Parametric inference for discrete observations of diffusion processes with mixed effects

Maud Delattre, Valentine Genon-Catalot, Catherine Larédo

Parametric inference for discrete observations of diffusion processes with mixed effects

Maud Delattre[1], Valentine Genon-Catalot[2] and Catherine Larédo[3*]

[1] AgroParisTech, France

[2] UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, Sorbonne Paris Cité, France

[3] Laboratoire MaIAGE, I.N.R.A. and LPMA, Université Denis Diderot, CNRS-UMR 7599.

Abstract

Stochastic differential equations with mixed effects provide means to model intraindividual and interindividual variability in biomedical experiments based on longitudinal data. We consider N *i.i.d.* stochastic processes $(X_i(t), t \in [0,T])$, $i = 1, \ldots, N$, defined by a stochastic differential equation with linear mixed effects. We consider a parametric framework with distributions leading to explicit approximate likelihood functions and investigate the asymptotic behaviour of estimators under the double asymptotic framework: the number N of individuals (trajectories) and the number n of observations per individual tend to infinity within the fixed time interval $[0,T]$. The estimation method is assessed on simulated data for various models comprised in our framework.

Key Words: discrete observations, estimating equations, mixed-effects models, parametric inference, stochastic differential equations.

1 Introduction

Stochastic differential equations with mixed effects (SDEMEs) have received increasing interest due to their ability to model the whole inherent variability of biomedical experiments which usually contain repeated measurements over time on several experimental units. This occurs especially in pharcokinetics/pharmacodynamics modelling. Mixed effects models use random variables for the individual specific parameters and thus provide means to model interindividual variability. Using stochastic differential equations allow in addition to take into account the inherent randomness of the intraindividual dynamics. Hence, SDEMEs provide a good framework to study population characteristics when longitudinal data are collected on multiple individuals ruled by the same intraindividual mechanisms (see *e.g.* Overgaard *et al.* (2005), Ditlevsen and De Gaetano (2005b), Picchini *et al.* (2008), Møller *et al.* (2010), Donnet *et al.* (2010), Berglund *et al.* (2001), Leander *et al.* (2015) for discussions and applications on real data sets).

The estimation of population parameters, *i.e.* the parameters of the random effects distributions, in SDEMEs is a delicate problem because there is in general no closed form for the likelihood function. In the abstract framework of non linear mixed effects models, theoretical properties of the exact maximum likelihood estimator (MLE) of parameters are studied in Nie and Yang (2005), Nie (2006),Nie (2007)

*Corresponding author

Email addresses: maud.delattre@agroparistech.fr (Maud Delattre), valentine.genon-catalot@parisdescartes.fr (Valentine Genon-Catalot), catherine.laredo@jouy.inra.fr (Catherine Larédo)

under several asymptotic frameworks: the number of subjects and/or the number of observations per subject goes to infinity. The results rely on many assumptions difficult to check in practice. Other approaches coupled with implementation estimation algorithms have been developped in *e.g.*Overgaard *et al.* (2005), Ditlevsen and De Gaetano (2005a) (exact likelihood for Brownian motion with drift), in Picchini *et al.* (2010), Picchini and Ditlevsen (2011) (approximation of the likelihood).

For SDEMEs continuously observed throughout a time interval $[0, T]$, with random effect in the drift term, estimation of the random effect itself is possible on the basis of one trajectory as T tends to to infinity. This allows to build, when N trajectories are available, plug-in type parametric or nonparametric estimators of the random effects distributions (see *e.g.* Comte *et al.* (2013), Dion and Genon-Catalot (2015), Genon-Catalot and Larédo (2015)).

In this paper, we study discretely observed SDEMEs with linear random effect in the drift and fixed effect in the diffusion coefficient or fixed effect in the drift and random effect in the diffusion coefficient. We consider here a parametric framework with distributions leading to explicit approximate likelihood functions and study the asymptotic behaviour of the associated estimators under the double asymptotic framework: the number N of trajectories and the number n of sample points per trajectory both tend to infinity while the time interval of observations is fixed.

More precisely, we consider N real valued stochastic processes $(X_i(t), t \geq 0)$, $i = 1, \ldots, N$, with dynamics ruled by the following SDEME:

$$d\,X_i(t) = \Phi_i' b(X_i(t))dt + \Psi_i \sigma(X_i(t))\, dW_i(t), \quad X_i(0) = x,\, i = 1, \ldots, N, \tag{1}$$

where $(W_1, \ldots, W_N)$ are N independent Wiener processes, $(\Phi_i, \Psi_i), i = 1, \ldots N$ are N *i.i.d.* $\mathbb{R}^d \times (0, +\infty)$-valued random variables, $((\Phi_i, \Psi_i), i = 1, \ldots, N)$ and $(W_1, \ldots, W_N)$ are independent and x is a known real value. The functions $\sigma(.) : \mathbb{R} \to \mathbb{R}$ and $b(.) = (b_1(.), \ldots, b_d(.))' : \mathbb{R} \to \mathbb{R}^d$ are known. Each process $(X_i(t))$ represents an individual and the $d+1$-dimensional random vector (Φ_i, Ψ_i) represents the (multivariate) random effect of individual i. We assume that each process $(X_i(t))$ is discretely observed on a fixed time interval $[0, T]$ with $T > 0$ at n times $t_j = jT/n$. We consider models with parametric distributions for the random effects. Our aim is to estimate the unknown parameters from the observations $\{X_i(t_j), j = 1, \ldots, n, i = 1, \ldots, N\}$, as n, N go to infinity. We consider the two following cases:

(1) We set $\Psi_i = \psi = \gamma^{-1/2}$ unknown and assume that Φ_i is a d-dimensional Gaussian vector $\mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$. We denote $\theta = (\gamma, \boldsymbol{\mu}, \boldsymbol{\Omega})$.

(2) We set $\Phi_i = \varphi$ unknown, $\Psi_i = \Gamma_i^{-1/2}$ and assume that Γ_i has Gamma distribution $G(a, \lambda)$ with density $(\lambda^a/\Gamma(a))\gamma^{a-1} \exp(-\lambda\gamma)$ on $(0, +\infty)$. We denote $\tau = (\lambda, a, \varphi)$.

Random effects are often modelled by normal laws when they are real valued or log-normal laws when they are positive. Gamma distributions are another classical way for the latter case. In addition, choices (1)-(2) lead to closed-form formulae for the likelihood function. If the sample paths are continously observed throughout the time interval $[0, T]$ and for known γ, the exact likelihood of $(X_i(t), t \in [0, T], i = 1, \ldots, N)$ for case (1) was obtained in Delattre *et al.* (2013). Here, we consider discrete observations and assume that the parameter γ in the diffusion coefficient is unknown. Model (1) with nul drift and discrete observations is considered in Delattre *et al.* (2015) with $\Psi_i = \Gamma_i^{-1/2}$ and Γ_i with Gamma distribution $G(a, \lambda)$. Case (2) extends this paper by considering a non null drift depending on an unknown parameter.

We present in Section 2 the model and assumptions and give preliminary lemmas linked with sample paths discretisations. Section 3 gives the approximate likelihood functions for cases (1)-(2) derived from the Euler-scheme approximation of equation (1). In Section 4, we study the asymptotic properties of estimators (Theorem 1, Theorem 2). In case (1), the fixed-effect parameter γ is estimated with rate $\sqrt{Nn}$ and the population parameters $(\boldsymbol{\mu}, \boldsymbol{\Omega})$ are estimated with rate $\sqrt{N}$. The case where $\boldsymbol{\Omega}$ is singular is included in Theorem 1 which allows to consider mixed effects as some components of Φ_i can be deterministic. In case (2), all parameters are estimated with the same rate $\sqrt{N}$. Moreover, the estimators of population parameters are asymptotically equivalent to the maximum likelihood estimators based on direct observations of N *i.i.d.* Gamma random variables. Section 5 provides numerical simulation results for various examples of SDEMEs and different values of N, n and illustrate the performances of the present framework for statistical inference. Section 6 gives concluding remarks. Proofs are gathered in Section 7 and auxiliary results in Section 8.

2 Model, assumptions and preliminaries

Consider N real valued stochastic processes $(X_i(t), t \geq 0)$, $i = 1, \ldots, N$, with dynamics ruled by (1). The processes $(W_1, \ldots, W_N)$ and the r.v.'s $(\Phi_i, \Psi_i), i = 1, \ldots, N$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the filtration $(\mathcal{F}_t = \sigma(\Phi_i, \Psi_i, W_i(s), s \leq t, i = 1, \ldots, N), t \geq 0)$.
The canonical space associated with one trajectory on $[0, T]$ is given by $((\mathbb{R}^d \times (0, +\infty) \times C_T))$ where C_T denotes the space of real valued continuous functions on $[0, T]$ endowed with the usual Borel σ-algebra. Let $\theta = (\gamma, \boldsymbol{\mu}, \boldsymbol{\Omega})$ in case (1) (resp. $\tau = (\lambda, a, \varphi)$ in case (2)) denote the unknown parameter. Let P_θ in case (1) (resp. P_τ in case (2)) be the distribution of $(\Phi_i, \Psi_i, (X_i(t), t \in [0, T]))$. For the N trajectories, the canonical space is $\prod_{i=1}^N ((\mathbb{R}^d \times (0, +\infty) \times C_T), \mathbb{P}_\theta = \otimes_{i=1}^N P_\theta)$ (resp. $\prod_{i=1}^N ((\mathbb{R}^d \times (0, +\infty) \times C_T), \mathbb{P}_\tau = \otimes_{i=1}^N P_\tau))$.
We introduce the following assumptions:

(H1) The real valued functions $x \to b_j(x)$, $j = 1, \ldots, d$ and $x \to \sigma(x)$ are C^2 on $\mathbb{R}$ with first and second derivatives bounded by L. The function $\sigma(.)$ is lower bounded : $\exists \sigma_0 > 0, \forall x \in \mathbb{R}, \sigma(x) \geq \sigma_0$.

(H2) There exists a constant K such that, $\forall x \in \mathbb{R}$, $\|b(x)\| + |\sigma(x)| \leq K$.
($\|.\|$ denotes the Euclidian norm of $\mathbb{R}^d$.)

(H3) The matrix $V_i(T)$ is positive definite $\mathbb{P}_\theta$-a.s. for all θ (resp. $\mathbb{P}_\tau$-a.s. for all τ), where

$$V_i(T) = \left(\int_0^T \frac{b_k(X_i(s)) b_\ell(X_i(s))}{\sigma^2(X_i(s))} ds \right)_{1 \leq k, \ell \leq d}, \tag{2}$$

Assumption (H1) is standard and ensures that, for $i = 1, \ldots, N$, for all deterministic $(\varphi, \psi) \in \mathbb{R}^d \times (0, +\infty)$, the stochastic differential equation

$$d X_i^{\varphi,\psi}(t) = \varphi' b(X_i^{\varphi,\psi}(t)) dt + \psi \sigma(X_i^{\varphi,\psi}(t))\, dW_i(t), \quad X_i^{\varphi,\psi}(0) = x \tag{3}$$

admits a unique strong solution process $(X_i^{\varphi,\psi}(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t)$. Moreover, the stochastic differential equation with random effects (1) admits a unique strong solution adapted to $(\mathcal{F}_t)$ such that the joint process $(\Phi_i, \Psi_i, X_i(t), t \geq 0)$ is strong Markov and the conditional distribution of $(X_i(t))$ given $\Phi_i = \varphi, \Psi_i = \psi$ is identical to the distribution of (3). The processes $(\Phi_i, \Psi_i, X_i(t), t \geq$

$0)), i=1,\ldots,N$ are *i.i.d.*. (see *e.g.* Delattre *et al.* (2013), Genon-Catalot and Larédo (2015), Comte *et al.* (2013)).

Assumption (H2) may appear strong. However, it ensures, in particular, that, for any distribution of (Φ_i, Ψ_i) such that $\mathbb{E}(\|\Phi_i\|^{2p} + \Psi_i^{2p}) < +\infty$, $\mathbb{E}(X_i(t)^{2p}) < +\infty$. If (H2) does not hold, this property may not be satisfied (see Section 8 for a discussion). Assumption (H3) ensures that all the components of Φ_i can be estimated. If the functions (b_k/σ^2) are not linearly independent, the dimension of Φ_i is not well defined and (H3) is not fulfilled. If $d=1$, (H3) is obviously verified.

As usual in mixed effects models, the likelihood of the i-th vector of observations $(X_i(t_j), j=1,\ldots,n)$ is computed in two steps. First, we consider the conditional likelihood given the random effects $\Phi_i = \varphi, \Psi_i = \psi$. Then, we integrate it with respect to the distribution of the random effects. As we have discrete observations, the exact conditional likelihood given the random effects, *i.e.* the exact likelihood of a discretized sample of (3), is untractable because the transition densities are not explicit. Therefore, we use an approximate conditional likelihood derived from the likelihood of the Euler scheme associated with equation (3).

Let us first introduce the notations that we need in the sequel, using the notation Y' for the transposition of a vector or a matrix Y.

$$\Delta_n = \Delta = \frac{T}{n},\ X_{i,n} = X_i = (X_i(t_j)),\ \ t_{j,n} = t_j = jT/n \text{ for } j=1,\ldots,n,$$

$$S_{i,n} = S_i = \frac{1}{\Delta}\sum_{j=1}^{n} \frac{(X_i(t_j) - X_i(t_{j-1}))^2}{\sigma^2(X_i(t_{j-1}))}, \tag{4}$$

$$V_{i,n} = V_i = \left(\sum_{j=1}^{n} \Delta \frac{b_k(X_i(t_{j-1}))b_\ell(X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))}\right)_{1\le k,\ell\le d}, \tag{5}$$

$$U_{i,n} = U_i = \left(\sum_{j=1}^{n} \frac{b_k(X_i(t_{j-1}))(X_i(t_j) - X_i(t_{j-1}))}{\sigma^2(X_i(t_{j-1}))}\right)_{1\le k\le d}, \tag{6}$$

$$U_i(T) = \left(\int_0^T \frac{b_k(X_i(s))}{\sigma^2(X_i(s))} dX_i(s)\right)_{1\le k\le d}. \tag{7}$$

Using that the Euler scheme increments are conditionally Gaussian random variables yields that the approximate likelihood obtained by the Euler scheme discretisation for (3) with fixed φ and $\psi = 1/\sqrt{\gamma}$ is (up to constants) given by:

$$L_n(X_{i,n}, \gamma, \varphi) = \gamma^{n/2} \exp\left[-\frac{\gamma}{2}(S_{i,n} + \varphi' V_{i,n}\varphi - 2\varphi' U_{i,n})\right], \quad \text{where} \tag{8}$$

$$S_{i,n} + \varphi' V_{i,n}\varphi - 2\varphi' U_{i,n} = \sum_{j=1}^{n} \frac{(X_i(t_j) - X_i(t_{j-1}) - \Delta\varphi' b(X_i(t_{j-1})))^2}{\Delta\sigma^2(X_i(t_{j-1}))}. \tag{9}$$

As n tends to infinity, $S_{i,n}/n$ converges to Γ_i^{-1} in probability, $V_{i,n}$ converges a.s. to $V_i(T)$ (defined in (2)) and $U_{i,n}$ converges in probability to $U_i(T)$. Hence, Assumption (H3) ensures that for n large enough $V_{i,n}$ is positive definite.

In the general case, let $\nu_\vartheta(d\gamma, d\varphi)$ denote the joint distribution (Ψ_i, Φ_i) depending on an unknown parameter ϑ, we propose to define an approximate likelihood function by:

$$\mathcal{L}_{N,n}(\vartheta) = \prod_{i=1}^{N} \mathcal{L}_n(X_{i,n}, \vartheta) \qquad \text{with} \qquad \mathcal{L}_n(X_{i,n}, \vartheta) = \int L_n(X_{i,n}, \gamma, \varphi)\nu_\vartheta(d\gamma, d\varphi).$$

This function could be studied theoretically by using tools analogous to the ones developped by Nie and Yang (2005), Nie (2006), Nie (2007). However, proofs are technically heavy. This is why we rather focus on distributions for the random effects leading to explicit formulae for $\mathcal{L}_{N,n}(\vartheta)$ and thus to explicit estimators of the unknown parameters.

Let us first state some preliminary lemmas. In the first two ones, we set $X_1(t) = X(t), \Phi_1 = \Phi, \Psi_1 = \Psi$.

Lemma 1. *Under (H1)-(H2), for $s \le t$ and $t - s \le 1$, $p \ge 1$,*

$$\mathbb{E}_\vartheta(|X(t) - X(s)|^p | \Phi = \varphi, \Psi = \psi) \lesssim K^p (t-s)^{p/2}(\|\varphi\|^p + \psi^p).$$

For $t \to H(t, X.)$ a predictable process, let $V(H;T) = \int_0^T H(s, X.)ds$ and $U(H;T) = \int_0^T H(s, X.)dX(s)$.

Lemma 2. *Assume (H1)-(H2) and $p \ge 1$. If H is bounded, $\mathbb{E}_\vartheta(|U(H;T)|^p | \Phi = \varphi, \Psi = \psi) \lesssim \|\varphi\|^p + \psi^p$. Consider $f : \mathbb{R} \to \mathbb{R}$ and set $H(s, X.) = f(X(s)), H_n(s, X.) = \sum_{j=1}^n f(X((j-1)\Delta))1_{((j-1)\Delta, j\Delta]}(s)$. If f is Lipschitz,*

$$\mathbb{E}_\vartheta(|V(H;T) - V(H_n;T))|^p | \Phi = \varphi, \Psi = \psi) \lesssim \Delta^{p/2}(\|\varphi\|^p + \psi^p). \tag{10}$$

If f is C^2 with f', f'' bounded

$$\mathbb{E}_\vartheta(|U(H;T) - U(H_n;T))|^p | \Phi = \varphi, \Psi = \psi) \lesssim \Delta^{p/2}(\|\varphi\|^{2p} + \|\varphi\|^p \psi^p + \psi^{2p} + \psi^{3p}). \tag{11}$$

We introduce the pivotal random variable which is needed to get the right constraint on N and n:

$$S_{i,n}^{(1)} = \frac{1}{\Gamma_i} C_{i,n}^{(1)}, \quad \text{where} \quad C_{i,n}^{(1)} = \sum_{j=1}^{n} \frac{(W_i(t_j) - W_i(t_{j-1})))^2}{\Delta} \sim \chi^2(n) \tag{12}$$

which corresponds to $S_{i,n}$ when $b(.) = 0, \sigma(.) = 1$ and has an explicit distribution as Γ_i and $C_{i,n}^{(1)}$ are independent.

Lemma 3. *Assume (H1)-(H2) and $p \ge 1$. Then,*

$$\mathbb{E}_\vartheta \left| \frac{S_{i,n}}{n} - \frac{S_{i,n}^{(1)}}{n} \right|^p |\Phi_i = \varphi, \Psi_i = \psi) \lesssim \Delta^p (1 + \psi^{2p}\|\varphi\|^{2p} + \psi^{4p} + \|\varphi\|^{2p})$$

In the sequel, when there is no ambiguity, the subscript n is omitted in notations.

3 Approximate Likelihoods with random effects.

In this paragraph, we compute the approximate likelihoods in cases (1)-(2). We denote $\Lambda_n(X_i, \theta)$ the likelihood associated to the sample path X_i observed with sampling interval T/n and $\Lambda_{N,n}(\theta)$ the likelihood of the N paths in case (1) (resp. $\widetilde{\Lambda}_n(X_i, \tau), \widetilde{\Lambda}_{N,n}(\tau)$ in case (2)).

Case (1): Φ_i random , $\Psi_i = \psi = 1/\sqrt{\gamma}$ deterministic and $\theta = (\gamma, \boldsymbol{\mu}, \boldsymbol{\Omega})$.
Let I_d denote the identity matrix of $\mathbb{R}^d$ and set, for $i = 1, \dots, N$,

$$R_i^{-1} = (I_d + V_i\boldsymbol{\Omega})^{-1}V_i. \tag{13}$$

Proposition 1. *Assume that for $i = 1, \dots, N, \Psi_i = \gamma^{-1/2}$ with $\gamma > 0$ deterministic and that Φ_i has distribution $\mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$. Then, under (H1) and (H3), an explicit approximate likelihood for $(X_{i,n})$ is*

$$\begin{aligned}\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}) &= \gamma^{n/2}(det(I_d + V_i\boldsymbol{\Omega}))^{-1/2} \\ &\quad \exp\Big[-\frac{\gamma}{2}\left(S_i + (\boldsymbol{\mu} - V_i^{-1}U_i)'R_i^{-1}(\boldsymbol{\mu} - V_i^{-1}U_i) - U_i'V_i^{-1}U_i\right)\Big].\end{aligned}$$

Proof. Assume first that $\boldsymbol{\Omega}$ is invertible and define

$$\boldsymbol{\Sigma}_i = (\boldsymbol{\Omega}^{-1} + V_i)^{-1}, \quad \mathbf{m}_i = \boldsymbol{\Sigma}_i(U_i + \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}). \tag{14}$$

$$T_{i,n}(\boldsymbol{\mu}, \boldsymbol{\Omega}) = T_i(\boldsymbol{\mu}, \boldsymbol{\Omega}) = \boldsymbol{\mu}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - \mathbf{m}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{m}_i. \tag{15}$$

Integrating (8) with respect to the distribution of Φ_i yields:

$$\begin{aligned}\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}) &= \gamma^{n/2}\exp(-\frac{\gamma}{2}S_i)\frac{\gamma^{d/2}}{(2\pi)^{d/2}(det(\boldsymbol{\Omega}))^{1/2}} \times \\ &\quad \int_{\mathbb{R}^d}\exp(\gamma(\varphi'U_i - \frac{1}{2}\varphi'V_i\varphi))\exp(-\frac{\gamma}{2}(\varphi - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\varphi - \boldsymbol{\mu}))d\varphi \\ &= \gamma^{n/2}\exp(-\frac{\gamma}{2}S_i)\left(\frac{det(\boldsymbol{\Sigma}_i)}{det(\boldsymbol{\Omega})}\right)^{1/2}\exp(-\frac{\gamma}{2}T_i(\boldsymbol{\mu}, \boldsymbol{\Omega})).\end{aligned}$$

Noting that $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}(I_d + V_i\boldsymbol{\Omega})^{-1}$, we get $\frac{det(\boldsymbol{\Sigma}_i)}{det(\boldsymbol{\Omega})} = (det(I_d + V_i\boldsymbol{\Omega}))^{-1}$. Some computations using matrices equalities and (5),(6), (13) yield the other expression for T_i

$$T_i(\boldsymbol{\mu}, \boldsymbol{\Omega}) = (\boldsymbol{\mu} - V_i^{-1}U_i)'R_i^{-1}(\boldsymbol{\mu} - V_i^{-1}U_i) - U_i'V_i^{-1}U_i. \tag{16}$$

Now, the formula for $\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega})$ is valid even for non invertible $\boldsymbol{\Omega}$. So the same conclusion holds as we can apply the Scheffé theorem to conclude. □

Afterwards, the independence of the X_i's yields that, for the observation $(X_i, i = 1, \dots, N)$,

$$\Lambda_{N,n}(\gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}) = \prod_{i=1}^{N}\Lambda_n(X_{i,n}, \gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}). \tag{17}$$

***Remark* 1.** *When γ is known and $\boldsymbol{\Omega}$ invertible, the exact likelihood associated with a continuous observation of $(X_i(t), t \in [0,T], i = 1, \ldots, N)$ is studied in Delattre* et al. *(2013). Proposition 1 considers the case of discrete observations and the presence of an unknown parameter γ in the diffusion coefficient;*

Case (2): $\Phi_i = \varphi$ unknown deterministic parameter and $\Psi_i = 1/\sqrt{\Gamma_i}$ with $\Gamma_i \sim G(a,\lambda)$ and $\tau = (\lambda, a, \varphi)$. Integrating (8) with respect to the distribution of Γ_i yields:

$$\widetilde{\Lambda}_n(X_{i,n},\tau) = \frac{\lambda^a \Gamma(a + (n/2))}{\Gamma(a)(\lambda + \frac{1}{2}(S_i - 2\varphi' U_i + \varphi' V_i \varphi))^{a+(n/2)}}, \tag{18}$$

This expression is always well defined since the denominator is positive by (9). For the N paths, we get the approximate likelihood,

$$\widetilde{\Lambda}_{N,n}(\tau) = \prod_{i=1}^{N} \widetilde{\Lambda}_n(X_{i,n},\tau). \tag{19}$$

4 Asymptotic properties of estimators.

In this section, we study the asymptotic behaviour of the estimators based on the approximate likelihood functions of the previous section. First, a natural question arises here about the comparison with direct observation of an *i.i.d.* sample (Φ_i) or (Γ_i). The case where φ, ψ are both deterministic is detailed at the end of Section 4.3.

4.1 Direct observation of the random effects

Assume that a sample $(\Phi_i, i = 1, \ldots, N)$ or $(\Gamma_i, i = 1, \ldots, N)$ is observed. Let $\tilde{\ell}_N(\mu, \omega^2)$ (resp. $\ell_N(\lambda, a)$) denote the likelihood of of the N-sample (Φ_i) (resp. (Γ_i)). For the Gaussian observations, we have

$$\frac{\partial}{\partial \mu}\tilde{\ell}_N(\mu,\omega^2)) = \gamma\omega^{-2}\sum_{i=1}^{N}(\mu - \Phi_i), \quad \frac{\partial}{\partial \omega^2}\tilde{\ell}_N(\mu,\omega^2)) = \frac{\gamma}{2\omega^4}\sum_{i=1}^{N}\left((\Phi_i - \mu)^2 - \gamma^{-1}\omega^2\right).$$

The Fisher information matrix is $J_0(\mu,\omega^2) = \begin{pmatrix} \frac{\gamma}{\omega^2} & 0 \\ 0 & \frac{1}{2\omega^4} \end{pmatrix}$. The parameter γ cannot be estimated (only $\gamma\omega^{-2}$ is identifiable).

If we observe a sample (Γ_i) with Gamma distribution $G(a,\lambda)$, setting $\psi(a) = \Gamma'(a)/\Gamma(a)$ (the di-Gamma function), the score function is given by

$$\frac{\partial}{\partial \lambda}\ell_N(\vartheta) = \sum_{i=1}^{N}\left(\frac{a}{\lambda} - \Gamma_i\right), \quad \frac{\partial}{\partial a}\ell_N(\vartheta) = \sum_{i=1}^{N}\left(-\psi(a) + \log\lambda + \log\Gamma_i\right)$$

and the Fisher information matrix is

$$I_0(\lambda, a) = \begin{pmatrix} \frac{a}{\lambda^2} & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \psi'(a) \end{pmatrix}. \tag{20}$$

Using properties of the di-gamma function $(a\psi'(a) - 1 \neq 0)$, $I_0(a,\lambda)$ is invertible for all $(a,\lambda) \in (0, +\infty)^2$ (see Section 4.1)).

4.2 Random effect in the drift and fixed effect in the diffusion coefficient

4.2.1 Univariate random effect in the drift coefficient

We assume $d = 1$, $\Phi_i \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$, $\Psi_i = \gamma^{-1/2}$ (fixed, unknown). So $\theta = (\gamma, \mu, \omega^2) \in (0, +\infty) \times \mathbb{R} \times (0, +\infty)$. The associated approximate likelihoood obtained in Proposition 1 is given by (17) with

$$\Lambda_n(X_{i,n}, \theta) = \frac{\gamma^{n/2}}{(1 + \omega^2 V_{i,n})^{1/2}} \exp\left[-\frac{\gamma}{2}(S_i + T_i(\mu, \omega^2))\right],$$

where U_i, V_i are defined in (5)-(6) and (see (15) and (16))

$$T_i(\mu, \omega^2) = \frac{V_i}{(1 + \omega^2 V_i)}\left(\mu - \frac{U_i}{V_i}\right)^2 - \frac{U_i^2}{V_i} = \frac{\mu^2}{\omega^2} - \frac{(U_i + \omega^{-2}\mu)^2}{(V_i + \omega^{-2})}. \tag{21}$$

Set $\ell_{N,n}(\theta) = \log \Lambda_{N,n}(\theta)$ and define the (pseudo-)score function associated with $\log \Lambda_{N,n}(\theta)$

$$G_{N,n}(\theta) = \left(\frac{\partial}{\partial \gamma}\ell_{N,n}(\theta) \;\; \frac{\partial}{\partial \mu}\ell_{N,n}(\theta) \;\; \frac{\partial}{\partial \omega^2}\ell_{N,n}(\theta)\right)'. \tag{22}$$

We study the estimators defined by the estimating equation:

$$G_{N,n}(\tilde{\theta}_{N,n}) = 0. \tag{23}$$

Define the random variables using (5),(6),(2) and (7),

$$A_{i,n} = A_i = \frac{U_{i,n} - \mu V_{i,n}}{1 + \omega^2 V_{i,n}} \quad \text{and} \quad B_{i,n} = B_i = \frac{V_{i,n}}{1 + \omega^2 V_{i,n}}, \tag{24}$$

$$A_i(T; \mu, \omega^2) = \frac{U_i(T) - \mu V_i(T)}{1 + \omega^2 V_i(T)} \quad \text{and} \quad B_i(T; \omega^2) = \frac{V_i(T)}{1 + \omega^2 V_i(T)}. \tag{25}$$

In the proofs, we use that $A_{i,n}$ (resp. $B_{i,n}$) converges to $A_i(T; \mu, \omega^2)$ (resp. $B_i(T; \omega^2)$ and control their differences (see Lemma 4 in the Appendix). The vector (22) is given by:

$$\begin{aligned}
\frac{\partial}{\partial \gamma}\ell_{N,n}(\theta) &= \frac{N\,n}{2\gamma} - \frac{1}{2}\sum_{i=1}^{N}\left(S_i + T_i(\mu, \omega^2)\right) \;,\; \frac{\partial}{\partial \mu}\ell_{N,n}(\theta) = -\frac{\gamma}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mu}T_i(\mu, \omega^2) = \gamma\sum_{i=1}^{N} A_i, \\
\frac{\partial}{\partial \omega^2}\ell_{N,n}(\theta) &= -\frac{\gamma}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \omega^2}T_i(\mu, \omega^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{V_i}{1 + \omega^2 V_i} = \frac{1}{2}\sum_{i=1}^{N}(\gamma A_i^2 - B_i).
\end{aligned}$$

The parameters γ and μ, ω^2 have different rates of convergence. Thus, we define the rate matrix $D_{N,n}$ and the Fisher information matrix $\mathcal{I}(\theta)$ by

$$D_{N,n} = \begin{pmatrix} \frac{1}{\sqrt{Nn}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{N}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{N}} \end{pmatrix}, \qquad \mathcal{I}(\theta) = \left(\begin{array}{c|c} \frac{1}{2\gamma^2} & \mathbf{0} \\ \hline \mathbf{0} & I(\theta) \end{array}\right) \quad \text{where} \tag{26}$$

$$I(\theta) = \begin{pmatrix} \gamma \mathbb{E}_\theta B_1(T;\omega^2) & \gamma \mathbb{E}_\theta A_1(T;\mu,\omega^2) B_1(T;\omega^2) \\ \gamma \mathbb{E}_\theta A_1(T;\mu,\omega^2) B_1(T;\omega^2) & \mathbb{E}_\theta \left(\gamma A_1^2(T;\mu,\omega^2) B_1(T;\omega^2) - \frac{1}{2} B_1^2(T;\omega^2)\right) \end{pmatrix}. \tag{27}$$

It is proved in Delattre *et al.* (2013) (see Section 8.3) that $I(\theta)$ is the covariance matrix of the vector

$$\begin{pmatrix} \gamma A_1(T;\mu,\omega^2) \\ \frac{1}{2}(\gamma A_1^2(T,\mu,\omega^2) - B_1(T;\omega^2)) \end{pmatrix}. \tag{28}$$

The following holds:

Theorem 1. *Assume (H1)-(H2) and that $I(\theta)$ is invertible. Then, if N, n tend to infinity with $N/n \to 0$, with probability tending to 1, there exists a solution to (23), $\tilde{\theta}_{N,n}$, which is consistent and such that $D_{N,n}^{-1}(\tilde{\theta}_{N,n} - \theta)$ converges in distribution under $\mathbb{P}_\theta$ to $\mathcal{N}_3(0, \mathcal{I}^{-1}(\theta))$ where $\mathcal{I}(\theta)$ is defined in (26).*

Note that, if the sample paths $(X_i(t)), t \in [0,T], i = 1,\ldots,N)$ are continuouly observed throughout $[0,T]$ and if γ is known, the strong consistency of the exact maximum likelihood $(\hat{\mu}_N, \hat{\omega}_N^2)$ of (μ,ω^2) is proved in Delattre *et al.* (2013) together with the convergence in distribution of $\sqrt{N}((\hat{\mu}_N - \mu), (\hat{\omega}_N^2 - \omega^2))$ to $\mathcal{N}_2(0, I^{-1}(\theta))$. Hence, there is no loss of efficiency due to the discrete observations under the constraint $N/n \to 0$.

The parameter γ can be estimated from each trajectory: it is well known that, for each i, $(\sqrt{n}((n/S_{i,n}) - \gamma)$ converges, as n tends to infinity, in distribution to $\mathcal{N}(0, 1/2\gamma^2)$. On the basis of N trajectories, $\tilde{\gamma}_{N,n}$ has the same asymptotic distribution with rate $\sqrt{Nn}$.

We can also compare the above result to the case of direct observation of (Φ_i). If ω^2 is known, the asymptotic variance of $\tilde{\mu}_{N,n}$ is (see (25) and Section 4.1)

$$\left(\gamma \mathbb{E}_\theta B_1(T;\omega^2)\right)^{-1} \geq \frac{\omega^2}{\gamma}.$$

So there is a loss of efficiency w.r.t. the direct observation of the random effects.

It is difficult to compare the asymptotic variance of $\tilde{\omega}_{N,n}^2$ to the corresponding one in case of direct observation of (Φ_i) (see the expression in $I(\theta)$). We only do it on a simple example.

Example. Consider the model $X_i(t) = \Phi_i t + \gamma^{-1/2} W_i(t)$ $(b = 1, \sigma = 1)$. Then, $\ell_{N,n}(\theta)$ is the exact likelihood, $U_i = X_i(T), V_i = T$, the estimation of (μ,ω^2) corresponds to the observation of $X_i(T) = \Phi_i T + \gamma^{-1/2} W_i(T)$. We have:

$$\frac{\partial}{\partial \gamma}\ell_{N,n}(\theta) = \frac{N\,n}{2\gamma} - \frac{1}{2}\sum_{i=1}^N [S_i + \frac{T}{1+\omega^2 T}(\frac{X_i(t)}{T} - \mu)^2]\;, \quad \frac{\partial}{\partial \mu}\ell_{N,n}(\theta) = \frac{\gamma}{1+\omega^2 T}\sum_{i=1}^N (X_i(T) - \mu T),$$

$$\frac{\partial}{\partial \omega^2}\ell_{N,n}(\theta) = \frac{1}{2}\sum_{i=1}^N [\gamma \frac{(X_i(T) - \mu T)^2}{(1+\omega^2 T)^2} - \frac{T}{1+\omega^2 T}], \quad I(\theta) = \begin{pmatrix} \gamma \frac{T}{1+\omega^2 T} & 0 \\ 0 & \frac{1}{2}\left(\frac{T}{1+\omega^2 T}\right)^2 \end{pmatrix}.$$

We can see that $2\left(\frac{1+\omega^2 T}{T}\right)^2 \geq 2\omega^4$, thus a larger variance than for direct observation.

4.2.2 Multivariate random effect in the drift coefficient

Let us now consider that $\Phi_i \sim \mathcal{N}_d(\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Omega})$, $\Psi_i = \gamma^{-1/2}$ (fixed, unknown) and $\theta = (\gamma, \boldsymbol{\mu}, \boldsymbol{\Omega}) \in (0, +\infty) \times \mathbb{R}^d \times M(d)$, where $M(d)$ is the set of nonnegative symmetric matrices of $\mathbb{R}^d$. The associated approximate likelihood given in Proposition 1 leads to the expression, using (13) and the expression for $T_i(\boldsymbol{\mu}, \boldsymbol{\Omega})$ given in (16)

$$\ell_{N,n}(\theta) = \frac{Nn}{2} \log \gamma - \frac{1}{2} \sum_{i=1}^{N} \log det(I_d + V_i \boldsymbol{\Omega}) - \frac{\gamma}{2} \sum_{i=1}^{N} (S_i + T_i(\boldsymbol{\mu}, \boldsymbol{\Omega})), \tag{29}$$

Let us introduce the quantities

$$\begin{aligned} B_{i,n} &= B_i = R_i^{-1} = (V_i^{-1} + \boldsymbol{\Omega})^{-1} && (30)\\ A_{i,n} &= A_i = (I_d + V_i \boldsymbol{\Omega})^{-1}(U_i - V_i \boldsymbol{\mu}) = B_i(V_i^{-1} U_i - \boldsymbol{\mu}) && (31) \end{aligned}$$

They converge as $n \to \infty$ to the random variables

$$A_i(T; \mu, \Omega) = B_i(T; \boldsymbol{\Omega})(V_i(T)^{-1} U_i(T) - \boldsymbol{\mu}) \text{ and } B_i(T; \boldsymbol{\Omega}) = (V_i(T)^{-1} + \boldsymbol{\Omega})^{-1} \tag{32}$$

Denote by $\nabla_{\boldsymbol{\mu}} f(.)$ and $\nabla_{\boldsymbol{\Omega}} f(.)$ the vector $(\frac{\partial}{\partial \mu_k} f(.))_{1 \le k \le d}$ and the matrix $(\frac{\partial}{\partial \omega_{k,l}} f(.))_{1 \le k,l \le d}$. The (pseudo-)score function associated with $\log \Lambda_{N,n}(\theta)$ is

$$G_{N,n}(\theta) = \left(\frac{\partial}{\partial \gamma} \ell_{N,n}(\theta) \;\; \nabla_{\boldsymbol{\mu}} \ell_{N,n}(\theta) \;\; \nabla_{\boldsymbol{\Omega}} \ell_{N,n}(\theta) \right)'. \tag{33}$$

In order to avoid multiple indexes, we denote by M^{lk} the (l,k) term of a matrix and $(Y^1, \ldots, Y^d)$ the coordinates of a vector $Y \in \mathbb{R}^d$. The vector (22) is now , using that $\partial \log det R_i = Tr(R_i^{-1} \partial R_i)$,

$$\begin{aligned} \frac{\partial}{\partial \gamma} \ell_{N,n}(\theta) &= \frac{N\,n}{2\gamma} - \frac{1}{2} \sum_{i=1}^{N} (S_i + T_i(\boldsymbol{\mu}, \boldsymbol{\Omega})) \;, \\ \nabla_{\boldsymbol{\mu}} \ell_{N,n}(\theta) &= -\frac{\gamma}{2} \sum_{i=1}^{N} \nabla_{\boldsymbol{\mu}} T_i(\boldsymbol{\mu}, \boldsymbol{\Omega}) = \gamma \sum_{i=1}^{N} A_i, \\ \nabla_{\boldsymbol{\Omega}} \ell_{N,n}(\theta) &= -\frac{1}{2} \sum_{i=1}^{N} B_i + \frac{\gamma}{2} \sum_{i=1}^{N} A_i A_i'. \end{aligned}$$

Similarly, we define the rate matrix $D_{N,n}$ and the Fisher information matrix $\mathcal{I}(\theta)$ by

$$D_{N,n} = \begin{pmatrix} \frac{1}{\sqrt{Nn}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{N}} I_d & 0 \\ 0 & 0 & \frac{1}{\sqrt{N}} I_{d \times d} \end{pmatrix}, \qquad \mathcal{I}(\theta) = \left(\begin{array}{c|c} \frac{1}{2\gamma^2} & \mathbf{0} \\ \hline \mathbf{0} & I(\theta) \end{array} \right) \quad \text{where}$$

where $I(\theta)$ is the covariance matrix of the vector

$$\begin{pmatrix} \gamma A_1(T; \boldsymbol{\mu}, \boldsymbol{\Omega}) \\ \frac{1}{2}(\gamma A_1(T; \boldsymbol{\mu}, \boldsymbol{\Omega})(A_1(T; \boldsymbol{\mu}, \boldsymbol{\Omega}))' - B_1(T; \boldsymbol{\Omega})) \end{pmatrix}. \tag{34}$$

The asymptotic study of the estimators of θ can be done similarly.

4.3 Fixed effect in the drift and random effect in the diffusion coefficient

The Φ_i's are deterministic equal to an unknown value φ and $\Psi_i = \Gamma_i^{-1/2}$ with $\Gamma_i \sim G(a, \lambda)$ and $\tau = (\lambda, a, \varphi)$. For sake of simplicity, we just consider $d = 1$ ($\varphi \in \mathbb{R}$). Let us set $\widetilde{\ell}_{N,n}(\tau) = \log \widetilde{\Lambda}_{N,n}(\tau)$ and

$$\widetilde{G}_{N,n}(\tau) = \left(\frac{\partial}{\partial \lambda} \widetilde{\ell}_{N,n}(\tau) \;\; \frac{\partial}{\partial a} \widetilde{\ell}_{N,n}(\tau) \;\; \frac{\partial}{\partial \varphi} \widetilde{\ell}_{N,n}(\tau) \right)'. \tag{35}$$

We consider the estimators $\widetilde{\tau}_{N,n}$ defined by the estimating equation

$$\widetilde{G}_{N,n}(\widetilde{\tau}_{N,n}) = 0. \tag{36}$$

Recall that $\psi(z) = \Gamma'(z)/\Gamma(z)$ and set

$$\zeta_i(\tau) = \zeta_i = \frac{\lambda + \frac{1}{2}\left(S_i - 2\varphi U_i + \varphi^2 V_i\right)}{a + (n/2)}. \tag{37}$$

$$\begin{aligned}
\frac{\partial}{\partial \lambda} \widetilde{\ell}_{N,n}(\tau) &= \sum_{i=1}^{N} \left(\frac{a}{\lambda} - \zeta_i^{-1} \right), \\
\frac{\partial}{\partial a} \widetilde{\ell}_{N,n}(\tau) &= \sum_{i=1}^{N} \left(\log \lambda - \psi(a) - \log \zeta_i \right) + N \left(\psi(a + (n/2)) - \log\left(a + (n/2)\right) \right), \\
\frac{\partial}{\partial \varphi} \widetilde{\ell}_{N,n}(\tau) &= \sum_{i=1}^{N} \zeta_i^{-1} (U_i - \varphi V_i).
\end{aligned}$$

The following holds.

Theorem 2. *Assume (H1)-(H2), $a > 5$ and that N, n tend to infinity with $N/n \to 0$. Then, a solution $\widetilde{\tau}_{N,n}$ to (36) exists with probability tending to 1 which is consistent and such that $\sqrt{N}(\widetilde{\tau}_{N,n} - \tau)$ converges in distribution under $\mathbb{P}_\tau$ to $\mathcal{N}_3(0, \mathcal{V}^{-1}(\tau))$ where $I_0(\lambda, a)$ is defined in (20), $V(t)$ in (2) and*

$$\mathcal{V}(\tau) = \left(\begin{array}{c|c} I_0(\lambda, a) & \mathbf{0} \\ \hline \mathbf{0} & \mathbb{E}_\tau\left(\Gamma V(T)\right) \end{array} \right).$$

For the first two components of $\widetilde{\tau}_{N,n}$, the constraint $N/n^2 \to 0$ is enough.

Let us stress that the estimator of (λ, a) based on the indirect observations $(X_i, i = 1, \ldots, N)$ is asymptotically equivalent to the exact maximum likelihood estimator of (λ, a) based on the direct observation of (Γ_i) under the constraint $N/n^2 \to 0$ (see Section 4.1). This result was obtained for $\varphi = 0$ (null drift) in Delattre *et al.* (2014) under the constraint $N/n \to 0$. Proposition 2 thus extends this study, improves the constraint on N, n and contains the additional result concerning the estimation of φ which is new. Note that, contrary to the previous section, all the components of τ are estimated with the same rate $\sqrt{N}$.

We can compare this result to the estimation of fixed φ, γ when observing the N-sample paths. The model is $dX_i(t) = \varphi b(X_i(t))dt + \gamma^{-1/2}\sigma(X_i(t))dW_i(t), X_i(0) = x, i = 1, \ldots, N$. Using (8), we find the estimators

$$\hat{\varphi}_{N,n} = \frac{\sum_{i=1}^N U_i}{\sum_{i=1}^N V_i}, \quad \hat{\gamma}_{N,n} = \frac{nN}{\sum_{i=1}^N S_i + \hat{\varphi}_{N,n}^2 V_i - 2\hat{\varphi}_{N,n} U_i}.$$

The limiting distribution of $\sqrt{Nn}(\hat{\gamma}_{N,n} - \gamma)$ is $\mathcal{N}(0, 2\gamma^2)$. Provided that $N/n \to 0$, the estimator $\hat{\varphi}_{N,n}$ converges in distribution to $\mathcal{N}(0, (\gamma \mathbb{E} V(T))^{-1})$. The result obtained for $\tilde{\varphi}_{N,n}$ in Theorem 2 is thus not surprising. Note that the estimator $\hat{\varphi}_{N,n}$ can be considered with random effects Γ_i and satisfies $\sqrt{N}(\hat{\varphi}_{N,n} - \varphi)$ converges in distribution to $\mathcal{N}(0, \mathbb{E}(\Gamma^{-1}V(T))/\mathbb{E}^2(V(T)))$.

5 Simulation study

We investigate both the cases of random effect in the drift and of random effect in the diffusion coefficient. Several models are simulated in each case. For each SDEME model, 100 data sets are generated. Different values for N, different numbers of observations per trajectory n and several sets of parameters are used. Each data set is simulated as follows. First, the random effect is drawn, then, the diffusion sample path is simulated using a Euler scheme with a very small discretization step-size $\delta = 0.001$. The time interval between consecutive observations is taken equal to $\Delta = 0.01$ with a resulting time interval $n\Delta$.The empirical mean and standard deviation of the estimates are computed from the 100 datasets and are compared with those of the estimates based on a direct observation of the random effects. It follows from this simulation design that N random effects variables are simulated for each value of n, resulting in different estimates.

5.1 Random effect in the drift and fixed effect in the diffusion coefficient

We consider models with univariate or bivariate random effect in the drift. Examples 1 and 2 concern a univariate random effect. Example 3 concerns a bivariate random effect with one deterministic component whereas Example 4 has a bivariate random effect with two random components. The value of γ is either 4 or 10 which respectively corresponds to $\gamma^{-1/2} = 0.5$, or $\gamma^{-1/2} \simeq 0.32$.

Example 1. $dX_i(t) = \Phi_i X_i(t)dt + \frac{1}{\sqrt{\gamma}} dW_i(t), X_i(0) = 0$, $\Phi_i \underset{i.i.d}{\sim} \mathcal{N}(\mu, \frac{\omega^2}{\gamma})$

Example 2. $dX_i(t) = \Phi_i X_i(t)^2/(1 + X_i(t)^2)dt + \frac{1}{\sqrt{\gamma}} dW_i(t), X_i(0) = 0$, $\Phi_i \underset{i.i.d}{\sim} \mathcal{N}(\mu, \frac{\omega^2}{\gamma})$

Example 3. $dX_i(t) = (\rho X_i(t) + \Phi_i)dt + \frac{1}{\sqrt{\gamma}} dW_i(t), X_i(0) = 0$, $\Phi_i \underset{i.i.d}{\sim} \mathcal{N}(\mu, \frac{\omega^2}{\gamma})$

Example 4. $dX_i(t) = (\Phi_{i1} X_i(t) + \Phi_{i2})dt + \frac{1}{\sqrt{\gamma}} dW_i(t), X_i(0) = 0$, $\Phi_{i1} \underset{i.i.d}{\sim} \mathcal{N}(\mu_1, \frac{\omega_1^2}{\gamma})$, $\Phi_{i2} \underset{i.i.d}{\sim} \mathcal{N}(\mu_2, \frac{\omega_2^2}{\gamma})$ with Φ_{i1}, Φ_{i2} independent random variables.

Note that, in Examples 1 and 3, which are classical, assumption (H1) is not satisfied. Nevertheless, the numerical results are quite good giving evidence that the method can be successfully applied even in these cases.

The results are given in Tables 1 to 4. For the three considered experimental designs $(N = 50, n = 500)$, $(N = 100, n = 500)$ and $(N = 100, n = 1000)$, the model parameters are estimated with very little bias, especially $\tilde{\gamma}$. This is consistent with the theoretical result that $\tilde{\gamma}$ has rate $\sqrt{Nn}$ (Theorem 1). The estimates of $(\tilde{\mu}, \tilde{\omega}^2)$ are almost as good as the estimations based on a direct observation of the random

		$N=50$		$N=100$	
		$n=500$	$n=1000$	$n=500$	$n=1000$
		$(\mu_0=0,\omega_0^2=0.1,\gamma_0=4)$			
X	$\tilde{\mu}$	0.00 (0.07)	0.00 (0.04)	0.00 (0.04)	0.00 (0.03)
	$\tilde{\omega}^2$	0.09 (0.08)	0.10 (0.04)	0.09 (0.05)	0.10 (0.03)
	$\tilde{\gamma}$	4.00 (0.03)	4.00 (0.03)	4.00 (0.02)	4.00 (0.02)
Φ	$\hat{\mu}$	0.00 (0.02)	0.00 (0.02)	0.00 (0.01)	0.00 (0.02)
	$\hat{\omega}^2$	0.10 (0.02)	0.10 (0.02)	0.10 (0.01)	0.10 (0.01)

Table 1: (Example 1) Empirical mean and standard deviation (in brackets) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Φ: estimates based on direct observation of the Φ_i's.

		$N=50$		$N=100$	
		$n=500$	$n=1000$	$n=500$	$n=1000$
		$(\mu_0=0,\omega_0^2=0.1,\gamma_0=4)$			
X	$\tilde{\mu}$	0.00 (0.10)	0.00 (0.06)	0.00 (0.06)	0.01 (0.03)
	$\tilde{\omega}^2$	0.10 (0.15)	0.07 (0.06)	0.10 (0.12)	0.09 (0.04)
	$\tilde{\gamma}$	4.00 (0.03)	4.01 (0.03)	4.00 (0.03)	4.00 (0.02)
Φ	$\hat{\mu}$	0.00 (0.02)	0.00 (0.02)	0.00 (0.01)	0.00 (0.02)
	$\hat{\omega}^2$	0.10 (0.02)	0.10 (0.02)	0.10 (0.01)	0.10 (0.01)

Table 2: (Example 2) Empirical mean and standard deviation (in brackets) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Φ: estimates based on direct observation of the Φ_i's.

effects. Increasing N and n reduces the bias and the standard deviation of the estimates. Note that the results are also very satisfactory in Example 3, where there are both random and fixed effects in the drift, showing the validity of our method even in situations where the covariance matrix of the random effects is not invertible.

5.2 Fixed effect in the drift and random effect in the diffusion coefficient

The following models are studied:
Example 5. $dX_i(t)=\rho dt+\Gamma_i^{-1/2}dW_i(t), X_i(0)=0\ ,\ \Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$
Example 6. $dX_i(t)=\rho X_i(t)dt+\Gamma_i^{-1/2}dW_i(t), X_i(0)=x\ ,\ \Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$
Example 7. $dX_i(t)=\rho X_i(t)dt+\Gamma_i^{-1/2}\sqrt{1+X_i(t)^2}dW_i(t), X_i(0)=x\ ,\ \Gamma_i \underset{i.i.d}{\sim} G(a,\lambda)$
The results are given in Tables 5 to 7. Estimations based on the processes and estimations based on direct observations of the random effects are compared. The maximisation of the likelihood of Gamma distributed random variables may be numerically difficult, which may explain that the results are a little bit biased with relatively large standard deviation. Moreover, the estimates vary according to the successive N-samples corresponding to each n. The quality of the estimations based on the processes

		$N=50$		$N=100$	
		$n=500$	$n=1000$	$n=500$	$n=1000$
		$(\mu_0=1, \omega_0^2=1, \gamma_0=10, \rho_0=-0.1)$			
X	$\tilde{\mu}$	1.00 (0.05)	1.01 (0.05)	1.00 (0.03)	1.00 (0.04)
	$\tilde{\omega}^2$	0.99 (0.26)	1.01 (0.26)	1.00 (0.21)	0.97 (0.15)
	$\tilde{\gamma}$	10 (0.10)	10.03 (0.06)	10.00 (0.05)	10.01 (0.05)
	$\tilde{\rho}$	-0.10 (0.02)	-0.10 (0.01)	-0.10 (0.01)	-0.10 (0.01)
ϕ	$\hat{\mu}$	1.00 (0.04)	1.00 (0.05)	1.00 (0.03)	1.00 (0.03)
	$\hat{\omega}^2$	1.00 (0.21)	1.00 (0.22)	1.00 (0.15)	0.99 (0.14)
		$(\mu_0=1, \omega_0^2=0.4, \gamma_0=4, \rho_0=-0.1)$			
X	$\tilde{\mu}$	1.01 (0.08)	1.02 (0.06)	1.01 (0.06)	1.00 (0.05)
	$\tilde{\omega}^2$	0.40 (0.14)	0.40 (0.12)	0.41 (0.10)	0.40 (0.09)
	$\tilde{\gamma}$	4.01 (0.04)	4.01 (0.02)	4.01 (0.03)	4.01 (0.02)
	$\tilde{\rho}$	-0.11 (0.02)	-0.10 (0.01)	-0.10 (0.02)	-0.10 (0.01)
Φ	$\hat{\mu}$	1.00 (0.05)	1.00 (0.04)	1.00 (0.03)	1.00 (0.03)
	$\hat{\omega}^2$	0.39 (0.06)	0.41 (0.09)	0.41 (0.07)	0.40 (0.06)

Table 3: (Example 3) Empirical mean and standard deviation (in brackets) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Φ: estimates based on direct observation of the Φ_i's.

is not always satisfactory, even for $N = 100$. Indeed, the value of N seems to have little impact while increasing n highly improves the estimation of (λ, a). On the whole, parameters a and λ are highly underestimated when n is small ($n = 500$ and $n = 1000$), and a high number of observations per trajectory (here $n = 10000$) is required to get unbiased estimations of parameters a and λ. When n is large, the variances of the parameter estimates is close to the variances of the estimates based on a direct observation of the random effects. The theoretical asymptotic variances of the estimators are explicit in Example 5, and we see from table 5 that the variances of the parameter estimates are also close to the theoretical variances when $n = 10000$. Nevertheless, let us point out that the estimates of $\mathbb{E}\Psi_i^2 = \mathbb{E}\Gamma_i^{-1} = \lambda/(a-1)$ and (S.D.)$\Psi_i^2 = \lambda/((a-1)\sqrt{a-2})$ are always satisfactory even for small n. Parameter ρ is correctly estimated in all (N, n) configurations. The standard deviation of the estimations decreases when n becomes larger, which is in accordance with the theory (see table 5 where the true theoretical values are known).

6 Concluding remarks

In this paper, we consider N SDEMEs ruled by (1) which are discretely observed on a fixed time interval. We study the parametric inference for the mixed effects when N and the number n of observations per sample path grow to infinity. We investigate the two cases of random effect in the drift and fixed effect in the diffusion coefficient or fixed effect in the drift and random effect in the diffusion coefficient. For specific distributions of the random effects, we obtain explicit approximation of the likelihood functions

		$N=50$		$N=100$	
		$n=500$	$n=1000$	$n=500$	$n=1000$
		$(\mu_{1,0}=-0.1, \mu_{2,0}=1, \omega^2_{1,0}=0.1, \omega^2_{1,0}=1, \gamma_0=10, \rho_0=-0.1)$			
X	$\tilde{\mu}_1$	-0.10 (0.03)	-0.10 (0.02)	-0.10 (0.02)	-0.10 (0.01)
	$\tilde{\omega}^2_1$	0.10 (0.04)	0.10 (0.02)	0.10 (0.03)	0.10 (0.02)
	$\tilde{\mu}_2$	1.00 (0.07)	1.00 (0.05)	1.00 (0.04)	1.00 (0.04)
	$\tilde{\omega}^2_2$	0.96 (0.28)	0.97 (0.28)	0.94 (0.19)	0.99 (0.20)
	$\tilde{\gamma}$	10.03 (0.09)	10.02 (0.07)	10.03 (0.07)	10.02 (0.05)
Φ	$\hat{\mu}_1$	-0.10 (0.02)	-0.10 (0.01)	-0.10 (0.01)	-0.10 (0.01)
	$\hat{\omega}^2_1$	0.10 (0.02)	0.10 (0.02)	0.10 (0.01)	0.10 (0.02)
	$\hat{\mu}_2$	1.00 (0.05)	1.00 (0.04)	1.00 (0.03)	1.00 (0.03)
	$\hat{\omega}^2_2$	1.00 (0.21)	1.01 (0.23)	0.98 (0.13)	1.00 (0.14)
		$(\mu_{1,0}=-0.1, \mu_{2,0}=1, \omega^2_{1,0}=0.04, \omega^2_{2,0}=0.4, \gamma_0=4)$			
X	$\tilde{\mu}_1$	-0.10 (0.04)	-0.11 (0.02)	-0.10 (0.02)	-0.10 (0.01)
	$\tilde{\omega}^2_1$	0.04 (0.02)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)
	$\tilde{\mu}_2$	0.98 (0.08)	1.01 (0.06)	1.00 (0.05)	1.00 (0.05)
	$\tilde{\omega}^2_2$	0.37 (0.15)	0.37 (0.13)	0.39 (0.11)	0.41 (0.10)
	$\tilde{\gamma}$	4.02 (0.03)	4.00 (0.03)	4.01 (0.02)	4.00 (0.08)
Φ	$\hat{\mu}_1$	-0.10 (0.02)	-0.10 (0.01)	-0.10 (0.01)	-0.10 (0.01)
	$\hat{\omega}^2_1$	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)
	$\hat{\mu}_2$	1.00 (0.04)	1.00 (0.04)	1.00 (0.03)	1.00 (0.03)
	$\hat{\omega}^2_2$	0.40 (0.07)	0.41 (0.07)	0.40 (0.07)	0.42 (0.06)

Table 4: (Example 4) Empirical mean and standard deviation (in brackets) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Φ: estimates based on direct observation of the Φ_i's.

		$N=50$			$N=100$		
		$n=500$	$n=1000$	$n=10000$	$n=500$	$n=1000$	$n=10000$
				$(a_0=8, \lambda_0=2, \rho_0=-1)$			
X	$\tilde{a}$	6.03	7.03	8.32	6.05	6.88	8.14
		(0.73 - 1.57)	(0.91 - 1.57)	(1.80 - 1.57)	(0.44 - 1.11)	(0.73 - 1.11)	(1.11 - 1.11)
	$\tilde{\lambda}$	1.50	1.76	2.08	1.51	1.71	2.04
		(0.20 - 0.40)	(0.24 - 0.40)	(0.44 - 0.40)	(0.12 - 0.29)	(0.19 - 0.29)	(0.30 - 0.29)
	$\tilde{\rho}$	-0.99	-1.00	-1.00	-1.00	-1.00	-1.00
		(0.03 - 0.03)	(0.02 - 0.02)	(0.01 - 0.01)	(0.02 - 0.02)	(0.02 - 0.02)	(0.01 - 0.01)
Ψ	$\hat{a}$	8.23	8.57	8.42	8.38	8.32	8.27
		(1.77 - 1.57)	(1.76 - 1.57)	(2.15 - 1.57)	(1.16 - 1.11)	(1.35 - 1.11)	(1.31 - 1.11)
	$\hat{\lambda}$	2.06	2.14	2.10	2.09	2.07	2.07
		(0.46 - 0.40)	(0.45 - 0.40)	(0.53 - 0.40)	(0.30 - 0.29)	(0.34 - 0.29)	(0.34 - 0.29)

Table 5: (Example 5) Empirical mean and, in brackets, (empirical standard deviation - theoretical standard deviation) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Ψ: Estimates based on the Ψ_i's.

		$N=50$			$N=100$		
		$n=500$	$n=1000$	$n=10000$	$n=500$	$n=1000$	$n=10000$
				$(a_0=8, \lambda_0=2, \rho_0=-1)$			
X	$\tilde{a}$	6.18 (0.66)	6.79 (0.99)	8.20 (1.61)	6.02 (0.50)	6.70 (0.63)	7.91 (1.04)
	$\tilde{\lambda}$	1.52 (0.17)	1.69 (0.27)	2.03 (0.42)	1.49 (0.14)	1.66 (0.16)	1.96 (0.25)
	$\tilde{\rho}$	-1.00 (0.10)	-1.01 (0.06)	-0.99 (0.02)	-1.00 (0.06)	-1.00 (0.04)	-1.00 (0.01)
Ψ	$\hat{a}$	8.71 (1.80)	8.12 (1.94)	8.20 (1.61)	8.16 (1.30)	8.10 (1.19)	8.10 (1.16)
	$\hat{\lambda}$	2.16 (0.45)	2.04 (0.51)	2.07 (0.47)	2.04 (0.33)	2.02 (0.30)	2.03 (0.29)

Table 6: (Example 6) Empirical mean and standard deviation (in brackets) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Ψ: Estimates based on the Ψ_i's.

and prove that the corresponding estimators are asymptotically Gaussian. The estimation method yields very good results on various simulated data sets. When the random effect is in the drift, N influences the quality of estimates. When the random effect is in the diffusion coefficient, we observe that the estimates are more depedent on the number n of observations per trajectory.

The case where functions b, σ depend on time $(b(t,x), \sigma(t,x))$ can be studied with little change. Multidimensional SDEMEs with linear random effects are often encountered in applications (see *e.g.* Leander *et al.* (2015), Berglund *et al.* (2001)) and could be studied by the same approach .

In a forthcoming paper, we complete this study and investigate the case where both Φ_i and Ψ_i are random.

An interesting extension is to include another source of variability by considering measurement errors in the observations. This requires an additional step of filtering which is not immediate.

		$N = 50$			$N = 100$		
		$n = 500$	$n = 1000$	$n = 10000$	$n = 500$	$n = 1000$	$n = 10000$
				$(a_0 = 8, \lambda_0 = 2, \rho_0 = -1)$			
X	$\tilde{a}$	6.06 (0.73)	6.90 (0.92)	8.03 (1.61)	6.00 (0.47)	6.81 (0.65)	8.08 (1.16)
	$\tilde{\lambda}$	1.50 (0.20)	1.70 (0.25)	2.01 (0.41)	1.49 (0.12)	1.68 (0.17)	2.03 (0.30)
	$\tilde{\rho}$	-1.01 (0.11)	-1.00 (0.07)	-1.00 (0.02)	-1.00 (0.08)	-0.99 (0.05)	-1.00 (0.02)
ψ	$\hat{a}$	8.44 (1.77)	8.41 (1.81)	8.17 (1.79)	8.16 (1.24)	8.25 (1.18)	8.08 (1.16)
	$\hat{\lambda}$	2.10 (0.45)	2.10 (0.47)	2.06 (0.45)	2.05 (0.32)	2.05 (2.31)	2.07 (0.37)

Table 7: (Example 7) Empirical mean and standard deviation (in brackets) of the parameter estimates from 100 datasets for different values of N and n. X: Estimates based on the $(X_{i,n})$'s; Ψ: Estimates based on the Ψ_i's.

References

Berglund, M., Sunnaker, M., Adiels, M., Jirstrand, M. and Wennberg, B. (2001). Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Mathematical Medecine and Biology* p. doi:10.1093/imammb/dqr021.

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

Delattre, M., Genon-Catalot, V. and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics* **40(2)**, 322–343.

Delattre, M., Genon-Catalot, V. and Samson, A. (2015). Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. *ESAIM Probability and Statistics* .

Delattre, M., Poursat, M. and Lavielle, M. (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics* **8**, 456–475.

Dion, C. and Genon-Catalot, V. (2015). Bidimensional random effect estimation in mixed stochastic differential model. *Statistical Inference for Stochastic processes* **18**.

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

Ditlevsen, S. and De Gaetano, A. (2005b). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.* **67**, 547–561.

Donnet, S., Foulley, J. and Samson, A. (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics* **66**, 733ï£¡741.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**, 119–151.

Genon-Catalot, V. and Larédo, C. (2015). Estimation for stochastic differential equations with mixed effects. *Hal-00807258* **V2**.

Jacod, J. (1997). On continuous conditional gaussian martingales and stable convergence in law. *Lecture Notes in Mathematics, 1655* pp. 232–246. Séminaire de Probabilités XXXI.

Kessler, M., Lindner, A. and Sørensen, M. E. (2012). *Statistical methods for stochastic differential equations*. Chapman & Hall. Monographs on Statistics and Applied Probability 124.

Leander, J., Almquist, J., Ahlstrom, 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. *AAPS Journal, DOI: 10.1208/s12248-015-9718-8* .

Møller, J., Overgaard, R., Madsen, H., Hansen, T., Pedersen, O. and Ingwersen, S. (2010). Predictive performance for population models using stochastic differential equations applied on data from an oral glucose tolerance test. *J. Pharmacokinet. Pharmacodyn.* **37**, 85–98.

Nie, L. (2006). Strong consistency of the maximum likelihood estimator in generalized linear 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., Jonsson, N., Tornøe, C. and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *J Pharmacokinet. Pharmacodyn.* **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). Practicle estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55**, 1426–1444.

Picchini, U., Ditlevsen, S. and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Math. Med. Biol.* **25**, 141–155.

The corresponding author is:
Catherine Larédo (catherine.laredo@jouy.inra.fr)
Laboratoire MaIAGE, I.N.R.A., Jouy-en-Josas, France.

7 Appendix

7.1 Proofs of Section 2

In what follows, we use the notation $\varphi.b = \sum_{\ell=1}^{d} \varphi_\ell b_\ell$ to avoid confusion with the derivatives.

Proof of Lemma 1. Using the Hölder inequality, we have:

$$|X(t) - X(s)|^p \leq 2^{p-1} \left(\|\Phi\|^p (\int_s^t \|b(X(u))\| d)^p + \Psi^p | \int_s^t \sigma(X(u)) dW(u)|^p \right).$$

Using (H2) and the Burkholder-Davis-Gundy (B-D-G) inequality, we obtain the result.□

Proof of Lemma 2.

We only prove (10)-(11). Denote L_f the Lipschitz constant of f. Applying the Hölder innequality twice, we get

$$(V(f;T) - V_n(f)))^{2p} \leq L_f^{2p} T^{2p-1} \sum_{j=1}^{n} \int_{(j-1)\Delta}^{j\Delta} (X(s) - X((j-1)\Delta))^{2p} ds.$$

Taking the conditional expectation w.r.t. Φ, Ψ, we obtain the first inequality using Lemma 1.

We split $U(g;T) - U_n(g)) = T_1 + T_2$ with

$$T_1 = \sum_{j=1}^{n} \int_{(j-1)\Delta}^{j\Delta} (g(X(s)) - g(X((j-1)\Delta)))\Phi.b(X(s) ds$$

$$T_2 = \Psi \sum_{j=1}^{n} \int_{(j-1)\Delta}^{j\Delta} (g(X(s)) - g(X((j-1)\Delta)))\sigma(X(s) dW(s).$$

The term T_1 can be treated as above and we find

$$\mathbb{E}_\vartheta(T_1^{2p} | \Phi = \varphi, \Psi = \psi) \lesssim \|\varphi\|^{2p} (\|\varphi\|^{2p} + \psi^{2p}) \Delta^p.$$

The Ito formula yields, setting $L_{\Phi,\Psi} g = \Phi.bg' + \frac{1}{2}\Psi^2\sigma^2 g''$

$$\begin{aligned} g(X(s)) - g(X((j-1)\Delta)) &= \int_{(j-1)\Delta}^{s} L_{\Phi,\Psi} g(X(u)) du \\ &+ \int_{(j-1)\Delta}^{s} g'(X(u)\Psi\sigma(X(u)) dW(u). \end{aligned}$$

Therefore, T_2 is split into $T_2 = T_{2,1} + T_{2,2}$ with

$$\begin{aligned} T_{2,1} &= \Psi \sum_{j=1}^{n} \int_{(j-1)\Delta}^{j\Delta} \int_{(j-1)\Delta}^{s} L_{\Phi,\Psi} g(X(u)) du \sigma(X(s)) dW(s) \\ T_{2,2} &= \Psi^2 \sum_{j=1}^{n} \int_{(j-1)\Delta}^{j\Delta} \int_{(j-1)\Delta}^{s} g'(X(u))\sigma(X(u)) dW(u) \sigma(X(s) dW(s). \end{aligned}$$

We write $T_{2,1} = \Psi \int_0^T H_s^n dW(s)$ with

$$H_s^n = \sum_{j=1}^n 1_{((j-1)\Delta, j\Delta]}(s)\sigma(X(s)) \int_{(j-1)\Delta}^s L_{\Phi,\Psi} g(X(u))du.$$

By the assumptions, $|H_s^n| \lesssim \Delta(\|\Phi\| + \Psi^2)$. By the B-D-G inequality, $\mathbb{E}_\vartheta(T_{2,1}^{2p}|\Phi,\Psi) \lesssim \Delta^p \Psi^{2p}(\|\Phi\|+\Psi^2)^{2p}$. Analogously, $T_{2,2} = \Psi^2 \int_0^T K_s^n dW(s)$ where

$$K_s^n = \sum_{j=1}^n 1_{((j-1)\Delta, j\Delta]}(s)\sigma(X(s)) \int_{(j-1)\Delta}^s g'(X(u))\sigma(X(u))dW(u).$$

Applying the Hölder inequality, we obtain

$$\left(\int_0^T (K_s^n)^2 ds\right)^p \lesssim \sum_{j=1}^n \int_{(j-1)\Delta}^{j\Delta} ds \left(\int_{(j-1)\Delta}^s g'(X(u))\sigma(X(u))dW(u)\right)^{2p}.$$

The B-D-G inequality implies $\mathbb{E}_\vartheta(T_{2,2}^{2p}|\Phi,\Psi) \lesssim \Delta^p \Psi^{4p}$. Joining all terms, we get (11). $\square$

Proof of Lemma 3.

For the proof, we omit the index i and first consider the case of exponent $2p$ with $p \geq 1$. We have

$$\begin{aligned}\frac{S_n^{(1)}}{n} &= \Psi^2 + \nu_n^{(1)}(W) = \Psi^2 + \Psi^2 \sum_{j=0}^{n-1} \frac{1}{n\Delta}\left[\left(W_{(j+1)\Delta} - W_{j\Delta}\right)^2 - \Delta\right] \\ &= \Psi^2 + \Psi^2 \sum_{j=0}^{n-1} \frac{2}{n\Delta} \int_{j\Delta}^{(j+1)\Delta} (W(s) - W(j\Delta))dW(s).\end{aligned}$$

Next, we decompose S_n/n using the following classical development (see Comte *et al.* 2007, p.522):

$$\begin{aligned}\frac{(X((j+1)\Delta)) - X(j\Delta))^2}{\Delta} &= \Psi^2\sigma^2(X(j\Delta)) + \Psi^2 V_{j\Delta}^{(1)}(X) + 2\Psi^3 V_{j\Delta}^{(2)}(X) \\ &+ 2\Psi V_{j\Delta}^{(3)}(X) + R_{j\Delta}(\Phi,\Psi,X),\end{aligned}$$

where

$$V_{j\Delta}^{(1)}(X) = \frac{1}{\Delta}\left[\left(\int_{j\Delta}^{(j+1)\Delta} \sigma(X(s))dW(s)\right)^2 - \int_{j\Delta}^{(j+1)\Delta} \sigma^2(X(s))ds\right] \tag{38}$$

$$V_{j\Delta}^{(2)}(X) = \frac{1}{\Delta}\int_{j\Delta}^{(j+1)\Delta} ((j+1)\Delta - u)\sigma'(X(u))\sigma^2(X(u)dW(u) \tag{39}$$

$$V_{j\Delta}^{(3)}(X) = \Phi.b(X(j\Delta)) \int_{j\Delta}^{(j+1)\Delta} \sigma(X(s))dW(s) \tag{40}$$

$$R_{j\Delta}(\Phi,\Psi,X) = \frac{1}{\Delta}\left(\int_{j\Delta}^{(j+1)\Delta}\Phi.b(X(s))ds\right)^2 + \frac{2\Psi}{\Delta}\int_{j\Delta}^{(j+1)\Delta}\Phi.(b(X(s)-b(X(j\Delta)))ds\int_{j\Delta}^{(j+1)\Delta}\sigma(X(s))dW(s) + \frac{1}{\Delta}\int_{j\Delta}^{(j+1)\Delta}((j+1)\Delta-u)K_{\Phi,\Psi}(X(u))du \tag{41}$$

where $K_{\Phi,\Psi}(.)) = \Psi^2[2\Phi.b\sigma\sigma' + \Psi^2\sigma^2(\sigma^2)'']$. So, $\frac{S_n}{n} = \Psi^2 + \nu_n^{(1)} + \nu_n^{(2)} + \nu_n^{(3)} + \nu_n^{(4)}$, with

$$\nu_n^{(1)} = \Psi^2\frac{1}{n}\sum_{j=0}^{n-1}\frac{V_{j\Delta}^{(1)}(X)}{\sigma^2(X(j\Delta))}, \qquad \nu_n^{(2)} = \frac{\Psi^3}{n}\int_0^{n\Delta}H_n(s)dW(s),$$

$$\nu_n^{(3)} = \frac{\Psi}{n}\int_0^{n\Delta}K_n(s)dW(s), \qquad \nu_n^{(4)} = \frac{1}{n}\sum_{j=0}^{n-1}\frac{R_{j\Delta}(\Phi,\Psi,X)}{\sigma^2(X(j\Delta))} \quad \text{with}$$

$$H_n(s) = \sum_{j=0}^{n-1}\frac{((j+1)\Delta-s)\sigma'(X(s))\sigma^2(X(s))}{\Delta\sigma^2(X(j\Delta))}1_{(j\Delta,(j+1)\Delta]}(s), \quad |H_n(s)| \le \frac{LK^2}{\sigma_0^2},$$

$$K_n(s) = \sum_{j=0}^{n-1}1_{(j\Delta,(j+1)\Delta]}(s)\frac{2\Phi.b(X(j\Delta))\sigma(X(s))}{\sigma^2(X(j\Delta))}, \quad |K_n(s)| \le 2\|\Phi\|\frac{K^2}{\sigma_0^2}.$$

Thus,

$$\frac{S_n}{n} - \frac{S_n^{(1)}}{n} = \nu_n^{(1)} - \nu_n^{(1)}(W) + \nu_n^{(2)} + \nu_n^{(3)} + \nu_n^{(4)}.$$

For $M_t = \int_0^t \sigma(X(s))dW(s)$, we have

$$\left(M_{(j+1)\Delta} - M_{j\Delta}\right)^2 - \int_{j\Delta}^{(j+1)\Delta}\sigma^2(X(s))ds = 2\int_{j\Delta}^{(j+1)\Delta}(M_s - M_{j\Delta})dM_s.$$

Therefore, (see (38)),

$$\begin{aligned}\frac{V_{j\Delta}^{(1)}(X)}{\sigma^2(X(j\Delta)} &= \frac{2}{\Delta\sigma^2(X(j\Delta)}\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}(\sigma(X(u)) - \sigma(X(j\Delta)) + \sigma(X(j\Delta)))dW(u) \\ &\quad (\sigma(X(s)) - \sigma(X(j\Delta)) + \sigma(X(j\Delta)))dW(s) \\ &= T_{1,j} + T_{2,j} + T_{3,j} + \frac{2}{\Delta}\int_{j\Delta}^{(j+1)\Delta}(W(s) - W(j\Delta))dW(s).\end{aligned}$$

It follows that: $\nu_n^{(1)} - \nu_n^{(1)}(W) = 2\Psi^2(n\Delta)^{-1}\int_0^{n\Delta}L_n(s)dW(s)$ where

$$\begin{aligned}L_n(s) &= \sum_{j=0}^{n-1}1_{(j\Delta,(j+1)\Delta]}(s)\frac{(\sigma(X(s)) - \sigma(X(j\Delta))}{\sigma^2(X(j\Delta))}\int_{j\Delta}^{s}(\sigma(X(u)) - \sigma(X(j\Delta)))dW(u) \\ &+ \frac{(\sigma(X(s) - \sigma(X(j\Delta)))(W(s) - W(j\Delta))}{\sigma(X(j\Delta))} \\ &+ \frac{1}{\sigma(X(j\Delta))}\int_{j\Delta}^{s}(\sigma(X(u) - \sigma(X(j\Delta))dW(u).\end{aligned}$$

We have: $\left(\int_0^{n\Delta} L_n^2(s)ds\right)^p \leq (n\Delta)^{p-1} \int_0^{n\Delta} L_n^{2p}(s)ds$ with

$$\begin{aligned}\int_0^{n\Delta} L_n^{2p}(s)ds &\lesssim \sum_{j=0}^{n-1} \int_{j\Delta}^{(j+1)\Delta} ds\ (X(s)-X(j\Delta))^{2p} \left(\int_{j\Delta}^{s} (\sigma(X(u))-\sigma(X(j\Delta)))dW(u)\right)^{2p} \\ &+ \sum_{j=0}^{n-1} \int_{j\Delta}^{(j+1)\Delta} ds\ (X(s)-X(j\Delta))^{2p}(W(s)-W(j\Delta))^{2p} \\ &+ \sum_{j=0}^{n-1} \int_{j\Delta}^{(j+1)\Delta} ds\ \left(\int_{j\Delta}^{s} (\sigma(X(u)-\sigma(X(j\Delta))dW(u)\right)^{2p}.\end{aligned}$$

We apply Lemma 1, the Cauchy-Schwarz, the Hölder and the B-D-G inequalities to obtain:

$$\mathbb{E}_\vartheta(\int_0^{n\Delta} L_n^{2p}(s)ds|\Phi_i=\varphi,\Psi_i=\psi)) \quad \lesssim \quad \Delta^{3p}(\|\varphi\|^{4p}+\psi^{4p})+2\Delta^{2p}(\|\varphi\|^{2p}+\psi^{2p}).$$

Consequently,

$$\mathbb{E}_\vartheta(\left(\nu_n^{(1)}-\nu_n^{(1)}(W)\right)^{2p}|\Phi_i=\varphi,\Psi_i=\psi) \lesssim \psi^{4p}\Delta^{2p}(\|\varphi\|^{2p}+\psi^{2p}+\|\varphi\|^{4p}+\psi^{4p}).$$

For $\nu_n^{(2)},\nu_n^{(3)}$, we have $\mathbb{E}_\vartheta((\nu_n^{(2)})^{2p}|\Phi_i=\varphi,\Psi_i=\psi) \lesssim \psi^{6p}n^{-2p}$ and $\mathbb{E}_\vartheta((\nu_n^{(3)})^{2p}|\Phi_i=\varphi,\Psi_i=\psi) \lesssim \psi^{2p}\|\varphi\|^{2p}n^{-2p}$.
For $\nu_n^{(4)}$, we have: $R_{j\Delta}^2(\Phi,\Psi,X) \lesssim \Delta^2\left(\|\Phi\|^4+(\|\Phi\|\Psi^2+\Psi^4)^2\right)+\Delta^{-2}4\Psi^2 I_j^2$ and

$$I_j = \int_{j\Delta}^{(j+1)\Delta} \Phi.(b(X(s)-b(X(j\Delta)))ds \int_{j\Delta}^{(j+1)\Delta} \sigma(X(s))dW(s) \tag{42}$$

Using Lemma 1, we find $\mathbb{E}_{\vartheta_0}(I_j^2|\Phi=\varphi,\Psi=\psi) \lesssim \Delta^4(\|\varphi\|^4+\psi^2)$.

$$\mathbb{E}_\vartheta((\nu_n^{(4)})^{2p}|\Phi_i=\varphi,\Psi_i=\psi) \lesssim \Delta^{2p}\left[\|\varphi\|^{4p}+\psi^{2p}(\|\varphi\|+\psi^2)^{2p}+(\|\varphi\|\psi)^{2p}(\|\varphi\|^{2p}+\psi^{2p})\right].$$

Joining terms, we finally get the inequality with exponent $2p \geq 2$. We conclude by application of the Cauchy-Schwarz inequality.□

7.2 Proofs of Section 4.2

Now, we assume for sake of clarity that $d=1$. The following lemma is easily obtained using Lemma 2.

Lemma 4. *Recall notations (24)-(25). Under (H1)-(H2),*

$$\begin{aligned}\mathbb{E}_\vartheta(|B_{i,n}-B_i(T;\omega^2)|^p|\Phi_i=\varphi,\Psi_i=\psi) &\lesssim \Delta^{p/2}(|\varphi|^p+\psi^p) \\ \mathbb{E}_\vartheta(|A_{i,n}-A_i(T;\mu,\omega^2)|^p|\Phi_i=\varphi,\Psi_i=\psi) &\lesssim \Delta^{p/2}(\varphi^{2p}+\psi^{3p}) \\ \mathbb{E}_\vartheta(|A_i(T;\mu,\omega^2)|^p|\Phi_i=\varphi,\Psi_i=\psi)) &\lesssim (|\varphi|^p+\psi^p).\end{aligned}$$

Proof of Theorem 1

We follow the classical scheme (see *e.g.* Kessler *et al.* (2012), Theorem 1.58-1.60, p.86-88). Let us set

$\theta = (\theta_1, \theta_2, \theta_3) = (\gamma, \mu, \omega^2)$ and denote by

$$-\mathcal{I}_{N,n}(\theta) = \left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ell_{N,n}(\theta) \right)_{1 \le i,j \le 3} \tag{43}$$

the Hessian matrix (opposite of the pseudo Fisher information matrix) containing the second order derivatives of $\ell_{N,n}(\theta)$ with respect to θ. The classical framework to study the asymptotic behaviour of the estimator $\tilde{\theta}_{N,n}$ relies on the following remarks. By the Taylor formula, we write for some $\theta^* \in (\theta, \tilde{\theta}_{N,n})$, $G_{N,n}(\tilde{\theta}_{N,n}) = 0 = G_{N,n}(\theta) - \mathcal{I}_{N,n}(\theta^*)(\tilde{\theta}_{N,n} - \theta)$. This yields

$$D_{N,n} G_{N,n}(\theta) = D_{N,n} \mathcal{I}_{N,n}(\theta^*) D_{N,n} \; D_{N,n}^{-1}(\tilde{\theta}_{N,n} - \theta).$$

Therefore, due to different rates of convergence, the limiting distribution of $D_{N,n}^{-1}(\tilde{\theta}_{N,n} - \theta)$ is obtained via the studies of the convergence in probability of $D_{N,n}\mathcal{I}_{N,n}(\theta^*)D_{N,n}$ and the convergence in distribution of $D_{N,n}G_{N,n}(\theta)$.

We first prove that, as N, n go to infinity, $D_{N,n}\mathcal{I}_{N,n}(\theta)D_{N,n}$ converges in probability under $\mathbb{P}_\theta$ to the 3×3 matrix $\mathcal{I}(\theta)$. The terms of $D_{N,n}\mathcal{I}_{N,n}(\theta)D_{N,n}$ are (see (22) and (24)):

$$\begin{aligned}
-\frac{1}{Nn}\frac{\partial^2}{\partial\gamma^2}\ell_{N,n}(\theta) &= \frac{1}{2\gamma^2}\ ; \quad -\frac{1}{N}\frac{\partial^2}{\partial\mu^2}\ell_{N,n}(\theta) = \frac{\gamma}{N}\sum_{i=1}^N B_{i,n}\ ;\\
-\frac{1}{N}\frac{\partial^2}{(\partial\omega^2)^2}\ell_{N,n}(\theta) &= \frac{1}{2N}\sum_{i=1}^N \left(2\gamma A_{i,n}^2 B_{i,n} - B_{i,n}^2\right)\ ;\\
-\frac{1}{N}\frac{\partial^2}{\partial\mu\partial\omega^2}\ell_{N,n}(\theta) &= \frac{1}{N}\sum_{i=1}^N \gamma A_{i,n} B_{i,n}\ ;\\
-\frac{1}{N\sqrt{n}}\frac{\partial^2}{\partial\gamma\partial\mu}\ell_{N,n}(\theta) &= -\frac{1}{\sqrt{n}}\frac{1}{N}\sum_{i=1}^N A_{i,n}\ ; \ -\frac{1}{N\sqrt{n}}\frac{\partial^2}{\partial\gamma\partial\omega^2}\ell_{N,n}(\theta) = -\frac{1}{2\sqrt{n}}\frac{1}{N}\sum_{i=1}^N A_{i,n}^2.
\end{aligned}$$

$$\begin{aligned}
-\frac{1}{Nn}\frac{\partial^2}{\partial\gamma^2}\ell_{N,n}(\theta) &= \frac{1}{2\gamma^2}\ ; \quad -\frac{1}{N}\frac{\partial^2}{\partial\mu^2}\ell_{N,n}(\theta) = \frac{\gamma}{N}\sum_{i=1}^N B_{i,n}\ ;\\
-\frac{1}{N}\frac{\partial^2}{(\partial\omega^2)^2}\ell_{N,n}(\theta) &= \frac{1}{2N}\sum_{i=1}^N \left(2\gamma A_{i,n}^2 B_{i,n} - B_{i,n}^2\right)\ ;\\
-\frac{1}{N}\frac{\partial^2}{\partial\mu\partial\omega^2}\ell_{N,n}(\theta) &= \frac{1}{N}\sum_{i=1}^N \gamma A_{i,n} B_{i,n}\ ;\\
-\frac{1}{N\sqrt{n}}\frac{\partial^2}{\partial\gamma\partial\mu}\ell_{N,n}(\theta) &= -\frac{1}{\sqrt{n}}\frac{1}{N}\sum_{i=1}^N A_{i,n}\ ; \ -\frac{1}{N\sqrt{n}}\frac{\partial^2}{\partial\gamma\partial\omega^2}\ell_{N,n}(\theta) = -\frac{1}{2\sqrt{n}}\frac{1}{N}\sum_{i=1}^N A_{i,n}^2.
\end{aligned}$$

In the above terms, replace $A_{i,n}, B_{i,n}$ respectively by $A_i(T;\mu,\omega^2), B_i(T;\omega^2)$. Denote the resulting matrix by $\mathcal{I}_N^{(T)}$. As the random variables appearing in the sums are *i.i.d.* and have finite moments (see Section 8.3), we apply the large law of numbers and obtain that, as N tends to infinity, $\mathcal{I}_N^{(T)}$ converges *a.s.* to $\mathcal{I}(\theta)$.

First note that the Φ_i's have moments of any order. Therefore, using Lemma 4, all terms

$$\mathbb{E}_\theta|B_{i,n} - B_i(T;\omega^2)|, \mathbb{E}_\theta|2\gamma A_{i,n}^2 B_{i,n} - B_{i,n}^2 - (2\gamma A_i^2(T;\mu,\omega^2)B_i(T;\mu,\omega^2) - B_i^2(T;\mu,\omega^2))|, \ldots,$$

tend to 0. Hence, $D_{N,n}\mathcal{I}_{N,n}(\theta)D_{N,n}$ converges in probability to $\mathcal{I}(\theta)$.
Similarly, for the multidimensional case, we have:

$$\begin{aligned}
\frac{\partial^2}{\partial\gamma^2}\ell_{N,n}(\theta) &= -\frac{Nn}{2\gamma^2}, \quad \frac{\partial^2}{\partial\gamma\partial\mu_k}\ell_{N,n}(\theta) = \sum_{i=1}^N A_{i,n}^k, \\
\frac{\partial^2}{\partial\gamma\partial\omega_{kl}}\ell_{N,n}(\theta) &= \sum_{i=1}^N A_{i,n}^k\ A_{i,n}^l, \quad \frac{\partial^2}{\partial\mu_k\partial\mu_l}\ell_{N,n}(\theta) = -\gamma\sum_{i=1}^N B_{i,n}^{lk}, \\
\frac{\partial^2}{\partial\mu_j\partial\omega_{kl}}\ell_{N,n}(\theta) &= \frac{\gamma}{2}\sum_{i=1}^N \frac{\partial}{\partial\mu_j}(A_{i,n}^k A_{i,n}^l) = -\gamma\sum_{i=1}^N B_{i,n}^{jk}\ A_{i,n}^l, \\
\frac{\partial^2}{\partial\omega_{kl}\partial\omega_{k'l'}}\ell_{N,n}(\theta) &= \frac{1}{2}\sum_{i=1}^N B_{i,n}^{l'k}B_{i,n}^{lk'} - \gamma\sum_{i=1}^N A_{i,n}^l\ A_{i,n}^{l'}B_{i,n}^{kk'}.
\end{aligned}$$

The second step is to prove that, under $\mathbb{P}_\theta$,

$$D_{N,n}G_{N,n}(\theta) \to \mathcal{N}_3(0,\mathcal{I}(\theta)) \text{ in distribution.} \tag{44}$$

We can write (see Lemma 3 with $\Gamma_i = \gamma$ deterministic)

$$\begin{aligned}
\frac{1}{\sqrt{Nn}}\frac{\partial}{\partial\gamma}\ell_{N,n}(\theta) &= \sqrt{Nn}\left(\frac{1}{2\gamma} - \frac{1}{2N}\sum_{i=1}^N \frac{S_{i,n}^{(1)}}{n}\right) + R_1 = \sum_{i=1}^N \chi_{i,n,N}^{(1)} + R_1, \\
\frac{1}{\sqrt{N}}\frac{\partial}{\partial\mu}\ell_{N,n}(\theta) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \gamma A_i(T;\mu,\omega^2) + R_2 = \sum_{i=1}^N \chi_{i,N}^{(2)} + R_2, \\
\frac{1}{\sqrt{N}}\frac{\partial}{\partial\omega^2}\ell_{N,n}(\theta) &= \frac{1}{2\sqrt{N}}\sum_{i=1}^N(\gamma A_i^2(T;\mu,\omega^2) - B_i(T,\omega^2)) + R_3 = \sum_{i=1}^N \chi_{i,N}^{(3)} + R_3,
\end{aligned}$$

with

$$\begin{aligned}
\chi_{i,n,N}^{(1)} &= \sqrt{n/N}\frac{1}{2\gamma}\left(1 - \frac{1}{n}\sum_{j=1}^n \frac{(W_i(t_j) - W_i(t_{j-1}))^2}{\Delta}\right), \quad R_1 = \sqrt{Nn}\frac{1}{2N}\sum_{i=1}^N \frac{S_{i,n}^{(1)} - S_{i,n} - T_i(\mu,\omega^2)}{n}, \\
\chi_{i,N}^{(2)} &= \frac{1}{\sqrt{N}}\gamma A_i(T;\mu,\omega^2), \quad R_2 = \frac{\gamma}{\sqrt{N}}\sum_{i=1}^N(A_{i,n} - A_i(T;\mu,\omega^2)), \\
\chi_{i,N}^{(3)} &= \frac{1}{2\sqrt{N}}(\gamma A_i^2(T;\mu,\omega^2) - B_i(T,\omega^2)), \\
R_3 &= \frac{1}{2\sqrt{N}}\sum_{i=1}^N\left((\gamma A_{i,n}^2 - B_{i,n}) - (\gamma A_i^2(T;\mu,\omega^2) - B_i(T,\omega^2))\right).
\end{aligned}$$

By the central limit theorem and the results recalled in Section 8.3, $(\sum_{i=1}^N \chi_{i,N}^{(2)}, \sum_{i=1}^N \chi_{i,N}^{(3)})'$ converges in distribution to $\mathcal{N}_2(0, I(\theta))$. Moreover, $\sum_{i=1}^N \chi_{i,N}^{(1)}$ converges in distribution to $\mathcal{N}(0, 1/(2\gamma^2))$. Now, let us prove that $\sum_{i=1}^N \mathbb{E}_\theta \chi_{i,N}^{(1)}\chi_{i,N}^{(2)}, \sum_{i=1}^N \mathbb{E}_\theta \chi_{i,N}^{(1)}\chi_{i,N}^{(3)}$ both tend to 0. We only treat the first term as the second one is analogous. We have

$$\sum_{i=1}^N \mathbb{E}_\theta \chi_{i,N}^{(1)}\chi_{i,N}^{(2)} = \mathbb{E}_\theta \left(\gamma A_1(T;\mu,\omega^2)\sqrt{n}\frac{1}{2\gamma}(1 - \frac{1}{n}\sum_{j=1}^n \frac{(W_1(t_j) - W_1(t_{j-1}))^2}{\Delta}) \right).$$

According to a theorem for stable convergence for discretized processes (see Genon-Catalot and Jacod (1993),Jacod (1997)), for all random variable Y, $\mathcal{F}_T$-measurable,

$$(Y, \sqrt{n}(1 - \frac{1}{n}\sum_{j=1}^n \frac{(W_1(t_j) - W_1(t_{j-1}))^2}{\Delta})) \to_{\mathcal{D}} (Y, \eta)$$

with η, Y independent and $\eta \sim \mathcal{N}(0, 2)$. We can apply this result for $Y = A_1(T;\mu,\omega^2)$. As $A_1(T;\mu,\omega^2)$ has moments of any order (see Section 8.3), we easily obtain that the sequence $(A_1(T;\mu,\omega^2)\sqrt{n}(1 - \frac{1}{n}\sum_{j=1}^n \frac{(W_1(t_j)-W_1(t_{j-1}))^2}{\Delta}))$ is uniformly integrable. Thus,

$$\mathbb{E}_\theta A_1(T;\mu,\omega^2)\sqrt{n}(1 - \frac{1}{n}\sum_{j=1}^n \frac{(W_1(t_j) - W_1(t_{j-1}))^2}{\Delta}) \to \mathbb{E}_\theta A_1(T;\mu,\omega^2)\mathbb{E}(\eta) = 0.$$

Hence, $(\sum_{i=1}^N \chi_{i,N}^{(\ell)}, \ell = 1, 2, 3)'$ converges in distribution to $\mathcal{N}_3(0, \mathcal{I}(\theta))$.

By Lemma 4, as Φ_i is Gaussian and $\Psi_i = 1/\sqrt{\gamma}$ is fixed, we have $\mathbb{E}_\theta|A_{i,n} - A_i(T;\mu,\omega^2)| = O(1/\sqrt{n})$, thus $\mathbb{E}_\theta|R_2| = O(\sqrt{N/n})$. The same holds for R_3. By Lemma 3, $\mathbb{E}_\theta|(S_{i,n}^{(1)} - S_{i,n})/n| = O(1/n)$ and $\mathbb{E}_\theta|T_i(\mu,\omega^2)/n| = O(1/n)$. Therefore, $\mathbb{E}_\theta|R_1| = O(\sqrt{N/n})$.
Hence, (44) is proved.
The last step to obtain weak consistency and asymptotic normality is to prove a uniformity condition, the convergence under $\mathbb{P}_\theta$ of $-D_{N,n}\mathcal{I}_{N,n}(\theta^*) - \mathcal{I}_{N,n}(\theta)D_{N,n}$ to 0 uniformly for θ^* such that $|\theta^* - \theta| \leq c/\sqrt{N}$. This last step can be obtained easily by standard computations.

7.3 Proofs of Section 4.3

We first state a lemma useful for the sequel.

Lemma 5. *We have (see (37)):*

$$\begin{aligned}
\mathbb{E}_\tau[|\zeta_i - \frac{S_{i,n}^{(1)}}{n}|^p|\Psi_i] &\lesssim \frac{1}{n^p}(1 + \Psi_i^p + \Psi_i^{2p} + \Psi_i^{4p}) = \frac{1}{n^p}(1 + \Gamma_i^{-p/2} + \Gamma_i^{-p} + \Gamma_i^{-2p})), \\
\mathbb{E}_\tau[|\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}}|^p|\Psi_i] &\lesssim \frac{1}{n^p}\left(1 + \Psi_i^{-2p} + \Psi_i^{-3p} + \Psi_i^{-4p} + \Psi_i^{4p}\right) \\
&\lesssim \frac{1}{n^p}\left(1 + \Gamma_i^p + \Gamma_i^{3p/2} + \Gamma_i^{2p} + \Gamma_i^{-2p}\right)
\end{aligned}$$

and $\mathbb{E}_\tau[|\zeta_i|^{-p}|\Psi_i] \lesssim \left(1 + \Gamma_i^p + \Gamma_i^{3p/2} + \Gamma_i^{2p} + \Gamma_i^{-2p}\right) \wedge \left(\frac{a+(n/2)}{\lambda}\right)^p$.

Proof of Lemma 5.
We can write

$$\zeta_i - \frac{S_{i,n}^{(1)}}{n} = \left(\frac{S_{i,n}}{n} - \frac{S_{i,n}^{(1)}}{n}\right)\frac{n}{2a+n} - \frac{S_{i,n}^{(1)}}{n}\frac{2a}{2a+n} + \frac{2\lambda}{2a+n} + \frac{2(\varphi^2 V_i - 2\varphi U_i)}{2a+n}. \quad (45)$$

By (12),

$$\mathbb{E}_\tau\left[\left(\frac{S_{i,n}^{(1)}}{n}\right)^p |\Psi_i\right] = \Psi_i^{2p}\mathbb{E}_\tau\left(\frac{C_{i,n}^{(1)}}{n}\right)^p = \Psi_i^{2p}0(1) = \Gamma_i^{-p}O(1).$$

Due to (H1)-(H2), V_i is bounded and by Lemma 2,

$$\mathbb{E}_\tau(|U_i|^p|\Psi_i) \lesssim (1+\Psi_i^p) = (1+\Gamma_i^{-p/2}).$$

Using Lemma 3 yields, as φ is deterministic, that :

$$\mathbb{E}_\tau\left[\left|\frac{S_{i,n}}{n} - \frac{S_{i,n}^{(1)}}{n}\right|^p |\Psi_i\right] \lesssim \frac{1}{n^p}(1+\Psi_i^{2p}+\Psi_i^{4p})) = \frac{1}{n^p}(1+\Gamma_i^{-p}+\Gamma_i^{-2p}).$$

Thus,

$$\mathbb{E}_\tau\left[\left|\zeta_i - \frac{S_{i,n}^{(1)}}{n}\right|^p |\Psi_i\right] \lesssim \frac{1}{n^p}(1+\Psi_i^p+\Psi_i^{2p}+\Psi_i^{4p}) = \frac{1}{n^p}(1+\Gamma_i^{-p/2}+\Gamma_i^{-p}+\Gamma_i^{-2p}). \quad (46)$$

Now, we write:

$$\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}} = (\frac{S_{i,n}^{(1)}}{n} - \zeta_i)\frac{1}{\frac{S_{i,n}^{(1)}}{n}}(\frac{1}{\zeta_i} - \frac{1}{\frac{S_{i,n}^{(1)}}{n}} + \frac{1}{\frac{S_{i,n}^{(1)}}{n}}).$$

Thus,

$$\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}} = \left(\frac{n}{S_{i,n}^{(1)}}\right)^2\zeta_i^{-1}\left(\frac{S_{i,n}^{(1)}}{n} - \zeta_i\right)^2 + \left(\frac{n}{S_{i,n}^{(1)}}\right)^2\left(\frac{S_{i,n}^{(1)}}{n} - \zeta_i\right).$$

Using that $\zeta_i \geq \lambda/(a+(n/2))$ yields

$$|\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}}| \leq \left(\frac{n}{C_{i,n}^{(1)}}\right)^2\left[\frac{a+(n/2)}{\lambda}\Psi_i^{-4}\left(\frac{S_{i,n}^{(1)}}{n} - \zeta_i\right)^2 + \Psi_i^{-4}\left|\frac{S_{i,n}^{(1)}}{n} - \zeta_i\right|\right]. \quad (47)$$

Therefore,

$$|\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}}|^p \lesssim \left(\frac{n}{C_{i,n}^{(1)}}\right)^{2p}\left[\left(\frac{a+(n/2)}{\lambda}\right)^p\Psi_i^{-4p}\left(\frac{S_{i,n}^{(1)}}{n} - \zeta_i\right)^{2p} + \Psi_i^{-4p}|\frac{S_{i,n}^{(1)}}{n} - \zeta_i|^p\right]. \quad (48)$$

For all $p < n/8$ (see Section 8.2),

$$\mathbb{E}_\tau\left(\frac{n}{C_{i,n}^{(1)}}\right)^{4p} = O(1). \quad (49)$$

We apply the Cauchy-Schwarz inequality and (46) to obtain:

$$\begin{aligned}
\mathbb{E}_\tau\left[\left|\zeta_i^{-1}-\frac{n}{S_{i,n}^{(1)}}\right|^p|\Psi_i\right] &\lesssim \frac{1}{n^p}\left(1+\Psi_i^{-4p}+\Psi_i^{-6p}+\Psi_i^{-8p}+\Psi_i^{8p}\right)^{1/2}\\
&\lesssim \frac{1}{n^p}\left(1+\Psi_i^{-2p}+\Psi_i^{-3p}+\Psi_i^{-4p}+\Psi_i^{4p}\right)\\
&\lesssim \frac{1}{n^p}\left(1+\Gamma_i^{p}+\Gamma_i^{3p/2}+\Gamma_i^{2p}+\Gamma_i^{-2p}\right).
\end{aligned}$$

We deduce $\mathbb{E}_\tau(|\zeta_i|^{-p}|\Psi_i)\lesssim 1+\Psi_i^{-2p}+\Psi_i^{-3p}+\Psi_i^{-4p}+\Psi_i^{4p}$. □

Proof of Theorem 2

We again follow the classical scheme. First, we study the convergence in distribution of $N^{-1/2}\widetilde{G}_{N,n}(\tau)$ (see (35)). We can write (see (37)):

$$\begin{aligned}
N^{-1/2}\frac{\partial}{\partial\lambda}\widetilde{\ell}_{N,n}(\tau) &= N^{-1/2}\sum_{i=1}^N\left(\frac{a}{\lambda}-\Gamma_i\right)+r_1, \qquad r_1=N^{-1/2}\sum_{i=1}^N\left(\Gamma_i-\zeta_i^{-1}\right)\\
N^{-1/2}\frac{\partial}{\partial a}\widetilde{\ell}_{N,n}(\tau) &= N^{-1/2}\sum_{i=1}^N(\log\lambda-\psi(a)+\log\Gamma_i)+r_2+r_2', \qquad r_2=N^{-1/2}\sum_{i=1}^N\left(\log(\Gamma_i^{-1})-\log\zeta_i\right)\\
N^{-1/2}\frac{\partial}{\partial\varphi}\widetilde{\ell}_{N,n}(\tau) &= N^{-1/2}\sum_{i=1}^N\Gamma_i(U_i(T)-\varphi V_i(T))+r_3+r_3', \quad r_3=N^{-1/2}\sum_{i=1}^N(\Gamma_i-\zeta_i^{-1})(U_i-\varphi V_i),
\end{aligned}$$

$$r_2'=N^{1/2}\left(\psi(a+(n/2))-\log(a+(n/2))\right), \qquad r_3'=N^{-1/2}\sum_{i=1}^N\Gamma_i\left(U_i-U_i(T)-\varphi(V_i-V_i(T))\right). \quad (50)$$

We will check that all r, r'-terms are negligible.

Before, we look at the main term of the decomposition of $N^{-1/2}\widetilde{G}_{N,n}(\tau)$. The first two components are exactly the score function associated with the observation of $(\Gamma_i, i=1,\dots,N)$ (see Section 4.1). The whole vector is ruled by the standard central limit theorem. We only need to check that the third component is not correlated to the first two ones. This follows from the fact that:

$$\mathbb{E}_\tau((U_i(T)-\varphi V_i(T))|\Gamma_i)=0.$$

Hence, the vector $N^{-1/2}\widetilde{G}_{N,n}(\tau)$ converges in distribution to $\mathcal{N}_3(0,\mathcal{V}(\tau))$ provided that all remainder terms are negligible.

Now we look at the r, r' terms. For the term r_2', we use that $(\psi(a+(n/2))-\log(a+(n/2)))=O(n^{-1})$ (see (58)). Thus $r_2'=O(\sqrt{N}n^{-1})$ tends to 0 under the constraint $N/n^2\to 0$. For r_3', we use

$$\mathbb{E}_\tau|r_3'|\le\sqrt{N}\mathbb{E}_\tau[\Gamma_i(|U_i-U_i(T)|+|\varphi(V_i-V_i(T)|)]$$

By Lemma 2, as φ is deterministic,

$$\begin{aligned}
\mathbb{E}_\tau\left[[\Gamma_i(|U_i-U_i(T)|+|\varphi(V_i-V_i(T)|)]|\Psi_i\right] &\lesssim \Gamma_i(\Delta^{1/2}(1+\Psi_i+\Psi_i^2+\Psi_i^3)+\Delta^{1/2}(1+\Psi_i)\\
&\lesssim \frac{1}{\sqrt{n}}(1+\Gamma_i+\Gamma_i^{1/2}+\Gamma_i^{-1/2}).
\end{aligned}$$

Note that Γ_i has moments of any order. Provided that $\mathbb{E}_\tau \Gamma_i^{-1/2} < +\infty$,

$$\mathbb{E}_\tau |r_3'| \lesssim \sqrt{\frac{N}{n}}.$$

Thus, $r_3' = O_P(\sqrt{N/n})$ tends to 0 under the constraints $N/n \to 0$ and $a > 1/2$.
Now, r_1, r_2 can be written as follows:

$$r_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\Gamma_i - \frac{n}{S_{i,n}^{(1)}} \right) + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{n}{S_{i,n}^{(1)}} - \zeta_i^{-1} \right) := r_{1,1} + r_{1,2}, \tag{51}$$

$$r_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\log \Gamma_i^{-1} - \log \frac{S_{i,n}^{(1)}}{n} \right) + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\log \frac{S_{i,n}^{(1)}}{n} - \log(\zeta_i) \right) := r_{2,1} + r_{2,2}. \tag{52}$$

For $r_{1,1}, r_{2,1}$, we use that (see (12))

$$\frac{S_{i,n}^{(1)}}{n} = \Gamma_i^{-1} \frac{C_{i,n}^{(1)}}{n} \qquad \text{where} \qquad C_{i,n}^{(1)} = \sum_{j=1}^n \Delta^{-1} (W_i(t_j) - W_i(t_{j-1}))^2$$

is independent of Γ_i and has distribution $\chi^2(n) = G(n/2, 1/2)$. By exact computations, using Gamma distributions (see Section 8)), we obtain, for $n > 4$,

$$\mathbb{E} \left(\frac{n}{C_{i,n}^{(1)}} - 1 \right) = \frac{2}{n-2}, \quad \mathbb{E} \left(\frac{n}{C_{i,n}^{(1)}} - 1 \right)^2 = \frac{2n+8}{(n-2)(n-4)} = O(n^{-1}),$$

$$\mathbb{E} \log C_{i,n}^{(1)}/2 - \log(n/2) = \psi(n/2) - \log(n/2) = O(n^{-1}), \mathrm{Var}(\log C_{i,n}^{(1)}/2) = \psi'(n/2) = O(n^{-1}).$$

Thus, $\mathbb{E} r_{1,1}^2 = O(N/n^2)$ and $\mathbb{E} r_{1,2}^2 = O(N/n^2)$ which implies

$$r_{1,1} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\Gamma_i - \frac{n}{S_{i,n}^{(1)}} \right) = O_P(\frac{\sqrt{N}}{n}), \quad r_{2,1} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\log \Gamma_i^{-1} - \log \frac{S_{i,n}^{(1)}}{n} \right) = O_P(\frac{\sqrt{N}}{n}).$$

Next, for $r_{1,2}, r_{2,2}$, we need to study (see (51) and (52))

$$\mathbb{E}_\tau \left| \frac{n}{S_{i,n}^{(1)}} - \zeta_i^{-1} \right|, \qquad \mathbb{E}_\tau \left| \log \frac{S_{i,n}^{(1)}}{n} - \log(\zeta_i) \right| \tag{53}$$

By Lemma 5,

$$\mathbb{E}_\tau(|\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}}||\Psi_i) \lesssim \frac{1}{n}(1 + \Gamma_i + \Gamma_i^2 + \Gamma_i^{3/2} + \Gamma_i^{-2}).$$

Finally, as Γ_i has moments of any order, if $\mathbb{E}_\tau \Gamma_i^{-2} < +\infty$, *i.e.* $a > 2$,

$$\mathbb{E}_\tau(|r_{1,2}|) \leq \sqrt{N} \mathbb{E}_\tau(|\zeta_i^{-1} - \frac{n}{S_{i,n}^{(1)}}| = O(\sqrt{N}/n).$$

For the second term of (53), we write

$$\begin{aligned}\log(\zeta_i) - \log\frac{S^{(1)}_{i,n}}{n} &= (\zeta_i - \frac{S^{(1)}_{i,n}}{n})\int_0^1 ds\left(\frac{1}{s\frac{S^{(1)}_{i,n}}{n} + (1-s)\zeta_i} - \frac{1}{\frac{S^{(1)}_{i,n}}{n}} + \frac{1}{\frac{S^{(1)}_{i,n}}{n}}\right)\\ &= \frac{1}{\frac{S^{(1)}_{i,n}}{n}}(\zeta_i - \frac{S^{(1)}_{i,n}}{n}) - \frac{1}{\frac{S^{(1)}_{i,n}}{n}}(\zeta_i - \frac{S^{(1)}_{i,n}}{n})^2\int_0^1 ds\frac{(1-s)}{s\frac{S^{(1)}_{i,n}}{n} + (1-s)\zeta_i}.\end{aligned}$$

This yields

$$\left|\log\zeta_i - \log\frac{S^{(1)}_{i,n}}{n}\right| \le \frac{n}{C^{(1)}_{i,n}}\left[\Gamma_i\left|\zeta_i - \frac{S^{(1)}_{i,n}}{n}\right| + \frac{a+(n/2)}{\lambda}\Gamma_i\left(\zeta_i - \frac{S^{(1)}_{i,n}}{n}\right)^2\right]. \tag{54}$$

Therefore, we get, applying the Cauchy-Swarz inequality and Lemma 37,

$$\mathbb{E}_\tau|\log\zeta_i - \log\frac{S^{(1)}_{i,n}}{n})| \lesssim \frac{1}{n}(1 + \mathbb{E}_\tau\Gamma_i^{-1} + \mathbb{E}_\tau\Gamma_i^{-3}).$$

We thus obtain that for $a > 3$,

$$\mathbb{E}_\tau(|r_{2,2}|) = O(\sqrt{N}/n).$$

Analogously, in r_3, we split $\Gamma_i - \zeta_i^{-1} = \Gamma_i - \frac{n}{S^{(1)}_{i,n}} + \frac{n}{S^{(1)}_{i,n}} - \zeta_i^{-1}$. This yields

$$r_3 = r_{3,1} + r_{3,2} = N^{-1/2}\sum_{i=1}^N(\Gamma_i - \frac{n}{S^{(1)}_{i,n}})(U_i - \varphi V_i) + N^{-1/2}\sum_{i=1}^N(\frac{n}{S^{(1)}_{i,n}} - \zeta_i^{-1})(U_i - \varphi V_i).$$

We proceed as above for r_1, r_2. Using Lemma 2 yields

$$\begin{aligned}\mathbb{E}_\tau(|(\Gamma_i - \frac{n}{S^{(1)}_{i,n}})(U_i - \varphi V_i)|\Psi_i) &\le \Gamma_i\left(\mathbb{E}_\tau(1 - \frac{n}{C^{(1)}_{i,n}})^2\mathbb{E}_\tau(U_i - \varphi V_i)^2|\Psi_i)\right)^{1/2}\\ &\lesssim \Gamma_i\, 0(\frac{1}{\sqrt{n}})(1 + \Gamma_i^{-1})^{1/2} \lesssim 0(\frac{1}{\sqrt{n}})(\Gamma_i + \Gamma_i^{1/2}).\end{aligned}$$

Hence, $\mathbb{E}_\tau(|r_{3,1}|) = O(\sqrt{N/n})$.
In the same way, using Lemma 37,

$$\begin{aligned}\mathbb{E}_\tau(|(\frac{n}{S^{(1)}_{i,n}} - \zeta_i^{-1})(U_i - \varphi V_i)|\Psi_i) &\le \left(\mathbb{E}_\tau((\frac{n}{S^{(1)}_{i,n}} - \zeta_i^{-1})^2|\Psi_i)\,\mathbb{E}_\tau(U_i - \varphi V_i)^2|\Psi_i)\right)^{1/2}\\ &\lesssim \frac{1}{n}(1 + \Gamma_i + \Gamma_i^2 + \Gamma_i^{3/2} + \Gamma_i^{-2})(1 + \Gamma_i^{-1/2})\\ &\lesssim \frac{1}{n}(1 + \Gamma_i^{1/2} + \Gamma_i + \Gamma_i^2 + \Gamma_i^{3/2} + \Gamma_i^{-2} + \Gamma_i^{-5/2}).\end{aligned}$$

Hence, $\mathbb{E}_\tau(|r_{3,2}|) = O(\sqrt{N}/n)$ provided that $\mathbb{E}_\tau\Gamma_i^{-5/2} < +\infty$, *i.e.* $a > 5/2$. We thus obtain $\mathbb{E}_\tau(|r_3|) = O(\sqrt{N/n})$ instead of $O(\sqrt{N/n})$ as for r_1, r_2.
Finally, $N^{-1/2}\widetilde{G}_{N,n}(\tau)$ converges in distribution to $\mathcal{N}_3(0, \mathcal{V}(\tau))$ if $N/n \to 0$ and $a > 3$. For the first two

components, the constraints $\sqrt{N}/n \to 0$ and $a > 3$ are enough.

Let us set $\tau = (\tau_1, \tau_2, \tau_3) = (\lambda, a, \varphi)$ and denote by

$$-\mathcal{V}_{N,n}(\tau) = \left(\frac{\partial^2}{\partial \tau_i \partial \tau_j} \widetilde{\ell}_{N,n}(\tau) \right)_{1 \le i,j \le 3} \tag{55}$$

the Hessian matrix $\widetilde{\ell}_{N,n}(\tau)$. By computations analogous to the ones above using assumptions (H1)-(H2), Lemma 2, Lemma 4 and 5, we prove that $-\mathcal{V}_{N,n}(\theta)/N$ converges in probability to $\mathcal{V}(\tau)$ as N, n tend to infinity under the constraint $a > 5$.
The last step of the proof of weak consistency is standard and omitted. $\square$

8 Auxiliary results

8.1 Discussion on Assumption (H2)

Assumption (H2) may appear strong. Usually, one would only assume that the functions b_k, σ have linear growth and impose moment assumptions on $X(0)$. When there are no random effects, as $X(0)$ is supposed to be deterministic, this implies that $X_i(t)$ has moments of any order. In the case of model(1), with unbounded random effects, the assumption of linear growth for drift and diffusion coefficient is not enough to ensure that $X_i(t)$ has finite moments. Let us illustrate this property on the example of the mixed effects Ornstein-Uhlenbeck process:

$$dX(t) = \Phi X(t)dt + \Psi dW(t), X(0) = x. \tag{56}$$

It has the explicit solution: $X(t) = xe^{\Phi t} + Y(t), \quad Y(t) = \Psi e^{\Phi t} \int_0^t e^{-\Phi s} dW(s)$.
Hence $\mathbb{E}|X(t)| < +\infty$ if $\mathbb{E} \exp(\Phi t) < +\infty$ and $\mathbb{E}(|Y(t)|) < +\infty$. If Φ is random, and Ψ deterministic, the conditions hold for Φ Gaussian but may not hold for another distribution. If Φ is deterministic and Ψ random, the second condition requires $\mathbb{E}|\Psi| < +\infty$. In the case of $\Psi = \Gamma^{-1/2}$ with $\Gamma \sim G(a, \lambda)$, this holds only for $a > 1/2$.
Nevertheless, on some models that we have implemented, Assumption (H2) is not verified and this does not seem to deteriorate results.

8.2 Properties of the Gamma distribution

The digamma function $\psi(a) = \Gamma'(a)/\Gamma(a)$ admits the following integral representation: $\psi(z) = -\gamma + \int_0^1 (1 - t^{z-1})/(1-t)dt$. (where $\gamma = \psi(1) = \Gamma'(1)$). For all positive a, we have $\psi'(a) = -\int_0^1 \frac{\log t}{1-t} t^{a-1} dt$. Consequently, using an integration by part, $-a\psi'(a) = -1 - \int_0^1 t^a g(t)dt$, where $g(t) = (\log t/(1-t))'$. A simple study yields that $t^a g(t)$ integrable on $(0, 1)$ and positive except at $t = 1$. Thus, $1 - a\psi'(a) \neq 0$. The following asymptotic expansions as a tends to infinity hold:

$$\log \Gamma(a) = (a - \frac{1}{2}) \log a - a + \frac{1}{2} \log 2\pi + O(\frac{1}{a}), \tag{57}$$

$$\psi(a) = \log a - \frac{1}{2a} + O(\frac{1}{a^2}), \quad \psi'(a) = \frac{1}{a} + O(\frac{1}{a^2}). \tag{58}$$

The following results are classical.
If X has distribution $G(a, \lambda)$, then λX has distribution $G(a, 1)$. For all integer k, $\mathbb{E}(\lambda X)^k = \frac{\Gamma(a+k)}{\Gamma(a)}$. For $a > k$, $\mathbb{E}(\lambda X)^{-k} = \frac{\Gamma(a-k)}{\Gamma(a)}$. Moreover, $\mathbb{E}\log(\lambda X) = \psi(a)$, Var $[\log(\lambda X)] = \psi'(a)$.
In particular, if $X = \sum_{j=1}^n \varepsilon_i^2$ where the ε_i's are *i.i.d.* $\mathcal{N}(0,1)$, then $X \sim \chi^2(n) = G(n/2, 1/2)$. Therefore, $\mathbb{E}X^{-p} < +\infty$ for $n > 2p$ and as $n \to +\infty$,

$$\mathbb{E}\left(\frac{X}{n}\right)^p = O(1), \qquad \mathbb{E}\left(\frac{n}{X}\right)^p = O(1). \tag{59}$$

Using the Rosenthal inequality, for all $p \geq 1$

$$\mathbb{E}|\frac{X}{n} - 1|^p \leq c_p n^{-p}\left(n\mathbb{E}|\varepsilon_i^2 - 1|^p + (n\mathbb{E}(\varepsilon_i^2-1)^2)^{p/2}\right) \lesssim O(\frac{1}{n^{p-1}} + \frac{1}{n^{p/2}}), \tag{60}$$

and for $n > 4p$, and $p \geq 2$,

$$\mathbb{E}|\frac{n}{X} - 1|^p \leq \left(\mathbb{E}\left(\frac{n}{X}\right)^{2p}\mathbb{E}\left(\frac{X}{n} - 1\right)^{2p}\right)^{1/2} \lesssim O(\frac{1}{n^{p/2}}). \tag{61}$$

8.3 Continuous observations of the sample paths

Assume that $\Psi_i = \gamma^{-1/2}$ is deterministic and known, that $\Phi_i \sim \mathcal{N}(\mu, \gamma^{-1}\omega^2)$ and that the sample paths $(X_i(t), t \in [0,T]), i = 1, \ldots, N$ are continuously observed throughout $[0,T]$. The exact likelihood associated with these observations is explicit and was studied in Delattre *et al.* (2013) together with the consistency and asymptotic normality of the maximum likelihood estimator of (μ, ω^2). Moment properties of the random variables $U_i(T), V_i(T)$ were derived (see Section 4.2 p.327 and Section 5, p.329). We recall those needed in the present paper. We use notations (7)-(2) with $d = 1$ and (25). The random variable $A_i(T;\mu,\omega^2)$ has moments of any order, $B_i(T;\mu,\omega^2)$ also as it is bounded and the following relations holds:

$$\begin{aligned}
\mathbb{E}_\theta(A_i(T;\mu,\omega^2)) &= 0, \quad \mathbb{E}_\theta(\gamma\, A_i^2(T;\mu,\omega^2)) - B_i(T;\mu,\omega^2)) = 0,\\
\mathbb{E}_\theta(A_i(T;\mu,\omega^2)B_i(T;\mu,\omega^2)) &= \frac{1}{2}\mathbb{E}_\theta(A_i(T;\mu,\omega^2)\left(\gamma\, A_i^2(T;\mu,\omega^2)) - B_i(T;\mu,\omega^2)\right)\\
\frac{1}{4}\mathbb{E}_\theta\left(\gamma\, A_i^2(T;\mu,\omega^2) - B_i(T;\mu,\omega^2)\right)^2 &= \frac{1}{2}\mathbb{E}_\theta(2\gamma\, A_i^2(T;\mu,\omega^2)B_i(T;\mu,\omega^2) - B_i^2(T;\mu,\omega^2)).
\end{aligned}$$

This shows that the matrix (27) is positive and is the covariance matrix of (28).