Parametric Estimation for Diffusion Processes from Discrete-time and Noisy Observations

Sophie Donnet, M. Samson

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Parametric Estimation for Diffusion Processes from Discrete-time and Noisy Observations

Sophie DONNET — Adeline SAMSON

N° 5809

Janvier 2005

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--5809--FR+ENG

INRIA
FUTURS

Parametric Estimation for Diffusion Processes from Discrete-time and Noisy Observations

Sophie DONNET*, Adeline SAMSON †

Thème COG — Systèmes cognitifs
Projet Select

Rapport de recherche n° 5809 — Janvier 2005 — 26 pages

Abstract: Noisy discretely observed diffusion processes with random drift function parameters are considered. Maximum likelihood and Bayesian estimation methods are extended to this model, respectively the Stochastic Approximation EM and the Gibbs sampler algorithms. They are based on the Euler-Maruyama approximation of the diffusion, achieved using latent auxiliary data introduced to complete the diffusion process between each pair of measurement instants. A tuned hybrid Gibbs algorithm based on conditional Brownian bridges simulations of the unobserved process paths is included in these two algorithms. Their convergence is proved. Errors induced on the likelihood and the posterior distribution by the Euler-Maruyama approximation are bounded as a function of the step size of the approximation. Results of a pharmacokinetic mixed model simulation study illustrate the accuracy of the maximum likelihood estimation method. The analysis of the Theophyllin real dataset illustrates the relevance of the SDE approach relative to the deterministic approach.

Key-words: Bayesian estimation, Brownian bridge, Diffusion process, Euler-Maruyama approximation, Gibbs algorithm, Incomplete data model, Maximum likelihood estimation, Non-linear mixed effects model, SAEM algorithm, Stochastic differential equation

* Equipe Select, Laboratoire de Mathématiques, Université Paris-Sud, Orsay, France
† INSERM U738, Paris, France; Universi y Paris 7, Paris, France

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 — Télécopie : +33 1 72 92 59 ??

Estimation paramétrique à partir d'observations bruitées et à temps discrets d'un processus de diffusion

Résumé : Nous considérons des processus de diffusion (solution d'une SDE) observés à temps discrets avec bruit de mesure, et dont les paramètres de la fonction de drift sont aléatoires. Des méthodes d'estimation standard par maximum de vraisemblance et par approche bayésienne (respectivement une version stochastique de l'algorithme EM et un échantillonneur de Gibbs) sont développées pour ces modèles. Elles sont basées sur une approximation de la diffusion par la méthode d'Euler-Maruyama, approximation obtenue en introduisant des temps auxiliaires entre les instants de mesure. La convergence de ces algorithmes est démontrée. Les erreurs induites par l'approximation d'Euler-Maruyama sur la vraisemblance et la distribution a posteriori des paramètres sont controlées par le pas de la discrétisation. La précision de la méthode d'estimation par maximum de vraisemblance est illustrée par une étude sur données simulées à partir d'un modèle non-linéaire à effets mixtes issu de la pharmacocinétique. L'analyse du jeu de données réelles Theophyllin illustre la pertinence de l'approche par SDE par rapport à l'approche déterministe (par ODE).

Mots-clés : Algorithme de Gibbs, Algorithme SAEM, Approximation de Euler-Maruyama, Estimation Bayésienne, Estimation par maximum de vraisemblance, Modèle à données incomplètes, Modèle non linéaire à effets mixtes, Pont Brownien, Processus de Diffusion, Equation différentielle stochastique

1 Introduction

Time-dependent dynamic processes that follow the laws of finance, physics, physiology or biology are usually described by differential systems. For example, stock price dynamics or short-term interest rates can be described using a wide class of financial differential systems. As another example, in biology, pharmacokinetics consists in the study of the evolution of a drug in an organism. It is described through dynamic systems, the human body being assimilated to a set of compartments within which the drug flows. In these contexts, diffusion models described by stochastic differential equations (SDEs) are natural extensions to the corresponding deterministic models (defined by ordinary differential equations, ODEs) to account for time-dependent or serial correlated residual errors and to handle real life variations in model parameters occurring over time. This variability in the model parameters is most often not predictable, not fully understood or too complex to be modeled deterministically. Thus the SDEs consider errors associated with misspecifications and approximations in the dynamic system.

The parametric estimation of such diffusion processes is a key issue. Estimation of continuously observed diffusion processes is widely studied (see for instance Kutoyants, 1984; Prakasa Rao, 1999). However, for obvious practical purposes, real longitudinal data are always gathered at discrete points in time (for example stock prices collected once a day, drug concentration measured every hour in patient blood, etc.). Within this framework, statistical inference of discretely observed diffusion processes is a critical question for both maximum likelihood and Bayesian approaches. When the transition probability of the diffusion process is explicitly known, Dacunha-Castelle and Florens-Zmirou (1986) propose a consistent maximum likelihood estimator. Classical Bayesian algorithm such as Gibbs sampling can also be directly applied in this particular case.

However, this transition density has generally no closed form and the estimation methods have to sidestep this difficulty. A short summary of such estimation methods is provided below (see Prakasa Rao, 1999; Sørensen, 2004, for complete reviews). Analytical methods include those of Bibby and Sørensen (1995), Sørensen (2000) – using estimating functions –, Poulsen (1999) – using a numerical solution of the Kolmogorov equation – or Aït-Sahalia (2002) – based on an analytical non-Gaussian approximation of the likelihood function. Other methods approximate the transition density via simulation. They consider the unobserved paths as missing data and introduce a set of auxiliary latent data points between every pair of observations. Along these auxiliary latent data points, the process can be

finely sampled using the Gaussian Euler-Maruyama approximation to evaluate the likelihood function via numerical integration as proposed by Pedersen (1995) and Elerian *et al.* (2001), or to evaluate the posterior distribution in a Bayesian analysis again via numerical integration, as discussed by Eraker (2001) and Roberts and Stramer (2001). In this context and for both maximum likelihood and Bayesian estimations, standard Markov Chain Monte-Carlo (MCMC) methods are used to sample the process with the conditional distributions. However, the convergence rate of these estimation methods decreases with the increase in number of latent data points. Different solutions are proposed to overcome this difficulty: Eraker (2001) suggests the sampling of only one element at a time, while Elerian *et al.* (2001) propose to sample block-wise with an importance sampling algorithm. Roberts and Stramer (2001) take a slightly different approach as they sample transformations of the diffusion process. To sidestep the Euler-Maruyama approximation, Beskos *et al.* (2005) develop an exact simulation method of the diffusion process, applicable even without any analytical form of the transition density. This algorithm can be included in a Monte-Carlo procedure to approximate the likelihood function for a classical estimation and in a Gibbs algorithm for a Bayesian inference. However, this exact simulation method is only adapted for time-homogeneous SDEs, which is frequently not the case when studying biological dynamical systems for example. Furthermore, even under the conditions defined by Beskos *et al.* (2005), this exact method requires the inclusion of accept-reject algorithms, which are difficult to implement in the general case of non-linear SDEs and often require a large computational time. Therefore an Euler-Maruyama approximation approach is considered in this paper.

The above-cited papers do not take into account the observation noise on the collected data, which is non-realistic in many cases. For example in the financial context, the daily evolution of an asset price depends on the price fluctuations within each business day. In the biological context, endpoints such as drug concentrations are generally measured with a certain variability due to experimental limits. To reflect this observation noise, we consider the following regression statistical model $\mathcal{M}$: the observed data $y = (y_1, \dots, y_J)$ are a realization of a random variable Y deduced from a scalar diffusion process Z, as stated by the following equation:

$$Y_j = Z(t_j) + \varepsilon_j, \qquad (\mathcal{M})$$

where $(\varepsilon_j)_{j=1,\dots,J}$ is a sequence of i.i.d Gaussian random variables of variance σ^2, representing the measurement errors. The diffusion process Z is defined as the solution of the SDE describing the observed dynamic process:

$$dZ(t) = F(Z, t, \phi)dt + \gamma dB(t),$$

driven by a Brownian motion $\{B_t, t_0 \leq t \leq T\}$, a drift function F depending on a parameter ϕ and a volatility coefficient γ. If the volatility coefficient γ is null, the SDE is an ODE, the SDE model parameter ϕ being evidently equivalent to the parameter of the corresponding ODE system, and therefore being interpreted in the same way. In such models, two fundamentally different types of noise have to be distinguished: the dynamic noise γ, reflecting the real random fluctuations around the corresponding theoretical dynamic model, and the measurement noise σ representing the uncorrelated part of the residual variability associated with assay, dosing and sampling errors, for instance, in a biological context. The problem of the parameter estimation of discretely observed diffusion processes with additive measurement noise is evoked in few papers and is not completely solved. In the particular case of

linear SDEs, the Kalman filter (Schweppe, 1965) or the EM algorithms (Singer, 1993) can be used. When the observed process is a Gaussian martingale, Jensen and Petersen (1999) and Gloter and Jacod (2001) exhibit estimators and study their theoretical properties. Unfortunately, these explicit forms of maximum likelihood estimates are limited to the linear SDEs case.

In this paper, we assume in addition that the parameter ϕ is a realization of a random variable Φ distributed with a probability π depending on a parameter β. This is the case for example in drug pharmacokinetics studies of which use will be detailed below. Basically, in order to estimate drug pharmacokinetic parameters, the drug concentration is sampled repeatedly among several individuals, the parameter ϕ being assumed different between the subjects and thus considered as individual non-observed random data.

The main objective of this paper is to develop methods to estimate the parameters vector $\theta = (\beta, \gamma^2, \sigma^2)$ in the general case of non-linear SDEs. Such a method is proposed by Overgaard *et al.* (2005) and Tornøe *et al.* (2005) in the particular case of non-linear mixed effects models. They combine an extended Kalman filter of the diffusion process with an approximated maximum likelihood estimation algorithm based on a linearization of the model. However, the convergence properties of this estimation algorithm based on linearization are not proved. A different point of view can be taken for the parameters estimation, the random quantities Z and Φ being considered as non-observed random data. In that case, the model $\mathcal{M}$ belongs to the framework of incomplete data models, for which several estimation methods are developed for both classical and Bayesian approaches. For classical inference, the Expectation-Maximization (EM) algorithm proposed by Dempster *et al.* (1977) is a broadly applied approach taking advantage of the incomplete data model structure. When the E-step has no closed form, Celeux and Diebolt (1985), Wei and Tanner (1990) and Delyon *et al.* (1999) propose different stochastic versions of this algorithm. These methods require the simulation of the non-observed data using Markov Chain Monte-Carlo (MCMC) algorithms, as proposed by Kuhn and Lavielle (2004). For the Bayesian approach, tuned Gibbs algorithms are developed to estimate the posterior distribution $p_{\theta|Y}(\cdot|y)$ of θ, a specified prior distribution $p_\theta(\cdot)$ for θ being given. When the simulation under the posterior distribution cannot be done in a closed form, hybrid Gibbs sampling algorithms are proposed in the literature, including Metropolis-Hastings procedures (Wakefield *et al.*, 1994; Bennet *et al.*, 1996). To our knowledge, these estimation methods are not yet extended to noisy discretely observed diffusion processes models considered in this paper.

Our objective is thus to propose efficient estimation methods of the vector of parameters θ for the model $\mathcal{M}$, together with theoretical convergence results for both classical and Bayesian inference. We consider an approximate statistical model, of which the regression term is the Euler-Maruyama discretized approximate diffusion process of the SDE. The parameter inference is then performed on this new model, using a stochastic version of the EM algorithm for the classical estimation approach, or using a hybrid version of the Gibbs sampling algorithm for the Bayesian approach.

Section 2 describes the setup of the problem which is considered in this paper, detailing the diffusion process and its Euler-Maruyama approximation. The estimation algorithms for the maximum likelihood and the Bayesian approaches are respectively presented in Sections 3 and 4. These sections detail a tuned MCMC procedure supplying both theoretical and computational convergence properties to these algorithms. The error on the estimation induced by the Euler-Maruyama scheme is quantified in Section 5. In Section 6, the

maximum likelihood algorithm is applied to a non-linear mixed effects model issued from pharmacokinetics. Section 7 concludes with some discussion.

2 Data and Model

2.1 Incomplete data model defined by SDEs

Let $y = (y_j)_{j=0..J}$ denote the vector of the observations measured at times $t_0 \leq t_1 \leq \ldots \leq t_J \leq T$. We consider that y is a realization of the random variable Y defined through the following statistical model $\mathcal{M}$:

$$\left.\begin{array}{rcll} Y_j & = & Z(t_j) + \varepsilon_j, & 0 \leq j \leq J \\ \varepsilon_j & \sim_{i.i.d} & \mathcal{N}(0, \sigma^2), & \\ dZ(t) & = & F(Z, t, \Phi)\, dt + \gamma dB(t)\ ,\ Z(t_0, \Phi) = Z_0(\Phi), & (1) \\ \Phi & \sim & \pi(\cdot, \beta) & \end{array}\right\} \quad (\mathcal{M})$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ represents the measurement error, with a residual variance σ^2. The regression term is a realization of the diffusion process $Z : \mathbb{R} \longrightarrow \mathbb{R}$ defined by the equation (2.1), with B a one-dimensional Brownian motion, γ is the volatility coefficient and the function $F : \mathbb{R} \times [t_0, T] \times \mathbb{R}^d \longrightarrow \mathbb{R}$ is the known measurable drift function, non-linearly depending on the non-observed parameter $\Phi \in \mathbb{R}^d$. We assume that Φ is a random variable, distributed with the density π, depending on the parameter $\beta \in \mathbb{R}^p$. The initial condition Z_0 of this process is a deterministic known function of the random parameter Φ (this deterministic function can be a constant).

Our objective is to propose a classical and a Bayesian estimation methods of the parameters vector θ, where $\theta = (\beta, \gamma^2, \sigma^2)$ belongs to some open subset Θ of the Euclidean space $\mathbb{R}^{p+2}$. As the random parameter Φ and the random trajectory Z are not observed, this statistical problem can be viewed as an incomplete data model. The observable vector Y is thus consider as part of a so-called complete vector (Y, Z, Φ).

***Remark* 1**

- *This work can be extended to a statistical model with a regression function being equal to $g(Z(t))$, with g a linear or non-linear function, i.e.*

$$Y_j = g(Z(t_j)) + \varepsilon_j, \qquad 0 \leq j \leq J.$$

However, for the simplicity's sake, we only consider the case $g(Z(t)) = Z(t)$ in this paper.

- *The identifiability of this model is a complex problem which is beyond the scope of this paper. However, for simple examples such as linear SDEs the parameters identifiability can be proved.*

2.2 Diffusion model

The diffusion process (2.1) is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. Statistical inference makes sense only if the existence and uniqueness of a solution of the SDE (2.1) for all $Z(t_0)$, Φ and γ is ensured. Sufficient conditions of existence and uniqueness are the following globally Lipschitz, linear growth and boundedness conditions:

Assumption **(A0)**:

1. For all $\phi \in \mathbb{R}^d$, for all $0 < R < \infty$, there exists $0 < K_R < \infty$ such that for all $t_0 \leq t \leq T$, for all $x, x' \in \mathbb{R}$ with $|x| \leq R$, $|x'| \leq R$

$$|F(x, t, \phi) - F(x', t, \phi)| \leq K_R |x - x'|.$$

2. For all $\phi \in \mathbb{R}^d$, for all $0 < T < \infty$, there exists a constant $0 < C_T < \infty$ such that for all $t_0 \leq t \leq T$, for all $x \in \mathbb{R}$

$$\gamma + |F(x, t, \phi)| \leq C_T(1 + |x|).$$

Under this assumption, for any $t_0 < t < T$, the distribution of $Z(t)$ conditioned by the filtration $\mathcal{F}_{t-}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ ($\mathcal{F}_{t-}$ being the filtration generated by $\{Z(s), s < t\}$). This distribution is denoted $p_{Z|\Phi}(\cdot|\phi; \gamma^2)$ in the following. As a consequence, both Y and (Y, Z, Φ) have density functions, denoted respectively $p_Y(y; \theta)$ and $p_{Y,Z,\Phi}(y, z, \phi; \theta)$ depending on the parameter θ.

2.3 Introduction of an approximate statistical model

For common SDEs, the diffusion density $p_{Z|\Phi}$ has generally no closed form. Consequently neither the likelihood of the observed data $p_Y(y; \theta)$ nor the likelihood of the complete data $p_{Y,Z,\Phi}(y, z, \phi; \theta)$ have analytical forms, which further complicates the parameters estimation. To overcome this difficulty, an approximate statistical model, based on the Euler-Maruyama approximation of the diffusion process is introduced.

2.3.1 Euler-Maruyama approximation of the diffusion process

The Euler-Maruyama scheme is one of the simplest discrete-time approximation of a diffusion process leading to Gaussian approximations of the transition densities. If the time intervals between the observation instants are too great to obtain a good approximation of the transition density, a natural approach is to introduce a set of auxiliary latent data points between every pair of observations, as first proposed by Pedersen (1995). Let $t_0 = \tau_0 < \tau_1 < \ldots < \tau_n < \ldots < \tau_N = t_J$ denote the deduced discretization of the time interval $[t_0, t_J]$. Let us assume that, for all $j = 0 \ldots J$, there exists an integer n_j verifying $t_j = \tau_{n_j}$, with $n_0 = 0$ by definition. Let $(h_n)_{1 \leq n \leq N}$ be the sequence of the step sizes defined as $h_n = \tau_n - \tau_{n-1}$. Let $h = \max_{1 \leq n \leq N} h_n$ be the maximal step size.

Then the diffusion process denoted W and supplied by the Euler-Maruyama approximation of the SDE is described by the following iterative scheme: for a fixed ϕ, $W_0 = Z_0(\phi)$, and for $n = 1 \ldots N$,

$$\begin{aligned} h_n &= \tau_n - \tau_{n-1}\,, \\ W_n &= W_{n-1} + h_n\, F(W_{n-1}, \tau_{n-1}, \phi) + \gamma\sqrt{h_n}\,\xi_n\,, \\ \xi_n &\sim_{i.i.d} \mathcal{N}(0, 1). \end{aligned}$$

Consequently, $(w_{n_0}, \ldots, w_{n_J})$ is an approximation of the original diffusion process at observations instants $(z(t_0), \ldots, z(t_J))$. In the following, let $w = (w_n)_{n=0 \cdots N}$ denote a realization vector of the process W at the discrete times $(\tau_n)_{n=0 \cdots N}$.

2.3.2 Approximate statistical model

Using this approximation of the diffusion process provided by the Euler-Maruyama scheme of step size h, an approximate statistical model denoted model $\mathcal{M}_h$ is defined as:

$$\left.\begin{array}{rcll} Y_j & = & W_{n_j} + \varepsilon_j, & 0 \leq j \leq J\,, \\ \varepsilon_j & \sim_{i.i.d} & \mathcal{N}(0, \sigma^2)\,, & \\ h_n & = & \tau_n - \tau_{n-1}\,, & \\ W_n & = & W_{n-1} + h_n\, F(W_{n-1}, \tau_{n-1}, \Phi) + \gamma\sqrt{h_n}\,\xi_n\,, & 1 \leq n \leq N, \\ \xi_n & \sim_{i.i.d} & \mathcal{N}(0,1)\,, & \\ \Phi & \sim & \pi(\cdot;\beta)\,, & \end{array}\right\} \quad (\mathcal{M}_h)$$

with $W_0 = Z_0(\Phi)$. On this model $\mathcal{M}_h$, Y results from the partial observation of the complete data (Y, W, Φ) where W is the process at the discrete times $(\tau_n)_{n=0\cdots N}$.

***Remark* 2** *In this data augmentation framework, the choice of the discretization grid $(\tau_n)_{0\leq n\leq N}$ is a central issue to guarantee the fast convergence of the estimation algorithms. Indeed, on the one hand, a small step size h ensures a fine Gaussian diffusion approximation. However, on the other hand, it increases the volume of missing data (W, Φ), which can lead to arbitrarily poor convergence properties of the algorithms when the missing data volume widely exceeds the volume of actually observed data Y. Furthermore, the time intervals between two observations can be strongly different. Therefore, for practical purposes and to prevent unbalanced volumes of missing data, we propose to adjust the step sizes for each single time interval.*

In the following, the distributions referring to the model $\mathcal{M}_h$ are denoted q while those referring to the model $\mathcal{M}$ are denoted p. On $\mathcal{M}_h$, the observation vector y is distributed with density distribution $q_Y(y;\theta)$, which has no closed form because of the SDE non-linearity with respect to ϕ. But by enriching the observed data with the missing data, and by the Markov property of the diffusion process, the complete data likelihood is analytically known:

$$\begin{aligned} q_{Y,W,\Phi}(y, w, \phi; \theta) &= q_{Y|W}(y|w;\sigma^2) \prod_{n=1}^{N} q_{W|\Phi}(w_n|w_{n-1}, \phi; \gamma^2)\; \pi(\phi;\beta) \\ &= q_{Y|W}(y|w;\sigma^2) \prod_{n=1}^{N} d(w_n;\; w_{n-1} + h_n\, F(w_{n-1}, \tau_{n-1}, \phi),\; \gamma^2 h_n)\; \pi(\phi;\beta), \end{aligned}$$

where $d(.; m, v)$ denotes the Gaussian density with mean m and variance v. As a consequence, the estimation of θ can be performed on the model $\mathcal{M}_h$, using a stochastic version of the EM algorithm for a Maximum Likelihood approach or a Gibbs algorithm for a Bayesian approach.

3 Maximum Likelihood Estimation on the model $\mathcal{M}_h$

In this section we propose a maximum likelihood estimation method, the vector of parameters θ being thus estimated as the maximizing value of the likelihood $q_Y(.\,;\theta)$.

3.1 Stochastic versions of the EM algorithm

The Expectation Maximization (EM) algorithm proposed by Dempster *et al.* (1977) takes advantage of the incomplete data model structure. We consider that the observed data Y are the partial observations of the complete data (Y, X) with X the vector of the non-observed data. The EM algorithm is useful in situations where the direct maximization of $\theta \to q_Y(.\,;\theta)$ is more complex than the maximization of $\theta \to Q(\theta|\theta')$, with:

$$Q(\theta|\theta') = E_{X|Y}\left[\log p_{Y,X}(y, x; \theta)|y; \theta'\right],$$

The EM algorithm is an iterative procedure: at the k-th iteration, the E-step is the evaluation of $Q_k(\theta) = Q(\theta \,|\, \theta_{k-1})$ while the M-step updates θ_{k-1} by maximizing $Q_k(\theta)$. For cases where the E-step has no closed form, Delyon *et al.* (1999) propose the Stochastic Approximation EM algorithm (SAEM) replacing the E-step by a stochastic approximation of $Q_k(\theta)$. The E-step is thus divided into a simulation step (S-step) of the non-observed data $x^{(k)}$ with the conditional distribution $p_{X|Y}(.\,|y;\theta_{k-1})$ and a stochastic approximation step (SA-step):

$$Q_k(\theta) = Q_{k-1}(\theta) + \alpha_k \left[\log \left(p_{Y,X}(y, x^{(k)}; \theta_{k-1}) \right) - Q_{k-1}(\theta) \right],$$

where $(\alpha_k)_{k\in\mathbb{N}}$ is a sequence of positive numbers decreasing to zero.

The distribution $p_{X|Y}(.|y;\theta_{k-1})$ is likely to be a complex distribution, as for the model $\mathcal{M}_h$, resulting in the impossibility of a direct simulation of the non-observed data x. For such cases, Kuhn and Lavielle (2004) suggest a MCMC scheme by constructing a Markov chain with an unique stationary distribution $p_{X|Y}(\cdot|y;\theta_{k-1})$ at the k-th iteration. They prove the convergence of the estimates sequence provided by this SAEM algorithm towards a maximum of the likelihood under general conditions and in the case where $p_{Y,X}$ belongs to a regular curved exponential family.

3.2 Extension of the SAEM algorithm to the model $\mathcal{M}_h$

In the particular case of the model $\mathcal{M}_h$, the non-observed data vector is equal to $X = (W, \Phi)$. As the simulation under the conditional distribution $q_{W,\Phi|Y}$ can not be performed directly, the SAEM algorithm combined with a MCMC procedure is applied to the model $\mathcal{M}_h$ to estimate the model parameter θ. To ensure the convergence of the SAEM algorithm, the model $\mathcal{M}_h$ is assumed to fulfill some regular conditions:

Assumption (**A1**):

1. $\pi(.\,;\beta)$ is such that $q_{Y,W,\Phi}$ belongs to the exponential family:

$$\log q_{Y,W,\Phi}(y, w, \phi; \theta) = -\psi(\theta) + \langle S(y, w, \phi), \nu(\theta) \rangle,$$

   where ψ and ν are two functions of θ, $S(y, w, \phi)$ is known as the minimal sufficient statistics of the complete model, taking its value in a subset $\widetilde{\mathcal{S}}$ of $\mathbb{R}^m$ and $\langle \cdot, \cdot \rangle$ is the scalar product on $\mathbb{R}^m$.

2. $\beta \longmapsto \pi(\phi;\beta)$ is of class $\mathcal{C}^m$ for all $\phi \in \mathbb{R}^d$.

Under the assumption (**A1**), the SA-step of the SAEM algorithm reduces to the approximation of $E\left[S(y, w, \phi)|y;\theta'\right]$. The k-th iteration of the SAEM algorithm is thus

- S-Step: a realization of the non-observed data $(w^{(k)}, \phi^{(k)})$ is generated through the succession of M iterations of a MCMC procedure providing an uniformly ergodic Markov chain with $q_{W,\Phi|Y}(\cdot|y;\theta_{k-1})$ as unique stationary distribution,

- SA-Step: s_{k-1} is updated using the following stochastic approximation scheme:

$$s_k = s_{k-1} + \alpha_k (S(y, w^{(k)}, \phi^{(k)}) - s_{k-1}),$$

- M-Step: θ_{k-1} is updated to maximize the complete log-likelihood:

$$\widehat{\theta}_k = \arg\max_{\theta} \left(-\psi(\theta) + \langle s_k, \nu(\theta) \rangle \right).$$

For example, the sufficient statistics corresponding to σ^2 and γ^2 are:

$$\begin{aligned} S^{(1)}(y,w,\phi) &= \frac{1}{J+1}\sum_{j=0}^{J}(y_j - w_{n_j})^2, \\ S^{(2)}(y,w,\phi) &= \frac{1}{N}\sum_{n=1}^{N}\frac{(w_n - h_n F(w_{n-1},\tau_{n-1},\phi))^2}{h_n}, \end{aligned}$$

and the M-step for σ^2 and γ^2 at iteration k reduces to $\widehat{\sigma^2}_k = s_k^{(1)}$ and $\widehat{\gamma^2}_k = s_k^{(2)}$. The sufficient statistics for β depend on the distribution $\pi(.\,;\beta)$.

3.3 Convergence of the SAEM algorithm on the model $\mathcal{M}_h$

Let denote Π_θ the transition probability of the Markov chain generated by the MCMC algorithm. Following Kuhn and Lavielle (2004), the convergence of the SAEM algorithm combined with MCMC is ensured under the following additional assumption:

Assumption (**A2**):

1. The chain $(w^{(k)},\phi^{(k)})_{k\geq 0}$ takes its values in a compact set $\mathcal{E}$ of $\mathbb{R}^N \times \mathbb{R}^d$.

2. For any compact subset V of Θ, there exists a real constant L such that for any (θ,θ') in V^2

$$\sup_{\{(w,\phi),(w',\phi')\}\in\mathcal{E}} |\Pi_\theta\,(w',\phi'|w,\phi) - \Pi_{\theta'}\,(w',\phi'|w,\phi)| \leq L\|\theta-\theta'\|_{\mathbb{R}^{p+2}}$$

3. The transition probability Π_θ supplies an uniformly ergodic chain of which invariant probability is the conditional distribution $q_{W,\Phi|Y}(\cdot\,;\theta)$, i.e.

$$\exists K_\theta \in \mathbb{R}^+, \quad \exists \rho_\theta \in]0,1[\quad | \quad \forall k \in \mathbb{N} \quad \|\Pi_\theta^k(\cdot|w,\phi) - q_{W,\Phi|Y}(\cdot;\theta)\|_{TV} \leq K_\theta \rho_\theta^k$$

   where $\|\cdot\|_{TV}$ is the total variation norm. Furthermore,

$$K = \sup_{\theta\in\Theta} K_\theta < \infty \quad \text{and} \quad \rho = \sup_{\theta\in\Theta}\rho_\theta < 1$$

4. The function S_h is bounded on $\mathcal{E}$.

Theorem 1 *Let assumptions (**A0-A1-A2**) hold. Let $q_{W,\Phi|Y}$ have finite moments of order 1 and 2. Let (α_k) be a sequence of positive numbers decreasing to 0 such that for all k in $\mathbb{N}$, $\alpha_k \in [0,1]$, $\sum_{k=1}^{\infty}\alpha_k = \infty$ and $\sum_{k=1}^{\infty}\alpha_k^2 < \infty$.*

Assuming the sequence $(s_k)_{k\geq 1}$ takes its values in a compact set, the sequence $(\widehat{\theta}_k)_{k\geq 1}$ obtained by the SAEM algorithm on the model $\mathcal{M}_h$ converges almost surely towards a (local) maximum of the likelihood q_Y.

Proof: Assuming (**A1-A2**) and the existence of finite moments for $q_{W,\Phi|Y}$, the assumptions of Kuhn and Lavielle (2004) are fulfilled and ensure the convergence of the estimates towards a local maximum of the likelihood function.

***Remark* 3** *If the compactness on $(s_k)_{k\geq 0}$ is not checked or difficult to check, the algorithm can be stabilized using the method of dynamic bounds proposed by Chen* et al. *(1988) and used by Delyon* et al. *(1999).*

A MCMC procedure fulfilling the assumption (**A2**) is proposed in the following part 3.4.

3.4 Simulation of the non-observed data using a MCMC procedure

At the k-th iteration of the SAEM algorithm, given en estimate $\widehat{\theta}_{k-1}$, a realization of the non-observed data $(w^{(k)}, \phi^{(k)})$ is generated through the succession of M iterations of a MCMC procedure. MCMC procedures construct a Markov chain with $q_{W,\Phi|Y}(.|y\ ;\widehat{\theta}_{k-1})$ as the invariant distribution, by proposing candidates (ϕ^c, w^c) with any proposal density Q. However, sampling all the missing data at the same time can lead to poor convergence properties. Therefore, a hybrid Gibbs algorithm is implemented and realized successively M times, the m-th iteration being written as:

1. generation of $\phi^{(m)}$, using a Metropolis-Hastings (M-H) procedure with Q_1 as proposal density and such that $q_{\Phi|Y,W}(.\ |y, w^{(m-1)}; \widehat{\theta}_{k-1})$ is the invariant distribution.

2. generation of $w^{(m)}$, using a M-H procedure with Q_2 as proposal distribution and such that $q_{W|Y,\Phi}(.\ |y, \phi^{(m)}; \widehat{\theta}_{k-1})$ is the invariant distribution.

A careful choice of the proposal densities Q_1 and Q_2 will help the algorithm to quickly explore the parameters space. In the following, some proposal densities of which efficiency is proved on numerical examples are detailed. To simplify the notation, the parameter $\widehat{\theta}_{k-1}$ is omitted since this simulation is performed for a fixed $\widehat{\theta}_{k-1}$.

3.4.1 Proposal distributions

1. Simulation of the candidate ϕ^c can be carried out with the prior density π which allows an efficient exploration of the space of parameters. This leads to an independent M-H algorithm. An alternative consists in generating a candidate in a neighborhood of $\phi^{(m-1)}$, $\phi^c = \phi^{(m-1)} + \eta$ with $\eta \sim \mathcal{N}(0, \delta)$ and where δ is a scaling parameter on which the algorithm convergence depends. This results in the so-called random-walk M-H algorithm (see for example Bennet *et al.*, 1996).

2. A trajectory candidate w^c can be generated using the Euler-Maruyama scheme which corresponds to the prior distribution. An alternative to simulate w^c consists in splitting the vector w into two parts $(w_{n_0}, \cdots, w_{n_J})$ and w_{aux}, the former being the process observed at times $(t_j)_{j=0\cdots J}$ and the latter being the process observed at the auxiliary latent times excluding the observation times. The simulation of $(w^c_{n_0}, \cdots, w^c_{n_J})$ can be performed with random walk distributions: $w^c_{n_j} = w^{(m-1)}_{n_j} + \eta'$ where $\eta' \sim \mathcal{N}(0, \delta')$ and δ' is a scaling parameter chosen to ensure good convergence properties. As proposed by Pedersen (1995), the trajectory at the auxiliary times w^c_{aux} can be generated using an unconditioned distribution but it would have poor convergence properties. A more appropriate strategy consists in generating a candidate w^c_{aux} using Brownian bridges, conditioning the proposed bridge on the events $(w^c_{n_j})_{j=0\cdots J}$, as suggested by Eraker (2001) or Roberts and Stramer (2001). More precisely, for $n_{j-1} < n < n_j$, w_{τ_n} is simulated with:

$$w^c_{\tau_n} = w^c_{n_{j-1}} + \frac{w^c_{n_j} - w^c_{n_{j-1}}}{t_j - t_{j-1}}(\tau_n - t_{j-1}) + \overline{B}_{\tau_n}$$

where $\overline{B}$ is a standard Brownian bridge on $[0, 1]$ equal to zero for $t = 0$ and $t = 1$, which can be easily simulated.

3.4.2 Uniform ergodicity of the MCMC procedure

For checking assumption (**A2**-3), it is possible to verify some minoration condition or Doeblin's condition for the transition probability Π_θ (see Chap. 16 of Meyn and Tweedie, 1993). Otherwise, each case has to be considered individually.

For the independent M-H algorithm, its uniform ergodicity is ensured as soon as the proposal distribution verifies:

$$\exists \lambda \in \mathbb{R}^+ \quad | \quad \forall (w,\phi) \in \mathcal{E}, \qquad Q(w,\phi) \geq \lambda q_{W,\Phi|Y}(w,\phi|y),$$

(see Th 2.1 in Tierney (1994) for more details). The proposal distribution equal to the prior distribution $Q(w,\phi) = q_{W,\Phi}(w,\phi)$ fulfills this condition.

Moreover, in case of a cyclic combination, the uniform ergodicity of the Markov Chain is ensured if one of the proposal distributions satisfies a minoration condition (Prop. 3 and 4 of Tierney (1994)). Thus the introduction of the prior distribution as proposal distribution is sufficient to ensure the uniform ergodicity of the Markov Chain.

4 Bayesian estimation of the model $\mathcal{M}_h$

For a fully Bayesian treatment of the estimation problem, we shall fix prior distributions on all unknown parameters $(\beta, \gamma^2, \sigma^2)$. We assume that β, γ^2 and σ^2 have continuous prior densities $p_\beta(\cdot)$, $p_\gamma(\cdot)$, $p_\sigma(\cdot)$ respectively on $\mathbb{R}^p$, $\mathbb{R}$ and $\mathbb{R}$. The Bayesian approach consists in the evaluation of the posterior distribution $p_{\theta|Y}$. According to the arguments developed in Section 2, the estimation procedure is applied to the model $\mathcal{M}_h$.

A simplistic approach would be to consider a basic Gibbs algorithm which simulates the non-observed data (ϕ, w) and then updates the parameter θ. However, as emphasized and illustrated by Roberts and Stramer (2001), the quadratic variation of the diffusion process satisfies, for almost surely all observation times t_j and t_{j+1}:

$$\lim_{h\to 0} \sum_{n=n_j}^{n_{j+1}-1} (w_{n+1} - w_n)^2 = (t_{j+1} - t_j)\gamma^2. \tag{2}$$

Therefore, conditional on any process satisfying (2), the posterior distribution of the volatility $q(\gamma^2|y,\phi,w,\sigma^2,\beta) \propto q_{W|\Phi}(w|\phi;\gamma^2)p_\gamma(\gamma^2)$ is just a point mass at γ^2. Consequently this data augmentation scheme is reducible. Roberts and Stramer (2001) propose a reparameterization to avoid this problem and consider the following transformation:

$$\dot{w}_n = \frac{w_n}{\gamma} \quad \text{and} \quad \dot{F}(x,t,\phi) = \frac{F(\gamma x, t, \phi)}{\gamma}.$$

Consequently, Bayesian inference is performed on the approximate model $\dot{\mathcal{M}}_h$ deduced from the model $\mathcal{M}_h$ using the same reparameterization:

$$\left.\begin{array}{rcll} Y_j & = & \gamma\, \dot{W}_{n_j} + \varepsilon_j, & 0 \leq j \leq J\,, \\ \varepsilon_j & \sim_{i.i.d} & \mathcal{N}(0,\sigma^2)\,, & \\ h_n & = & \tau_n - \tau_{n-1}\,, & \\ \dot{W}_n & = & \dot{W}_{n-1} + h_n\, \dot{F}(\dot{W}_{n-1}, \tau_{n-1}, \Phi) + \sqrt{h_n}\, \xi_n\,, & 1 \leq n \leq N, \\ \xi_n & \sim_{i.i.d} & \mathcal{N}(0,1)\,, & \\ \Phi & \sim & \pi(\cdot;\beta)\,. & \end{array}\right\} \quad (\dot{\mathcal{M}}_h)$$

A Gibbs algorithm based on this reparameterization is described below. The posterior distributions can be written as:

- $q_{\Phi,\dot{W}|Y,\theta}(\phi,\dot{w}|y,\theta) \propto q_{Y|\dot{W}}(y|\dot{w};\gamma,\sigma^2)q_{\dot{W}|\Phi}(\dot{w}|\phi,\gamma^2)p(\phi;\beta)$,
- $q(\sigma^2|y,\phi,\dot{w},\beta,\gamma^2) \propto q_{Y|\dot{W}}(y|\dot{w};\gamma^2,\sigma^2)p_\sigma(\sigma^2)$,
- $q(\gamma^2|y,\phi,\dot{w},\sigma^2,\beta) \propto q_{Y|\dot{W}}(y|\dot{w};\gamma^2,\sigma^2)q_{\dot{W}|\Phi}(\dot{w}|\phi,\gamma^2)p_\gamma(\gamma^2)$,
- $q(\beta|y,\phi,\dot{w},\sigma^2,\gamma^2) \propto p(\phi;\beta)p_\beta(\beta)$.

These conditional distributions provide the basis for the algorithm, alternating between updating $(\phi,\dot{w})$, β, γ^2 and σ^2 according to their conditional posterior distributions. Updating β, γ^2 and σ^2 can be carried out using standard M-H algorithms and is not discussed in detail here. Updating $(\phi,\dot{w})$ is less straightforward and is detailed in Section 3.4 in the case of the basic data augmentation. This procedure is easily adjustable to the reparameterization case by using the conditional distributions detailed previously. For a practical implementation, we recommend the paper of Roberts and Stramer (2001) which can be adapted to the model $\dot{\mathcal{M}}_h$.

The convergence of this Gibbs algorithm is proved in the following theorem:

Theorem 2 *Let p_θ and $q_{\theta|Y}$ be respectively the prior and the posterior distributions of θ on the model $\mathcal{M}_h$.*

Assuming the proposal distributions specified above, the hybrid Gibbs algorithm detailed previously converges and provides an ergodic Markov Chain generated with the posterior distribution $q_{\theta|Y}$.

Proof: As previously detailed in Section 3.4, the ergodicity of the Markov Chain is ensured if one of the proposal distributions of the cyclic combination fulfills a minoration condition detailed in Tierney (1994), which is the case with the proposal distributions used to generate $(\phi,\dot{w})$.

It is well known that prior distributions must be properly defined and that their choice may have a considerable impact on the posterior distribution evaluation. Classically, standard non-informative prior distributions are assumed. Following Gilks *et al.* (1996), Gamma distributions can be chosen for σ and γ, and a multivariate Gaussian distribution for β.

5 Survey of the error induced by the Euler-Maruyama approximation

Both estimation methods proposed in this paper, respectively the maximum likelihood and the Bayesian schemes, generate two distinct types of errors on the parameter estimates that have to be controlled.

The first type of error is induced by the estimation method itself. For the maximum likelihood approach, the estimation algorithm produces a sequence $(\widehat{\theta}_k)_{k\geq 0}$ of estimates which converges towards θ_h^*, the maximum of the $\mathcal{M}_h$-likelihood $q_Y(y;\cdot)$ function. Delyon *et al.* (1999) prove an asymptotic normal result of convergence of an averaged SAEM procedure. The variance of this estimate $\widehat{\theta}_k$ is classically controlled by the standard error evaluated through the Fisher information matrix of the estimates. Kuhn and Lavielle (2004) propose to estimate this Fisher information matrix by using the stochastic approximation procedure and the Louis' missing information principle (Louis, 1982). For the Bayesian approach, the equivalent of the problem of obtaining the standard errors is to obtain estimated variances for the posterior mean $E(\theta|y)$. Gilks *et al.* (1996) propose MCMC convergence diagnostics

tools in their book, the simplest and more generally used one is independent parallel simulations mixed in together. Different variance estimates are proposed such as the effective sample size estimate or the batching approach, that provide 95% confidence interval for $E(\theta|y)$ (see Carlin and Louis, 2000, for a review of such methods). Because this type of error is not specific to the situation exposed in this paper, it is not further discussed here.

A second type of error is induced on the estimates by the Euler-Maruyama scheme. Indeed, for the reasons evoked in Section 2, the estimation algorithms are applied to the model $\mathcal{M}_h$ instead of to the model $\mathcal{M}$. In the maximum likelihood approach, the algorithm maximizes the $\mathcal{M}_h$-likelihood function q_Y instead of the $\mathcal{M}$-likelihood function p_Y; in the Bayesian framework, the parameters are generated under the posterior distribution $q_{\theta|Y}$ instead of the posterior distribution $p_{\theta|Y}$.

The aim of this section is to study this second type of error induced by the Euler-Maruyama scheme on the conditional distribution $q_{W,\Phi|Y}$, on the likelihood function q_Y and on the posterior distribution $q_{\theta|Y}$. In Theorem 3 we propose bounds of these three errors as functions of the maximal step size of the Euler-Maruyama scheme h. In the following, some additional assumptions hold:

Assumption (**A3**):

The function $F : \mathbb{R} \times [t_0, T] \times \mathbb{R}^d \longrightarrow \mathbb{R}$ is infinitely differentiable in the variable space and its partial derivatives of any order are uniformly bounded with respect to x and ϕ.

Assumption (**A4**):

The assumption **(UH)** of Bally and Talay (1996) is satisfied. More precisely, let A_0 and A_1 denote the vector fields defined respectively by $A_0 = F(\cdot)\partial_z$ and $A_1 = \gamma\partial_z$. For multiindices $a = (a_1, \ldots, a_\ell) \in \{0,1\}^\ell$, let the vector fields A_1^a be defined by induction: $A_1^\emptyset := A_1$ and for $j = 0$ or 1 $A_1^{(a,j)} := [A_1^j, A_1^a]$ where $[\cdot,\cdot]$ denotes the Lie bracket. For $L \leq 1$, let define the quadratic form $V_L(\xi, \eta) := \sum_{|a|\leq L-1} \langle A_1^a(\xi), \eta\rangle^2$. We assume that

$$V_L(\xi) := 1 \wedge \inf_{\|\eta\|=1} V_L(\xi, \eta) \geq 0$$

Theorem 3 *Let the assumptions (**A0-A4**) hold.*

1. *Let Z and W be the diffusion processes of the models $\mathcal{M}$ and $\mathcal{M}_h$ respectively, at the observation times: $Z = (Z(t_0), \cdots, Z(t_J))$ and $W = (W(t_0), \cdots, W(t_J))$.*

   Let $p_{Z,\Phi|Y}$ and $q_{W,\Phi|Y}$ be the conditional distributions on the models $\mathcal{M}$ and $\mathcal{M}_h$ respectively. There exists a constant $C(y)$ such that, for any $0 < h < H_0$,

   $$\left\|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\right\|_{TV} \leq C(y)h,$$

   where $\|\cdot\|_{TV}$ denotes the total variation distance.

2. *Let p_Y and q_Y be the likelihoods of the models $\mathcal{M}$ and $\mathcal{M}_h$ respectively. There exists a constant $C_2(y)$ such that for all $0 < h < H_0$*

   $$\sup_{\{\theta=(\beta,\gamma^2,\sigma^2),\sigma^2>\sigma_0^2,\gamma^2>\gamma_0^2\}} |p_Y(y;\theta) - q_Y(y;\theta)| \leq C_2(y)h.$$

3. *In the Bayesian approach, let p_θ denote the prior distribution. Let $p_{\theta|Y}$ and $q_{\theta|Y}$ be the posterior distributions of the models $\mathcal{M}$ and $\mathcal{M}_h$ respectively. There exists a constant $C_3(y)$ such that for all $0 < h < H_0$*

   $$\left\|q_{\theta|Y} - p_{\theta|Y}\right\|_{TV} \leq C_3(y)h.$$

Theorem 3 is proved in Appendix A. These results are based on the convergence rate of the transition densities proposed by Bally and Talay (1996).

***Remark* 4** *Assumption (**A3**) requires only the derivatives of the function F to be bounded and not F itself. Assumption (**A4**) is easily satisfied for linear drift functions F: $F(x,t,\phi) = A(\phi,t) + xB(\phi,t)$.*

6 Theophyllin pharmacokinetic example

The maximum likelihood estimation method developed in Section 3 is applied below to a pharmacokinetics example.

6.1 Pharmacokinetics and Non-linear mixed effects models

Pharmacokinetics (PK) studies the time course of drug substances in the organism. This can be described through dynamic systems, the human body being assimilated to a set of compartments within which the drug evolves with time. In general, these systems are considered in their deterministic version. However, in a recent book on PK modeling, Krishna (2004) claims that the fluctuations around the theoretical pharmacokinetic dynamic model may be appropriately modeled by using SDEs rather than ODEs. Overgaard *et al.* (2005) suggest the introduction of SDEs to consider serial correlated residual errors due for example to erroneous dosing, sampling history or structural model misspecification. This new variability is distinct from the standard measurement noise representing the experimental uncertainty such as assay error.

Generally, several patients are followed up in a clinical trial, their drug concentration being measured along time repeatedly. Longitudinal data are thus gathered at discrete times and classically analyzed using non-linear mixed-effects models. Indeed, the mixed-effects models are a means to discriminate the intra-subject variability from the inter-subject variability, the parameter ϕ being a random parameter proper to each subject. The non-linear mixed-effects model can be written as follows:

$$\left.\begin{array}{rcl} y_{ij} & = & Z(t_{ij},\phi_i) + \varepsilon_{ij} \\ \varepsilon_{ij} & \sim & \mathcal{N}(0,\sigma^2) \\ \phi_i & \sim & \mathcal{N}(\mu,\Omega) \end{array}\right\} \qquad (\mathcal{M}_{mix})$$

where y_{ij} is the observation for subject i, $i = 1,\ldots,I$ at time t_{ij}, $j = 1,\ldots,J_i$ and ϕ_i is the vector of individual and non-observed parameters of subject i.

In a deterministic approach, the regression function Z is defined as the solution of a PK ordinary differential system: $dZ(t)/dt = F(Z(t),t,\phi)$ with $Z(t_0) = Z_0$, each component of the vector ϕ having a PK meaning. For example, a classic one compartment PK model with first order absorption and first order elimination is described by the following dynamic equation: $Z_0 = Dose$ and

$$\frac{dZ(t,\phi)}{dt} = \frac{Dose \cdot K_a K_e}{Cl} e^{-K_a t} - K_e Z(t,\phi), \tag{3}$$

where Z is the drug concentration, $Dose$ is the known drug oral dose received by the subject, K_e is the elimination rate constant, K_a is the absorption rate constant and Cl is the clearance of the drug. A stochastic differential system can be deduced from the ODE:

$$dZ(t,\phi) = \left(\frac{Dose \cdot K_a K_e}{Cl} e^{-K_a t} - K_e Z(t,\phi)\right) dt + \gamma dB_t \tag{4}$$

where B_t is a Brownian motion and γ is the volatility coefficient of the SDE.

In its SDE version, the non-linear mixed-effects model ($\mathcal{M}_{mix}$) is a particular case of the model $\mathcal{M}$ previously presented i.e. a diffusion process is observed at discrete times with noise measurement and its drift function parameters are random.

6.2 Simulation study

The aim of this simulation study is to illustrate the accuracy (bias and root mean square errors) of the extended SAEM algorithm developed in Section 3.2 on a PK application.

We use the previous PK model to mimic the Theophyllin drug pharmacokinetic. To prevent the parameters from taking unrealistic negative values, the vector $\phi \in \mathbb{R}^3$ is classically composed of the log parameters $\phi = (\log(K_e), \log(K_a), \log(Cl))$. The individual parameters ϕ are thus simulated with Gaussian distributions $\mathcal{N}(\mu, \Omega)$, with μ equal to $(-2.52, 0.40, -3.22)$ as proposed by Pinheiro and Bates (1995). A diagonal variance-covariance matrix Ω is assumed for the Gaussian distribution of ϕ. Let $\omega^2 = (\omega_1^2, \omega_2^2, \omega_3^2)$ denote the vector of these variances. The inter-subject variability is set equal for the three parameters: $\omega_1^2 = \omega_2^2 = \omega_3^2 = 0.01$, corresponding to a variation coefficient of 10%. We set a volatility coefficient equal to $\gamma^2 = 0.2$ and an additive Gaussian measurement error $\sigma^2 = 0.1$. We generate 100 datasets with $I = 36$ subjects and with nine blood samples per patient ($J = 8$), taken at 15 minutes, 30 minutes, 1, 2, 3.5, 5, 7, 9, 12 hours after dosing. The drug oral dose (*Dose*) received by the subject is chosen arbitrarily between 3 and 6 mg.

To evaluate the accuracy of the estimates of $\theta = (\mu, \omega^2, \gamma^2, \sigma^2)$ produced by the SAEM algorithm, the estimation of the parameters is performed on the 100 simulated datasets using the extension of the SAEM algorithm presented in Section 3.2.

The Euler-Maruyama scheme included in the SAEM algorithm is implemented on a grid with auxiliary latent data points introduced between each pair of observation instants as detailed in Section 2.3.1. The number of auxiliary points has to be chosen carefully because a volume of missing data too large can induce arbitrarily poor convergence properties of the Gibbs algorithm. In this example, we divide each time interval $[t_{i,j}, t_{i,j+1}]$ into 20 sub-intervals of equal length. This choice supplies a reasonable volume of missing data, avoids unbalance between the observation-time intervals and proves its numerical efficiency in accurately approximating the solution of the SDE.

The implementation of the Gibbs procedure included in the SAEM algorithm requires subtle tuning in practice. In particular, the simulation of the diffusion process w on the auxiliary grid is highly critical. An unconditioned trajectory simulation with $q(w_{n_j}|w_{n_{j-1}}; \theta)$ as proposed by Pedersen (1995) provides poor numerical results in the case of this example. Indeed, a great number of these simulated trajectories produce large jumps $(w_{\tau_{n_j}} - w_{\tau_{n_j}-1})$. The probability of such trajectories being close to zero, it induces too low an acceptance rate. As suggested by Eraker (2001) or Roberts and Stramer (2001) and detailed in Section 3.4, a conditioned trajectory simulation through Brownian bridge distributions is preferred. Moreover, we update the missing trajectories at once for each subject, as recommended by Elerian *et al.* (2001) to avoid a high level of rejection. In this example, we obtain acceptance rates in the neighborhood of 25%.

The implementation of the SAEM algorithm requires initial values and the choice of the stochastic approximation sequence $(\alpha_k)_{k\geq 0}$. The initial values of the parameters are chosen arbitrarily and set to $\theta_0 = (-3, 1, -3, 0.1, 0.1, 0.1, 2, 1)$. The step of the stochastic approximation scheme is chosen as recommended by Kuhn and Lavielle (2005): $\alpha_k = 1$

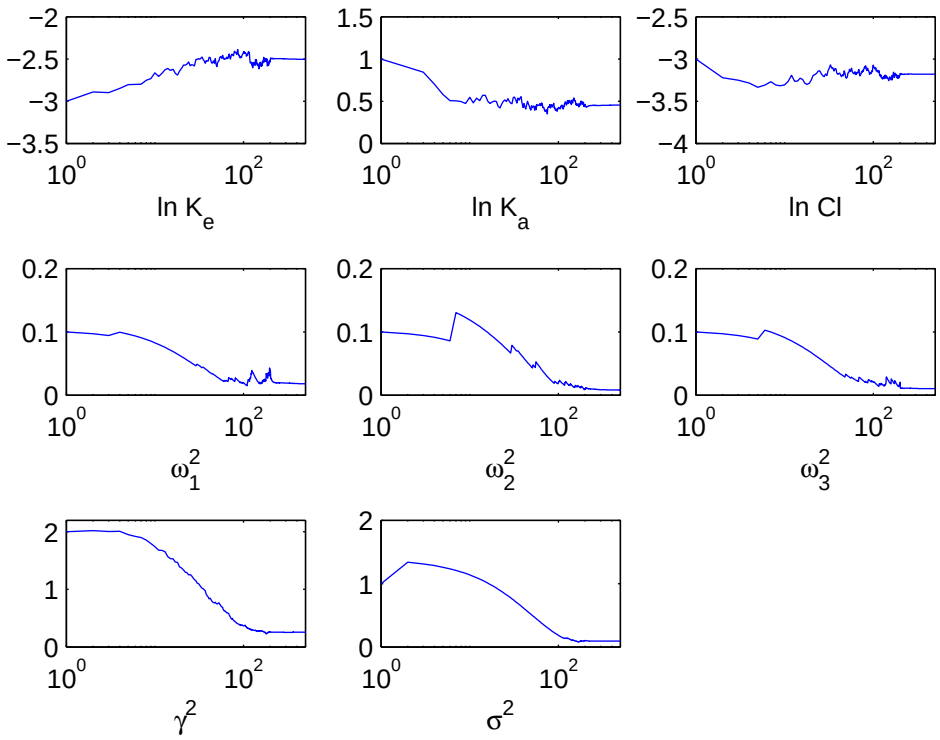

Figure 1: Evolution of the SAEM parameter estimates function of the iteration number in a logarithmic scale

during the first iterations $1 \leq k \leq K_1$, and $\alpha_k = (k-K_1)^{-1}$ during the subsequent iterations. Indeed, the initial guess θ_0 might be far from the maximum likelihood value and the first iterations with $\alpha_k = 1$ allow the sequence of estimates to converge to a neighborhood of the maximum likelihood estimate. Subsequently, smaller step sizes during $K - K_1$ additional iterations ensure the almost sure convergence of the algorithm to the maximum likelihood estimate. We implement the extended SAEM algorithm with $K_1 = 200$ and $K = 500$ iterations. Figure 1 illustrates the convergence of the parameter estimates provided by the extended SAEM algorithm as a function of the iteration number in a logarithmic scale. During the first iterations of SAEM, the parameter estimates fluctuate, reflecting the Markov chain construction. After 200 iterations, the curves smooth out but still continue to converge towards a neighborhood of the likelihood maximum. Convergence is obtained after 500 iterations.

The relative bias and relative root mean square error (RMSE) for each component of θ are computed and presented in Table 1.

The estimates of the mean parameter μ have very low bias ($<5\%$). The variance parameters have small bias ($<9\%$) except γ^2, this variance parameter being slightly over-estimated (13%). The RMSE are very satisfactory for the mean parameter ($<9\%$). The RMSE for the variance parameters are greater but still satisfactory ($\leq 40\%$) in comparison to the small number of subjects ($I = 36$). The RMSE of σ^2 is particularly satisfactory ($<20\%$) considering the complexity of the variability model.

In conclusion, even if this simulation study is performed on a complex model, the convergence of the extended SAEM algorithm towards the maximum likelihood neighborhood is

Table 1: Relative bias (%) and relative root mean square error (RMSE) (%) of the estimated parameters evaluated by the SAEM algorithm from 100 simulated trials with $I = 36$ subjects.

Parameters	Bias (%)	RMSE (%)
log K_e	0.42	-3.19
log K_a	4.14	8.95
log Cl	-0.23	-2.27
ω_1^2	3.83	40.03
ω_2^2	8.49	36.76
ω_3^2	-8.81	37.52
γ^2	13.02	21.31
σ^2	-4.44	18.79

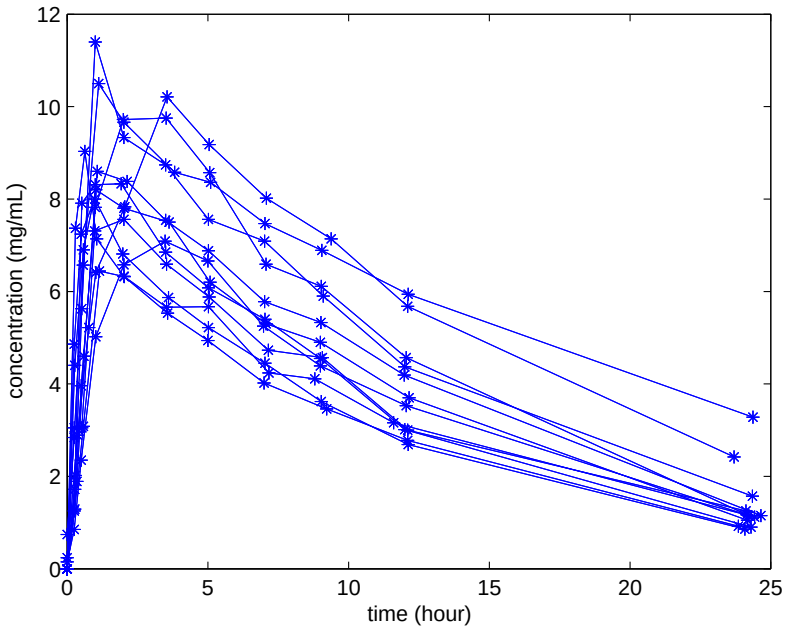

Figure 2: Individual concentrations for the pharmacokinetics of Theophyllin for 12 subjects.

computationally efficient. In addition despite the fact that the number of subjects is small, the extended SAEM algorithm all in all supplies accurate estimations of the parameters. Furthermore, the accuracy is comparable to that obtained with the classic SAEM algorithm for an ODE version of a mixed model ($\mathcal{M}_{mix}$) i.e. for a model with one less variability level.

6.3 A real data example

The extended SAEM algorithm is used to estimate the PK parameters of the Theophyllin drug PK real dataset. This new analysis of the Theophyllin dataset aims at illustrating the advantage of the SDE approach over the ODE approach.

In this clinical trial, twelve subjects received a single oral dose of 3 to 6 mg of Theophyllin. Ten blood samples were taken 15 minutes, 30 minutes, 1, 2, 3.5, 5, 7, 9, 12 and 24 hours after dosing. The individual data are displayed in Figure 2. The Theophyllin PK is classically described by the one compartment model with first order absorption and first order elimination presented previously. We fit the Theophyllin data with the regression term successively defined as the solution (3) and then as that of the SDE (4).

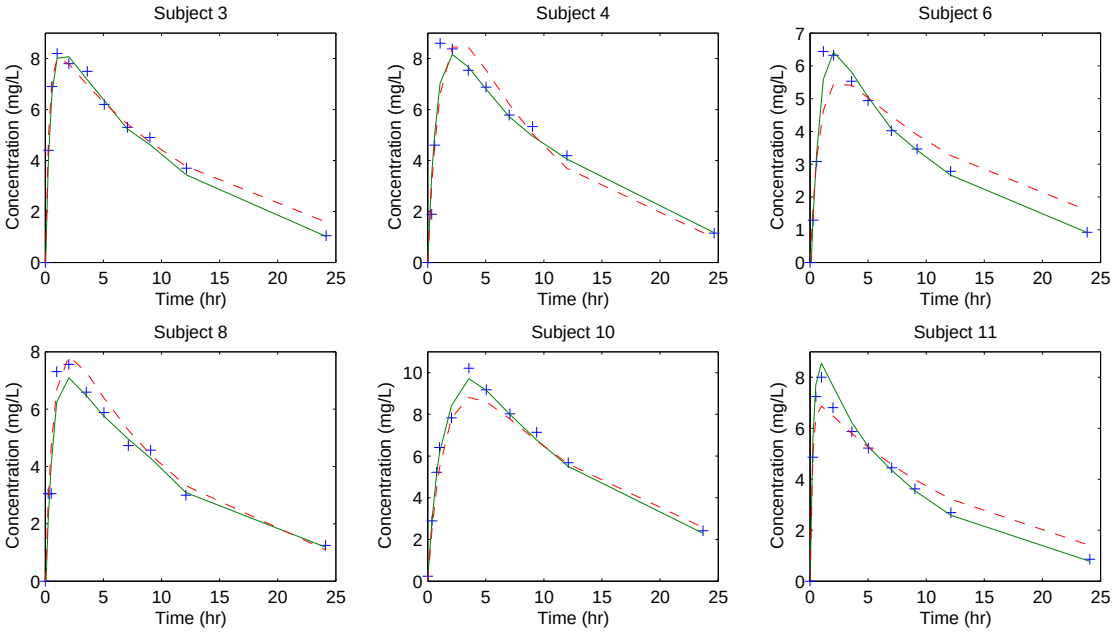

Figure 3: Six individual concentration cruves predictec by SAEM with the ODE approach (dotted line) and the SDE approach (plain line) overlaid on the data points for the pharmacokinetic of Theophyllin

In the ODE approach, the differential equation (3) has an explicit solution. Thus, the parameters estimates are obtained using the SAEM algorithm combined with a MCMC procedure proposed by Kuhn and Lavielle (2004). The individual concentration profiles are predicted by $\widehat{Z}_{ij} = Z(t_{ij}, \hat{\phi}_i)$ for all i and j where Z is the solution of (3) and $\hat{\phi}_i$ is the posterior mean evaluated during the last iterations of the SAEM algorithm. In the SDE approach, the same implementation of the extended SAEM algorithm as the one detailed for the simulation study is used. The individual concentration predictions $E(Z(t_{ij}, \phi_i)|y_i; \hat{\theta})$ for all i and j are evaluated by $\widehat{Z}_{ij} = 1/100 \sum_{k=K-99}^{K} Z^{(k)}(t_{ij}, \phi_i^{(k)})$ where $Z^{(k)}(t_{ij}, \phi_i^{(k)})$ is simulated under the conditional distribution $q_{W,\Phi|Y}(\ .\ |y_i; \widehat{\theta})$ during the 100 last iterations of the extended SAEM algorithm.

The ODE and SDE predictions are overlaid on the data in Figure 3 for six typical subjects. Both ODE and SDE predicted curves for the other six subjects are satisfactory and thus not presented here. For these six subjects, the ODE predicted curves miss some of the observed data, particularly the last one or the initial concentration peak. The SDE predicted curves improve almost all of these individual profiles.

In conclusion, in this real dataset case study, the individual predictions supplied by the SDE model fit the data better than those obtained by the ODE model. Consequently, in this case, the SDE approach has to be preferred to the ODE approach.

7 Discussion

This paper proposes estimation methods for models defined by a discretely observed diffusion process including additive measurement noise and with random drift function parameters. To that end, an approximate model $\mathcal{M}_h$ is introduced, of which the regression term is evaluated using a Gaussian Euler-Maruyama approximation of maximal step size h. A Gibbs

sampler based on the reparameterization of the model suggested by Roberts and Stramer (2001) in the Bayesian framework and the SAEM algorithm in the Maximum Likelihood approach are extended to this model. These two estimation algorithms require the simulation of the missing data (w, ϕ) with the conditional distribution $q_{W,\Phi|Y}$. The choice of the proposal distributions governs the convergence properties of the algorithm and thus is a key issue. A tuned MCMC procedure to perform this simulation is thus proposed, combining a hybrid Gibbs algorithm with independent or random walk Metropolis-Hastings schemes.

Moreover, the error induced by the Euler-Maruyama Gaussian approximation of the diffusion process on the conditional distribution, the likelihood and the posterior distribution of the model $\mathcal{M}_h$ are controlled by the step size h of the numerical scheme. This error is distinct from the error on the estimates induced by the estimation algorithms.

In the maximum likelihood approach, the stochastic version of the EM algorithm SAEM proposed by Kuhn and Lavielle (2004) is preferred to the Monte-Carlo EM (MCEM) developed by Wei and Tanner (1990) or Wu (2004) because of its computational properties. Indeed, SAEM requires the generation of only one realization of the non-observed data at each iteration. In a context where the missing data have to be simulated by a MCMC method, decreasing the size of these missing data is a key issue to ensure acceptable computational times.

The accuracy of the extended SAEM algorithm is illustrated on a pharmacokinetic simulation study using a non-linear mixed effect model defined by SDEs. The relevance of the SDEs approach with respect to the deterministic one is exemplified on a real dataset.

The estimation of such models is mentioned in few papers and is not completely solved. In the general case of non-linear SDE, the only method proposed is exclusively adapted to the particular case of mixed models and for a maximum likelihood approach (Overgaard *et al.*, 2005). Furthermore, this method relies on the linearization of the model and its convergence is not established. In this paper, we propose estimation methods not only for classic but also Bayesian inference that are adapted to more general missing data models. In addition, the convergence of the proposed algorithms is demonstrated.

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* **70**, 223–262. ISSN 0012-9682.

Bally, V. and Talay, D. (1995). The law of the Euler Scheme for Stochastic Differential Equations: I. Convergence Rate of the Density. Tech. Rep. 2675, INRIA.

Bally, V. and Talay, D. (1996). The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density. *Monte Carlo Methods Appl.* **2**, 93–128.

Bennet, J. E., Racine-Poon, A. and Wakefield, J. C. (1996). *MCMC for nonlinear hierarchical models*, pp. 339–358. Chapman & Hall, London.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. (2005). Retrospective exact simulation of diffusion sample paths with applications. *submitted* .

Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*, vol. 69 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Celeux, G. and Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational. Statistics Quaterly* **2**, 73–82.

Chen, H. F., Lei, G. and Gao, A. J. (1988). Convergence and robustness of the robbins-monroe algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27**, 217–231.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19**, 263–284.

Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27**, 94–128.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1–38. With discussion.

Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69**, 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econ. Statist.* **19**, 177–191.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Interdisciplinary Statistics. Chapman & Hall, London.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.* **5**, 225–242.

Jensen, J. and Petersen, N. (1999). Asympotical normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27**, 514–535.

Krishna, R. (2004). *Applications of Pharmacokinetic principles in drug development.* Kluwer Academic/Plenum Publishers, New York.

Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM Probab. Stat.* **8**, 115–131.

Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Statist. Data Anal.* **49**, 1020–1038.

Kusuoka, S. and Stroock, D. (1985). Applications of the malliavin calculus, part II. *J. Fac. Sci. Univ. Tokyo. Sect. IA, Math.* **32**, 1–76.

Kutoyants, T. (1984). *Parameter estimation for stochastic processes.* Helderman Verlag Berlin.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44**, 226–233.

Meyn, S. and Tweedie, R. (1993). *Markov chains and stochastic stability.* Comm. Control. Engrg. Ser. Springer-Verlag London Ltd., London.

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.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.* **22**, 55–71.

Pinheiro, J. and Bates, D. (1995). Approximations to the log-likelihood function in the non-linear mixed-effect models. *J. Comput. Graph. Statist.* **4**, 12–35.

Poulsen, R. (1999). Approximate maximum likelihood estimation of discretely observed diffusion process. *Center for Analytical Finance* **Working paper 29**.

Prakasa Rao, B. (1999). *Statistical Inference for Diffusion Type Processes.* Arnold Publisher.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88**, 603–621.

Schweppe, F. (1965). Evaluation of likelihood function for gaussian signals. *IEEE Trans. Inf. Theory* **11**, 61–70.

Singer, H. (1993). Continuous-time dynamical systems with sampled data, error of measurement and unobserved components. *J. Time Series Anal.* **14**, 527–545.

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: a survey. *Int. Stat. Rev* **72**, 337–354.

Sørensen, M. (2000). Prediction-based estimating functions. *Econom. J.* **3**, 123–147.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701–1762.

Tornøe, C., Overgaard, R., Agersø, H., Nielsen, H., Madsen, H. and Jonsson, E. (2005). Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharm. Res.* **22**, 1247–58.

Wakefield, J., Smith, A., Racine-Poon, A. and Gelfand, A. (1994). Bayesian analysis of linear and non-linear population models using the gibbs sampler. *Appl. Statist* **43**, 201–221.

Wei, G. C. G. and Tanner, M. A. (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika* **77**, 649–652.

Wu, L. (2004). Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *J. Amer. Statist. Assoc.* **99**, 700–709.

A Proof of theorem 3

1. The aim is to bound

$$\left\|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\right\|_{TV} = \int \left|p_{Z,\Phi|Y}(x,\phi|y;\theta) - q_{W,\Phi|Y}(x,\phi|y;\theta)\right| dxd\phi,$$

Using the fact that the conditional distributions $p_{Y|Z}(x;\sigma^2)$ and $q_{Y|W}(x;\sigma^2)$ are equal, the Bayes theorem application provides:

$$\begin{aligned}&\left|p_{Z,\Phi|Y}(x,\phi|y;\theta) - q_{W,\Phi|Y}(x,\phi|y;\theta)\right| \leq \\ &\frac{p_{Y|Z}(x;\sigma^2)\pi(\phi;\beta)}{p_Y(y;\theta)}\left[\left|p_{Z|\Phi}(x|\phi;\gamma^2) - q_{W|\Phi}(x|\phi;\gamma^2)\right| + \frac{q_{W|\Phi}(x|\phi;\gamma^2)}{q_Y(y;\theta)}\left|p_Y(y;\theta) - q_Y(y;\theta)\right|\right]\end{aligned}$$

As a consequence, the total variation distance is bounded by:

$$\begin{aligned}\left\|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\right\|_{TV} \quad &\leq \frac{\int p_{Y|Z}(x;\sigma^2)\pi(\phi;\beta)dxd\phi}{p_Y(y;\theta)}\left[\sup_{x,\phi}\left|p_{Z|\Phi}(x|\phi;\gamma^2) - q_{W|\Phi}(x|\phi;\gamma^2)\right|\right. \\ &\left. + \frac{|p_Y(y;\theta) - q_Y(y;\theta)|}{q_Y(y;\theta)}\sup_{x,\phi} q_{W|\Phi}(x|\phi;\gamma^2)\right] \qquad (5)\end{aligned}$$

   (a) The quantity $\sup_{x,\phi}\left|p_{Z|\Phi}(x|\phi;\gamma^2) - q_{W|\Phi}(x|\phi;\gamma^2)\right|$ is bounded using a result demonstrated by Bally and Talay (1995). This result, based on the Malliavin Calculus, controls the density convergence rate in the case of the Euler-Maruyama scheme.

   - By the assumption (**A3**) and because the volatility function is constant, the Hörmander's condition detailed in Bally and Talay (1995) is verified. Thus, there exists a constant $C(\phi,\gamma^2,t_j - t_{j-1})$ independent of h, x_j and x_{j-1} such that

$$|p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right) - q_{W|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right)| \leq C(\phi,\gamma^2,t_j - t_{j-1})h.$$

     The constant depends on the bounds of the derivatives of the drift function, independent of ϕ under assumption (**A3**). Besides, if γ^2 is contained in $[\gamma_0,\Gamma_0]$, there exists C_1 independent of γ^2 such that, for all $j = 1\cdots J$,

$$|p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right) - q_{W|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right)| \leq C_1 h \qquad (6)$$

   - Under the assumption (**A4**) and using a result of Kusuoka and Stroock (1985) based on the Malliavin calculus (corollary 3.25), for all $j = 1\cdots J$, there exists a constant $C_2(\phi,\gamma^2,t_j - t_{j-1})$ such that

$$p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right) \leq C_2(\phi,\gamma^2,t_j - t_{j-1})$$

     By the same arguments as before, this constant is bounded independently of ϕ and γ^2. Hence, there exists C_2 such that, for all $j = 1\cdots J$,

$$p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right) \leq C_2 \qquad (7)$$

   - In addition, we can write:

$$\begin{aligned}&|q_{W|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right)| \\ &\leq |q_{W|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right) - p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right)| + |p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right)| \\ &\leq hC_1 + C_2. \qquad (8)\end{aligned}$$

- Finally, the Markov property provides:

$$\left|p_{Z|\Phi}(x|\phi;\gamma^2) - q_{W|\Phi}(x|\phi;\gamma^2)\right| = \left|\prod_{j=1}^{J} p_{Z|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right) - \prod_{j=1}^{J} q_{W|\Phi}\left(x_j|x_{j-1},\phi;\gamma^2\right)\right| \tag{9}$$

for any $j = 1 \cdots J$. So, by combining the (9), (6), (7) and (8), there exists a bound C_3 independent of h, j and γ^2 such that

$$\sup_{x,\phi} \left|p_{Z|\Phi}(x|\phi;\gamma^2) - q_{W|\Phi}(x|\phi;\gamma^2)\right| \leq C_3 h \tag{10}$$

(b) By the Markov decomposition of the probability $q_{W|\Phi}(x|\phi;\gamma^2)$, and using the inequality (8), there exists C_4 such that

$$\sup_{x,\phi} q_{W|\Phi}(x|\phi;\gamma^2) \leq C_4 \tag{11}$$

By integration and using the inequality (10), we have:

$$|p_Y(y;\theta) - q_Y(y;\theta)| \leq C_3 h \int p_{Y|Z}(x;\sigma^2)\pi(\phi;\beta)dxd\phi = C_3 h \tag{12}$$

(c) The quantity $q_{W|\Phi}(x|\phi;\gamma^2)$ can be down-bounded. Indeed,

$$\begin{aligned} q_Y(y;\theta) &\geq p_Y(y;\theta) - |p_Y(y;\theta) - q_Y(y;\theta)| \\ &\geq p_Y(y;\theta) - C_3 h \qquad \text{following the inequality (12)} \\ &\geq p_Y(y;\theta) - C_3 H_0 \qquad \text{for } h < H_0 \text{ and } H_0 \text{ small enough.} \end{aligned}$$

Hence there exists $C_5(y)$ such that

$$q_Y(y;\theta) \geq C_5(y) \tag{13}$$

(d) Finally, the inequalities (5) , (11), (12) and (13) provide the final result:

$$\left\|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\right\|_{TV} \leq \frac{1}{p_Y(y;\theta)}\left[C_3 h + \frac{C_3 h}{C_5(y)} C_4\right]$$

□

2. The proof of the part 2 of Theorem 3 directly derives from (12).

3. By Bayes theorem, we have:

$$p_{\theta|Y}(\theta|y) = \frac{p_{y|\theta}(y|\theta)p(\theta)}{p_Y(y)}$$

where $p_Y(y) = \int p_{y|\theta}(y|\theta)p(\theta)d\theta$. From (12), there exists a constant C_3, independent of θ such that $|p_{y|\theta}(y|\theta) - q_{Y|\theta}(y|\theta)| \leq hC_3$. Consequently $|p_Y(y) - q_Y(y)| \leq C_3\ h$ and

$$\begin{aligned} |p_{\theta|Y}(\theta|y) - q_{\theta|Y}(\theta|y)| &\leq \frac{p(\theta)}{p_Y(y)}\left||p_{y|\theta}(y|\theta) - q_{Y|\theta}(y|\theta)| + \frac{q_{Y|\theta}(y)}{q_Y(y)}|q_Y(y) - p_Y(y)|\right| \\ &\leq \frac{C_3 h}{p_Y(y)} p(\theta)\left|1 + \frac{q_{Y|\theta}(y|\theta)}{p_Y(y)}\right| = C_6(y)\ h\left[p(\theta) + q_{\theta|Y}(\theta|y)\right]. \end{aligned}$$

The final result can be directly deduced:

$$\begin{aligned} \left\|q_{\theta|Y} - p_{\theta|Y}\right\|_{TV} &= \int |p_{\theta|Y}(\theta|y) - q_{\theta|Y}(\theta|y)|d\theta \\ &\leq C_6(y)\ h \int (p(\theta) + q_{\theta|Y}(\theta|y))d\theta \leq 2\ C_6(y)\ h \end{aligned}$$

□

Contents

1 Introduction **3**

2 Data and Model **6**
2.1 Incomplete data model defined by SDEs 6
2.2 Diffusion model 6
2.3 Introduction of an approximate statistical model 7
2.3.1 Euler-Maruyama approximation of the diffusion process 7
2.3.2 Approximate statistical model 7

3 Maximum Likelihood Estimation on the model $\mathcal{M}_h$ **8**
3.1 Stochastic versions of the EM algorithm 8
3.2 Extension of the SAEM algorithm to the model $\mathcal{M}_h$ 9
3.3 Convergence of the SAEM algorithm on the model $\mathcal{M}_h$ 10
3.4 Simulation of the non-observed data using a MCMC procedure 11
3.4.1 Proposal distributions 11
3.4.2 Uniform ergodicity of the MCMC procedure 12

4 Bayesian estimation of the model $\mathcal{M}_h$ **12**

5 Survey of the error induced by the Euler-Maruyama approximation **13**

6 Theophyllin pharmacokinetic example **15**
6.1 Pharmacokinetics and Non-linear mixed effects models 15
6.2 Simulation study 16
6.3 A real data example 18

7 Discussion **19**

Bibliography **20**

A Proof of theorem 3 **24**

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399