Lakner (2003); see also Kailath and Zakai (1971). Also, observe that our dominating measure $Q_{T,Y}$ is not the Wiener measure, unlike the aforementioned papers, albeit it reduces to the Wiener measure if $\sigma_Y \equiv 1$ and $\sigma_X \equiv 1$.

5.1.1 *Asymptotic approximation of* $p_T(\theta_0)$. Using (5.5) and (5.13) we obtain

$$B_{L,T}(\theta_0) \le p_T(\theta_0) \le B_{U,T}(\theta_0),$$

where

$$B_{L,T}(\theta_0) = E_{T,X}\big(Z_{L,T,\theta_0}(X)\big); \tag{5.15}$$

$$B_{U,T}(\theta_0) = E_{T,X}\big(Z_{U,T,\theta_0}(X)\big), \tag{5.16}$$

where

$$Z_{L,T,\theta_0}(X) = \exp\left(\phi_{Y,0}^2 K_{Y,\xi,T,1} + \phi_{Y,0} I_{Y,X,T} - \frac{\phi_{Y,0}^2}{2} K_{Y,\xi,T,2}\right) \times \exp\left(\phi_{X,0}^2 K_{X,\xi,T,1} + \phi_{X,0} I_{X,T} - \frac{\phi_{X,0}^2}{2} K_{X,\xi,T,2}\right)$$

and

$$Z_{U,T,\theta_0}(X) = \exp\left(\phi_{Y,0}^2 K_{Y,\xi,T,2} + \phi_{Y,0} I_{Y,X,T} - \frac{\phi_{Y,0}^2}{2} K_{Y,\xi,T,1}\right) \times \exp\left(\phi_{X,0}^2 K_{X,\xi,T,2} + \phi_{X,0} I_{X,T} - \frac{\phi_{X,0}^2}{2} K_{X,\xi,T,1}\right).$$

The expressions (5.15) and (5.16) have the same asymptotic form. We first provide the intuitive idea and then rigorously prove our result on asymptotic approximation. Note that, by (H3) and (H7), (5.7), (5.8), (5.9), (5.10), the facts that $\frac{1}{b_T - a_T}\int_{a_T}^{b_T} \lambda^2(s)\,ds \to 0$ as $T \to \infty$, and ξ is a finite random variable, that $K_{Y,\xi,T,1} \overset{\text{a.s.}}{\sim} (b_T - a_T)K_Y$, $K_{Y,\xi,T,2} \overset{\text{a.s.}}{\sim} (b_T - a_T)K_Y$, $K_{X,\xi,T,1} \overset{\text{a.s.}}{\sim} (b_T - a_T)K_X$ and $K_{X,\xi,T,2} \overset{\text{a.s.}}{\sim} (b_T - a_T)K_X$, where, for any two random sequences $\{A_T : T \ge 0\}$ and $\{B_T : T \ge 0\}$, $A_T \overset{\text{a.s.}}{\sim} B_T$ stands for $A_T/B_T \to 1$, almost surely, as $T \to \infty$. Also, as we show, the distributions of $(b_T - a_T)^{-\frac{1}{2}} I_{Y,X,T}$ and $(b_T - a_T)^{-\frac{1}{2}} I_{X,T}$ are asymptotically normal with zero means and variances K_Y and K_X, respectively. Heuristically substituting these in (5.15) and (5.16) yields the form

$$\hat{p}_T(\theta_0) = \exp\left(\frac{(b_T - a_T)K_Y\phi_{Y,0}^2}{2} + \phi_{Y,0}\sqrt{K_Y}\big(W_Y(b_T) - W_Y(a_T)\big) + (b_T - a_T)K_X\phi_{X,0}^2\right).$$

5.2 Modeled likelihood and its asymptotic approximation

Our modeled likelihood associated with the state space model described by (2.3) and (2.4) is given by:

$$L_T(\theta) = \int \exp\left(\phi_Y u_{Y|X,T} - \frac{\phi_Y^2}{2} v_{Y|X,T}\right) \times \exp\left(\phi_X u_{X,T} - \frac{\phi_X^2}{2} v_{X,T}\right) dQ_{T,X}. \tag{5.17}$$

Using the same method of obtaining bounds of $p_T(\theta_0)$, we obtain the following bounds for $L_T(\theta)$:

$$\tilde{B}_{L,T}(\theta) \leq L_T(\theta) \leq \tilde{B}_{U,T}(\theta),$$

where

$$\tilde{B}_{L,T}(\theta) = E_{T,X}\big(\tilde{Z}_{L,T,\theta}(X)\big);$$
$$\tilde{B}_{U,T}(\theta) = E_{T,X}\big(\tilde{Z}_{U,T,\theta}(X)\big),$$

where

$$\tilde{Z}_{L,T,\theta}(X) = \exp\left(\phi_Y \phi_{Y,0} K_{Y,\xi,T,1} + \phi_Y I_{Y,X,T} - \frac{\phi_Y^2}{2} K_{Y,\xi,T,2}\right) \times \exp\left(\phi_X \phi_{X,0} K_{X,\xi,T,1} + \phi_X I_{X,T} - \frac{\phi_X^2}{2} K_{X,\xi,T,2}\right)$$

and

$$\tilde{Z}_{U,T,\theta}(X) = \exp\left(\phi_Y \phi_{Y,0} K_{Y,\xi,T,2} + \phi_Y I_{Y,X,T} - \frac{\phi_Y^2}{2} K_{Y,\xi,T,1}\right) \times \exp\left(\phi_X \phi_{X,0} K_{X,\xi,T,2} + \phi_X I_{X,T} - \frac{\phi_X^2}{2} K_{X,\xi,T,1}\right).$$

It follows as before that the modeled likelihood can be approximated as

$$\hat{L}_T(\theta) = \exp\bigg((b_T - a_T) K_Y \phi_Y \phi_{Y,0} + \phi_Y \sqrt{K_Y}\big(W_Y(b_T) - W_Y(a_T)\big) - \frac{(b_T - a_T) K_Y \phi_Y^2}{2} + (b_T - a_T) K_X \phi_X \phi_{X,0}\bigg).$$

5.3 A briefing on the formal results on the asymptotic approximations

Formal proof of the results $p_T(\theta_0) \overset{\text{a.s.}}{\sim} \hat{p}_T(\theta_0)$ and $L_T(\theta) \overset{\text{a.s.}}{\sim} \hat{L}_T(\theta)$ requires the following two additional assumptions:

(H9) There exists an integer $k_0 \geq 1$ such that $\sum_{T=1}^{\infty} \delta_T^{-2k_0} (b_T - a_T)^{k_0 - 1} \int_{a_T}^{b_T} \lambda^2(s)\, ds < \infty$, where $\delta_T \downarrow 0$ as $T \to \infty$ is a specific sequence decreasing fast enough so that it satisfies, because of continuity of the exponential function, the following: for any $\varepsilon > 0$,

$$\sum_{T=1}^{\infty} P\big(\big|I_{Y,X,T} - \sqrt{K_Y}\big(W_Y(b_T) - W_Y(a_T)\big)\big| \leq \delta_T, \big|\exp(I_{Y,X,T}) - \exp\big(\sqrt{K_Y}\big(W_Y(b_T) - W_Y(a_T)\big)\big)\big| > \varepsilon\big) < \infty. \tag{5.18}$$

Also assume that $E|\xi|^{2k_0} < \infty$.

(H10)

$$\sup_{T>0} E\left(\frac{Z_{L,T,\theta_0}(X)}{\hat{p}_T(\theta_0)}\right) < \infty, \qquad \sup_{T>0} E\left(\frac{Z_{U,T,\theta_0}(X)}{\hat{p}_T(\theta_0)}\right) < \infty,$$

$$\sup_{T>0,\theta\in\Theta} E\left(\frac{\tilde{Z}_{L,T,\theta}(X)}{\hat{L}_T(\theta)}\right) < \infty \quad \text{and} \quad \sup_{T>0,\theta\in\Theta} E\left(\frac{\tilde{Z}_{U,T,\theta}(X)}{\hat{L}_T(\theta)}\right) < \infty.$$

The following lemma shows that under assumptions (H1)–(H9), $\exp(I_{Y,X,T})$ and $\exp(I_{X,T})$ are asymptotically independent of X.

Lemma 8. *Under assumptions* (H1)–(H9),

$$\left|\exp(I_{Y,X,T}) - \exp\left(\sqrt{K_Y}\left(W_Y(b_T) - W_Y(a_T)\right)\right)\right| \xrightarrow{\text{a.s.}} 0; \tag{5.19}$$

$$\left|\exp(I_{X,T}) - \exp\left(\sqrt{K_X}\left(W_X(b_T) - W_X(a_T)\right)\right)\right| \xrightarrow{\text{a.s.}} 0. \tag{5.20}$$

The following corollary of Lemma 8 shows asymptotic normality of the relevant quantities involved in the asymptotic approximations.

Corollary 9. *Since* $(b_T - a_T)^{-\frac{1}{2}}\sqrt{K_Y}(W_Y(b_T) - W_Y(a_T))$ *and* $(b_T - a_T)^{-\frac{1}{2}}\sqrt{K_X} \times (W_X(b_T) - W_X(a_T))$ *are normally distributed with mean zero and variances* K_Y *and* K_X, *respectively*, *it follows that*

$$(b_T - a_T)^{-\frac{1}{2}} I_{Y,X,T} \xrightarrow{\text{a.s.}} N(0, K_Y);$$

$$(b_T - a_T)^{-\frac{1}{2}} I_{X,T} \xrightarrow{\text{a.s.}} N(0, K_X).$$

Finally, our asymptotic approximation result is given by the following theorem, which requires assumptions (H1)–(H10).

Theorem 10. *Assume* (H1)–(H10). *Then*

$$p_T(\theta_0) \overset{\text{a.s.}}{\sim} \hat{p}_T(\theta_0); \tag{5.21}$$

$$L_T(\theta) \overset{\text{a.s.}}{\sim} \hat{L}_T(\theta) \quad \textit{for all } \theta \in \Theta. \tag{5.22}$$

The proofs of Lemma 8 and Theorem 10 are presented in the supplement (Maitra and Bhattacharya (2019)).

6 Convergence of the posterior distribution of θ

In order to prove convergence of our posterior distribution, we verify the conditions of the theorem proved in Shalizi (2009) which take account of dependence setups and misspecifications. The detailed assumptions in our state space *SDE* context and Shalizi's theorem is provided in Section S-1 of the supplement (Maitra and Bhattacharya (2019)).

6.1 Further assumptions

Before proceeding further, we make the following assumptions regarding ψ_Y and ψ_X:

(H11) (i) For every $\theta \in \Theta \cup \{\theta_0\}$, $\psi_Y(\theta)$ and $\psi_X(\theta)$ are finite and satisfy $(\psi_Y(\theta_1), \psi_X(\theta_1)) = (\psi_Y(\theta_2), \psi_X(\theta_2))$ implies $\theta_1 = \theta_2$.

(ii) $|\psi_Y|$ is coercive, that is, for every sequence $\{\theta_T : T > 0\}$ such that $\|\theta_T\| \to \infty$, $|\psi_Y(\theta_T)| \to \infty$.

(iii) For every sequence $\{\theta_T : T > 0\}$ such that $\|\theta_T\| \to \infty$, $|\psi_Y(\theta_T)|^2(b_T - a_T)|K_{Y,j,T} - K_Y| \to 0$ and $|\psi_X(\theta_T)|^2(b_T - a_T)|K_{X,j,T} - K_X| \to 0$, for $j = 1, 2$, and $C_1(b_T - a_T) \leq (\psi_Y(\theta_T) - \psi_Y(\theta_0))^8 \leq C_2(b_T - a_T)$, for some constants $C_1, C_2 > 0$, as $T \to \infty$.

(iv) $|\psi_Y(\theta)|$ is assumed to have finite expectation with respect to the prior $\pi(\theta)$.

(v) $|\psi_X(\theta)| \leq |\psi_X(\theta_0)|$, for all $\theta \in \Theta$.

(vi) The first and second derivatives of ψ_X vanish at $\theta = \theta_0$.

(vii) ψ_Y and ψ_X are at least thrice continuously differentiable.

6.2 Verification of the assumptions of Shalizi

6.2.1 *Verification of* (A1). Recall that our likelihood $L_T(\theta)$ is given by (5.17). In the same way as the proof of the second part of Proposition 2 of Delattre, Genon-Catalot and Samson (2013), it can be proved that the first factor of the integrand of (5.17) is a measurable function of $(\{Y(s); s \in [a_T, b_T]\}, \{X(s); s \in [a_T, b_T]\}, \theta)$. Also, by the same result of Delattre, Genon-Catalot and Samson (2013) the second factor of the integrand is a measurable function of $(\{X(s); s \in [a_T, b_T]\}, \theta)$. Thus, the integrand is a measurable function of $(\{Y(s); s \in [a_T, b_T]\}, \{X(s); s \in [a_T, b_T]\}, \theta)$. Since the associated measure spaces are σ-finite, $L_T(\theta)$ is clearly $\mathcal{F}_T \times \mathcal{T}$-measurable for all $T > 0$.

6.2.2 *Verification of* (A2). We consider the likelihood ratio $R_T(\theta)$ given by (S-1.1). Using Theorem 10 we obtain that

$$\begin{aligned}\frac{1}{b_T - a_T} \log R_T(\theta) \overset{\text{a.s.}}{\sim} & -\frac{K_Y}{2}(\phi_Y - \phi_{Y,0})^2 \\ & + \sqrt{K_Y}(\phi_Y - \phi_{Y,0})\frac{(W_Y(b_T) - W_Y(a_T))}{b_T - a_T} \\ & - \frac{K_X}{2}(\phi_X - \phi_{X,0})^2 + \frac{K_X}{2}(\phi_X^2 - \phi_{X,0}^2).\end{aligned} \tag{6.1}$$

Since $\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T} \xrightarrow{\text{a.s.}} 0$, it follows that, almost surely,

$$\begin{aligned}&\frac{1}{b_T - a_T} \log R_T(\theta) \\ &\quad \to -\frac{1}{2}\left[K_Y(\phi_Y - \phi_{Y,0})^2 + K_X(\phi_X - \phi_{X,0})^2 + K_X(\phi_{X,0}^2 - \phi_X^2)\right].\end{aligned} \tag{6.2}$$

Let

$$\begin{aligned}h(\theta) &= \frac{1}{2}\left[K_Y(\phi_Y - \phi_{Y,0})^2 + K_X(\phi_X - \phi_{X,0})^2 + K_X(\phi_{X,0}^2 - \phi_X^2)\right] \\ &= \frac{1}{2}\left[K_Y(\psi_Y(\theta) - \psi_Y(\theta_0))^2 + K_X(\psi_X(\theta) - \psi_X(\theta_0))^2 \right. \\ &\quad \left. + K_X(\psi_X^2(\theta_0) - \psi_X^2(\theta))\right].\end{aligned} \tag{6.3}$$

Note that due to (H11)(v), $h(\theta) \geq 0$, for all $\theta \in \Theta$. Thus (A2) holds.

6.2.3 *Verification of* (A3). We now obtain the limit of the quantity

$$\frac{1}{b_T - a_T} E_{\theta_0}\left(\log \frac{p_T(\theta_0)}{L_T(\theta)}\right) = -\frac{1}{b_T - a_T} E_{\theta_0}(\log R_T(\theta)),$$

where E_{θ_0} is the expectation with respect to the true likelihood $p_T(\theta_0)$. Proceeding in the same way as in the case of $R_T(\theta)$ and noting that $E_{\theta_0}(W_Y(b_T) - W_Y(a_T)) = 0$, it is easy to see that

$$\frac{1}{b_T - a_T} E_{\theta_0}\left(\log \frac{p_T(\theta_0)}{L_T(\theta)}\right) \to h(\theta),$$

as $T \to \infty$.

6.2.4 *Verification of* (A4). To verify (A4) we reformulate the original parameter space Θ as $\Theta \setminus I$. Abusing notation, we continue to denote $\Theta \setminus I$ as Θ. Hence, the prior π on Θ clearly satisfies $\pi(I) = 0$.

6.2.5 *Verification of* (A5)(i). Now consider $\mathcal{G}_T = \{\theta \in \Theta : |\psi_Y(\theta)| \leq \exp(\beta(b_T - a_T))\}$, where β is chosen such that $\beta > 2h(\Theta)$. Coerciveness of $\|\psi_Y\|$ implies compactness of $\mathcal{G}_T$, for every $T > 0$.

The above definition of $\mathcal{G}_T$ clearly implies $\mathcal{G}_T \to \Theta$. Also,

$$\begin{aligned}\pi(\mathcal{G}_T) &> 1 - E\big(|\psi_Y(\theta)|\big)\exp\big(-\beta(b_T - a_T)\big)\\ &= 1 - \alpha \exp\big(-\beta(b_T - a_T)\big),\end{aligned}$$

where the first inequality is due to Markov's inequality and $\alpha = E(|\psi_Y(\theta)|) > 0$. The expectation, which is with respect to the prior π, exists by (H11)(iv).

6.2.6 *Verification of* (A5)(ii). We now show that convergence of (6.2) is uniform in θ over $\mathcal{G}_T \setminus I$. First note that $\mathcal{G}_T \setminus I = \mathcal{G}_T$, since we have already removed I from Θ. Now note that, because of compactness of $\mathcal{G}_T$ and continuity of $|\frac{1}{b_T - a_T} \log R_T(\theta) + h(\theta)|$ in θ, there exists $\theta_T \in \mathcal{G}_T$ such that

$$\sup_{\theta \in \mathcal{G}_T \setminus I} \left|\frac{1}{b_T - a_T} \log R_T(\theta) + h(\theta)\right| = \left|\frac{1}{b_T - a_T} \log R_T(\theta_T) + h(\theta_T)\right|. \tag{6.4}$$

Note that θ_T depends upon the data. However, under the additional condition (H11)(iii), it is clear from the proof of Theorem 10 (see Section S-4 of the supplement (Maitra and Bhattacharya (2019))) that our asymptotic approximation of $L_T(\theta_T)$ remains valid even in this case. Formally,

Theorem 11. *Assume* (H1)–(H10) *and* (H11)(iii). *Consider any, perhaps, data-dependent sequence* $\{\theta_T : T > 0\}$, *where either* $\|\theta_T\|$ *remains finite almost surely or* $\|\theta_T\| \to \infty$, *almost surely, as* $T \to \infty$. *Then* $L_T(\theta_T) \overset{\text{a.s.}}{\sim} \hat{L}_T(\theta_T)$.

The above theorem guarantees that (6.4) admits the following approximation:

$$\begin{aligned}&\left|\frac{1}{b_T - a_T} \log R_T(\theta_T) + h(\theta_T)\right|\\ &\quad \overset{\text{a.s.}}{\sim} \sqrt{K_Y}\left|\frac{(\psi_Y(\theta_T) - \psi_Y(\theta_0))}{\sqrt{b_T - a_T}} \times \frac{W_Y(b_T) - W_Y(a_T)}{\sqrt{b_T - a_T}}\right|.\end{aligned} \tag{6.5}$$

By Corollary 9 and (H11)(iii), the right hand side of (6.5) goes to zero almost surely, as $T \to \infty$. Hence, the convergence of (6.2) is uniform in θ over $\mathcal{G}_T \setminus I$.

6.2.7 *Verification of* (A5)(iii). We now show that $h(\mathcal{G}_T) \to h(\Theta)$, as $T \to \infty$. Due to compactness of $\mathcal{G}_T$ and continuity of $h(\theta)$, it follows that there exists $\tilde{\theta}_T \in \mathcal{G}_T$ such that $h(\mathcal{G}_T) = h(\tilde{\theta}_T)$. Also, since $\mathcal{G}_T$ is a non-decreasing sequence of sets, $h(\tilde{\theta}_T)$ is non-increasing in T. Since $\mathcal{G}_T \to \Theta$, it follows that $h(\mathcal{G}_T) \to h(\Theta)$, as $T \to \infty$.

6.2.8 *Verification of* (A6). Under (A1)–(A3), which we have already verified, it holds that (see equation (18) of Shalizi (2009)) for any fixed $\mathcal{G}$ of the sequence $\mathcal{G}_T$, for any $\varepsilon > 0$ and for sufficiently large T,

$$\frac{1}{b_T - a_T} \log \int_{\mathcal{G}} R_T(\theta)\pi(\theta)\, d\theta \le -h(\mathcal{G}) + \varepsilon + \frac{1}{b_T - a_T} \log \pi(\mathcal{G}).$$

It follows that $\tau(\mathcal{G}_T, \delta)$ is almost surely finite for all T and δ. We now argue that for sufficiently large T, $\tau(\mathcal{G}_T, \delta) > (b_T - a_T)$ only finitely often with probability one. By equation (41) of Shalizi (2009),

$$\begin{aligned}
&\sum_{T=1}^{\infty} P\big(\tau(\mathcal{G}_T, \delta) > (b_T - a_T)\big) \\
&\quad \le \sum_{T=1}^{\infty} \sum_{m=T+1}^{\infty} P\left(\frac{1}{b_m - a_m} \log \int_{\mathcal{G}_T} R_m(\theta)\pi(\theta)\, d\theta > \delta - h(\mathcal{G}_T)\right).
\end{aligned} \tag{6.6}$$

Now, by compactness of $\mathcal{G}_T$, $h(\mathcal{G}_T) = h(\tilde{\theta}_T)$, for $\tilde{\theta}_T \in \mathcal{G}_T$, and by the mean value theorem for integrals,

$$\frac{1}{b_m - a_m} \log \int_{\mathcal{G}_T} R_m(\theta)\pi(\theta)\, d\theta = \frac{1}{b_m - a_m} \log R_m(\hat{\theta}_T)\pi(\mathcal{G}_T),$$

for $\hat{\theta}_T \in \mathcal{G}_T$ depending upon the data, so that

$$\frac{1}{b_m - a_m} \log \int_{\mathcal{G}_T} R_m(\theta)\pi(\theta)\, d\theta > \delta - h(\mathcal{G}_T)$$

implies, since $h(\hat{\theta}_T) \ge h(\tilde{\theta}_T)$, that

$$\frac{1}{b_m - a_m} \log R_m(\hat{\theta}_T) + h(\hat{\theta}_T) > \delta - \frac{1}{b_m - a_m} \log \pi(\mathcal{G}_T) > \delta.$$

Thus, it follows from (6.6) and Chebychev's inequality, that

$$\begin{aligned}
&\sum_{T=1}^{\infty} P\big(\tau(\mathcal{G}_T, \delta) > (b_T - a_T)\big) \\
&\quad \le \sum_{T=1}^{\infty} \sum_{m=T+1}^{\infty} P\left(\left|\frac{1}{b_m - a_m} \log R_m(\hat{\theta}_T) + h(\hat{\theta}_T)\right| > \delta\right) \\
&\quad \le \sum_{T=1}^{\infty} \sum_{m=T+1}^{\infty} \delta^{-8} E\left(\frac{1}{b_m - a_m} \log R_m(\hat{\theta}_T) + h(\hat{\theta}_T)\right)^8.
\end{aligned} \tag{6.7}$$

From (6.1) and (6.3) it is clear that

$$\begin{aligned}
&\frac{1}{b_m - a_m} \log R_m(\hat{\theta}_T) + h(\hat{\theta}_T) \\
&\quad \overset{\text{a.s.}}{\sim} \sqrt{K_Y} \frac{(\psi_Y(\hat{\theta}_T) - \psi_Y(\theta_0))}{\sqrt{b_m - a_m}} \times \frac{W_Y(b_m) - W_Y(a_m)}{\sqrt{b_m - a_m}}.
\end{aligned} \tag{6.8}$$

Now, let $Z_m = \frac{1}{b_m-a_m}\log R_m(\hat{\theta}_T) + h(\hat{\theta}_T)$ and $\tilde{Z}_m = \sqrt{K_Y}\frac{(\psi_Y(\hat{\theta}_T)-\psi_Y(\theta_0))}{\sqrt{b_m-a_m}} \times \frac{W_Y(b_m)-W_Y(a_m)}{\sqrt{b_m-a_m}}$. Then

$$\frac{Z_m^8-\tilde{Z}_m^8}{E(\tilde{Z}_m^8)} = \frac{Z_m^8-\tilde{Z}_m^8}{\tilde{Z}_m^8} \times \frac{\tilde{Z}_m^8}{E(\tilde{Z}_m^8)} \xrightarrow{\text{a.s.}} 0 \quad \text{as } m \to \infty, \tag{6.9}$$

because, due to 6.8 the first factor on the right-hand side of (6.9) tends to zero almost surely, while by (H11)(iii) the second factor is bounded above by a constant times standard normal distribution raised to the power 6. It can be easily verified using (H11)(iii) that $\sup_{m\geq 1} E[\frac{Z_m^8-\tilde{Z}_m^8}{E(\tilde{Z}_m^8)}]^2 < \infty$, so that $\frac{Z_m^8-\tilde{Z}_m^8}{E(\tilde{Z}_m^8)}$ is uniformly integrable. Hence, it follows from (6.9) that

$$\frac{E(Z_m^8)-E(\tilde{Z}_m^8)}{E(\tilde{Z}_m^8)} \to 0 \quad \text{as } m \to \infty.$$

In other words, as $m \to \infty$,

$$E(Z_m^8) \stackrel{\text{a.s.}}{\sim} E(\tilde{Z}_m^8). \tag{6.10}$$

Now note that for studying convergence of the double sum (6.7), it is enough to investigate convergence of

$$S_{T_0} = \sum_{T=T_0}^{\infty}\sum_{m=T+1}^{\infty} E\left(\frac{1}{b_m-a_m}\log R_m(\hat{\theta}_T) + h(\hat{\theta}_T)\right)^8,$$

for some sufficiently large T_0. By virtue of (6.10) it is then enough to study convergence of

$$\begin{aligned} S_{T_0} &= \sum_{T=T_0}^{\infty}\sum_{m=T+1}^{\infty} E(\tilde{Z}_m^8) \\ &= \tilde{c}\sum_{T=T_0}^{\infty}\sum_{m=T+1}^{\infty} \frac{(\psi_Y(\tilde{\theta}_T)-\psi_Y(\theta_0))^8}{(b_m-a_m)^4}, \end{aligned}$$

where $\tilde{c}$ (> 0) is a constant. By (H11)(iii), for sufficiently large T, $(\psi_Y(\tilde{\theta}_T)-\psi_Y(\theta_0))^8 \leq C_2(b_T-a_T)$, for some $C_2 > 0$. Hence,

$$S_{T_0} \leq C_Y \sum_{T=T_0}^{\infty}\sum_{m=T+1}^{\infty} \frac{b_T-a_T}{(b_m-a_m)^4},$$

where C_Y (> 0) is a constant. Now note that, since (b_T-a_T) is increasing in T, $(b_{T_0+j}-a_{T_0+j}) < (b_{T_0+j+1}-a_{T_0+j+1})$ for $j \geq 0$, so that

$$\begin{aligned} &\sum_{T=T_0}^{\infty}\sum_{m=T+1}^{\infty} \frac{b_T-a_T}{(b_m-a_m)^4} \\ &= \frac{(b_{T_0}-a_{T_0})}{(b_{T_0+1}-a_{T_0+1})^4} + \frac{(b_{T_0}-a_{T_0})+(b_{T_0+1}-a_{T_0+1})}{(b_{T_0+2}-a_{T_0+2})^4} \\ &\quad + \frac{(b_{T_0}-a_{T_0})+(b_{T_0+1}-a_{T_0+1})+(b_{T_0+2}-a_{T_0+2})}{(b_{T_0+3}-a_{T_0+3})^4} + \cdots \\ &\leq \sum_{k=1}^{\infty} \frac{k}{(b_{T_0+k}-a_{T_0+k})^3} \end{aligned}$$

$$\leq \sum_{k=1}^{\infty} \frac{k}{(T_0+k)^3} \leq \sum_{k=1}^{\infty} \frac{(T_0+k)}{(T_0+k)^3} = \sum_{k=1}^{\infty} \frac{1}{(T_0+k)^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2}$$
$$< \infty.$$

That is, $S_{T_0} < \infty$ for sufficiently large T_0. In other words, (A6) holds.

6.2.9 *Verification of* (A7). For any set $A \subseteq \Theta$ with $\pi(A) > 0$, it follows that $\mathcal{G}_T \cap A \to \Theta \cap A = A$. Since $h(\mathcal{G}_T \cap A)$ is non-increasing as T increases, it follows that $h(\mathcal{G}_T \cap A) \to h(A)$, as $T \to \infty$.

To summarize, we have the following theorem on posterior convergence of θ.

Theorem 12. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Assume* (H1)–(H10) *and* (H11)(i)–(v). *For the prior π on θ, consider any set $A \in \mathcal{T}$ with $\pi(A) > 0$ and $h(A) > h(\Theta)$. Then, almost surely,*

$$\lim_{T\to\infty} \pi(A|\mathcal{F}_T) = 0.$$

Moreover, if $\beta > 2h(A)$ or $A \subset \cap_{k=T}^{\infty} \mathcal{G}_k$ for some T, then almost surely,

$$\lim_{T\to\infty} \frac{1}{b_T - a_T} \log \pi(A|\mathcal{F}_T) = -J(A).$$

7 Consistency and asymptotic normality of the maximum likelihood estimator

Now we make the following further assumption:

(H12) The parameter space Θ is compact.

Let

$$g_{Y,T}(\theta) = -\frac{K_Y}{2}\big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)^2 + \sqrt{K_Y}\big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T}; \tag{7.1}$$

$$g_{X,T}(\theta) = -\frac{K_X}{2}\big(\psi_X(\theta) - \psi_X(\theta_0)\big)^2 + \frac{K_X}{2}\big(\psi_X^2(\theta) - \psi_X^2(\theta_0)\big). \tag{7.2}$$

Then note that

$$\sup_{\theta\in\Theta}\left|\frac{1}{b_T - a_T}\log R_T(\theta) - g_{Y,T}(\theta) - g_{X,T}(\theta)\right| = \left|\frac{1}{b_T - a_T}\log R_T(\theta_T^*) - g_{Y,T}(\theta_T^*) - g_{X,T}(\theta_T^*)\right|, \tag{7.3}$$

for some $\theta_T^* \in \Theta$ where θ_T^* is dependent on data. Proceeding in the same way as in Section 6.2.6 it is easily seen that (7.3) tends to zero almost surely with respect to both Y and X, as $T \to \infty$. Hence, the maximum likelihood estimator (*MLE*) can be approximated by maximizing the function

$$\tilde{g}_T(\theta) = g_{Y,T}(\theta) + g_{X,T}(\theta)$$

with respect to θ.

7.1 Strong consistency of the maximum likelihood estimator of θ

Observe that for $k = 1, \ldots, d$,

$$\frac{\partial \tilde{g}_T(\theta)}{\partial \theta_k} = -K_Y\big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)\frac{\partial \psi_Y(\theta)}{\partial \theta_k} - K_X\big(\psi_X(\theta) - \psi_X(\theta_0)\big)\frac{\partial \psi_X(\theta)}{\partial \theta_k} + K_X\psi_X(\theta)\frac{\partial \psi_X(\theta)}{\partial \theta_k} + \sqrt{K_Y}\frac{\partial \psi_Y(\theta)}{\partial \theta_k}\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T}.$$

Let

$$\tilde{g}'_T(\theta) = \left(\frac{\partial \tilde{g}_T(\theta)}{\partial \theta_1}, \ldots, \frac{\partial \tilde{g}_T(\theta)}{\partial \theta_d}\right)^T.$$

Also, let

$$\tilde{g}''_T(\theta) = \begin{pmatrix} \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_1^2} & \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_1 \partial \theta_d} \\ \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_2 \partial \theta_d} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_d \partial \theta_1} & \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_d \partial \theta_2} & \cdots & \frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_d^2} \end{pmatrix}$$

denote the matrix with (j,k)th element given by

$$\begin{aligned}\frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_j \partial \theta_k} = &-K_Y\left[\frac{\partial \psi_Y(\theta)}{\partial \theta_j}\frac{\partial \psi_Y(\theta)}{\partial \theta_k} + \big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)\frac{\partial^2 \psi_Y(\theta)}{\partial \theta_j \partial \theta_k}\right] \\ &- K_X\left[\frac{\partial \psi_X(\theta)}{\partial \theta_j}\frac{\partial \psi_X(\theta)}{\partial \theta_k} + \big(\psi_X(\theta) - \psi_X(\theta_0)\big)\frac{\partial^2 \psi_X(\theta)}{\partial \theta_j \partial \theta_k}\right] \\ &+ K_X\left[\frac{\partial \psi_X(\theta)}{\partial \theta_j}\frac{\partial \psi_X(\theta)}{\partial \theta_k} + \psi_X(\theta)\frac{\partial^2 \psi_X(\theta)}{\partial \theta_j \partial \theta_k}\right] \\ &+ \sqrt{K_Y}\frac{\partial^2 \psi_Y(\theta)}{\partial \theta_j \partial \theta_k}\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T}.\end{aligned}$$

Note that by (H11)(vi),

$$\left[\frac{\partial \tilde{g}_T(\theta)}{\partial \theta_k}\right]_{\theta=\theta_0} = \sqrt{K_Y}\left[\frac{\partial \psi_Y(\theta)}{\partial \theta_k}\right]_{\theta=\theta_0}\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T}; \tag{7.4}$$

$$\begin{aligned}\left[\frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_j \partial \theta_k}\right]_{\theta=\theta_0} = &-K_Y\left[\frac{\partial \psi_Y(\theta)}{\partial \theta_j}\frac{\partial \psi_Y(\theta)}{\partial \theta_k}\right]_{\theta=\theta_0} \\ &+ \sqrt{K_Y}\left[\frac{\partial^2 \psi_Y(\theta)}{\partial \theta_j \partial \theta_k}\right]_{\theta=\theta_0}\frac{W_Y(b_T) - W_Y(a_T)}{b_T - a_T}.\end{aligned} \tag{7.5}$$

Letting $\hat{\theta}_T$ denote the *MLE*, note that

$$0 = \tilde{g}'_T(\hat{\theta}_T) = \tilde{g}'_T(\theta_0) + \tilde{g}''_T(\theta^*_T)(\hat{\theta}_T - \theta_0), \tag{7.6}$$

where θ^*_T lies between θ_0 and $\hat{\theta}_T$. From (7.5) it is clear that

$$\left[\frac{\partial^2 \tilde{g}_T(\theta)}{\partial \theta_j \partial \theta_k}\right]_{\theta=\theta_0} \xrightarrow{\text{a.s.}} -K_Y\left[\frac{\partial \psi_Y(\theta)}{\partial \theta_j}\frac{\partial \psi_Y(\theta)}{\partial \theta_k}\right]_{\theta=\theta_0},$$

as $T \to \infty$. Let $\mathcal{I}(\theta)$ denote the matrix with (j,k)th element given by

$$\{\mathcal{I}(\theta)\}_{jk} = K_Y\left[\frac{\partial \psi_Y(\theta)}{\partial \theta_j}\frac{\partial \psi_Y(\theta)}{\partial \theta_k}\right].$$

From (7.4) it is obvious that $\{\mathcal{I}(\theta_0)\}_{jk}$ is the covariance between the jth and the kth components of $\sqrt{b_T - a_T}\tilde{g}'_T(\theta_0)$, and so $\mathcal{I}(\theta_0)$ is non-negative definite. We make the following assumptions:

(H13) The true value $\theta_0 \in \mathrm{int}(\Theta)$, where by $\mathrm{int}(\Theta)$ we mean the interior of Θ.
(H14) The matrix $\mathcal{I}(\theta)$ is positive definite for $\theta \in \mathrm{int}(\Theta)$.

Hence, from (7.6) we obtain, after pre-multiplying both sides of the relevant equation with $\mathcal{I}^{-1}(\theta_T^*)$, the following:

$$-\mathcal{I}^{-1}(\theta_T^*)\tilde{g}''_T(\theta_T^*)(\hat{\theta}_T - \theta_0) = \mathcal{I}^{-1}(\theta_T^*)\tilde{g}'_T(\theta_0). \tag{7.7}$$

Since as $T \to \infty$, $\tilde{g}'_T(\theta_0) \xrightarrow{\text{a.s.}} 0$ and $-\mathcal{I}^{-1}(\theta_T^*)\tilde{g}''_T(\theta_T^*) \xrightarrow{\text{a.s.}} \mathfrak{I}_d$, $\mathfrak{I}_d$ being the identity matrix of order d, it hold that

$$\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta_0, \tag{7.8}$$

as $T \to \infty$, showing that the *MLE* is strongly consistent. The result can be formalized as the following theorem.

Theorem 13. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Assume conditions* (H1)–(H14). *Then the MLE of* θ *is strongly consistent in the sense that* (7.8) *holds.*

7.2 Asymptotic normality of the maximum likelihood estimator of θ

Since $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta_0$ and θ_T^* lies between θ_0 and $\hat{\theta}_T$, it follows that $\theta_T^* \xrightarrow{\text{a.s.}} \theta_0$ as $T \to \infty$. This, and the fact that $(W_Y(b_T) - W_Y(a_T))/\sqrt{b_T - a_T} \sim N(0,1)$, guarantee that

$$-\sqrt{b_T - a_T}\mathcal{I}^{-1}(\theta_T^*)\tilde{g}'_T(\theta_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathcal{I}^{-1}(\theta_0)),$$

where "$\xrightarrow{\mathcal{L}}$" denotes convergence in distribution. From (7.7) it then follows, using the fact $\mathcal{I}^{-1}(\theta_T^*)\tilde{g}''_T(\theta_T^*) \xrightarrow{\text{a.s.}} \mathfrak{I}_d$, that

$$\sqrt{b_T - a_T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} N_d\big(0, \mathcal{I}^{-1}(\theta_0)\big). \tag{7.9}$$

Thus, we can present the following theorem.

Theorem 14. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Assume conditions* (H1)–(H14). *Then the MLE of* θ *is asymptotically normal in the sense that* (7.9) *holds.*

8 Asymptotic posterior normality

Let $\ell_T(\theta) = \log L_T(\theta)$ stand for the log-likelihood, and let

$$\Sigma_T^{-1} = \begin{cases} -\ell''_T(\hat{\theta}_T) & \text{if the inverse and } \hat{\theta}_T \text{ exist,} \\ \mathfrak{I}_d & \text{if not,} \end{cases}$$

where for any z,

$$\ell''_T(z) = \left(\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_T(\theta)\bigg|_{\theta=z}\right)\right).$$

Thus, Σ_T^{-1} is the observed Fisher's information matrix.

8.1 Regularity conditions and a theorem of Schervish (1995)

(1) The parameter space is $\Theta \subseteq \mathbb{R}^d$ for some finite d.
(2) θ_0 is a point interior to Θ.
(3) The prior distribution of θ has a density with respect to Lebesgue measure that is positive and continuous at θ_0.
(4) There exists a neighborhood $\mathcal{N}_0 \subseteq \Theta$ of θ_0 on which $\ell_T(\theta) = \log L_T(\theta)$ is twice continuously differentiable with respect to all co-ordinates of θ, a.s. $[P_{\theta_0}]$.
(5) The largest eigenvalue of Σ_T goes to zero in probability.
(6) For $\delta > 0$, define $\mathcal{N}_0(\delta)$ to be the open ball of radius δ around θ_0. Let ρ_T be the smallest eigenvalue of Σ_T. If $\mathcal{N}_0(\delta) \subseteq \Theta$, there exists $K(\delta) > 0$ such that

$$\lim_{T\to\infty} P_{\theta_0}\Big(\sup_{\theta\in\Theta\backslash\mathcal{N}_0(\delta)} \rho_T\big[\ell_T(\theta) - \ell_T(\theta_0)\big] < -K(\delta)\Big) = 1. \tag{8.1}$$

(7) For each $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that

$$\lim_{T\to\infty} P_{\theta_0}\Big(\sup_{\theta\in\mathcal{N}_0(\delta(\varepsilon)),\|\gamma\|=1} \big|1 + \gamma^T \Sigma_T^{\frac{1}{2}} \ell''_T(\theta) \Sigma_T^{\frac{1}{2}} \gamma\big| < \varepsilon\Big) = 1. \tag{8.2}$$

Theorem 15 (Schervish (1995)). *Assume the above seven regularity conditions. Then denoting $\Psi_T = \Sigma_T^{-1/2}(\theta - \hat{\theta}_T)$, for each compact subset B of $\mathbb{R}^d$ and each $\varepsilon > 0$, the following holds*:

$$\lim_{T\to\infty} P_{\theta_0}\Big(\sup_{\Psi_T\in B} \big|\pi(\Psi_T|\mathcal{F}_T) - \varrho(\Psi_T)\big| > \varepsilon\Big) = 0,$$

where $\varrho(\cdot)$ denotes the density of the standard normal distribution.

8.2 Verification of the seven regularity conditions for posterior normality

Also we assume that Θ is compact (assumption (H11)) which enables us to uniformly approximate $\frac{1}{b_T - a_T}\log R_T(\theta)$ by $g_{Y,T}(\theta) + g_{X,T}(\theta)$ for $\theta \in \Theta$; see Section 7. As a consequence, $\frac{1}{b_T - a_T}\ell_T(\theta)$ can be uniformly approximated by $g_{Y,T}(\theta) + g_{X,T}(\theta) + \frac{1}{b_T - a_T}\log p_T(\theta_0)$, for $\theta \in \Theta$. Let

$$\frac{1}{b_T - a_T}\tilde{\ell}_T(\theta) = g_{Y,T}(\theta) + g_{X,T}(\theta) + \frac{1}{b_T - a_T}\log p_T(\theta_0).$$

Henceforth, we shall be working with $\frac{1}{b_T - a_T}\tilde{\ell}_T(\theta)$ whenever convenient. With this, the first four regularity conditions presented in Section 8.1 trivially hold.

To verify regularity condition (5), note that, since $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta_0$,

$$\frac{1}{b_T - a_T}\tilde{\ell}''_T(\hat{\theta}_T) \xrightarrow{\text{a.s.}} -\mathcal{I}(\theta_0).$$

Hence, almost surely,

$$\Sigma_T^{-1} \sim (b_T - a_T) \times \mathcal{I}(\theta_0),$$

so that

$$\Sigma_T \xrightarrow{\text{a.s.}} 0,$$

as $T \to \infty$. Thus, regularity condition (5) holds.

For verifying condition (6), observe that

$$\rho_T\big[\ell_T(\theta) - \ell_T(\theta_0)\big] = \rho_T(b_T - a_T) \times \frac{1}{b_T - a_T}\log R_T(\theta),$$

where $\rho_T(b_T - a_T) \to c$, for some $c > 0$ and, due to (6.2),

$$\begin{aligned} &\rho_T\big[\ell_T(\theta) - \ell_T(\theta_0)\big] \\ &\quad \xrightarrow{\text{a.s.}} -\frac{c}{2}\big[K_Y\big(\psi_Y(\theta) - \psi_Y(\theta_0)\big)^2 + K_X\big(\psi_X(\theta) - \psi_X(\theta_0)\big)^2 \\ &\qquad + K_X\big(\psi_X^2(\theta_0) - \psi_X^2(\theta)\big)\big], \end{aligned} \tag{8.3}$$

for all $\theta \in \Theta \setminus \mathcal{N}_0(\delta)$. Now note that

$$\begin{aligned} &\lim_{T\to\infty} P_{\theta_0}\Big(\sup_{\theta\in\Theta\setminus\mathcal{N}_0(\delta)} \rho_T\big[\ell_T(\theta) - \ell_T(\theta_0)\big] < -K(\delta)\Big) \\ &\quad \geq \lim_{T\to\infty} P_{\theta_0}\bigg(\big(\rho_T(b_T - a_T)\big) \times \frac{1}{b_T - a_T} \log R_T(\theta) < -K(\delta)\ \forall\theta \in \Theta \setminus \mathcal{N}_0(\delta)\bigg) \\ &\quad = 1, \end{aligned} \tag{8.4}$$

the last step following due to (8.3). Thus, regularity condition (6) is verified.

For verifying condition (7), we note that $\theta \in \mathcal{N}_0(\delta(\varepsilon))$ can be represented as $\theta = \theta_0 + \delta_2\frac{\theta_0}{\|\theta_0\|}$, where $0 < \delta_2 \leq \delta(\varepsilon)$. Hence, Taylor's series expansion around θ_0 yields

$$\frac{\tilde{\ell}_T''(\theta)}{b_T - a_T} = \frac{\tilde{\ell}_T''(\theta_0)}{b_T - a_T} + \delta_2\frac{\tilde{\ell}_T'''(\theta^*)\theta_0}{(b_T - a_T)\|\theta_0\|}, \tag{8.5}$$

where θ^* lies between θ_0 and θ. As $T \to \infty$, $\frac{\tilde{\ell}_T''(\theta_0)}{b_T - a_T}$ tends to $-\mathcal{I}(\theta_0)$, almost surely. Now notice that

$$\frac{\|\tilde{\ell}_T'''(\theta^*)\theta_0\|}{(b_T - a_T)\|\theta_0\|} \leq \frac{\|\tilde{\ell}_T'''(\theta^*)\|}{b_T - a_T}.$$

Because of (H11)(vii) and compactness of Θ it follows that $\frac{\|\tilde{\ell}_T'''(\theta^*)\|}{b_T - a_T} \to 0$ as $T \to \infty$. Hence, it follows that $\tilde{\ell}_T''(\theta) = O(-(b_T - a_T) \times \mathcal{I}(\theta_0) + (b_T - a_T)\delta_2)$, almost surely. Since $\Sigma_T^{\frac{1}{2}}$ is asymptotically almost surely equivalent to $(b_T - a_T)^{-\frac{1}{2}}\mathcal{I}^{-\frac{1}{2}}(\theta_0)$, condition (7) holds. We summarize our result in the form of the following theorem.

Theorem 16. *Assume that the data was generated by the true model given by* (2.1) *and* (2.2), *but modeled by* (2.3) *and* (2.4). *Assume* (H1)–(H14). *Then denoting* $\Psi_T = \Sigma_T^{-1/2}(\theta - \hat{\theta}_T)$, *for each compact subset B of $\mathbb{R}^d$ and each* $\varepsilon > 0$, *the following holds*:

$$\lim_{T\to\infty} P_{\theta_0}\Big(\sup_{\Psi_T\in B}\big|\pi(\Psi_T|\mathcal{F}_T) - \varrho(\Psi_T)\big| > \varepsilon\Big) = 0,$$

where $\varrho(\cdot)$ denotes the density of the standard normal distribution.

9 Random effects models based on state space *SDE*s and a brief overview of the asymptotic results

9.1 True and postulated systems of state space *SDE*s with random effects

We now consider the following "true" random effects models based on state space *SDE*s: for $i = 1, \ldots, n$, and for $t \in [0, b_T]$,

$$dY_i(t) = \phi_{Y_i,0} b_Y\big(Y_i(t), X_i(t), t\big)\,dt + \sigma_Y\big(Y_i(t), X_i(t), t\big)\,dW_{Y,i}(t); \tag{9.1}$$

$$dX_i(t) = \phi_{X_i,0} b_X\big(X_i(t), t\big)\,dt + \sigma_X\big(X_i(t), t\big)\,dW_{X,i}(t). \tag{9.2}$$

In the above, $\phi_{Y_i,0} = \psi_{Y_i}(\theta_0)$ and $\phi_{X_i,0} = \psi_{X_i}(\theta_0)$, where ψ_{Y_i} and ψ_{X_i} are known functions; θ_0 is the true set of parameters.

Our modeled state space *SDE* is given, for $t \in [0, b_T]$ by:

$$dY_i(t) = \phi_{Y_i} b_Y\big(Y_i(t), X_i(t), t\big)\,dt + \sigma_Y\big(Y_i(t), X_i(t), t\big)\,dW_{Y,i}(t); \tag{9.3}$$

$$dX_i(t) = \phi_{X_i} b_X\big(X_i(t), t\big)\,dt + \sigma_X\big(X_i(t), t\big)\,dW_{X,i}(t), \tag{9.4}$$

where $\phi_{Y_i} = \psi_{Y_i}(\theta)$ and $\phi_{X_i} = \psi_{X_i}(\theta)$. As before, we wish to learn about the set of parameters θ. Note that for simplicity of our asymptotic analysis we assumed the same time interval $[0, b_T]$ for $i = 1, \ldots, n$. We assume that $\psi_{Y_i}(\theta) \to \bar{\psi}_Y(\theta)$ and $\psi_{X_i}(\theta) \to \bar{\psi}_X(\theta)$, as $i \to \infty$, for all $\theta \in \Theta$. Also, let $K_{Y,i}$ and $K_{X,i}$ be the relevant constants associated with (9.3) and (9.4), analogous to K_Y and K_X associated with (2.3) and (2.4), respectively. We assume that $K_{Y,i} \to \bar{K}_Y$ and $K_{X,i} \to \bar{K}_X$, as $i \to \infty$. Let $p_{T,i}(\theta_0)$ and $L_{T,i}(\theta)$ be the true and modeled likelihoods associated with the ith state space *SDE*.

9.2 A brief overview of the main asymptotic results

9.2.1 *Posterior convergence of* θ. Here the true likelihood on $[a_T, b_T]$ is of the form $\bar{p}_{n,T}(\theta_0) = \prod_{i=1}^n p_{T,i}(\theta_0) \overset{\text{a.s.}}{\sim} \prod_{i=1}^n \hat{p}_{T,i}(\theta_0)$, where

$$\hat{p}_{T,i}(\theta_0) = \exp\bigg(\frac{(b_T - a_T)K_{Y_i}\phi^2_{Y_i,0}}{2} + \phi_{Y_i,0}\sqrt{K_{Y_i}}\big(W_{Y_i}(b_T) - W_{Y_i}(a_T)\big) + (b_T - a_T)K_{X_i}\phi^2_{X_i,0}\bigg).$$

The modeled likelihood on $[a_T, b_T]$ is $\bar{L}_{n,T}(\theta) = \prod_{i=1}^n L_{T,i}(\theta) \overset{\text{a.s.}}{\sim} \prod_{i=1}^n \hat{L}_{T,i}(\theta)$, where

$$\hat{L}_{T,i}(\theta) = \exp\bigg((b_T - a_T)K_{Y_i}\phi_{Y_i}\phi_{Y_i,0} + \phi_{Y_i}\sqrt{K_{Y_i}}\big(W_{Y_i}(b_T) - W_{Y_i}(a_T)\big) - \frac{(b_T - a_T)K_{Y_i}\phi^2_{Y_i}}{2} + (b_T - a_T)K_{X_i}\phi_{X_i}\phi_{X_i,0}\bigg).$$

Let $\bar{R}_{n,T}(\theta) = \frac{\bar{L}_{n,T}(\theta)}{\bar{p}_{n,T}(\theta_0)}$. Then the following asymptotic equipartition property holds for the systems of state space *SDE*s:

$$\lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{n(b_T - a_T)}\log \bar{R}_{n,T}(\theta) = -\bar{h}(\theta),$$

almost surely, where

$$\bar{h}(\theta) = \frac{1}{2}\big[\bar{K}_Y\big(\bar{\psi}_Y(\theta) - \bar{\psi}_Y(\theta_0)\big)^2 + \bar{K}_X\big(\bar{\psi}_X(\theta) - \bar{\psi}_X(\theta_0)\big)^2 + \bar{K}_X\big(\bar{\psi}^2_X(\theta_0) - \bar{\psi}^2_X(\theta)\big)\big].$$

We define, in our current context, the following:

$$\bar{h}(A) = \operatorname*{ess\,inf}_{\theta\in A}\bar{h}(\theta);$$

$$\bar{J}(\theta) = \bar{h}(\theta) - \bar{h}(\Theta);$$

$$\bar{J}(A) = \operatorname*{ess\,inf}_{\theta\in A}\bar{J}(\theta).$$

We summarize our Bayesian convergence result in the form of the following theorem.

Theorem 17. *Let the true, data-generating model be given by* (9.1) *and* (9.2), *but let the data be modeled by* (9.3) *and* (9.4). *Consider any set* $A \in \mathcal{T}$ *with* $\pi(A) > 0$ *and* $\bar{h}(A) > \bar{h}(\Theta)$. *Then under appropriate assumptions*,

$$\lim_{n\to\infty, T\to\infty} \pi(A|\bar{\mathcal{F}}_{n,T}) = 0,$$

where $\bar{\mathcal{F}}_{n,T} = \sigma(\{Y_i(s); i = 1, \ldots, n; s \in [a_T, b_T]\})$. *If the set* A *satisfies a technical condition, then we further have*

$$\lim_{n\to\infty, T\to\infty} \frac{1}{n(b_T - a_T)} \log \pi(A|\bar{\mathcal{F}}_{n,T}) = -\bar{J}(A). \tag{9.5}$$

Sketch of the proof. The proof easily follows using the asymptotic equipartition property for the systems of state space *SDE*s and construction of appropriate sieves of the form $\bar{\mathcal{G}}_{n,T} = \{\theta : |\bar{\psi}_Y(\theta)| \leq \exp(\bar{\beta} n(b_T - a_T))\}$, which have the desired properties. Here $\bar{\beta} > 2\bar{h}(\Theta)$. □

9.2.2 *Strong consistency of the MLE of* θ.

Theorem 18. *Let the true, data-generating model be given by* (9.1) *and* (9.2), *but let the data be modeled by* (9.3) *and* (9.4). *Then, under suitable regularity conditions, the MLE of* θ, *denoted by* $\hat{\theta}_{n,T}$, *is strongly consistent in the sense that* $\hat{\theta}_{n,T} \xrightarrow{\text{a.s.}} \theta_0$.

Sketch of the proof. In this case, the *MLE* can be approximated by maximizing

$$\bar{g}_{n,T}(\theta) = \bar{g}_{Y,T}(\theta) + \bar{g}_{X,T}(\theta)$$

with respect to θ, where

$$\begin{aligned} \bar{g}_{Y,T}(\theta) &= -\frac{\bar{K}_Y}{2}(\bar{\psi}_Y(\theta) - \bar{\psi}_Y(\theta_0))^2 \\ &\quad + \sqrt{\bar{K}_Y}(\bar{\psi}_Y(\theta) - \bar{\psi}_Y(\theta_0))\frac{1}{n}\sum_{i=1}^{n} \frac{W_{Y_i}(b_T) - W_{Y_i}(a_T)}{b_T - a_T}; \qquad (9.6) \\ \bar{g}_{X,T}(\theta) &= -\frac{\bar{K}_X}{2}(\bar{\psi}_X(\theta) - \bar{\psi}_X(\theta_0))^2 + \frac{\bar{K}_X}{2}(\bar{\psi}_X^2(\theta) - \bar{\psi}_X^2(\theta_0)). \end{aligned}$$

The rest of the proof follows in the same lines as that of Theorem 3. □

9.2.3 *Asymptotic normality of the MLE of* θ.

Theorem 19. *Let the true, data-generating model be given by* (9.1) *and* (9.2), *but let the data be modeled by* (9.3) *and* (9.4). *Then, under suitable regularity conditions*,

$$\sqrt{n(b_T - a_T)}(\hat{\theta}_{n,T} - \theta_0) \xrightarrow{\mathcal{L}} N_d(0, \mathcal{I}^{-1}(\theta_0)),$$

as $n \to \infty$, $T \to \infty$. *In this case, the* (j, k)*th element of the matrix* $\mathcal{I}(\theta_0)$ *is given by*

$$\{\mathcal{I}(\theta_0)\}_{jk} = \bar{K}_Y\left[\frac{\partial\bar{\psi}_Y(\theta)}{\partial\theta_j}\frac{\partial\bar{\psi}_Y(\theta)}{\partial\theta_k}\right]_{\theta=\theta_0}.$$

Sketch of the proof. The proof of this result follows in the same way as that of Theorem 5. □

9.2.4 *Asymptotic posterior normality.* We summarize our result on asymptotic posterior normality for systems of state space *SDE*s in the form of the following theorem.

Theorem 20. *Let the true, data-generating model be given by* (9.1) *and* (9.2), *but let the data be modeled by* (9.3) *and* (9.4). *Then denoting* $\bar{\Psi}_{n,T} = \bar{\Sigma}_{n,T}^{-1/2}(\theta - \hat{\theta}_{n,T})$, *for each compact subset B of $\mathbb{R}^d$ and each $\varepsilon > 0$, the following holds under appropriate regularity conditions*:

$$\lim_{n\to\infty,T\to\infty} P_{\theta_0}\Big(\sup_{\bar{\Psi}_{n,T}\in B} \left|\pi(\bar{\Psi}_{n,T}|\bar{\mathcal{F}}_{n,T}) - \varrho(\bar{\Psi}_T)\right| > \varepsilon\Big) = 0.$$

Sketch of the proof. In this case, $\ell_{n,T}(\theta) = \log L_{n,T}(\theta)$, can be uniformly approximated by

$$\frac{1}{n(b_T - a_T)}\bar{\ell}_{n,T}(\theta) = \bar{g}_{Y,T}(\theta) + \bar{g}_{X,T}(\theta) + \frac{1}{n(b_T - a_T)}\log \bar{p}_{nT}(\theta_0),$$

for $\theta \in \Theta$. The rest of the proof follows in the same way as that of Theorem 6. □

10 Summary and discussion

In this paper, we have investigated the asymptotic properties of the *MLE* and the posterior distribution of the set of parameters associated with state space *SDE*s and random effects state space *SDE*s. In particular, we have established posterior consistency based on Shalizi (2009) and asymptotic posterior normality based on Schervish (1995). In addition, we have also established strong consistency and asymptotic normality of the *MLE* associated with our state space *SDE* models. Acknowledging the importance of discretization in practical scenarios, we have shown (in Section S-7 of the supplement (Maitra and Bhattacharya (2019))) that our results go through even with discretized data.

In the case of our random effects *SDE* models, we only required independence of the state space models for different individuals. That is, our approach and the results remain intact if the initial values for the processes associated with the individuals are different. This is in contrast with the asymptotic works of Maitra and Bhattacharya (2016) and Maitra and Bhattacharya (2015) in the context of independent but non-identical random effects models for the individuals. Although not based on state space *SDE*s, their approach required the simplifying assumption that the sequence of initial values is a convergent subsequence of some sequence in some compact space.

In fact, the relative simplicity of our current approach is due to the assumption of stochastic stability of the latent processes of our models, the key concept that we adopted in our approach to alleviate the difficulties of the asymptotic problem at hand. Specifically, we adopted the conditions of Theorem 6.2 provided in Mao (2011), as sufficient conditions of our results. Indeed, there is a large literature on stochastic stability of solutions of *SDE*s, with very many existing examples (see, for example, Mao (2011) and the references therein), which indicate that the assumption of stochastic stability is not unrealistic.

In our work, we have assumed stochastic stability of X only. If, in addition, asymptotic stability of Y is also assumed, then our results hold good by replacing (H3) in Section 4 with the following assumptions:

(H3(i)) $b_Y(0,0,t) = 0 = \sigma_Y(0,0,t)$ for all $t \geq 0$.

(H3(ii)) For every $T > 0$, there exist positive constants $K_{1,T}$, $K_{2,T}$, α_1, α_2, β_1 and β_2 such that for all $(x,t) \in \mathbb{R} \times [0,T]$,

$$K_{Y,1,T}\big(1 - \alpha_1 x^2 - \beta_1 y^2\big) \leq \frac{b_Y^2(y,x,t)}{\sigma_Y^2(y,x,t)} \leq K_{Y,2,T}\big(1 + \alpha_2 x^2 + \beta_2 y^2\big),$$

where $K_{Y,1,T} \to K_Y$ and $K_{Y,2,T} \to K_Y$ and as $T \to \infty$; K_Y being a positive constant as mentioned in (H3).

In this case the bounds of $\frac{b_Y^2(y,x,t)}{\sigma_Y^2(y,x,t)}$ are somewhat more general than in (H3) in that they depend upon both x and y, while in (H3) the bounds are independent of y.

To our knowledge, our work is the first time effort towards establishing asymptotic results in the context of state space *SDE*s, and the results we obtained are based on relatively general assumptions which are satisfied by a large class of models. Since the notion of stochastic stability is valid for any dimension of the associated *SDE*, it follows that our results admit straightforward extension to high-dimensional state space *SDE*s. Corresponding results in the multidimensional extension of the random effects is provided briefly in Section S-6 of the supplement (Maitra and Bhattacharya (2019)).

As we mentioned in the introduction, our random effects state space *SDE* model can not be interpreted as a *bona fide* random effects model from the classical perspective, and that introduction of actual random effects would complicate our method of asymptotic investigation. Also, in this article we have assumed that the diffusion coefficients are free of parameters, which is not a very realistic assumption. We are working on these issues currently, and will communicate our findings subsequently.

Acknowledgments

We are sincerely grateful to the reviewer whose constructive suggestions have led to improved quality and presentation of the manuscript. The first author gratefully acknowledges her NBHM Fellowship, Govt. of India.

Supplementary Material

Supplement to "On classical and Bayesian asymptotics in state space stochastic differential equations" (DOI: 10.1214/19-BJPS439SUPP; .pdf). Our supplement consists of the following additional details. Section S-1 provides the assumptions and the main theorem of Shalizi in the context of our state space SDE. Sections S-2 and S-3 provide the proofs of Lemma 7 and Lemma 8, respectively, while Section S-4 furnishes the proof of Theorem 10. The detailed asymptotic theory of random effects models based on state space SDEs is provided in Section S-5, while in Section S-6, the asymptotic theory for multidimensional linear random effects is established. Finally, in Section S-7, the asymptotic theory in the case of discrete data is discussed.

References

Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*. New York: John Wiley & Sons, Inc. MR0443083

Bucy, R. S. (1965). Nonlinear filtering theory. *IEEE Transactions on Automatic Control* **10**, 198.

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. New York: Springer. MR2159833

Chopin, N., Jacob, P. E. and Papaspiliopoulos, O. (2013). SMC2: An efficient algorithm for sequential analysis of state-space models. Preprint. Available at arXiv:1101.1528. MR3065473 https://doi.org/10.1111/j.1467-9868.2012.01046.x

Crisan, D. and Miguez, J. (2013). Nested particle filters for online parameter estimation in discrete-time state-space Markov models. Preprint. Available at arXiv:1308.1883. MR3779710 https://doi.org/10.3150/17-BEJ954

Crisan, D. and Rozovskii, B., eds. (2011). *The Oxford Handbook of Nonlinear Filtering*. New York: Oxford University Press. MR2882749

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. MR3066417 https://doi.org/10.1111/j.1467-9469.2012.00813.x

Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM P&S* **12**, 196–218. MR2374638 https://doi.org/10.1051/ps:2007045

Donnet, S. and Samson, A. (2013). A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. In *Advanced Drug Delivery Reviews*, 1–25. Amsterdam: Elsevier. https://doi.org/10.1018/j.addr.2013.03.005. Available at https://hal.archives-ouvertes.fr/hal-00777774.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press. MR1856951

Elliott, R. J., Aggoun, L. and Moore, J. B. (1995). *Hidden Markov Models*. *Applications of Mathematics (New York)* **29**. New York: Springer. MR1323178

Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed Ornstein–Uhlenbeck process with biological application. *Scandinavian Journal of Statistics* **7**, 200–220. MR2682296 https://doi.org/10.1111/j.1467-9469.2009.00679.x

Frydman, H. and Lakner, P. (2003). Maximum likelihood estimation of hidden Markov processes. *The Annals of Applied Probability* **13**, 1296–1312. MR2023878 https://doi.org/10.1214/aoap/1069786500

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Kailath, T. and Zakai, M. (1971). Absolute continuity and Radon–Nikodym derivatives for certain measures relative to Wiener measure. *The Annals of Mathematical Statistics* **42**, 130–140. MR0279887 https://doi.org/10.1214/aoms/1177693500

Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering* **83**, 95–108. MR0234760

Kushner, H. J. (1964). On the differential equations satisfied by conditional probability densities of Markov processes, with applications. *Journal of the Society for Industrial and Applied Mathematics Series A, on Control* **2**, 106–119. MR0180407

Leander, J., Almquist, J., Ahlström, C., Gabrielsson, J. and Jirstrand, M. (2015). Mixed effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of nicotinic acid in obese Zucker rats. *The AAPS Journal* **17**, 586–596.

Lipster, R. and Shiryaev, A. (2001). *Statistics of Random Processes I: General Theory*. New York: Springer. MR1800857

Maitra, T. and Bhattacharya, S. (2015). On Bayesian asymptotics in stochastic differential equations with random effects. *Statistics & Probability Letters* **103**, 148–159. Also available at arXiv:1407.3971. MR3350875 https://doi.org/10.1016/j.spl.2015.04.009

Maitra, T. and Bhattacharya, S. (2016). On asymptotics related to classical inference in stochastic differential equations with random effects. *Statistics & Probability Letters* **110**, 278–288. Also available at arXiv:1407.3968. MR3474768 https://doi.org/10.1016/j.spl.2015.10.001

Maitra, T. and Bhattacharya, S. (2019). Supplement to "On classical and Bayesian asymptotics in state space stochastic differential equations." https://doi.org/10.1214/19-BJPS439SUPP.

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

Martino, L., Read, J., Elvira, V. and Louzada, F. (2017). Cooperative parallel particle filters for on-line model selection and applications to urban mobility. *Digital Signal Processing* **60**, 172–185.

Maybeck, P. (1979). *Stochastic Models, Estimation and Control, Vol. 1*. London: Academic Press. MR0539145

Maybeck, P. (1982). *Stochastic Models, Estimation and Control, Vol. 2*. London: Academic Press. MR0690417 https://doi.org/10.1016/S0076-5392(08)62171-2

Møller, J. K., Bergmann, K. R., Christiansen, L. E. and Madsen, H. (2012). Development of a restricted state space stochastic differential equation model for bacterial growth in rich media. *Journal of Theoretical Biology* **305**, 78–87. MR2923641 https://doi.org/10.1016/j.jtbi.2012.04.015

Øksendal, B. (2003). *Stochastic Differential Equations*. New York: Springer. MR2001996 https://doi.org/10.1007/978-3-642-14394-6

Overgaard, R. V., Jonsson, N., Tornœ, C. W. and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics* **32**, 85–107.

Särkkä, S. (2006). Recursive Bayesian inference on stochastic differential equations. Doctoral thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology. MR2715937

Särkkä, S. (2007). On unscented Kalman filtering for state estimation of continuous-time nonlinear systems. *IEEE Transactions on Automatic Control* **52**, 1631–1641. MR2352439 https://doi.org/10.1109/TAC.2007.904453

Särkkä, S. (2012). Bayesian estimation of time-varying systems: Discrete-time systems. Technical report, Lectures notes, Aalto University.

Särkkä, S. and Sarmavuori, J. (2013). Gaussian filtering and smoothing for continuous-discrete dynamic systems. *Signal Processing* **93**, 500–510. MR3154309 https://doi.org/10.1017/CBO9781139344203

Schervish, M. J. (1995). *Theory of Statistics*. New York: Springer. MR1354146 https://doi.org/10.1007/978-1-4612-4250-5

Shalizi, C. R. (2009). Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics* **3**, 1039–1074. MR2557128 https://doi.org/10.1214/09-EJS485

Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. New York: Springer. MR2721825 https://doi.org/10.1007/978-1-4419-7865-3

Stratonovich, R. L. (1968). *Conditional Markov Processes and Their Application to the Theory of Optimal Control*. New York: American Elsevier Publishing Company, Inc. MR0221860

Urteaga, I., Bugallo, M. F. and Djuric, P. M. (2016). Sequential Monte Carlo methods under model uncertainty. In *IEEE Statistical Signal Processing Workshop (SSP)*, 15 pp.

Yan, F.-R., Zhang, P., Liu, J.-L., Tao, Y.-X., Lin, X., Lu, T. and Lin, J.-G. (2014). Parameter estimation of population pharmacokinetic models with stochastic differential equations: Implementation of an estimation algorithm. *Journal of Probability and Statistics* **2014**, 1–8. MR3280908 https://doi.org/10.1155/2014/836518

Zakai, M. (1969). On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **11**, 230–243. MR0242552 https://doi.org/10.1007/BF00536382

Interdisciplinary Statistical Research Unit
Indian Statistical Institute
203, B. T. Road
Kolkata 700108
India
E-mail: sourabh@isical.ac.in