Estimation for stochastic differential equations with mixed effects.

Valentine Genon-Catalot, Catherine Larédo

ESTIMATION FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH MIXED EFFECTS.

V. GENON-CATALOT[1] AND C. LARÉDO[2]

Abstract. We consider the long term behaviour of a one-dimensional mixed effects diffusion process $(X(t))$ with a multivariate random effect ϕ in the drift coefficient. We first study the estimation of the random variable ϕ based on the observation of one sample path on the time interval $[0,T]$ as T tends to infinity. The process $(X(t))$ is not Markov and we characterize its invariant distributions. We build moment and maximum likelihood-type estimators of the random variable ϕ which are consistent and asymptotically mixed normal with rate $\sqrt{T}$. Moreover, we obtain non asymptotic bounds for the moments of these estimators. Examples with a bivariate random effect are detailed. Afterwards, the estimation of parameters in the distribution of the random effect from N *i.i.d.* processes $(X_j(t), t \in [0,T]), j = 1, \ldots, N$ is investigated. Estimators are built and studied as both N and $T = T(N)$ tend to infinity. We prove that the convergence rate of estimators differs when deterministic components are present in the random effects. For true random effects, the rate of convergence is $\sqrt{N}$ whereas for deterministic components, the rate is $\sqrt{NT}$. Illustrative examples are given.

AMS 2000 subject classification: Primary: 62M05, 62F12 ; secondary: 60J60.

Keywords and phrases: Asymptotic mixed normality, invariant distributions, mixed-effects stochastic differential equations, parametric inference, random effects estimation.

Running title: Mixed effects estimation for SDEs.

1. Introduction

Stochastic differential equations constitute a well-established tool for modelling physical phenomena whose dynamics are affected by random noise. Some situations require a hierarchical modelling of the dynamics of X, with first a random variable ϕ (the random effect) and second, given that $\phi = \varphi$, an SDE with drift and diffusion coefficient depending on (x, φ) that rules the $X(.)$'s dynamics. This occurs in particular when it is of interest to model simultaneously the performance of several experiments or when adding this hierarchical modelling is needed to estimate population parameters. Usually, X represents the behaviour of an individual, the random components of ϕ describe the individual specificity and, given $\phi = \varphi$, the model describes a general evolution for all individuals with individual value φ (see *e.g.* Davidian and Giltinan (1995), Pinheiro and Bates (2000), Kuhn and Lavielle (2004), Nie and Yang (2005), Nie (2006), Nie (2007)). Stochastic differential equations (SDE) with mixed effects have been introduced to generalize classical ordinary differential equations models (Ditlevsen and De Gaetano (2005), Overgaard *et al.* (2005), Donnet and Samson (2008)) and also to model neuronal data (Picchini *et al.* (2010)).

The continuous-time stochastic process $X = (X(t), t \geq 0)$ with dynamics ruled by a stochastic differential equation with mixed effects satisfies

$$d\,X(t) = b(X(t), \phi)dt + \sigma(X(t), \phi)\, dW(t), \quad X(0) = \eta, \tag{1.1}$$

[1] Université Paris Descartes, MAP5, UMR CNRS 8145.
[2] Laboratoire MIA, I.N.R.A. and LPMA, Université Denis Diderot, CNRS-UMR 7599.

where W is a Brownian motion, ϕ is a random variable with distribution $\nu(d\varphi)$, (ϕ, η) is a random variable independent of W.
Recently, the estimation of the unknown distribution $\nu(d\varphi)$ of ϕ whether parametric or nonparametric based on the observation of N *i.i.d.* processes $(X_j(t), t \in [0,T]), j = 1, \ldots, N)$ distributed as (1.1) has been the subject of several contributions. Parametric estimation is investigated in Ditlevsen and De Gaetano (2005) for the specific model of mixed effects Brownian motion with drift and in Donnet and Samson (2008). In Delattre *et al.* (2012), the maximum likelihood estimator (MLE) for random effects SDEs is studied more generally for fixed T and N tending to infinity (see also Maitra and Bhattacharya (2014) for non i.i.d. sample paths). In particular, for $\phi \in \mathbb{R}$ having a Gaussian distribution with unknown mean μ and unknown variance ω^2, $b(x,\varphi) = \varphi b(x)$, $\sigma(x,\varphi) = \sigma(x)$ and $b(.), \sigma(.)$ known functions, an explicit formula for the exact likelihood is obtained and a complete asymptotic study of the exact MLE of (μ, ω^2) is given. Approximations of the likelihood are also proposed for general mixed SDEs in Picchini *et al.* (2010). For ϕ real-valued and $\nu(d\varphi) = f(\varphi)d\varphi$, nonparametric estimation of the density f is investigated in Comte *et al.* (2013), under the asymptotic framework that both N and $T = T(N)$ tend to infinity in such a way that $T(N)/N$ tends to infinity. Only two specific models are studied: $b(x,\varphi) = \varphi b(x), b(x,\varphi) = \varphi + b(x), \sigma(x,\varphi) = \sigma(x)$. Nonparametric adaptive estimation of a mixed-effect ϕ in the drift coefficient of an Ornstein-Uhlenbeck process has been investigated in Dion (2014). The general case of mixed-effects both in the drift and diffusion coefficient of multidimensional diffusions is also addressed from a practical view point in Picchini and Ditlevsen (2011) with implementation of estimators. For an SDE with a linear mixed effect in the diffusion coefficient, the estimation of population parameters is investigated in Delattre *et al.* (2014).
Our focus here is first to study the long term behaviour of a one-dimensional mixed effects SDE,

$$dX(t) = b(X(t), \phi)dt + \sigma(X(t))\, dW(t), \quad X(0) = \eta, \tag{1.2}$$

where $\sigma(x,\varphi) = \sigma(x)$ is a known real-valued function defined on $\mathbb{R}$, W is a standard Wiener process and (ϕ, η) is a random variable independent of W taking values in $\Phi \times \mathbb{R}$ with $\Phi \subset \mathbb{R}^d$. We study the properties of model (1.2). The process $(X(t))$ is not Markov and has infinitely many invariant distributions when ϕ is random. We are interested in the estimation of the random variable ϕ based on the observation of one trajectory $(X(t), t \in [0,T])$ as T tends to infinity. Let us point out that it is not a statistical problem in the usual sense since ϕ is not a parameter but a random variable. Nevertheless, several recent papers now deal with the estimation of random variables. One of the most popular examples concerns the estimation of the stochastic integrated volatility in semimartingales models based on high frequency observations of the sample path (see Kessler *et al.* (2012), Chapters 2 and 3). To build estimators of ϕ, two methods are investigated, a moment type method and a maximum likelihood approach. Both lead to estimators which are consistent and asymptotically mixed normal at rate $\sqrt{T}$. These two methods are applied on classical examples: the Ornstein-Uhlenbeck and the C.I.R. processes with a bivariate random effect.
Second, we investigate the framework where N *i.i.d.* processes $(X_j(t), t \in [0,T], j = 1, \ldots, N)$ are observed and each sample path X_j is associated with a random effect ϕ_j. We are interested in estimating unknown parameters in the distribution $\nu(d\varphi)$ of the ϕ_j's (the population parameters). The results of the first part lead to consider a two-stage procedure. First, build for each j, an estimator of ϕ_j from the trajectory $(X_j(t), t \leq T)$. Then, use a plug-in technique to derive estimators of unknown parameters of $\nu(d\varphi)$ and study their asymptotic properties as both N and T tend to infinity. Using moment type estimators, we prove that the convergence rate of estimators differs when deterministic components (fixed effects) are present in the ϕ_j's. For true

random effects, the rate of convergence is $\sqrt{N}$ whereas for deterministic components, the rate is $\sqrt{NT}$. This is consistent with Nie's results (Nie (2007)). The case of the Ornstein-Uhlenbeck process with two random effects is especially detailed, *i.e.* the observed processes are given by $dX_j(t) = -\phi_{2,j}(X_j(t) - \phi_{1,j})dt + dW_j(t), X_j(0) = \eta_j$ where $\phi_j = (\phi_{1,j}, \phi_{2,j}) \in \mathbb{R} \times (0, +\infty)$. Estimators of $\mathbb{E}\phi_1$ and $\mathbb{E}\phi_2^{-1}$ are studied both when the two components of ϕ_j are random or when only one is random while the other one is deterministic. The C.I.R. process (square root diffusion) with two random effects is also studied.
Next, we choose a specific model to develop a maximum likelihood type method. We concentrate on the Ornstein-Uhlenbeck model with one multiplicative random effect in the drift, *i.e.* $dX_j(t) = -\phi_j X_j(t)dt + dW_j(t), X_j(0) = \eta_j$ where $\phi_j \in (0, +\infty)$. Assuming that ϕ_j follows a Gamma distribution with unknown parameters (a, λ), we define appropriate estimators of ϕ_j and $\log \phi_j$ that we plug in the likelihood function associated with *i.i.d.* Gamma random variables. This defines a contrast. We prove that, as both $N, T = T(N)$ tend to infinity, the associated maximum contrast estimator is asymptotically equivalent to the maximum likelihood estimator based on the observation of $(\phi_j, j = 1, \dots, N)$ provided that N/T tends to 0.
The paper is organized as follows. Section 2 contains notations, assumptions and results on the invariant distributions of (1.2) (see Proposition 2.1). Section 3 is concerned with the estimation of the random effect ϕ when only one sample path is continuously observed on $[0, T]$ as $T \to \infty$. Section 3.1 is devoted to moment type estimators. Classical theorems for ergodic diffusions with no random effects are interpreted and restated for the case of diffusions with random effects. We study the asymptotic behaviour of $\frac{1}{T}\int_0^T g(X(s))ds$ for g a given function and obtain non asymptotic moment bounds (Propositions 3.1, 3.2, 3.3). These results are applied to the Ornstein-Uhlenbeck process and the square-root process with two random effects (Propositions 3.4, 3.5, 3.6). The maximum likelihood type estimators are presented in Section 3.2 (Proposition 3.7). We present a general truncation technique to study the moments of these estimators (Proposition 3.8). The Ornstein-Uhlenbeck example illustrates these results (Corollary 3.1). In Section 4, we consider that observations are N *i.i.d.* trajectories on $[0, T]$. It is therefore possible to estimate unknown parameters in the distribution of random effects (the population parameters). Moment type estimators are studied (Theorem 4.1). These results are then applied to the Ornstein-Uhlenbeck process with two random effects (Proposition 4.1). The maximum likelihood approach is detailed for the Ornstein-Uhlenbeck process with one multiplicative random effect with Gamma distribution (Proposition 4.2). Section 5 contains extensions and concluding remarks. Proofs are gathered in Section 6.

2. Notations, assumptions and preliminary result.

Consider a real valued stochastic process $(X(t), t \geq 0)$, with dynamics ruled by (1.2). The Wiener process W and the r.v. (ϕ, η) are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_t, t \geq 0)$ satisfying the usual conditions. The couple (ϕ, η) is assumed $\mathcal{F}_0$-measurable and W is a $(\mathcal{F}_t, t \geq 0)$-Brownian motion. We use the classical notations for random effects models, *i.e.* ϕ for the random variable and φ for the fixed value. We denote by X^{φ,x_0} the process solution of the SDE with fixed φ and initial condition x_0:

$$d\,X^{\varphi,x_0}(t) = b(X^{\varphi,x_0}(t), \varphi)dt + \sigma(X^{\varphi,x_0}(t))\,dW(t), \quad X^{\varphi,x_0}(0) = x_0 \tag{2.1}$$

We introduce now assumptions on model (1.2).
[**H1**] The real valued function $(x, \varphi) \to b(x, \varphi)$ (resp. $x \to \sigma(x)$) is continuous on $\mathbb{R} \times \mathbb{R}^d$ (resp. $\mathbb{R}$) and such that the process $(\phi, X(t))$ where $X(t)$ is given by (1.2) is well defined and adapted to the filtration $(\mathcal{F}_t)$ for all initial condition (ϕ, η).

[**H2**] (i) There exists an open subset Φ of $\mathbb{R}^d$ and an interval $(\ell, r) \subset \mathbb{R}$ such that $\forall x \in (\ell, r), \sigma^2(x) > 0$, and $\forall \varphi \in \Phi$, the scale density of (2.1)

$$s_\varphi(x) = \exp\left(-2\int_{x_0}^x \frac{b(u,\varphi)}{\sigma^2(u)}du\right), \quad x_0 \in (\ell, r)$$

satisfies $\int_\ell s_\varphi(x)dx = \int^r s_\varphi(x)dx = +\infty$ and the speed density $m_\varphi(x) = 1/(\sigma^2(x)s_\varphi(x))$ of (2.1) satisfies $M(\varphi) := \int_\ell^r m_\varphi(x)dx < +\infty$.
[**H2**] (ii) The function $(x, \varphi) \to b(x, \varphi)$ is C^1 on $(\ell, r) \times \mathbb{R}^d$ and $x \to \sigma(x)$ is C^2 on (ℓ, r).
[**H2**] (iii) The random variable ϕ takes values in Φ. We denote its distribution by $\nu(d\varphi)$.
[**H3**] The random couple (ϕ, η) has distribution $\pi^\nu(d\varphi, dx) = \nu(d\varphi) \otimes \pi_\varphi(x)dx$ where

$$\pi_\varphi(x) = 1_{(\ell,r)}(x)\frac{m_\varphi(x)}{M(\varphi)}. \tag{2.2}$$

Standard regularity assumptions on $b(x, \varphi)$ and $\sigma(x)$ ensure (H1) (see *e.g.* Karatzas and Shreve (2000) for general results and Comte *et al.* (2013) for the specific context of SDEs with mixed effects). Under [**H1**], the process $(\phi, X(t))$ is strong Markov. The Markov property is enlighted by looking at (1.2) as a system:

$$\begin{aligned} d\,\phi(t) &= 0, \quad \phi(0) = \phi, \\ d\,X(t) &= b(X(t), \phi(t))dt + \sigma(X(t))\,dW(t), \quad X(0) = \eta. \end{aligned}$$

Thus, $(\phi, X(t))$ is a two dimensional diffusion process with initial distribution

$$\mu^\nu(d\varphi, dx) = \nu(d\varphi) \otimes \mu_\varphi(dx) \tag{2.3}$$

where ν is the distribution of ϕ and μ_φ is a distribution on (ℓ, r) corresponding to the conditional distribution of η given $\phi = \varphi$. Assumption [**H2**] is classical for existence and uniqueness of invariant distributions for SDEs with non random parameters (see *e.g.* Kessler *et al.* (2012)). The unique invariant density of model (2.1) is (2.2). Denote by $p_t^\varphi(x, y)$ the transition density of model (2.1). Assumptions [**H1**]-[**H2**] imply the measurability of

$$(x, y, \varphi) \to p_t^\varphi(x, y) \quad \text{and} \quad (x, \varphi) \to \pi_\varphi(x). \tag{2.4}$$

For $g : (\ell, r) \to \mathbb{R}$ a positive or π_φ-integrable Borel function, $\varphi \to \pi_\varphi(g) := \displaystyle\int_\ell^r g(x)\pi_\varphi(x)dx$ is measurable on Φ. Therefore, we can define the random variables

$$\pi_\phi(g) = \int_\ell^r g(x)\pi_\phi(x)dx \qquad \text{and the sets} \tag{2.5}$$

$$L^1(\Phi) = \{g : (\ell, r) \to \mathbb{R},\ \text{Borel}\ \ \forall \varphi \in \Phi,\ \pi_\varphi(|g|) < +\infty\}, \tag{2.6}$$

$$L^2(\Phi) = \{g : (\ell, r) \to \mathbb{R},\ \text{Borel}\ \ \forall \varphi \in \Phi,\ \pi_\varphi(g^2) < +\infty\}. \tag{2.7}$$

For $f \in C^2((\ell, r))$, let us introduce the infinitesimal generator of (2.1):

$$\mathcal{L}_\varphi f(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x, \varphi)f'(x) = \frac{1}{2m_\varphi(x)}\left(\frac{f'}{s_\varphi}\right)'(x). \tag{2.8}$$

To $g \in L^2(\Phi)$, we associate a function $F_{\varphi,g}$ such that $(x, \varphi) \to F_{\varphi,g}(x) \in C^2((\ell, r) \times \Phi)$, for all $\varphi \in \Phi$, $\pi_\varphi F_{\varphi,g}^2 < +\infty$, the limit of $F'_{\varphi,g}/s_\varphi$ as $x \to \ell, r$ is nul, and

$$\mathcal{L}_\varphi F_{\varphi,g} = -(g - \pi_\varphi(g)). \tag{2.9}$$

From (2.8), we have $F'_{\varphi,g}(x) = -2s_\varphi(x)\int_\ell^x (g(u)-\pi_\varphi(g))m_\varphi(u)du$. If moreover g satisfies

$$\int_\ell^r (F'_{\varphi,g}(x)\sigma(x))^2\pi_\varphi(x)dx := V_\varphi(g) < +\infty, \tag{2.10}$$

then, for $<.,.>_\varphi$ denoting the scalar product of $L^2(\pi_\varphi)$,

$$-2 < \mathcal{L}_\varphi F_{\varphi,g}, F_{\varphi,g} >_\varphi = V_\varphi(g), \tag{2.11}$$

(see Genon-Catalot *et al.* (2000)) for details on the existence of such function $F_{\varphi,g}$ and (2.9)-(2.11)). In the case of two functions g, h such that $V_\varphi(g) < +\infty, V_\varphi(h) < +\infty$, we set

$$V_\varphi(g) = V_\varphi(g,g) \quad \text{and} \quad V_\varphi(g,h) = \int_\ell^r F'_{\varphi,g}(x)F'_{\varphi,h}(x)\sigma^2(x)\pi_\varphi(x)dx. \tag{2.12}$$

Proposition 2.1. *Assume* [**H1**] *and consider the process* $(\phi, X(t))$ *given by (1.2) where the initial variable* (ϕ,η) *has distribution* μ^ν *(see (2.3)) with* ν *a distribution on* Φ *and* μ_φ *a distribution on* (ℓ, r)*. Then, given that* $\phi = \varphi$*, the conditional distribution of* $(X(t))$ *is identical to the distribution of* X^φ *given by*

$$dX^\varphi(t) = b(X^\varphi(t)),\varphi)dt + \sigma(X^\varphi(t))dW(t), \quad X^\varphi(0) \sim \mu_\varphi,\ X^\varphi(0) \ \textit{independent of } W. \tag{2.13}$$

Assume [**H1**]-[**H2**]*. Then, the distribution* $\pi^\nu(d\varphi, dx) = \nu(d\varphi)\otimes\pi_\varphi(x)dx$ *on* $\Phi\times(\ell,r)$ *is an invariant distribution for the Markov process* $(\phi, X(t))$*.*

The proof of Proposition 2.1 is easily obtained by comparing the finite-dimensional distributions and using that, by the Markov property of $(\phi,(X(t))$, the conditional distribution of $(X(t))$ given $\phi = \varphi, X(0) = x_0$ is identical to the distribution of X^{φ,x_0} satisfying (2.1). When (ϕ,η) has distribution π^ν (assumption [**H3**]), the process $(\phi, X(t))$ is strictly stationary with marginal distribution π^ν. The Markov process $(\phi, X(t))$ having infinitely many invariant distributions is not ergodic. The marginal distribution of $X(0)$ when $(\phi, X(0))$ has distribution π^ν is a mixture distribution with density given by $p^\nu(x) = \int_\Phi \nu(d\varphi)\pi_\varphi(x)$. The conditional distribution of X given $\phi = \varphi$ is the distribution of the stationary and ergodic solution of the SDE with fixed φ.

3. Estimation of a random effect in the drift.

We investigate two methods to estimate the random variable ϕ based on the observation of one trajectory $(X(t), t\in[0,T])$ in the asymptotic framework $T\to+\infty$. For this, we need to extend some classical limit theorems for ergodic diffusions with no random effects to the case of diffusions with random effects.

3.1. Moment method.

We consider the estimation of the random variable $\pi_\phi(g)$ (see (2.5)) by

$$\widetilde{\pi_\phi(g)} = \frac{1}{T}\int_0^T g(X(s))ds. \tag{3.1}$$

For a vector of functions $\underline{g} = (g_1,\dots,g_k)'$, we define the vector

$$\widetilde{\pi_\phi(\underline{g})} = (\frac{1}{T}\int_0^T g_i(X(s))ds, i = 1,\dots,k)'. \tag{3.2}$$

Proposition 3.1. *Assume* [**H1**] – [**H2**] *and consider the process* $(\phi, X(t))$ *defined in (1.2) where the initial variable* (ϕ, η) *has distribution* μ^ν *(see (2.3)). For* $g \in L^1(\Phi)$*, almost surely, as* T *tends to infinity,*

$$\widetilde{\pi_\phi(g)} \to \pi_\phi(g). \tag{3.3}$$

For the following two propositions, we consider $g, g_1, \ldots, g_k \in L^2(\Phi)$ and let $F_{\varphi,g}, F_{\varphi,g_i}, i = 1, \ldots k$ be defined by (2.9) and satisfy (2.10) for all $\varphi \in \Phi$.

Proposition 3.2. *Assume* [**H1**] – [**H2**] *and consider the process* $(\phi, X(t))$ *given by (1.2) where the initial variable* (ϕ, η) *has distribution* μ^ν *(see (2.3)). Then,*
$\sqrt{T}(\widetilde{\pi_\phi(g)} - \pi_\phi(g))$ *converges stably in law to the mixed normal law* $M\mathcal{N}(0, V_\phi(g))$, i.e. *the law of* $\varepsilon\,(V_\phi(g))^{1/2}$ *with* $\varepsilon \sim \mathcal{N}(0,1)$ *independent of* $V_\phi(g)$.
Assume that the matrix $V_\varphi(\underline{g}) := (V_\varphi(g_i, g_j))_{1\le i,j\le k}$ *is invertible for all* $\varphi \in \Phi$ *(see (2.12)). Then, the random vector*

$$\sqrt{T}(\widetilde{\pi_\phi(\underline{g})} - \pi_\phi(\underline{g}))$$

converges stably in law to the mixed normal law $M\mathcal{N}_k(0, V_\phi(\underline{g}))$.

From now on, we assume that [**H1**] – [**H2**] – [**H3**] hold *i.e.* that the process $(X(t))$ is in stationary regime. We thus consider $(X(t))$ be given by (1.2) where the initial variable (ϕ, η) has distribution π^ν with π_φ defined in (2.2) (stationary regime). Note that conditionally on ϕ, $(X(t))$ is also stationary.
Consider the following condition for $g \in L^2(\Phi)$ and $F_{\varphi,g}$ defined by (2.9) and satisfying (2.10) for all $\varphi \in \Phi$. :
• (**K**γ): For $\gamma \ge 1$ and for all $\varphi \in \Phi$,

$$K_\gamma(\varphi, g) := \pi_\varphi(F_{\varphi,g}^{2\gamma}) + \pi_\varphi((F'_{\varphi,g}\sigma)^{2\gamma}) < +\infty. \tag{3.4}$$

The previous results (Propositions 3.1-3.2) show that the random variable $\pi_\phi(g)$ for a given function g is consistently estimated by $\widetilde{\pi_\phi(g)}$. As the process is in stationary regime, we have $\mathbb{E}(g(X(s))|\phi) = \pi_\phi(g)$ for all s, so the estimators are conditionally unbiased:

$$\mathbb{E}(\widetilde{\pi_\phi(g)}|\phi) = \pi_\phi(g). \tag{3.5}$$

Define, using (2.10):

$$\rho(g, \phi) = 2\pi_\phi(F_{\phi,g}^2) + 2\left(2V_\phi(g)\pi_\phi(F_{\phi,g}^2)\right)^{1/2}. \tag{3.6}$$

Proposition 3.3. *Assume* [**H1**] – [**H2**] – [**H3**]. *For* g *satisfying condition* (**K**γ),

$$\mathbb{E}\left([\sqrt{T}(\widetilde{\pi_\phi(g)} - \pi_\phi(g))]^{2\gamma}|\phi\right) \le c_\gamma K_\gamma(\phi, g), \tag{3.7}$$

$$\mathbb{E}\left([\sqrt{T}(\widetilde{\pi_\phi(g)} - \pi_\phi(g))]^{2}|\phi\right) = V_\phi(g) + Z_T(g, \phi), \tag{3.8}$$

where c_γ *is a numerical constant and* $|Z_T(g,\phi)| \le \frac{1}{\sqrt{T}}\rho(g,\phi)$.
For two functions g, h *such that* $V_\varphi(g) < +\infty, V_\varphi(h) < +\infty$ *for all* $\varphi \in \Phi$*, we have*

$$\mathbb{E}\left([T(\widetilde{\pi_\phi(g)} - \pi_\phi(g))(\widetilde{\pi_\phi(h)} - \pi_\phi(h))]|\phi\right) = V_\phi(g,h) + Z_T(g,h,\phi), \tag{3.9}$$

where $|Z_T(g,h,\phi)| \le \frac{1}{\sqrt{T}}(\rho(g+h,\phi) + \rho(g-h,\phi))$.

Remark 3.1. *Note that the upper bounds for $Z_T(g,\phi), Z_T(g,h,\phi)$ can be improved on examples (see below) where these quantities are found of order $1/T$ by exact computation.*

We illustrate the method on examples.

3.1.1. *Ornstein-Uhlenbeck process.* Consider the Ornstein-Uhlenbeck process with two random effects:

$$dX(t) = -\phi_2(X(t)-\phi_1)dt + dW(t), \quad X(0)=\eta. \tag{3.10}$$

Let $\phi=(\phi_1,\phi_2)\in\Phi=\mathbb{R}\times(0,+\infty)$. The invariant distribution of the model with fixed φ is $\pi_\varphi(dx)=\mathcal{N}(\varphi_1,1/(2\varphi_2))(dx)$. Assume that (ϕ,η) has distribution $\nu(d\varphi)\otimes\pi_\varphi(dx)$ with ν a distribution on Φ. Define the estimators of ϕ_1, $\tau=\phi_2^{-1}$, ϕ_2 by

$$\widetilde{\phi}_1=\frac{1}{T}\int_0^T X(s)ds, \quad \widetilde{\tau}=2(\frac{1}{T}\int_0^T X^2(s)ds-(\widetilde{\phi}_1)^2), \quad \widetilde{\phi}_2=\frac{1}{\widetilde{\tau}}\,.$$

Note that τ is a natural parameter here as it is proportional to the variance of π_ϕ.

Proposition 3.4. *As T tends to infinity, $(\widetilde{\phi}_1,\widetilde{\phi}_2)$ converges* a.s. *to (ϕ_1,ϕ_2) and $\sqrt{T}(\widetilde{\phi}_1-\phi_1,\widetilde{\phi}_2-\phi_2)'$ is asymptotically $\mathcal{MN}_2(0,J(\phi))$ where*

$$J(\phi)=\begin{pmatrix}\phi_2^{-2} & 0\\ 0 & 2\phi_2\end{pmatrix}.$$

It is interesting to look at the case when one component of ϕ is deterministic. If ϕ_2 is deterministic, so is $J(\phi)$ and the asymptotic distribution is Gaussian. On the contrary, if $\phi_1=m$ is deterministic, $\sqrt{T}(\widetilde{\phi}_1-m)$ is still asymptotically mixed normal $\mathcal{MN}(0,\phi_2^{-2})$.
For the conditional moments, the following holds ($\lesssim$ means $\leq$ up to a constant):

Proposition 3.5. *Let $\gamma\geq 1$, then, with $\tau=\phi_2^{-1}$:*

$$\mathbb{E}((\widetilde{\phi}_1-\phi_1)^2|\phi)=\frac{\tau^2}{T}+Z_T, \qquad |Z_T|\leq\frac{\tau^3}{T^2}. \tag{3.11}$$

$$\mathbb{E}(\widetilde{\phi}_1-\phi_1)^{2\gamma}|\phi)\lesssim T^{-\gamma}(\tau^{3\gamma}+\tau^{2\gamma}), \tag{3.12}$$

$$\mathbb{E}(\widetilde{\tau}-\tau|\phi)=-\frac{2\tau^2}{T}+\frac{2\tau^3(1-e^{-T/\tau})}{T^2}, \qquad \mathbb{E}((\widetilde{\tau}-\tau)^2|\phi)=\frac{2\tau^3}{T}+\zeta_T, \tag{3.13}$$

$$\mathbb{E}((\widetilde{\tau}-\tau)^{2\gamma}|\phi)\lesssim T^{-\gamma}(\tau^{2\gamma}+\tau^{3\gamma}+\tau^{4\gamma}+\tau^{6\gamma}+\phi_1^{2\gamma}\tau^{3\gamma}+\phi_1^{2\gamma}\tau^{2\gamma}), \tag{3.14}$$

where $|\zeta_T|\leq\frac{1}{T\sqrt{T}}C(\phi)$ and $C(\phi)$ depends on powers of ϕ_1,τ.

Note that $\widetilde{\tau}$ is not conditionally unbiased. If $\mathbb{E}(\tau^{3\gamma}+\tau^{2\gamma})<+\infty$, we have $\mathbb{E}(\widetilde{\phi}_1-\phi_1)^{2\gamma})\leq CT^{-\gamma}$. Thus, the constraint only depends on the distribution of ϕ_2. If the r.h.s. of (3.14) has a finite expectation, $\mathbb{E}(\widetilde{\tau}-\tau)^{2\gamma}\leq CT^{-\gamma}$.

3.1.2. *The C.I.R. process.* Consider the C.I.R. process with two random effects:

$$dX(t) = -\phi_2(X(t) - \phi_1)dt + \sqrt{X(t)}dW(t), \quad X(0) = \eta$$

Let $\phi = (\phi_1, \phi_2) \in \Phi = \{2\varphi_1\varphi_2 \geq 1, \varphi_2 > 0\}$, where we standardly apply **[H2]** to obtain Φ. Assume that (ϕ, η) has distribution $\nu(d\varphi) \otimes \pi_\varphi(dx)$ for ν a distribution on Φ and $\pi_\varphi = G(2\varphi_1\varphi_2, 2\varphi_2)$ the Gamma distribution with tail parameter $2\varphi_1\varphi_2$ and scale parameter $2\varphi_2$. Consider the estimators of $\phi_1, \beta = \phi_1/(2\phi_2), \phi_2$ given by

$$\widetilde{\phi}_1 = \frac{1}{T}\int_0^T X(s)ds, \quad \widetilde{\beta} = \frac{1}{T}\int_0^T X^2(s)ds - (\widetilde{\phi}_1)^2, \quad \widetilde{\phi}_2 = \frac{\widetilde{\phi}_1}{2\widetilde{\beta}}.$$

Proposition 3.6. *As T tends to infinity, $(\widetilde{\phi}_1, \widetilde{\phi}_2)$ converges* a.s. *to (ϕ_1, ϕ_2) and $\sqrt{T}(\widetilde{\phi}_1 - \phi_1, \widetilde{\phi}_2 - \phi_2)'$ is asymptotically $\mathcal{MN}_2(0, J_{cir}(\phi))$ where*

$$J_{cir}(\phi) = \begin{pmatrix} \phi_1\phi_2^{-2} & -\phi_2^{-1} \\ -\phi_2^{-1} & 4\phi_2 + 2\phi_1^{-1} \end{pmatrix}.$$

For the conditional moments, the following holds, for $\gamma \geq 1$, :

$$\mathbb{E}((\widetilde{\phi}_1 - \phi_1))^2|\phi) = \frac{\phi_1}{T\phi_2^2} + Z_T, \qquad |Z_T| \leq \frac{\phi_1}{T^2\phi_2^3}, \tag{3.15}$$

$$\mathbb{E}(\widetilde{\phi}_1 - \phi_1)^{2\gamma}|\phi) \lesssim T^{-\gamma}\phi_2^{-2\gamma}(\phi_1^\gamma + \phi_1^{2\gamma}). \tag{3.16}$$

Even if either ϕ_1 or ϕ_2 is deterministic, $J_{cir}(\phi)$ is still stochastic. Note that, as $2\phi_1\phi_2 \geq 1$, $\phi_1\phi_2^{-3} \leq 2\phi_1^2\phi_2^{-2}$. Thus, (3.15) and (3.16) are not contradictory. The conditional moments of $\widetilde{\beta}$ could also be obtained with additional computations.

3.2. **Maximum likelihood method.** We can also use a maximum likelihood approach.

3.2.1. *General drift.* Assumption **[H2]** implies that $\mathbb{P}(\forall s \geq 0, \sigma(X(s)) > 0|\phi = \varphi) = 1$ for all $\varphi \in \Phi$. Therefore, $\mathbb{P}(\forall s \geq 0, \sigma(X(s)) > 0) = 1$ and the function

$$L_T(\psi) = \exp\left(\int_0^T \frac{b(X(s), \psi)}{\sigma^2(X(s))}dX(s) - (1/2)\int_0^T \frac{b^2(X(s), \psi)}{\sigma^2(X(s))}ds\right)$$

is well-defined for all $\psi \in \Phi$. Thus, we can introduce:

$$\hat{\phi}_T = \hat{\phi}_T(X(s), s \leq T) = \text{Argsup}_\psi L_T(\psi). \tag{3.17}$$

The functional $\hat{\varphi}_T = \hat{\phi}_T(X^{\varphi,x_0}(s), s \leq T)$ is the exact maximum likelihood of the true value φ based on the observation $(X^{\varphi,x_0}(s), s \leq T)$. Consider the additional assumptions:
[**H4**] The function $(x, \varphi) \to b(x, \varphi)$ is twice continuously differentiable on $(\ell, r) \times \mathbb{R}^d$.
[**H5**] The following matrix $I(\varphi) = (I_{jk}(\varphi))$ is invertible for $\varphi \in \Phi$:

$$I_{jk}(\varphi) = \int_{(\ell,r)} \frac{(\partial b/\partial\varphi_j)(x, \varphi)(\partial b/\partial\varphi_k)(x, \varphi)}{\sigma^2(x)}\pi_\varphi(x)dx.$$

Proposition 3.7. *Assume* **[H1]**-**[H5]**, *that the maximum likelihood estimator $\hat{\varphi}_T$ based on the observation $(X^{\varphi,x_0}(s), s \leq T)$ with fixed φ is consistent and that $\sqrt{T}(\hat{\varphi}_T - \varphi)$ converges in distribution to $\mathcal{N}(0, I^{-1}(\varphi))$ for all $\varphi \in \Phi$. Then, $\hat{\phi}_T$ converges in probability to ϕ and $\sqrt{T}(\hat{\phi}_T - \phi)$ converges in distribution to the mixed normal law $\mathcal{MN}(0, I^{-1}(\phi))$.*

The consistency and the asymptotic normality of the estimator $\hat{\varphi}_T$ is detailed in *e.g.* in Kutoyants (2004). The proof of Proposition 3.7 is obtained by conditioning and omitted.
To illustrate Proposition 3.7, consider a multidimensional linear random effect in the drift, $b(.,\varphi) = \sum_{i=1}^d \varphi^j b_j(.)$ with $\varphi = (\varphi^j, j = 1, \dots, d) \in \mathbb{R}^d$ and where the b_j's are $C^1(\mathbb{R})$ functions. Define the vector

$$U_T = \left(\int_0^T \frac{b_1(X(s))}{\sigma^2(X(s))} dX(s) \quad \dots \quad \int_0^T \frac{b_d(X(s))}{\sigma^2(X(s))} dX(s) \right)'$$

and the matrix $\mathcal{I}_T = (\mathcal{I}_T^{k,\ell})_{1\le k,\ell\le d})$ with $\mathcal{I}_T^{k,l} = \displaystyle\int_0^T \frac{b_k(X(s))b_\ell(X(s))}{\sigma^2(X(s))} ds$.
The estimator $\hat{\phi}_T$ of ϕ is defined by the equation: $\mathcal{I}_T\, \hat{\phi}_T = U_T$.
Assume that, for $k = 1, \dots, d$, $\sigma^{-1} b_k$ belongs to $L^2(\phi)$ (see (2.7)) and set $\mathcal{I}(\phi) = \left(\pi_\phi(b_k\, b_\ell/\sigma^2)\right)$. Then, $\mathbb{P}(\mathcal{I}_T/T \to \mathcal{I}(\phi)) = 1$. Denote by Φ_0 the subset of Φ such that $\mathcal{I}(\varphi)$ invertible for all $\varphi \in \Phi_0$. On Φ_0, $\sqrt{T}(\hat{\phi}_T - \phi)$ converges in distribution to $\mathcal{MN}(0, \mathcal{I}(\phi)^{-1})$.

In particular, for the Ornstein-Uhlenbeck example (section 3.1.1), set $\psi_1 = \phi_1\phi_2, \psi_2 = \phi_2$, $b_1(x) = 1, b_2(x) = -x$, $\phi = F(\psi) = ((\psi_1/\psi_2) \quad \psi_2)'$. The computation of $\mathcal{I}(\psi)$ and $J(\phi) = DF(\psi)\mathcal{I}^{-1}(\psi)(DF(\psi))'$ with $\psi = F^{-1}(\phi)$ yields

$$J(\phi) = \begin{pmatrix} \phi_2^{-2} & 0 \\ 0 & 2\phi_2 \end{pmatrix}.$$

Thus, $\sqrt{T}(\hat{\phi}_T - \phi)$ is asymptotically $\mathcal{MN}(0, J(\phi))$ with the same matrix $J(\phi)$ as in Proposition 3.4.

3.2.2. *Truncation technique.* For Section 4, we need non asymptotic bounds for the moments of estimators of ϕ. The exact moments of $\hat{\phi}_T$ are untractable. So we present a truncation technique which allows to study these moments. We detail it on the example of one linear random effect in the drift. Let $b(.,\varphi) = \varphi b(.)$ with $\varphi \in \Phi \subset \mathbb{R}$ with $b \in C^1(\mathbb{R})$. The estimator of ϕ is:

$$\hat{\phi}_T = \frac{\int_0^T \frac{b(X(s))}{\sigma^2(X(s)))} dX(s)}{\int_0^T \frac{b^2(X(s))}{\sigma^2(X(s))} ds}. \tag{3.18}$$

To obtain simple bounds, we introduce a truncated version of $\hat{\phi}_T$ (see (3.18)) and define, for k a constant:

$$\hat{\phi}_T^{(k)} = \hat{\phi}_T\, 1_{(\frac{V_T}{T} \ge \frac{k}{\sqrt{T}})} \quad \text{where} \quad V_T = \int_0^T h(X(s))ds, \quad h(x) = \frac{b^2(x)}{\sigma^2(x)}. \tag{3.19}$$

Note that the event $(\frac{V_T}{T} < \frac{k}{\sqrt{T}})$ has small probability (see below) and we can prove that the two estimators $\hat{\phi}_T, \hat{\phi}_T^{(k)}$ are asymptotically equivalent in the sense that $\sqrt{T}(\hat{\phi}_T - \hat{\phi}_T^{(k)}) = o_P(1)$.

Proposition 3.8. *Assume* [**H1**] – [**H3**]. *Assume that, for* $\gamma \ge 1$, *for all* $\varphi \in \Phi$, $\pi_\varphi h^{2\gamma} < +\infty$. *Assume that* $F_{\varphi,h}$ *defined in (2.9) satisfies Condition* (**K**2γ). *Then,*

$$\mathbb{E}(\hat{\phi}_T^{(k)} - \phi)^{2\gamma} | \phi) \le\ T^{-\gamma} C(\gamma, \phi),$$

where

$$C(\gamma,\phi) \lesssim \frac{1}{[\pi_\phi(h)]^{2\gamma}} \left(\pi_\phi(h^\gamma) + (\pi_\phi(h^{2\gamma}) K_{2\gamma}(\phi,h))^{1/2} + \phi^{2\gamma}(1 + K_\gamma(\phi,h)) \right).$$

Moreover, for all $p \geq 1$,

$$\mathbb{P}(\frac{V_T}{T} < \frac{k}{\sqrt{T}}|\phi) \lesssim T^{-p}\,(\pi_\phi h)^{-2p}\,(K_p(\phi,h)+1)\,. \tag{3.20}$$

We can illustrate the above conditions for the Ornstein-Uhlenbeck process.

Corollary 3.1. *Consider the Ornstein-Uhlenbeck process with one random effect* $dX(t) = -\phi X(t)dt + dW(t), \quad X(0) = \eta$, *with* $\phi \in (0,+\infty)$ *and* $(\phi,\eta) \sim d\nu(\varphi) \otimes \mathcal{N}(0,(2\varphi)^{-1})$ *where* ν *is a distribution on* $(0,+\infty)$. *Then, for all* $\gamma \geq 1$,

$$\mathbb{E}((\hat{\phi}_T^{(k)} - \phi)^{2\gamma}|\phi) \lesssim\ T^{-\gamma}(1+\phi^{\gamma}+\phi^{4\gamma}+\phi^{-2\gamma}+\phi^{-3\gamma}).$$

with $\hat{\phi}_T^{(k)}$ *defined by (3.18)-(3.19) with* $b(x) = -x, \sigma(x) = 1, h(x) = x^2$.

The interest of Proposition 3.8 and its corollary above is that we obtain an explicit and non asymptotic bound for conditional moments of $\hat{\phi}_T^{(k)} - \phi$. By taking expectation of both sides, this provides a clear link between the unconditional moments of $\hat{\phi}_T^{(k)} - \phi$ and moments of positive and negative order of the random effect ϕ.
In the spirit of Proposition 3.5, we can also consider the moment type estimator of ϕ^{-1}. In the above model (one multiplicative random effect), for $h(x) = x^2$, we have $\pi_\phi(h) = (2\phi)^{-1}$, $F_{\varphi,h}(x) = (2\varphi)^{-1}(x^2 - (2\varphi)^{-1}))$, and we easily obtain:

$$K_\gamma(\phi,h) \lesssim \phi^{-3\gamma} + \phi^{-4\gamma}.$$

The random variable ϕ^{-1} admits the conditionally unbiased estimator $(2/T)\int_0^T X^2(s)ds$ and:

$$\mathbb{E}(\frac{2}{T}\int_0^T X^2(s)ds - \phi^{-1})^{2\gamma}|\phi) \lesssim T^{-\gamma}(\phi^{-3\gamma} + \phi^{-4\gamma}), \quad \gamma \geq 1. \tag{3.21}$$

For $V_T = \int_0^T X^2(s)ds$ and $p \geq 1$,

$$\mathbb{P}(\frac{V_T}{T} < \frac{k}{\sqrt{T}}|\phi) \lesssim T^{-p}\left(\phi^{2p} + \phi^{-2p} + \phi^{-p}\right). \tag{3.22}$$

4. Estimation of the distribution of the random effect from N *i.i.d.* trajectories.

The previous sections show that the random effect ϕ can be estimated from one trajectory only as T tends to infinity. To estimate parameters in the distribution of ϕ requires additional information. In this section, we consider N *i.i.d.* sample paths $(X_j(t))$, $j = 1,\ldots,N$ satisfying

$$d\,X_j(t) = b(X_j(t),\phi_j)dt + \sigma(X_j(t))\,dW_j(t), \quad X_j(0) = \eta_j,\ j = 1,\ldots,N, \tag{4.1}$$

where $(W_1,\ldots,W_N)$ are N independent Wiener processes, $(\phi_1,\eta_1)\ldots,(\phi_N,\eta_N))$ are N *i.i.d.* random variables taking values in $\mathbb{R}^d \times \mathbb{R}$, $((\phi_1,\eta_1)\ldots,(\phi_N,\eta_N))$ and $(W_1,\ldots,W_N)$ are independent. We assume that $(X_j(t), t \leq T)$, $j = 1,\ldots,N$ are observed. We denote by $\phi_j = (\phi_{k,j}, k = 1,\ldots,d)$ the components of ϕ_j.
We assume that [**H1**] – [**H3**] hold,*i.e.* (ϕ_j,η_j) has distribution $\nu(d\varphi) \otimes \pi_\varphi(x)dx$ for some unknown distribution ν. To build estimators of the moments of ν, we rely on the double asymptotic framework $N \to +\infty$ and $T \to +\infty$.

4.1. **Moment method.** We first use the moment estimators described in Section 3. Let $g \in L^1(\Phi)$ (see (2.6)) be such that $\mathbb{E}|\pi_\phi(g)| < +\infty$, and set

$$m_g = \mathbb{E}(\pi_\phi(g)) = \int_{(\ell,r)\times\Phi} \pi_\varphi(x)g(x)d\nu(\varphi)dx, \quad \widetilde{m_g} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\int_0^T g(X_j(s))ds. \tag{4.2}$$

In general, $\pi_{\phi_j}(g)$ is a random variable. However, if ϕ_j contains deterministic components (fixed effects), then, $\pi_{\phi_j}g$ can be deterministic ($\pi_{\phi_j}g = \mathbb{E}(\pi_\phi(g)) = m_g$). The behaviour of $\widetilde{m_g}$ differs according to these two cases.

Theorem 4.1. *Assume that $\mathbb{E}(V_\phi(g)) + \rho(g,\phi)) < +\infty$ (see (2.10) and (3.6)).*
• *If $\pi_{\phi_j}(g)$ is random and $\mathbb{E}(\pi_\phi^2(g)) < +\infty$, then $\sqrt{N}(\widetilde{m_g} - m_g)$ converges in distribution to $\mathcal{N}(0, Var(\pi_\phi(g)))$, where $Var(\pi_\phi(g)) = \mathbb{E}(\pi_\phi^2(g)) - m_g^2$.*
• *Let $\pi_{\phi_j}(g) = m_g = \mathbb{E}(\pi_\phi(g)$ be deterministic. Assume that condition $\mathbf{K}(1+\varepsilon)$ holds with $\varepsilon > 0$ and that*

$$\mathbb{E}K_{1+\varepsilon}(\phi, g) < +\infty.$$

Then $\sqrt{NT}(\widetilde{m_g} - m_g)$ converges in distribution to $\mathcal{N}(0, \mathbb{E}(V_\phi(g)))$.

Let us come back to the Ornstein-Uhlenbeck example. Assume that X_j satisfies $dX_j(t) = -\phi_{2,j}(X_j(t) - \phi_{1,j})dt + dW_j(t), X_j(0) = \eta_j$ where $\phi_j = (\phi_{1,j}, \phi_{2,j}) \in \mathbb{R}\times(0,+\infty)$ and that given $\phi_j = \varphi = (\varphi_1, \varphi_2)$, η_j has distribution $\mathcal{N}(\varphi_1, (2\varphi_2)^{-1})$. For the Proposition below, we assume that $\mathbb{E}\phi_{1,j}^2 < +\infty, \mathbb{E}\phi_{2,j}^{-3} < +\infty$ and define $m = \mathbb{E}\phi_{1,j}$, $v = \mathbb{E}((\phi_{2,j})^{-1})$ and

$$\tilde{m} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T}\int_0^T X_j(s)ds, \quad \tilde{v} = \frac{1}{N}\sum_{i=1}^{N}\tilde{\tau}_j \quad \text{with}$$

$$\tilde{\tau}_j = 2(\frac{1}{T}\int_0^T X_j^2(s)ds - (\frac{1}{T}\int_0^T X_j(s)ds)^2).$$

Proposition 4.1. • *If $\phi_{1,j}$ is random, $\sqrt{N}(\tilde{m} - m)$ converges in distribution as both T, N tend to infinity, to $\mathcal{N}(0, Var(\phi_{1,j}))$.*
• *If $\phi_{1,j} = m$ is deterministic, $\sqrt{NT}(\tilde{m} - m)$ converges in distribution to $\mathcal{N}(0, \mathbb{E}(\frac{1}{\phi_{2,j}^2}))$.*
• *If $\phi_{2,j}$ is random and $\sqrt{N}/T$ tends to 0 as both $N, T = T(N)$ tend to infinity, then $\sqrt{N}(\tilde{v} - v)$ converges in distribution to $\mathcal{N}(0, Var(\tau_j))$ where $\tau_j = (\phi_{2,j})^{-1}$.*
• *If $\phi_{2,j} = v^{-1}$ is deterministic and N/T tends to 0 as both $N, T = T(N)$ tend to infinity, then $\sqrt{NT}(\tilde{v} - v)$ converges in distribution to $\mathcal{N}(0, 2v^3 + 12v^2)$.*

The additional constraint linking N and $T = T(N)$ comes from the fact that $\tilde{\tau}_j$ is a conditionally biased estimator of $\tau_j = (\phi_{2,j})^{-1}$.

4.2. **Maximum likelihood approach.** Suppose that a parametric model rules the ϕ_j's, *i.e.* $\nu(d\varphi) = f(\theta,\varphi)d\alpha(\varphi)$ where α is a dominating measure on Φ and $\theta \in \Theta$ is unknown. Consider the log-likelihood associated with $(\phi_j, j = 1,\dots,N)$, $\ell_N(\theta) = \sum_{j=1}^N \log f(\theta, \phi_j)$. As the random variables ϕ_j's are not directly observed, a natural strategy consists in plugging in $\ell_N(\theta)$ estimators of the ϕ_j's based on the results of Section 3. An analogous method is applied in Delattre *et al.* (2014) for a different mixed effect model and different observations. Such an approach has to rely on a precise model for the diffusion process and the parametric model for ν. We choose to develop the method for the Ornstein-Uhlenbeck process with one multiplicative random effect

$$dX_j(t) = -\phi_j X_j(t)dt + dW_j(t), \quad X_j(0) = \eta_j, j = 1,\dots,N.$$

where ϕ_j belongs to $\Phi = (0, \infty)$ (see Corollary 3.1). For the parametric model on ν, we assume that ν is a Gamma distribution with parameter $\theta = (a, \lambda)$, *i.e.*

$$f(\theta, \varphi) = \frac{\lambda^a}{\Gamma(a)} \varphi^{a-1} e^{-\lambda \varphi} 1_{\varphi > 0}.$$

Consider the loglikelihood associated with the observation of $(\phi_j, j = 1, \ldots, N)$

$$\ell_N(\theta) = Na \log \lambda - N \log \Gamma(a) + (a-1) \sum_{j=1}^{N} \log \phi_j - \lambda \sum_{j=1}^{N} \phi_j$$

and define $\theta_N = \mathrm{Argmax}_\theta \ell_N(\theta)$ the maximum likelihood estimator based on the direct observation of $(\phi_i, i = 1, \ldots, N)$. The asymptotic behaviour of θ_N is well known. As N tends to infinity, θ_N is consistent and $\sqrt{N}(\theta_N - \theta)$ converges in distribution to the Gaussian distribution $\mathcal{N}_2(0, I^{-1}(\theta))$ where

$$I(\theta) = \begin{pmatrix} \psi'(a) & -\lambda^{-1} \\ -\lambda^{-1} & a\lambda^{-2} \end{pmatrix}$$

and

$$\psi(a) = \frac{\Gamma'(a)}{\Gamma(a)} = -\gamma + \int_0^1 \frac{1 - t^{a-1}}{1-t} dt$$

is the di-Gamma function where $\gamma = -\Gamma'(1)$ is the Euler constant. Using that $\psi'(a) = -\displaystyle\int_0^1 t^{a-1} (\log t/(1-t)) dt$, yields $a\psi'(a) - 1 \neq 0$ so that $I(\theta)$ is invertible.

For the plug-in, we must define estimators of both ϕ_j and $\log \phi_j$ with appropriate moments properties. In this model, we can use two different estimators. The moment type estimators given by

$$\left(\frac{2}{T} \int_0^T X_j^2(s) ds \right)^{-1}$$

or the estimator based on a maximum likelihood approach

$$\hat{\phi}_j = -\frac{\int_0^T X_j(s) dX_j(s)}{\int_0^T X_j^2(s) ds}.$$

For both estimators, getting explicit bounds for their moments is not straightforward. Introducing some appropriate truncations is a good tool to obtain these bounds. In Section 3.2, we have applied the truncation method to $\hat{\phi}_j$ (see Corollary 3.1) . So we use it for the plug-in device. Let us set:

$$\hat{\phi}_i^{(k)} = \hat{\phi}_i 1_{(\frac{V_{i,T}}{T} > \frac{k}{\sqrt{T}})} \quad \text{with} \quad V_{i,T} = \int_0^T X_i^2(s) ds.$$

This estimator may be negative and nul. Therefore we define a specific estimator for $L_i = \log \phi_i$ and set, for k a constant:

$$\widehat{L}_i^{(k)} = (\log \hat{\phi}_i)\, 1_{(\hat{\phi}_i \geq \frac{k}{\sqrt{T}}, \frac{V_{i,T}}{T} > \frac{k}{\sqrt{T}})} = (\log (\hat{\phi}_i^{(k)}))\, 1_{(\hat{\phi}_i \geq \frac{k}{\sqrt{T}}, \frac{V_{i,T}}{T} > \frac{k}{\sqrt{T}})}.$$

Now, set

$$V_N(\theta) = Na \log \lambda - N \log \Gamma(a) - \lambda \sum_{i=1}^{N} \hat{\phi}_i^{(k)} + (a-1) \sum_{i=1}^{N} \widehat{L}_i^{(k)}$$

and define the estimator of θ as any solution of:

$$\tilde{\theta}_N = \mathrm{Argmax}_\theta V_N(\theta).$$

Proposition 4.2. *Assume that $a > 8$ and that both N, T tend to infinity. Then, $\tilde{\theta}_N$ is consistent. If moreover, N/T tends to 0, then $\sqrt{N}(\tilde{\theta}_N - \theta_N) = o_P(1)$ where θ_N is the maximum likelihood estimator based on the observation of $(\phi_1, \ldots, \phi_N)$.*

5. Concluding remarks

In this paper, we investigate estimation problems linked with the observation of processes defined by an SDE with mixed effects from a continuous observation on a time interval $[0, T]$. Generally, authors are interested in estimating both the fixed effects and unknown parameters in the distribution of the random effects when N *i.i.d.* processes $(X_j(t), t \leq T), j = 1, \ldots, N)$ are observed and N tends to infinity. Two cases are investigated: Either T is fixed or T tends to infinity (see references in the introduction). In the latter case, constraints linking N and $T = T(N)$ are required.

In the case T tends to infinity, the interest of our contribution is to consider separately the inference based on one sample path and the inference based on N *i.i.d.* sample paths. This leads to study the long term behaviour of a one-dimensional SDE with mixed effects. We provide sufficient conditions for the existence of invariant distributions for a process $(X(t))$ defined by an SDE with a random effect ϕ (see (1.2)). There is no uniqueness of the invariant distributions. We interpret and restate classical results for ergodic diffusions in the case of diffusions with random effects in the drift term. Then, we apply these results to derive estimators of the random variable ϕ itself based on functionals of the continuous sample path $(X(t), t \leq T)$. This is not a statistical problem in the usual sense as ϕ is a random variable. Nevertheless, the estimation of random variables has recently received a lot of attention in the field of statistics of random processes. Here, two kinds of estimators of ϕ are studied and proved to be consistent and asymptotically mixed normal as T goes to infinity with rate $\sqrt{T}$. An important point is that we also obtain precise non asymptotic conditional moment bounds for the estimators. The results are applied to various classical examples. We stress that assumption [**H3**] (stationary regime) is only required for the conditional moments bounds.

To estimate the common distribution of the ϕ_j's when N *i.i.d.* processes $(X_j(t), t \leq T), j = 1, \ldots, N)$ are observed, we proceed as follows. From each sample path $(X_j(t), t \leq T)$ we build an estimator of ϕ_j, say $\tilde{\phi}_j$ and replace ϕ_j by $\tilde{\phi}_j$ to obtain estimators of parameters in the distribution of ϕ_j. The method relies on a double asymptotic framework where both N and T go to infinity. We obtain different asymptotic behaviour of estimators: for fixed effects parameters, the rate is $\sqrt{NT}$ and for parameters in the distribution of the random effects, the rate is $\sqrt{N}$. In some cases, to control bias, a constraint linking N and $T = T(N)$ is needed.

The distribution of the random effects is often assumed to be normal or log-normal. Our approach can be applied as well to general diffusion models and to general distributions for the random effects up to some moments constraints. In this respect, our conditional moment bounds are particularly useful.

In practice, only discrete time observations are usually available. As it is usual, for small sampling interval, we can extend our results to estimators obtained by changing the integrals over $[0, T]$ into their one-step discretizations.

6. Appendix: proofs

Proof of Proposition 3.1

The process X^φ defined in (2.13) satisfies the ergodic theorem for all $\varphi \in \Phi$:

$$\mathbb{P}(\frac{1}{T}\int_0^T g(X^\varphi(s))ds \to_{T\to+\infty} \pi_\varphi(g)) = 1.$$

By Proposition 2.1,

$$\mathbb{P}(\frac{1}{T}\int_0^T g(X(s))ds \to \pi_\phi(g)|\phi=\varphi) = \mathbb{P}(\frac{1}{T}\int_0^T g(X^\varphi(s))ds \to \pi_\varphi(g)).$$

Thus, $\mathbb{P}(T^{-1}\int_0^T g(X(s))ds \to \pi_\phi(g)) = 1$.

Proof of Proposition 3.2
We consider for fixed φ the process X^φ given by (2.13) and $F_{\varphi,g}$ given by (2.9). By the Ito formula, we have:

$$F_{\varphi,g}(X^\varphi(T)) = F_{\varphi,g}(X^\varphi(0)) + \int_0^T \mathcal{L}_\varphi F_{\varphi,g}(X^\varphi(s))ds + \int_0^T F'_{\varphi,g}(X^\varphi(s))\sigma(X^\varphi(s))dW(s).$$

Therefore,

$$-\frac{1}{\sqrt{T}}\int_0^T \mathcal{L}_\varphi F_{\varphi,g}(X^\varphi(s))ds = \frac{1}{\sqrt{T}}\int_0^T F'_{\varphi,g}(X^\varphi(s))\sigma(X^\varphi(s))dW(s) + R_T^\varphi, \tag{6.1}$$

with

$$R_T^\varphi = T^{-1/2}(F_{\varphi,g}(X^\varphi(0)) - F_{\varphi,g}(X^\varphi(T)). \tag{6.2}$$

The ergodic properties of X^φ imply that $X^\varphi(T)$ converges in distribution to the stationary distribution $\pi_\varphi(x)dx$ as T tends to infinity. Hence, R_T^φ tends to 0 in probability. We can conclude that,

$$-\frac{1}{\sqrt{T}}\int_0^T \mathcal{L}_\varphi F_{\varphi,g}(X^{\varphi,g}(s))ds \to \mathcal{N}(0, V_\varphi(g)).$$

Setting $g_\varphi = g - \pi_\varphi g$, we have

$$\frac{1}{\sqrt{T}}\int_0^T g_\phi(X(s))ds = \sqrt{T}\widetilde{(\pi_\phi(g)} - \pi_\phi(g)) \tag{6.3}$$

By conditioning, we obtain, for $u \in \mathbb{R}, v \in \mathbb{R}^d$,

$$\mathbb{E}(\exp{(iu\frac{1}{\sqrt{T}}\int_0^T g_\phi(X(s))ds + iv.\phi)}) \to \mathbb{E}(\exp{(-\frac{u^2}{2}V_\phi(g) + iv.\phi)}),$$

which gives the convergence in distribution result.
The multidimensional result is obtained by applying the above result to $\sum_{i=1}^k u_j g_j = \underline{u}'\underline{g}$ with the Cramer-Wold device. □

Proof of Proposition 3.3
We use the same notation X^φ for the stationary and ergodic process with fixed $\varphi \in \Phi$:

$$dX^\varphi(t) = b(X^\varphi(t)), \varphi)dt + \sigma(X^\varphi(t))dW(t), \quad X^\varphi(0) \sim \pi_\varphi \tag{6.4}$$

with $X^\varphi(0)$ independent of W. The process X^φ has marginal distribution π_φ. Recall the notation $g_\varphi = g - \pi_\varphi g$. Using (6.1)-(6.2) and the Hölder and the Burkholder-Davis-Gundy inequalities yields:

$$\mathbb{E}\left(\int_0^T g_\varphi(X^\varphi(s))ds\right)^{2\gamma} \le c_\gamma\left(T^\gamma \pi_\varphi((F'_{\varphi,g}\sigma(.))^{2\gamma} + \pi_\varphi((F_{\varphi,g})^{2\gamma})\right).$$

Now,

$$\mathbb{E}(\left(\int_0^T g_\phi(X(s))ds\right)^{2\gamma}|\phi=\varphi) = \mathbb{E}\left(\int_0^T g_\varphi(X^\varphi(s))ds\right)^{2\gamma}.$$

Hence, the result for all $\gamma \geq 1$ using (6.3) provided that $T \geq 1$.
For $\gamma = 1$,

$$\mathbb{E}\left(\frac{1}{\sqrt{T}}\int_0^T g_\varphi(X^\varphi(s))ds\right)^2 = \pi_\varphi((F'_{\varphi,g}\sigma(.))^2 + Z_T(\varphi) = V_\varphi(g) + Z_T(\varphi)$$

with (see (6.2))

$$Z_T(\varphi) = \mathbb{E}(R_T^\varphi)^2 + 2\mathbb{E}\left(R_T^\varphi \frac{1}{\sqrt{T}}\int_0^T F'_{\varphi,g}(X^\varphi(s))\sigma(X^\varphi(s))dW(s)\right).$$

We have:

$$\mathbb{E}(R_T^\varphi)^2 \leq \frac{2}{T}\mathbb{E}(F_{\varphi,g}(X^\varphi(0)))^2 = \frac{2}{T}\pi_\varphi[(F_{\varphi,g})^2].$$

And

$$\begin{aligned}|\mathbb{E}(R_T^\varphi \frac{1}{\sqrt{T}}\int_0^T F'_{\varphi,g}(X^\varphi(s))\sigma(X^\varphi(s))dW(s))| &\leq \left(\mathbb{E}(R_T^\varphi)^2\mathbb{E}(\frac{1}{T}\int_0^T (F'_{\varphi,g}(X^\varphi(s))\sigma(X^\varphi(s)))^2 ds)\right)^{1/2} \\ &\leq \left(\frac{2}{T}\pi_\varphi[(F_{\varphi,g})^2]\ V_\varphi(g)\right)^{1/2}\end{aligned}$$

For considering the two functions g, h, we just write $gh = ((g+h)^2 - (g-h)^2)/4$ and apply the previous result. □

Proof of Proposition 3.4
Consider $g(x) = x$, $h(x) = x^2$. Then, $\pi_\phi(g) = \phi_1$, $\pi_\phi(h) := \Psi = \phi_1^2 + 1/(2\phi_2)$,

$$\widetilde{\pi_\phi(g)} = \widetilde{\phi}_1 = \frac{1}{T}\int_0^T X(s)ds, \quad \widetilde{\pi_\phi(h)} = \widetilde{\Psi} = \frac{1}{T}\int_0^T X^2(s)ds$$

and $(\widetilde{\phi}_1, \widetilde{\Psi})$ converges *a.s.* to (ϕ_1, Ψ). For (2.8)- (2.10), we check that

$$F_{\varphi,g}(x) = (x-\varphi_1)/\varphi_2, \qquad F_{\varphi,h}(x) = \frac{(x-\varphi_1)^2}{2\varphi_2} + 2\frac{\varphi_1}{\varphi_2}(x-\varphi_1), \tag{6.5}$$

$$V_\varphi(g,g) = \frac{1}{\varphi_2^2}, \quad V_\varphi(h,h) = \frac{1}{2\varphi_2^3} + 4\left(\frac{\varphi_1}{\varphi_2}\right)^2, \quad V_\varphi(g,h) = 2\frac{\varphi_1}{\varphi_2^2}. \tag{6.6}$$

The matrix $V(\varphi) = V_\varphi((g\ \ h)')$ satisfies $\det V(\varphi) = 1/(2\varphi_2^5) > 0$. Consequently,

$$\sqrt{T}(\widetilde{\phi}_1 - \phi_1, \widetilde{\Psi} - \Psi)' \to_{\mathcal{D}} \mathcal{MN}_2(0, V(\phi)).$$

Using that $\widetilde{\phi}_2 = (2(\widetilde{\Psi} - (\widetilde{\phi}_1)^2))^{-1}$ and the δ-method, we deduce after some computations that $\sqrt{T}(\widetilde{\phi}_1 - \phi_1, \widetilde{\phi}_2 - \phi_2)' \to_{\mathcal{D}} \mathcal{MN}_2(0, J(\phi))$.□

Proof of Proposition 3.5
In this model, the conditional distribution of $\widetilde{\phi}_1$ is explicit and conditionally Gaussian with mean ϕ_1. To compute the conditional variance, we use the solution of the SDE:

$$X(t) = \phi_1 + e^{-\phi_2 t}(\eta - \phi_1) + e^{-\phi_2 t}\int_0^t e^{\phi_2 s}dW(s).$$

As $(X(t))$ is stationary, $\mathbb{E}(X(t)X(t+h)|\phi) = \mathbb{E}(X(0)X(h)|\phi) = \phi_1^2 + e^{-\phi_2 h}(2\phi_2)^{-1}$. Therefore,

$$\mathbb{E}((\widetilde{\phi}_1)^2)|\phi) = 2T^{-2}\int_{0<s<s'<T} \phi_1^2 + \frac{e^{-\phi_2(s'-s)}}{2\phi_2}dsds' = \phi_1^2 + \Sigma_T(\phi_2),$$

where

$$\Sigma_T(\phi_2) = \frac{1}{T\phi_2^2} - \frac{1-e^{-\phi_2 T}}{T^2\phi_2^3}. \tag{6.7}$$

Conditionally on ϕ, $\widetilde{\phi}_1$ has distribution $\mathcal{N}(\phi_1, \Sigma_T(\phi_2))$. In particular,

$$\mathbb{E}((\widetilde{\phi}_1 - \phi_1))^{2\gamma}|\phi) = C(2\gamma)\Sigma_T^{\gamma}(\phi_2) \qquad \text{and} \tag{6.8}$$

$$\mathbb{E}(\sqrt{T}((\widetilde{\phi}_1 - \phi_1))^2|\phi) = \frac{1}{\phi_2^2} + Z_T, \qquad |Z_T| \leq \frac{1}{T\phi_2^3}. \tag{6.9}$$

Note that $\rho(g,\phi) = \phi_2^{-3} + 2\phi_2^{-5/2}$. Thus, the bound for $Z_T = Z_T(g,\phi)$ given in Proposition 3.3 is really improved by exact computations. We can alsoo check condition **(Kγ)** for g:

$$K_\gamma(\varphi, g) = \pi_\varphi(F_{\varphi,g}^{2\gamma}) + (\pi_\varphi((F'_{\varphi,g}\sigma)^{2\gamma}) = \frac{C(2\gamma)}{2^\gamma\varphi_2^{3\gamma}} + \frac{1}{\varphi_2^{2\gamma}}, \tag{6.10}$$

where $C(2\gamma) = \mathbb{E}X^{2\gamma}$ for X a standard Gaussian variable. Therefore, by (3.7),

$$\mathbb{E}(\widetilde{\phi}_1 - \phi_1)^{2\gamma}|\phi) \lesssim T^{-\gamma}(\phi_2^{-3\gamma} + \phi_2^{-2\gamma}). \tag{6.11}$$

Comparing with the exact result (6.8), we see that this bound is accurate.

Next, we study the estimation of $\tau = \phi_2^{-1}$ by $\widetilde{\tau} = \widetilde{\phi_2}^{-1} = 2(\widetilde{\Psi} - (\widetilde{\phi}_1)^2)$. Note that, using the Cauchy-Schwarz inequality and that $t \to X(t)$ is not constant, $(\widetilde{\phi}_1)^2 < \widetilde{\Psi}$, so that $\widetilde{\tau} > 0$. We have

$$\mathbb{E}(\widetilde{\tau}|\phi) = \tau - 2\Sigma_T(\phi_2). \tag{6.12}$$

We now use the decomposition

$$\widetilde{\tau} - \tau = 2[(\widetilde{\Psi} - \Psi) - (\widetilde{\phi}_1 - \phi_1)^2 - 2\phi_1(\widetilde{\phi}_1 - \phi_1)]. \tag{6.13}$$

By (3.7)-(3.8) with g replaced by $h(x) = x^2$ (see also (6.6)), we obtain

$$\mathbb{E}\left((\widetilde{\Psi} - \Psi)^{2\gamma}|\phi\right) \lesssim T^{-\gamma}(\phi_2^{-4\gamma} + \phi_2^{-3\gamma} + \phi_1^{2\gamma}\phi_2^{-3\gamma} + \phi_1^{2\gamma}\phi_2^{-2\gamma}), \tag{6.14}$$

$$\mathbb{E}(\sqrt{T}(\widetilde{\Psi} - \Psi))^2|\phi) = \frac{1}{2\phi_2^3} + 4(\frac{\phi_1}{\phi_2})^2 + Z_T(h,\phi). \tag{6.15}$$

where $|Z_T(h,\phi)| \leq (1/\sqrt{T})\rho(h,\phi)$ and $\rho(h,\phi)$ is given by (3.6). Using (6.5)-(6.6), we obtain

$$\pi_\phi F_{\phi,h}^2 = \frac{3}{16\phi_2^4} + 2\frac{\phi_1^2}{\phi_2^3} \text{ and } \rho(h,\phi) \lesssim (\frac{1}{\phi_2^4} + \frac{1}{\phi_2^{7/2}} + \frac{\phi_1^2}{\phi_2^3} + \frac{|\phi_1|}{\phi_2^3} + \frac{\phi_1^2}{\phi_2^{5/2}}).$$

By (6.13), we have

$$(\widetilde{\tau} - \tau)^{2\gamma} \lesssim (\widetilde{\Psi} - \Psi)^{2\gamma} + \phi_1^{2\gamma}(\widetilde{\phi}_1 - \phi_1)^{2\gamma} + (\widetilde{\phi}_1 - \phi_1)^{4\gamma}). \tag{6.16}$$

Therefore, combining (6.11)-(6.14) yields:

$$\mathbb{E}((\widetilde{\tau} - \tau)^{2\gamma}|\phi) \lesssim T^{-\gamma}(\phi_2^{-2\gamma} + \phi_2^{-4\gamma} + \phi_2^{-3\gamma} + \phi_1^{2\gamma}\phi_2^{-3\gamma} + \phi_1^{2\gamma}\phi_2^{-2\gamma} + \phi_2^{-6\gamma}). \tag{6.17}$$

Moreover,

$$\mathbb{E}((\widetilde{\tau} - \tau)^2|\phi) = 4\sum_{i=1}^{6} A_i \quad \text{where}$$

$A_1 = \mathbb{E}((\widetilde{\Psi} - \Psi)^2|\phi)$ is given in (6.15), $A_2 = 4\mathbb{E}((\phi_1^2(\widetilde{\phi}_1 - \phi_1)^2|\phi) = 4\phi_1^2\Sigma_T(\phi_2)$,

$$A_3 = \mathbb{E}((\widetilde{\phi}_1 - \phi_1)^4|\phi) = 3(\Sigma_T(\phi_2))^2, \quad A_4 = 4\mathbb{E}(\phi_1(\widetilde{\phi}_1 - \phi_1)^3|\phi) = 0.$$

Using (3.9) and $V_\phi(h,g) = 2\phi_1/\phi_2^2$ yields

$$A_5 = -4\mathbb{E}(\phi_1(\widetilde{\Psi} - \Psi)(\widetilde{\phi}_1 - \phi_1)|\phi) = -\frac{4\phi_1}{T}(\frac{2\phi_1}{\phi_2^2}) + Z_T(h,g,\phi)).$$

Finally, by the Cauchy-Schwarz inequality,

$$|A_6| = |-2\mathbb{E}((\widetilde{\Psi} - \Psi)((\widetilde{\phi}_1 - \phi_1)^2|\phi)| \leq \Big(\mathbb{E}((\widetilde{\Psi} - \Psi)^2|\phi)\mathbb{E}((\widetilde{\phi}_1 - \phi_1)^4|\phi)\Big)^{1/2}.$$

We obtain

$$\mathbb{E}(\sqrt{T}(\widetilde{\tau} - \tau)^2|\phi) = \frac{2}{\phi_2^3} + \zeta_T \quad \text{with} \quad |\zeta_T| \leq \frac{1}{\sqrt{T}}C(\phi). \tag{6.18}$$

The exact expression of $C(\phi)$ can be derived by some cumbersome computations. If, for $\gamma = 1$, the r.h.s. of (6.17) has finite expectation, then $\mathbb{E}(C(\phi)) < +\infty$. This yields Proposition 3.5 substituting ϕ_2^{-1} by τ. □

Proof of Proposition 3.6
We again set $g(x) = x$ and $h(x) = x^2$. Then,

$$\pi_\phi(g) = \phi_1, \quad \pi_\phi(h) = \phi_1^2 + \beta \quad \text{with} \quad \beta = \frac{\phi_1}{2\phi_2}.$$

Some computations lead to

$$F_{\varphi,g}(x) = \frac{x - \varphi_1}{\varphi_2}, F_{\varphi,h}(x) = \frac{(x-\varphi_1)^2}{2\varphi_2} + (x - \varphi_1)(\frac{1}{2\phi_2^2} + 2\frac{\phi_1}{\phi_2}).$$

We need the moments and the centered moments of a Gamma $G(a,\lambda)$ distribution (recall that $a = 2\varphi_1\varphi_2, \lambda = 2\varphi_2$): for k integer and $a \geq 1$

$$\int x^k dG(a,\lambda)(x) = \frac{(a+k-1)\dots(a+1)a}{\lambda^k} \leq \lambda^{-k}k!a^k,$$

$$\int (x - \frac{a}{\lambda})^k dG(a,\lambda)(x) = \lambda^{-k}\sum_{j=0}^{k}\binom{k}{j}(-a)^{k-j}(a+j-1)\dots(a+1)a.$$

For $a \geq 1$,

$$|\int (x - \frac{a}{\lambda})^k dG(a,\lambda)(x)| \leq a^k\lambda^{-k}\sum_{j=0}^{k}\frac{k!}{(k-j)!}.$$

This leads to:

$$V_\varphi(g,g) = \frac{\varphi_1}{\varphi_2^2}, \quad V_\varphi(g,h) = \frac{\varphi_1}{\varphi_2^3} + 2\frac{\varphi_1^2}{\varphi_2^2}, \quad V_\varphi(h,h) = \frac{5\varphi_1}{4\varphi_2^4} + \frac{9\varphi_1^2}{2\varphi_2^3} + \frac{4\varphi_1^3}{\varphi_2^2}.$$

The matrix $V := V_\varphi((g\ h)')$ satisfies $\det V = \varphi_1^2/(4\varphi_2^6) + \varphi_1^3/(2\varphi_2^5) > 0$. We apply Propositions 3.1 and 3.2 together with the δ-method to compute $J_{cir}(\phi)$ and obtain the first part of Proposition 3.6.
Now, we look at conditional moments. Using

$$X(t) = \phi_1 + e^{-\phi_2 t}(\eta - \phi_1) + e^{-\phi_2 t}\int_0^t e^{\phi_2 s}\sqrt{X(s)}dW(s),$$

we find $\mathbb{E}(X(t)X(t+h)|\phi) = \mathbb{E}(X(0)X(h)|\phi) = \phi_1^2 + \frac{\phi_1}{2\phi_2}e^{-\phi_2 h}$. Therefore,

$$\mathbb{E}((\widetilde{\phi}_1)^2)|\phi) = \phi_1^2 + \phi_1\Sigma_T(\phi_2),$$

with $\Sigma_T(\phi_2)$ given in (6.7).
Finally, we get, for γ integer, $K_\gamma(\varphi, g) \lesssim \varphi_2^{-2\gamma}(\varphi_1^{2\gamma} + \varphi_1^\gamma)$. □

Proof of Proposition 3.8
We have: $\sqrt{T}(\hat{\phi}_T^{(k)} - \phi) = A_1 + A_2$, with

$$A_1 = -\sqrt{T}\phi\, 1_{(\frac{V_T}{T} < \frac{k}{\sqrt{T}})}, \qquad A_2 = \frac{M_T/\sqrt{T}}{V_T/T}\, 1_{(\frac{V_T}{T} \geq \frac{k}{\sqrt{T}})}, \tag{6.19}$$

where $M_T = \int_0^T (b(X(s))/\sigma(X(s)))dW(s)$. Recall that $h = b^2/\sigma^2$. We first study the term A_2. Note that:

$$\mathbb{E}[\frac{V_T}{T}|\phi] = \frac{1}{T}\int_0^T \mathbb{E}[h(X(s))|\phi]ds,$$

and $\mathbb{E}[h(X(s))|\phi = \varphi] = \mathbb{E}[h(X^\varphi(s))] = \pi_\varphi(h)$. Thus,

$$L = \pi_\phi(h) = \mathbb{E}(\frac{V_T}{T}|\phi) = \mathbb{E}h(X(0))|\phi).$$

We have: $A_2 = A_2' + A_2''$ with

$$A_2' = \frac{M_T}{\sqrt{T}}\left(\frac{1}{V_T/T} - \frac{1}{L}\right) 1_{(\frac{V_T}{T} \geq \frac{k}{\sqrt{T}})}, \qquad A_2'' = \frac{M_T}{L\sqrt{T}} 1_{(\frac{V_T}{T} \geq \frac{k}{\sqrt{T}})}. \tag{6.20}$$

Using that L is $\mathcal{F}_0$-measurable, $< M >_T = V_T$, the Burkholder-Davis-Gundy and the Hölder inequalities, we get:

$$\mathbb{E}((A_2'')^{2\gamma}|\phi) \leq C_{2\gamma}\mathbb{E}(\frac{V_T^\gamma}{L^{2\gamma}T^\gamma}|\phi) \leq C_{2\gamma}\mathbb{E}(\frac{1}{L^{2\gamma}T}\int_0^T h^\gamma(X(s))ds|\phi) = C_{2\gamma}\frac{\pi_\phi(h^\gamma)}{L^{2\gamma}},$$

where $C_{2\gamma}$ is the constant of the B-D-G inequality. We need $\pi_\phi(h^\gamma) < +\infty$.
For $\gamma = 1$, we have $\mathbb{E}((A_2'')^2|\phi) \leq L^{-1}$. We look at A_2':

$$\mathbb{E}((A_2')^{2\gamma}|\phi) \leq \frac{1}{k^{2\gamma}}\left(\mathbb{E}\left((\frac{M_T}{L\sqrt{T}})^{4\gamma}|\phi\right)\mathbb{E}\left((\sqrt{T}(L - V_T/T))^{4\gamma}|\phi)\right)\right)^{1/2}.$$

We have:

$$\mathbb{E}\left((\frac{M_T}{L\sqrt{T}})^{4\gamma}|\phi\right) \leq C_{4\gamma}\mathbb{E}(\frac{1}{L^{4\gamma}}\left(\frac{V_T}{T}\right)^{2\gamma}|\phi) \leq C_{4\gamma}\frac{\pi_\phi(h^{2\gamma})}{L^{4\gamma}}.$$

Therefore, we need $\pi_\phi(h^{2\gamma}) < +\infty$. Then, we apply Proposition 3.3 and get:

$$\mathbb{E}(\left(\sqrt{T}(L - V_T/T)\right)^{4\gamma}|\phi) \lesssim K_{2\gamma}(\phi, h).$$

There remains the term A_1. We have $A_1^{2\gamma} \leq T^\gamma\phi^{2\gamma}1_{(\frac{V_T}{T} < \frac{k}{\sqrt{T}})}$. Therefore, setting $\ell = \pi_\varphi h$, we have, for all $p \geq 1$,

$$\begin{aligned}
\mathbb{P}(\frac{V_T}{T} < \frac{k}{\sqrt{T}}|\phi = \varphi) &\leq \mathbb{P}(|\ell - \frac{V_T}{T}| > \ell - \frac{k}{\sqrt{T}}|\phi = \varphi)\, 1_{(\ell > \frac{2k}{\sqrt{T}})} + 1_{(\ell \leq \frac{2k}{\sqrt{T}})} \\
&\leq \mathbb{P}(|\ell - \frac{V_T}{T}| > \frac{\ell}{2}|\phi = \varphi) + 1_{(\ell^{-1} \geq \frac{\sqrt{T}}{2k})} \\
&\leq (\frac{2}{\ell})^{2p}\mathbb{E}(\ell - \frac{V_T}{T})^{2p}|\phi = \varphi) + (\frac{2k}{\sqrt{T}})^{2p}\ell^{-2p}.
\end{aligned}$$

Thus, by Proposition 3.3, we get

$$\begin{aligned}\mathbb{E}(A_1^{2\gamma}|\phi) &\leq T^{\gamma-p}\left[\phi^{2\gamma}(\frac{2}{L})^{2p}K_p(\phi,h)\right] + T^{\gamma-p}(2k)^{2p}\ \phi^{2\gamma}L^{-2p}) \\ &\lesssim T^{\gamma-p}\phi^{2\gamma}L^{-2p}(1+K_p(\phi,h))\end{aligned}$$

We need $p \geq \gamma$ and that the left-hand side be finite. We take $p = \gamma$.
□
Proof of Corollary 3.1
With $h(x) = x^2$, in this model, we have $F_{\varphi,h}(x) = (2\varphi)^{-1}(x^2 - (2\varphi)^{-1}), F'_{\varphi,h}(x) = x/\varphi$. This yields for $p \geq 1$:

$$V_\varphi(h) = \frac{1}{2\varphi^3}, \quad \pi_\varphi(h^p) = C(2p)(2\varphi)^{-p}, \quad H(p,\varphi,h) \lesssim (\varphi^{-3p} + \varphi^{-4p}).$$

We compute $C(\gamma,\phi)$ (Proposition 3.8) and get the result. □

Proof of Theorem 4.1 Assume that $\pi_{\phi_j}(g)$ is random. We write $\sqrt{N}(\widetilde{m_g} - m_g) = T_1 + T_2$ with

$$T_1 = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\frac{1}{T}\int_0^T g(X_j(s))ds - \pi_{\phi_j}(g)), \ \ T_2 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\pi_{\phi_j}(g)) - m_g).$$

The term T_2 is ruled by the standard central limit theorem and converges in distribution to $\mathcal{N}(0, \mathrm{Var}(\pi_\phi(g)))$.
For T_1, we write:

$$T_1 = \sum_{j\leq N} A_{N,j}, \quad A_{N,j} = \frac{1}{\sqrt{N}}(\frac{1}{T}\int_0^T g(X_j(s))ds - \pi_{\phi_j}(g)).$$

We use the simple Lemma.

Lemma 6.1. *Let $\xi_N = \sum_{j=1}^N \xi_{N,j}$ where conditionally on $\underline{\phi} = (\phi_1,\ldots,\phi_N)$, the $\xi_{N,j}$'s are independent. If $\sum_{j=1}^N \mathbb{E}(\xi_{N,j}|\underline{\phi}) \to \xi$ and $\sum_{j=1}^N \mathbb{E}(\xi_{N,j}^2|\underline{\phi}) \to 0$ in probability, then $\xi_N \to \xi$ in probability.*

Proof of Lemma 6.1

$$\xi_N - \xi = \sum_{j=1}^N(\xi_{N,j} - \mathbb{E}(\xi_{N,j}|\underline{\phi})) + \sum_{j=1}^N \mathbb{E}(\xi_{N,j}|\underline{\phi}) - \xi := D_N + \sum_{j=1}^N \mathbb{E}(\xi_{N,j}|\underline{\phi}) - \xi.$$

We only need to prove that D_N tends to 0. Using the conditional independence, we have:

$$\mathbb{E}[D_N^2|\underline{\phi}] = \sum_{j\leq N}(\mathbb{E}(\xi_{N,j}^2|\underline{\phi}) - [\mathbb{E}(\xi_{N,j}|\underline{\phi})]^2) \leq \sum_{j\leq N}\mathbb{E}(\xi_{N,j}^2|\underline{\phi}) \to 0.$$

Hence, for $h > 0$, $\mathbb{P}(|D_N| \geq h|\underline{\phi}) \leq h^{-2}\mathbb{E}[D_N^2|\underline{\phi}] \to 0$. This implies $\mathbb{P}(|D_N| \geq h) \to 0$.□

Let us apply Lemma 6.1. Conditionally to $\underline{\phi}$, the $A_{N,j}$'s are independent, centered and by Proposition 3.3,

$$\sum_{j\leq N}\mathbb{E}(A_{N,j}^2|\underline{\phi}) = \frac{1}{T}\left(\frac{1}{N}\sum_{j\leq N}(V_{\phi_j}(g) + Z_T(g,\phi_j))\right),$$

where $|Z_T(g,\phi_j)| \leq (1/\sqrt{T})\rho(g,\phi_j)$. By assumption, $\mathbb{E}(V_{\phi_j}(g)+\rho(g,\phi_j)) < +\infty$. This implies, as both N and T tend to infinity, $\sum_{j\leq N}\mathbb{E}(A_{N,j}^2|\underline{\phi}) \to_{a.s.} 0$. Thus, $T_1 \to 0$.
Assume now that $\pi_{\phi_j}(g) = m_g$ is deterministic. Then,

$$\sqrt{NT}(\widetilde{m_g} - m_g) = \sum_{j\leq N} B_{N,j}, \quad B_{N,j} = \sqrt{T}A_{N,j}.$$

As above, given $\underline{\phi}$, the $B_{N,j}$'s are independent, centered and

$$\sum_{j\leq N}\mathbb{E}(B_{N,j}^2|\underline{\phi}) = \left(\frac{1}{N}\sum_{j\leq N}(V_{\phi_j}(g) + Z_T(g,\phi_j))\right) \to_{a.s.} \mathbb{E}V_\phi(g).$$

To complete the proof of the convergence in distribution, we check a Lyapounov condition (see *e.g.* Hall and Heyde (1980)). For $\varepsilon > 0$, using Proposition 3.3 ($\gamma = 1+\varepsilon$),

$$\sum_{j\leq N}\mathbb{E}(B_{N,j}^{2(1+\varepsilon)}|\underline{\phi}) \leq \frac{c_{1+\varepsilon}}{N^\varepsilon}\frac{1}{N}\sum_{j\leq N}K_{1+\varepsilon}(\phi_j,g) \to_{a.s.} 0.$$

Therefore, conditionally on $\underline{\phi}$, $\sqrt{NT}(\widetilde{m_g} - m_g)$ converges in distribution to $\mathcal{N}(0,\mathbb{E}V_\phi(g))$. Thus, the result. □

Proof of Proposition 4.1 Let $m = \mathbb{E}\phi_{1,j}$. We apply the result of Theorem 4.1 to the estimation of m with $g(x) = x$. We have $\pi_{\phi_j}(g) = \phi_{1,j}$. We use formulae (6.5)-(6.10):

$$\mathbb{E}V_{\phi_j}(g) = \mathbb{E}(\frac{1}{\phi_{2,j}^2}), \qquad \mathbb{E}\pi_{\phi_j}(F_{\phi_j,g})^2 = \mathbb{E}(\frac{1}{2\phi_{2,j}^3}).$$

From (3.6), $\mathbb{E}\rho(g,\phi_j) < +\infty$ holds if $\mathbb{E}V_{\phi_j}(g) < +\infty$ and $\mathbb{E}\pi_{\phi_j}(F_{\phi_j,g}^2) < +\infty$. If $\phi_{1,j}$ is random and $\mathbb{E}(\frac{1}{\phi_{2,j}^3}) < +\infty$, $\sqrt{N}(\tilde{m} - m)$ converges in distribution as both T, N tend to infinity, to $\mathcal{N}(0, \mathrm{Var}(\phi_{1,j}))$.
If $\phi_{1,j} = m$ is deterministic and $\mathbb{E}(\frac{1}{\phi_{2,j}^3}) < +\infty$ (as $\tilde{m}$ is Gaussian, the condition $\mathbb{E}(\frac{1}{\phi_{2,j}^{3(1+\varepsilon)}}) < +\infty$ for some $\varepsilon > 0$ is not required), $\sqrt{NT}(\tilde{m} - m)$ converges in distribution to $\mathcal{N}(0, \mathbb{E}(\frac{1}{\phi_{2,j}^2}))$. This is consistent with relation (6.9) obtained for one trajectory.
Now, we consider the estimation of $v = \mathbb{E}((\phi_{2,j})^{-1})$. Assume that

$$\tau_j = (\phi_{2,j})^{-1}$$

is random. We use the decomposition $\sqrt{N}(\tilde{v} - v) = T_1 + T_2$ with

$$T_1 = \sum_{j\leq N} A_{N,j}, \quad A_{N,j} = \frac{1}{\sqrt{N}}(\tilde{\tau}_j - \tau_j), \quad T_2 = \frac{1}{\sqrt{N}}\sum_{j=1}^N(\tau_j - v).$$

The term T_2 is ruled by the classical central limit theorem. Provided that $\mathbb{E}(\tau_j^2) < +\infty$, T_2 converges in distribution to $\mathcal{N}(0, \mathrm{Var}(\tau_j))$. For the term T_1, we have:

$$\mathbb{E}(A_{N,j}|\underline{\phi}) = -\frac{2}{\sqrt{N}}\Sigma_T(\phi_{2,j}) = -2\frac{\sqrt{N}}{T}\frac{1}{N}\sum_{j\leq N}(\frac{1}{\phi_{2,j}^2} + \frac{1-e^{-\phi_{2,j}T}}{T\phi_{2,j}^3}).$$

If $\sqrt{N}/T$ tends to 0 and $\mathbb{E}(\phi_2^{-3}) < +\infty$, $\sum_{j\leq N}\mathbb{E}(A_{N,j}|\underline{\phi})$ tends *a.s.* to 0. Moreover,

$$\sum_{j\leq N}\mathbb{E}(A_{N,j}^2|\underline{\phi}) \to 0.$$

Therefore, $\sqrt{N}(\tilde{v}-v)$ converges in distribution, as T, N tend to infinity and $\sqrt{N}/T$ tends to 0, to $\mathcal{N}(0, \mathrm{Var}(\tau_j))$.

Assume that $\tau_j = v$ is deterministic. Then $\sqrt{NT}(\tilde{v}-v) = \frac{\sqrt{T}}{\sqrt{N}} \sum_{j\leq N}(\tilde{\tau}_j - v) = \sqrt{T}\sum_{j\leq N} A_{N,j}$. If N/T tends to 0,

$$\sqrt{T}\mathbb{E}(A_{N,j}|\underline{\phi}) = -\frac{2\sqrt{T}}{\sqrt{N}} \sum_{j\leq N} \Sigma_T(\phi_{2,j}) = -2\frac{\sqrt{N}}{\sqrt{T}}\frac{1}{N}\sum_{j\leq N}(\frac{1}{\phi_{2,j}^2} - \frac{1-e^{-\phi_{2,j}T}}{T\phi_{2,j}^3}) \to 0.$$

Moreover, (see Proposition 3.5),

$$T\sum_{j\leq N} \mathbb{E}(A_{N,j}^2|\underline{\phi}) \to 2v^3.$$

In this case, $\sqrt{NT}(\tilde{v}-v)$ converges in distribution to a centered Gaussian random variable with variance $2v^3$. □

Proof of Proposition 4.2 We have

$$\frac{1}{N}(V_N(\theta) - \ell_N(\theta)) = \frac{a-1}{N}\sum_{i=1}^N \left(\widehat{L}_i^{(k)} - \log\phi_i\right) - \frac{\lambda}{N}\sum_{i=1}^N \left(\hat{\phi}_i^{(k)} - \phi_i\right),$$

$$\begin{aligned}
\frac{1}{\sqrt{N}}(\frac{\partial V_N}{\partial a}(\theta) - \frac{\partial \ell_N}{\partial a}(\theta)) &= \frac{1}{\sqrt{N}}(\sum_{i=1}^N \left(\widehat{L}_i^{(k)} - \log\phi_i\right) \\
\frac{1}{\sqrt{N}}(\frac{\partial V_N}{\partial \lambda}(\theta) - \frac{\partial \ell_N}{\partial \lambda}(\theta)) &= \frac{1}{\sqrt{N}}(\sum_{i=1}^N \left(\hat{\phi}_i^{(k)} - \phi_i\right),
\end{aligned}$$

and

$$\frac{\partial V_N}{\partial a^2}(\theta) = \frac{\partial \ell_N}{\partial a^2}(\theta), \frac{\partial V_N}{\partial a\partial\lambda}(\theta) = \frac{\partial \ell_N}{\partial a\partial\lambda}(\theta), \frac{\partial V_N}{\partial \lambda^2}(\theta) = \frac{\partial \ell_N}{\partial \lambda^2}(\theta),$$

where

$$-\frac{1}{N}\partial^2\ell_n(\theta) = I(\theta).$$

To obtain the announced result, it is enough to prove that, as both N, T tend to infinity,

$$\frac{1}{N}\sum_{i=1}^N \left(\widehat{L}_i^{(k)} - \log\phi_i\right) = o_P(1), \quad \frac{1}{N}\sum_{i=1}^N \left(\hat{\phi}_i^{(k)} - \phi_i\right) = o_P(1),$$

and that, if moreover N/T tends to 0,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\widehat{L}_i^{(k)} - \log\phi_i\right) = o_P(1), \quad \frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\hat{\phi}_i^{(k)} - \phi_i\right) = o_P(1).$$

We have, applying Corollary 3.1 with $\gamma = 1$:

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^N(\hat{\phi}_i^{(k)} - \phi_i)\right)^2 &= N\mathbb{E}(\hat{\phi}_1^{(k)} - \phi_1)^2 + N(N-1)\left(\mathbb{E}(\hat{\phi}_1^{(k)} - \phi_1)\right)^2 \\
&\lesssim \frac{N + N(N-1)}{T}(1 + \mathbb{E}(\phi_1 + \phi_1^4 + \phi_1^{-2} + \phi_1^{-3}).
\end{aligned}$$

Thus, $\frac{1}{N}\sum_{i=1}^N \left(\hat\phi_i^{(k)} - \phi_i\right) = o_P(1)$ if both T, N tend to infinity. Under the additional constraint $N/T \to 0$, $\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\hat\phi_i^{(k)} - \phi_i\right) = o_P(1)$.

To complete the proof, it remains to study $\mathbb{E}\left(\widehat{L}_1^{(k)} - \log\phi_1\right)^2$.

Lemma 6.2. *If $a > 8$, $\mathbb{E}\left(\widehat{L}_1^{(k)} - \log\phi_1\right)^2 \lesssim T^{-1}$.*

Proof of Lemma 6.2. As we deal with only one trajectory, we drop the index 1 in all notations. Set $A = (\hat\phi \geq k/\sqrt{T}, V_T/T \geq k/\sqrt{T})$. We write:

$$\widehat{L}^{(k)} - \log\phi = (\log\hat\phi - \log\phi)1_A - \log\phi 1_{A^c} := T_1 + T_2. \tag{6.21}$$

On the set A, $\hat\phi = \hat\phi^{(k)}$. Applying the Taylor formula yields:

$$T_1 = 1_A(\hat\phi^{(k)} - \phi)\int_0^1 \frac{ds}{\phi + s(\hat\phi^{(k)} - \phi)} = 1_A\left(\frac{\hat\phi^{(k)} - \phi}{\phi} + \frac{(\hat\phi^{(k)} - \phi)^2}{\phi}\int_0^1 \frac{-s\,ds}{s(\hat\phi^{(k)} + (1-s)\phi)}\right).$$

On A, $s(\hat\phi^{(k)} + (1-s)\phi) \geq s\ k/\sqrt{T}$. Therefore,

$$|T_1| \leq \frac{|\hat\phi^{(k)} - \phi|}{\phi} + \frac{(\hat\phi^{(k)} - \phi)^2}{\phi}\frac{\sqrt{T}}{k}.$$

So,

$$\mathbb{E}(T_1^2|\phi) \leq \frac{2}{\phi^2}\left(\mathbb{E}((\hat\phi^{(k)} - \phi)^2|\phi) + \frac{T}{k^2}\mathbb{E}((\hat\phi^{(k)} - \phi)^4|\phi)\right).$$

By Corollary 3.1 with both $\gamma = 1$ and $\gamma = 2$,

$$\mathbb{E}(T_1^2|\phi) \lesssim \frac{1}{T\,\phi^2}\left(1 + \phi + \phi^4 + \phi^{-2} + \phi^{-3} + \phi^2 + \phi^8 + \phi^{-4} + \phi^{-6}\right).$$

Thus, provided that $a > 8$, $\mathbb{E}(T_1^2) \lesssim \frac{1}{T}$.
Next, we study T_2:

$$\begin{aligned}\mathbb{P}(A^c|\phi) &\leq \mathbb{P}(\hat\phi < \frac{k}{\sqrt{T}}|\phi) + \mathbb{P}(V_T/T < \frac{k}{\sqrt{T}}|\phi) \leq \mathbb{P}(\hat\phi < \frac{k}{\sqrt{T}}, V_T/T \geq \frac{k}{\sqrt{T}}|\phi) + 2\mathbb{P}(V_T/T < \frac{k}{\sqrt{T}}|\phi)\\ &\leq \mathbb{P}(\hat\phi^{(k)} < \frac{k}{\sqrt{T}}|\phi) + 2\mathbb{P}(V_T/T < \frac{k}{\sqrt{T}}|\phi).\end{aligned}$$

The bound for the second term above is given in (3.22): For all $p \geq 1$,

$$\mathbb{P}(V_T/T < k/\sqrt{T}|\phi) \lesssim T^{-p}\left(\phi^{2p} + \phi^{-2p} + \phi^{-p}\right).$$

Next, as $(\phi - k/\sqrt{T} \geq \phi/2) = (\phi \geq 2k/\sqrt{T})$, we have the bound for all $\gamma, s \geq 1$

$$\begin{aligned}\mathbb{P}(\hat\phi^{(k)} < \frac{k}{\sqrt{T}}|\phi) &\leq \mathbb{P}(|\phi - \hat\phi^{(k)}| \geq \frac{\phi}{2}) + 1_{(\phi^{-1} \geq \frac{\sqrt{T}}{2k})}\\ &\leq \left(\frac{2}{\phi}\right)^{2\gamma}\mathbb{E}(|\hat\phi^{(k)} - \phi|^{2\gamma}|\phi) + \left(\frac{2k}{\sqrt{T}}\right)^{2s}\phi^{-2s}.\end{aligned}$$

Finally, taking $\gamma = p = s = 1$,

$$\mathbb{E}(T_2^2|\phi) \lesssim T^{-1}(\log\phi)^2(\phi^{-1} + \phi^{-2} + \phi^2 + \phi^{-4} + \phi^{-5}).$$

Provided that $a > 5$, $\mathbb{E}(T_2^2) \lesssim T^{-1}$. The Lemma follows from the two inequalities for T_1 and T_2. □

The proof of Proposition 4.2 is now complete. □

References

Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equations with random effects. *Stoch. Proc. and Appl.* **123**, 2522-2551.

Davidian, M. and Giltinan, D. (1995). *Non linear models to repeated measurement data*. Chapman and Hall.

Delattre, M., Genon-Catalot, V. and Samson, A. (2012). Maximum likelihood estimation for stochastic differential equations with random effects. *Scand. J. Statist.* **40**, 322-343.

Delattre, M., Genon-Catalot, V. and Samson, A. (2014). Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. Preprint MAP5 2014-07.

Dion, C. (2014). Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model. *Preprint MAP5* 2014.

Ditlevsen, S. and De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *Statistical Journal* **3**, 137-153.

Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM P & S* **12**, 196-218.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051-1079.

Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application.* Academic press.

Karatzas, I. and Shreve, S.E. (2000). *Brownian motion and stochastic calculus*. Second edition. Springer-Verlag, New-York. q

Kessler, M., Lindner, A. and Sorensen, M. (editors) (2012). *Statistical methods for stochastic differential equations*. CRC Press. Chapman & Hall.

Kuhn, E. and Lavielle, M. (2004). Coupling an approximation version of em with an mcmc procedure. *ESAIM P & S* **8**, 115-131.

Kutoyants, Y.A. (2004). *Statistical inference for ergodic diffusion processes*. Springer Verlag.

Maitra, T. and Bhattacharya, S. (2014). On asymptotics related to classical inference in stochastic differential equations with random effects. *arXiv:1407.3968v1.* 1-12.

Nie, L. (2006).Strong consistency of the maximum likelihood estimator in generalizedlinear and nonlinear mixed-effects models. *Metrika* **63**,123-143.

Nie, L. (2007). Convergence rate of the MLE in generalized linear and nonlinear mixed-effects models: theory and applications. *Journal of Statistical Planning and Inference* **137**, 1787-1804.

Nie, L. and Yang, M. (2005). Strong consistency of the MLE in nonlinear mixed-effects models with large cluster size. *Sankhya: The Indian Journal of Statistics* **67**, 736-763.

Overgaard, R.V., Jonsson, N., Tornoe, 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.

Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Statist.* **37**, 67-90.

Picchini, U. and Ditlevsen, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55**, 1426-1444.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, Springer.

Corresponding author: C. Larédo, e-mail: catherine.laredo@jouy.inra.fr