Likelihood estimation for stochastic differential equations with mixed effects

Fernando Baltazar-Larios
Facultad de Ciencias
Universidad Nacional Autónoma de México
A.P. 20-726, 01000 Mexico, D.F., Mexico
fernandobaltazar@ciencias.unam.mx

Mogens Bladt
Deptartment of Mathematical Sciences
University of Copenhagen
Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
bladt@math.ku.dk

Michael Sørensen
Deptartment of Mathematical Sciences
University of Copenhagen
Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
michael@math.ku.dk

Abstract

Stochastic differential equations provide a powerful tool for modelling dynamic phenomena affected by random noise. In case of repeated observations of time series for several experimental units, it is often the case that some of the parameters vary between the individual experimental units, which has motivated a considerable interest in stochastic differential equations with mixed effects, where a subset of the parameters are random. These models enable simultaneous representation of randomness in the dynamics and variability between experimental units. When the data are observations at discrete time points, the likelihood function is only rarely explicitly available, so for likelihood-based inference to be feasible, numerical methods are needed. We present Gibbs samplers and stochastic EM-algorithms based on augmented data obtained by the simple method for simulation of diffusion bridges in Bladt and Sørensen (2014). This method is easy to implement and has no tuning parameters. The method is, moreover, computationally efficient at low sampling frequencies because the computing time increases linearly with the time between observations. The algorithms can be extended to models with measurement errors. The Gibbs sampler as well as the EM-algorithm are shown to simplify considerably for exponential families of diffusion processes, including many models used in practice. In a simulation study, the estimation methods are shown to work well for Ornstein-Uhlenbeck processes and t-diffusions with mixed effects. Finally, the Gibbs sampler is applied to neuronal data.

1 Introduction

Stochastic differential equations provide a useful and versatile tool for modelling dynamic phenomena affected by random noise, which has been applied in several sciences. The data are time series of observations at discrete time points. In case of repeated observations of time series for several experimental units, for instance observations for different individuals or at different locations, it is often the case that some of the parameters vary between the individual experimental units. This is particularly the case for biological data, but is also common in the social sciences. Various types of mixed models have for a long time been used widely in biostatistics and in econometrics, where they are called models for panel data.

As a statistical tool for such repeated measurements of dynamical data, there has been considerable interest in stochastic differential equations with mixed effects, i.e. where a subset of the parameters are random, with independent values drawn for each experimental unit, while the rest are population parameters that are common to all observations. This approach has several advantages. The random effects model intersubject/unit variations, which may be of interest in themselves, but also implies that several time series can be analysed simultaneously, because the random effects take care of model uncertainty and environmental variation. This can substantially increase the statistical power, and therefore, for instance, provides a way around the well known problem that estimators of parameters in the drift coefficient are often highly imprecise because the time series is too short. These parameters can be well determined by observing a sufficient number of repeated time series, which is often much easier to obtain than longer time series. Stochastic differential equations with mixed effects have been applied to several areas of the biological sciences. Applications to neuro science, electroencephalography, pharmacokinetics and growth of animals and tumors can, for instance, be found in Picchini *et al.* (2008), Wiqvist *et al.* (2021), Ruse *et al.* (2020), Donnet and Samson (2008), Donnet *et al.* (2010), Picchini and Forman (2019) and Jamba *et al.* (2024).

The price for the advantages of using stochastic differential equations with mixed effects is the computational effort needed to obtain maximum likelihood or Bayesian estimators. While the dynamics is formulated in continuous time, the observations are made at discrete points in time, which implies that there is rarely an explicit expression for the likelihood function, even in the case of fixed effects. A further complication for random effects models is that the fixed effects likelihood function must be integrated with respect to the distribution of the random effects. Even when the fixed effects likelihood function is explicit this can rarely be done analytically. Therefore numerical methods or analytical approximations must be used to obtain estimators.

Several approaches have been studied. For high frequency data relatively simple approximations are available. One approach is to use the fixed effects likelihood function (conditional on the random effects) obtained from the Girsanov formula in the hypothetical situation, where the diffusion sample paths have been observed continuously in an interval. Estimators and theory are then developed for continuous time data, while in practice the estimators are approximated by replacing the integrals in the Girsanov formula by Riemann and Ito sums. Some of the papers give bounds on the error made by these approximations in terms of the sampling frequency. This approach was taken by Delattre *et al.* (2013, 2015, 2016) and Ruse *et al.* (2020), who considered exponential families of diffusion processes in the sense of Küchler and Sørensen (1997), where the distribution of the random effects is

the conjugate prior or a finite mixture of such distributions. In this case the likelihood function for continuous time data can be calculated explicitly. A disadvantage of the approach based on the continuous time likelihood function is that it only works when there are neither fixed parameters nor random effects in the diffusion coefficient. This limitation is avoided in another method for high frequency data proposed by Delattre *et al.* (2018a,b), where the conditional fixed effects likelihood function is approximated by the pseudo-likelihood obtained from the Euler approximation to the distribution of the discrete time observations. Like in the previous papers, these authors considered exponential families of diffusion processes with conjugate priors as the distribution of the random effects in order to obtain an explicit pseudo-likelihood. Nonparametric estimation for high frequency data was investigated by Comte *et al.* (2013), Dion (2016) and Dion and Genon-Catalot (2016), while high frequency asymptotics was studied by Maitra and Bhattacharya (2020) and Delattre (2021) and in papers referenced in these papers. A different approach to asymptotics based on local asymptotic normality and L^2-differentiability was taken in Ruse *et al.* (2020).

When the sampling frequency is not sufficiently high, better approximations to the likelihood function are needed. Picchini *et al.* (2010) and Picchini and Ditlevsen (2011) used the Hermite polynomial expansions of the fixed effect likelihood function by Aït-Sahalia (2008) combined with Gaussian quadrature or Laplace's method to approximate the likelihood function numerically. This can be rather time-consuming, so a number of other methods have been proposed. Overgaard *et al.* (2005) obtained an approximation to the likelihood function via the extended Kalman filter, while Jamba *et al.* (2022) used the Delta method (i.e. a second order Taylor expansion) to obtain an approximation to the fixed effect likelihood function that can be integrated analytically with respect to many distributions of the random effects.

More precise approximations can be obtained by thinking of the data as a missing data problem, where the missing data are the random effects for the individual experimental units, the sample paths between the observation times and possibly measurement errors. For the full augmented data set the likelihood function is explicitly available. This suggests application of the EM algorithm or the Gibbs sampler to obtain the maximum likelihood estimator or to simulate from the posterior. Donnet and Samson (2008, 2014) used the stochastic approximation EM algorithm (SAEM) combined with the Metropolis-Hastings algorithm or particle filters with sample paths simulated by the Euler scheme on a fine grid. Delattre and Lavielle (2013) moreover used the extended Kalman filter to reduce the amount of simulation needed at the cost of loosing control of the error resulting from the linearization needed for non-linear diffusion models. The Gibbs sampler was used by Donnet *et al.* (2010), Whitaker *et al.* (2017), Picchini and Forman (2019) and Wiqvist *et al.* (2021) in combination with a generalisation of the approximate diffusion bridge proposed by Durham and Gallant (2002), the pseudo-marginal Metropolis-Hastings algorithm and particle filters or synthetic likelihoods.

In recent years random-effects models based on stochastic differential equations driven by fractional Brownian motion and partial stochastic differential equations have been studied, see El Omari *et al.* (2019), Dai *et al.* (2021), Bishwal (2022), Prakasa Rao (2023) and El Maroufy *et al.* (2024).

In the present paper, we also augment of the data and apply both a stochastic EM algorithm and the Gibbs sampler. We propose to combine these algorithms with the simple method for simulation of diffusion bridges introduced by Bladt and Sørensen (2014); see also

the corrigendum Bladt *et al.* (2021). Advantages over other methods is that it is easy to implement and that there are no tuning parameter that must be set appropriately (except the discretization of the Euler scheme, but this is the case for all methods and is not difficult to tune). Another advantage is that our method is efficient for low frequency observations, as the computing time increase only linearly with the time between observations. To simplify the presentation we consider only one-dimensional diffusions proceses, but our algorithms can be directly generalised to multivariate diffusions the applying the methods for simulating multivariate diffusion bridges in Bladt *et al.* (2016, 2022).

The paper is organized as follows. The model and the data augmentation are presented in Section 2, and the Gibbs sampler and the stochastic EM algorithm are introduced for general models in Section 3. In Section 4 it is outlined how the algorithms can be straightforwardly extended to include the case of measurement errors, and in Section 5 it is investigated how the algorithms simplify when the fixed effects model obtained conditionally on the random effects is an exponential family of diffusion processes, which is the case for many models used in practice. It is investigated how well the new methods work for the Ornstein-Uhlenbeck process and for a t-diffusion in simulation studies in Section 6, and in Section 7 the Gibbs sampler is applied to neuronal data. The simple method for diffusion bridge simulation is briefly explained in Appendix A.

2 Model and augmented data

Consider N diffusion processes

$$dX_t^i = d_{\boldsymbol{\alpha},\boldsymbol{a}^i}(X_t^i)dt + \sigma_{\boldsymbol{\beta},\boldsymbol{b}^i}(X_t^i)dW_t^i, \quad i = 1, \ldots, N \tag{2.1}$$

where W^i, $i = 1, \cdots, N$ are independent standard Wiener processes. The vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are parameters to be estimated, while the vectors $\boldsymbol{a}^i$ and $\boldsymbol{b}^i$ are random effects. The random vectors $(\boldsymbol{a}^i, \boldsymbol{b}^i)$, $i = 1, \ldots, N$, are independent, identically distributed random vectors with density $p_{\boldsymbol{\gamma}}(\boldsymbol{a}, \boldsymbol{b})$ with respect to some dominating measure, and they are independent of the Wiener processes W^i, $i = 1, \ldots, N$. Thus the parameters to be estimated are $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. We assume that $\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x) > 0$ for all x in the state interval and for all values of $(\boldsymbol{\beta}, \boldsymbol{b})$, that $d_{\boldsymbol{\alpha},\boldsymbol{a}}(x)$ is continuously differentiable w.r.t. x for all values of $(\boldsymbol{\alpha}, \boldsymbol{a})$, and that $\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x)$ is twice continuously differentiable w.r.t. x for all $(\boldsymbol{\beta}, \boldsymbol{b})$. Moreover, we assume that for any given value of $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ the stochastic differential equation (2.1) has a unique weak solution, and that the speed measure is finite.

We consider discrete time data $\boldsymbol{X}_{\text{obs}} = (\boldsymbol{X}_{\text{obs}}^1, \ldots, \boldsymbol{X}_{\text{obs}}^N)$, where

$$\boldsymbol{X}_{\text{obs}}^i = \left(x_1^i, \ldots x_{n_i}^i\right),$$

and $x_j^i = X_{t_j^i}^i$, with $t_1^i < t_2^i < ... < t_{n_i}^i$.

There is usually no explicit expression for the likelihood function for the discrete time data $\boldsymbol{X}_{\text{obs}}$ conditional on the random effects $(\boldsymbol{a}^i, \boldsymbol{b}^i)$, $i = 1, \ldots, N$, but the data $\boldsymbol{X}_{\text{obs}}$ are partial observations of an augmented data set consisting of the random effects $(\boldsymbol{a}^i, \boldsymbol{b}^i)$, $i = 1, \ldots, N$ and the complete continuous time observations of the diffusion processes $X^1, \ldots, X^N$, i.e. observations of X^i in the time interval $[t_1^i, t_{n_i}^i]$, $i = 1, \ldots, N$. To augment the data, we must simulate the missing data. In particular, we need to simulate each of the processes X^i

conditionally on $\boldsymbol{X}^i_{\text{obs}}$ and on the random effects. This we can do by simulating independent $(t^i_{j-1}, x^i_{j-1}, t^i_j, x^i_j)$-bridges, $j = 1, \ldots, n_i$, for the diffusion (2.1) with fixed values of $\boldsymbol{a}^i$ and $\boldsymbol{b}^i$. The process obtained by conditioning on $X_{t_2} = x_2$ for a solution of (2.1) in the interval $[t_1, t_2]$ with $X_{t_1} = x_1$ is called a (t_1, x_1, t_2, x_2)-bridge. We propose to simulate the diffusion bridges by the simple method introduced in Bladt and Sørensen (2014) and the corrigendum Bladt *et al.* (2021). The algorithm is briefly described in Appendix A. Advantages of this method are is that it is easy to understand and implement, and that it is efficient at low sampling frequencies, because the computing time increases linearly with $t_2 - t_1$.

If the diffusion coefficient depends on $\boldsymbol{\beta}$ and $\boldsymbol{b}$, then the probability measures corresponding to continuous time observations of the diffusion model given by (2.1) are singular (for different values of the $\boldsymbol{\beta}$ and $\boldsymbol{b}$), so that the likelihood function for this augmented data set does not exist. Therefore, we use the Lamperti transformation

$$h_{\boldsymbol{\beta},\boldsymbol{b}}(x) = \int_{x^*}^{x} \frac{1}{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(y)} dy, \tag{2.2}$$

where x^* is a point in the state-interval. In the following we assume that $\boldsymbol{b}^i$ is fixed (i.e. we argue conditionally on the random vector $\boldsymbol{b}^i$). By Ito's formula, the process $Y^i_t = h_{\boldsymbol{\beta},\boldsymbol{b}^i}(X^i_t)$ solves

$$dY^i_t = \mu_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a}^i,\boldsymbol{b}^i}(Y^i_t)dt + dW^i_t, \tag{2.3}$$

where the drift coefficient is

$$\mu_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(y) = \frac{d_{\boldsymbol{\alpha},\boldsymbol{a}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(y))}{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(y))} - \frac{1}{2}\sigma'_{\boldsymbol{\beta},\boldsymbol{b}}\left(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(y)\right), \tag{2.4}$$

with $\sigma'_{\boldsymbol{\beta},\boldsymbol{b}}(x) = \partial_x \sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x)$. In (2.3) the diffusion coefficient does not depend on the parameters and random effects, which is a necessary condition for the existence of the likelihood function. The basic dominating measure for the likelihood function is the Wiener measure induced on $C([t_1, t_{n_i}])$ by the standard Wiener process. It is an assumption in the rest of the paper that the measures induced on $C([t_1, t_{n_i}])$ by Y^i (for all values of the parameters and the random effects) are dominated by the Wiener measure, and that the Radon-Nikodym derivatives are given by Girsanov's theorem.

The Lamperti transformation (2.2), and hence the transformed data, depends on the random effect $\boldsymbol{b}^i$ and the parameter $\boldsymbol{\beta}$. To reduce the Fisher information in the missing data and thus make the MCMC and EM algorithms more efficient, we follow Roberts and Stramer (2001) and augment the discrete time data in the following way that involves an extra path transformation, see also Beskos *et al.* (2006). Our augmented "full" data set is $(\boldsymbol{X}_{\text{obs}}, \boldsymbol{X}_{\text{mis}})$, where $\boldsymbol{X}_{\text{mis}} = (\boldsymbol{X}^1_{\text{mis}}, \ldots, \boldsymbol{X}^N_{\text{mis}})$ and

$$\boldsymbol{X}^i_{\text{mis}} = \{\boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i\}.$$

Here $\boldsymbol{Y}^{*i} = \{Y^{*ij}_t,\ t \in [t^i_{j-1}, t^i_j],\ j = 2, \ldots, n_i\}$ with

$$Y^{*ij}_t = Z^{ij}_t - \ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(t), \quad t \in [t^i_{j-1}, t^i_j],$$

where

$$\ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(t) = \frac{(t^i_j - t)h_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_{j-1}) + (t - t^i_{j-1})h_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_j)}{t^i_j - t^i_{j-1}}, \quad t \in [t^i_{j-1}, t^i_j], \tag{2.5}$$

$j = 2, \ldots, n_i$, interpolates linearly between the Lamperti transformed discrete time observations, and where, conditionally on $\boldsymbol{X}^i_{\text{obs}}$ and $(\boldsymbol{a}^i, \boldsymbol{b}^i)$, the processes Z^{ij}_t, $j = 2, \ldots, n_i, i = 1, \ldots, N$ are independent $(t^i_{j-1}, h_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_{j-1}), t^i_j,\ h_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_j))$-bridges for the diffusion Y^i given by (2.3) with parameter values $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and $(\boldsymbol{a}^i, \boldsymbol{b}^i)$. The bridge Z^{ij} can be obtain from a $(t^i_{j-1}, x^i_{j-1}, t^i_j, x^i_j)$-bridge, X^{ij}, of the basic diffusion given by (2.1) by the transformation $Z^{ij}_t = h_{\boldsymbol{\beta},\boldsymbol{b}^i}(X^{ij}_t)$. The process Y^{*ij} is not in general a $(t^i_{j-1}, 0, t^i_j, 0)$-bridge. Under the dominating Wiener measure, however, Y^i is a Brownian motion, so Y^{*ij} is a Brownian $(t^i_{j-1}, 0, t^i_j, 0)$-bridge. This is needed in the derivation of the following expression for the likelihood function.

The likelihood function for the augmented data set is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{X}_{\text{obs}}, \boldsymbol{X}_{\text{mis}}) = \prod_{i=1}^{N} L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{X}^i_{\text{mis}}) \prod_{i=1}^{N} p_{\boldsymbol{\gamma}}(\boldsymbol{a}^i, \boldsymbol{b}^i),$$

where

$$\log L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{X}^i_{\text{mis}}) = H_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a}^i,\boldsymbol{b}^i}(x^i_1, x^i_{n_i})$$

$$- \sum_{j=2}^{n_i} \left[\frac{(h_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_j) - h_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_{j-1}))^2}{2(t^i_j - t^i_{j-1})} + \log(\sigma_{\boldsymbol{\beta},\boldsymbol{b}^i}(x^i_j)) + \frac{1}{2} \int_{t^i_{j-1}}^{t^i_j} \phi_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a}^i,\boldsymbol{b}^i}(Y^{*ij}_s + \ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(s)) ds \right].$$

Here $h_{\boldsymbol{\beta},\boldsymbol{b}}$ is given by (2.2),

$$\begin{aligned}
\phi_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(x) &= \mu'_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(x) + \mu_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(x)^2 \\
&= d'_{\boldsymbol{\alpha},\boldsymbol{a}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x)) - 2 d_{\boldsymbol{\alpha},\boldsymbol{a}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x)) \frac{\sigma'_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))}{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))} \\
&\qquad - \tfrac{1}{2} \sigma''_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x)) \sigma_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x)) + \tfrac{1}{4} \left(\sigma'_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x)) \right)^2 + \frac{d^2_{\boldsymbol{\alpha},\boldsymbol{a}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))}{\sigma^2_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))},
\end{aligned}$$

and

$$H_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(x, y) = \int_{h_{\boldsymbol{\beta},\boldsymbol{b}}(x)}^{h_{\boldsymbol{\beta},\boldsymbol{b}}(y)} \mu_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(y) dy = \int_x^y \frac{d_{\boldsymbol{\alpha},\boldsymbol{a}}(y)}{\sigma^2_{\boldsymbol{\beta},\boldsymbol{b}}(y)} dy - \tfrac{1}{2} \log \left(\frac{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(y)}{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x)} \right). \tag{2.6}$$

The expression for $L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{X}^i_{\text{mis}})$ follows from Girsanov's theorem and Ito's formula by arguments in Roberts and Stramer (2001).

3 MCMC and EM algorithms

In order to use the likelihood function for the augmented data to find the maximum likelihood estimator, or an analogous Bayesian estimator, for the discrete time data, we can apply the EM-algoritm or the Gibbs sampler. In this section we present a Gibbs sampler and a stochastic EM-algorithm. To clarify the general structure of the algorithms, they are presented for the general model defined above. In section 5 we consider a particularly tractable sub-class of models for which the algorithms simplify considerably.

3.1 The Gibbs sampler

For MCMC-estimation, we first specify a suitable prior $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Then we can simulate from the posterior distribution of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ by means of the following Gibbs sampler with $2N + 1$ sites. The sites are $\boldsymbol{\theta}$ plus $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ and $\boldsymbol{Y}^{*i}$ for $i = 1, \ldots, N$. As we shall see later, it can be preferable to have more sites, for instance, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{a}^i$ and $\boldsymbol{b}^i$, $i = 1, \ldots, N$. As a bi-product we also obtain a sample of values of the random effects conditional on the observations.

To start the sampler, draw $\boldsymbol{\theta}$ from the prior π, and given the value of $\boldsymbol{\gamma}$, draw $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ from the distribution with density function $p_{\boldsymbol{\gamma}}$, independently for $i = 1, \ldots, N$. Finally, simulate independent sample paths Y^{*ij} conditionally on $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}^i, \boldsymbol{b}^i$ and $\boldsymbol{X}^i_{\text{obs}}$ for $j = 2, \ldots, n_i,\ i = 1, \ldots, N$. Then repeat the following algorithm.

1. Draw $\boldsymbol{\theta}$ conditionally on $(\boldsymbol{X}_{\text{mis}}, \boldsymbol{X}_{\text{obs}})$
2. For $i = 1, \ldots, N$, draw independent values of $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ conditionally on $(\boldsymbol{\theta}, \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i})$
3. Simulate independent sample paths Y^{*ij} conditionally on $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{a}^i, \boldsymbol{b}^i, \boldsymbol{X}^i_{\text{obs}})$ for $j = 2, \ldots, n_i,\ i = 1, \ldots, N$
4. GO TO 1

In step 1 and 2 the conditional densities of $\boldsymbol{\theta}$ and $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ are proportional to $\pi(\boldsymbol{\theta}) \cdot L(\boldsymbol{\theta}; \boldsymbol{X}_{\text{obs}}, \boldsymbol{X}_{\text{mis}})$ and $L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i) p_{\boldsymbol{\gamma}}(\boldsymbol{a}^i, \boldsymbol{b}^i)$, respectively. In Section 5 we shall se that these steps simplify for an important class of diffusion models, but for more complicated models, it is usually necessary to use Metropolis within Gibbs.

To simulate Y^{*ij} in step 3, we must simulate a $(t^i_{j-1}, h_{\boldsymbol{\beta}, \boldsymbol{b}_i}(x^i_{j-1}), t^i_j, h_{\boldsymbol{\beta}, \boldsymbol{b}_i}(x^i_j))$-bridge for the diffusion Y^i given by (2.3) (or equivalently a $(t^i_{j-1}, x^i_{j-1}, t^i_j, x^i_j)$-bridge of the diffusion given by (2.1)). We do this using the method by Bladt and Sørensen (2014), see also Bladt *et al.* (2021), which is briefly presented in the appendix. The main reasons for this choice are easy implementation and the fact that the computing time for this method is linear in the length of the interval $[t^i_{j-1}, t^i_j]$ and hence works for all sampling frequencies. The simplest method is to use the approximate diffusion bridges obtain by a simple rejection sampler, which usually has a high acceptance rate. In our experience this approximation is sufficiently accurate to obtain good estimates. Exact diffusion bridges can be obtained by Metropolis within Gibbs, as explained in the appendix. In the mth iteration we simulate, as a proposal, an approximate $(t^i_{j-1}, x^i_{j-1}, t^i_j, x^i_j)$-bridge, $X^{m,ij}$, of the diffusion (2.1)) supplemented by its associated geometric variable $S_{m,ij}$. The proposed bridge is accepted with probability $\min\{1, S_{m,ij}/S_{m-1,ij}\}$. Otherwise we keep the bridge used in iteration $m - 1$.

3.2 The EM algorithm

An alternative to MCMC-estimation is the EM-algorithm, which under weak conditions converges to a (possibly local) maximum of the likelihood function for the discrete time data, see e.g. McLachlan and Krishnan (1997).

The EM-algorithm works as follows.

0. Choose initial values $\hat{\boldsymbol{\alpha}}_0, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\gamma}}_0$, $k := 0$

1. (E–step) Calculate the function

$$Q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\gamma}}_k} \left[\log L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{X}^i_{\text{mis}}) + \log p_{\boldsymbol{\gamma}}(\boldsymbol{a}^i, \boldsymbol{b}^i) \,\middle|\, \boldsymbol{X}^i_{\text{obs}} \right].$$

2. (M–step) $(\hat{\boldsymbol{\alpha}}_{k+1}, \hat{\boldsymbol{\beta}}_{k+1}, \hat{\boldsymbol{\gamma}}_{k+1}) = \text{argmax}_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}} Q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

3. k:=k+1; GO TO 1.

In the E–step it is almost always impossible to calculate the conditional expectations of the log-likelihood functions explicitly. Therefore we calculate the ith conditional expectation numerically by generating MC-samples $\boldsymbol{X}^{m,i}_{\text{mis}} = \{Y^{*m,i}, \boldsymbol{a}^{m,i}, \boldsymbol{b}^{m,i}\}$, $m = 1, \ldots, M$ (conditionally on the observed data $\boldsymbol{X}^i_{\text{obs}}$) such that we can approximate $Q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ by $\hat{Q}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{N} \hat{Q}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, where

$$\hat{Q}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) := \frac{1}{M} \sum_{m=1}^{M} \left\{ \log L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*m,i}, \boldsymbol{a}^{m,i}, \boldsymbol{b}^{m,i}) + \log p_{\boldsymbol{\gamma}}(\boldsymbol{a}^{m,i}, \boldsymbol{b}^{m,i}) \right\}.$$

This can be done by means of the following simplified version of the previous Gibbs sampler.

0. Draw $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ from $p_{\hat{\boldsymbol{\gamma}}_k}$

1. Simulate independent sample paths Y^{*ij}, $j = 2, \ldots, n_i$, conditionally on $\boldsymbol{a}^i, \boldsymbol{b}^i, \boldsymbol{X}^i_{\text{obs}}$ (with the parameter values $\hat{\boldsymbol{\alpha}}_k$ and $\hat{\boldsymbol{\beta}}_k$)

2. Draw $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ conditionally on $\boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i}$ (with the parameter values $\hat{\boldsymbol{\alpha}}_k$, $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\gamma}}_k$)

3. GO TO 1

In step 1 of the Gibbs sampler, we use again the simple method by Bladt and Sørensen (2014), see also Bladt *et al.* (2021), to simulate diffusion bridges, cf. the appendix. Again there is a choice between using approximate bridges or using Metropolis within Gibbs to obtain exact diffusion bridges, as explained previously. In step 2, the conditional density of $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ is proportional to $L_i(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i) p_{\hat{\boldsymbol{\gamma}}_k}(\boldsymbol{a}^i, \boldsymbol{b}^i)$. As for the previous Gibbs sampler, it is often necessary to use Metropolis within Gibbs for complicated models.

When the conditional expectation of the log-likelihood function for the full data set is calculated by Monte Carlo methods, as we do, the algorithm is often referred to Monte Carlo EM (MCEM), see Wei and Tanner (1990). In our experience this works well without computational problems. If desired, the MCEM algorithm can be replaced by a stochastic EM algorithm, see e.g. Diebolt and Ip (1996), or a stochastic approximation EM (SAEM), see Delyon *et al.* (1999), both of which are easily implemented using the simple method for diffusion bridge simulation.

4 Measurement errors

In several applications, e.g. pharmacokinetics, measurement errors are expected. In this section we will outline how it is straightforward to extend our methods to include measurement errors.

Assume that the observations are $\boldsymbol{U}_{\text{obs}} = (\boldsymbol{U}^1_{\text{obs}}, \ldots, \boldsymbol{U}^N_{\text{obs}})$, where $\boldsymbol{U}^i_{\text{obs}} = \left(u^i_1, \ldots u^i_{n_i}\right)$, and

$$u^i_j = X^i_{t^i_j} + \varepsilon^i_j,$$

with $t^i_1 < t^i_2 < \ldots < t^i_{n_i}$. Here the process X^i_t is defined as in Section 2, and the random variables $\varepsilon^i_j \sim N(0, \tau^2)$ are mutually independent and independent of $X^i, \boldsymbol{a}^i, \boldsymbol{b}^i$, $i = 1, \ldots, N$. We consider the case of additive Gaussian measurement errors, but the general case where $u^i_j = k(X^i_{t^i_j}, \varepsilon^i_j)$ for some function k with the variables ε^i_j non-Gaussian, can be treated in exactly the same way – apart from a more complicated notation.

Here the unobserved data include the measurement errors, i.e. the augmented data set is $(\boldsymbol{U}_{\text{obs}}, \boldsymbol{U}_{\text{mis}})$ with $\boldsymbol{U}_{\text{mis}} = (\boldsymbol{U}^1_{\text{mis}}, \ldots, \boldsymbol{U}^N_{\text{mis}})$ and $\boldsymbol{X}^i_{\text{mis}} = \{\boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i, \boldsymbol{\varepsilon}^i\}$. Here $\boldsymbol{\varepsilon}^i = (\varepsilon^i_1, \ldots, \varepsilon^i_{n_i})$, $Y^{*ij}_t = \tilde{Z}^{ij}_t - \tilde{\ell}^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i,\boldsymbol{\varepsilon}^i}(t)$, $t \in [t^i_{j-1}, t^i_j]$ and

$$\tilde{\ell}^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i,\boldsymbol{\varepsilon}^i}(t) = \frac{(t^i_j - t)h_{\boldsymbol{\beta},\boldsymbol{b}^i}(u^i_{j-1} - \varepsilon^i_{j-1}) + (t - t^i_{j-1})h_{\boldsymbol{\beta},\boldsymbol{b}^i}(u^i_j - \varepsilon^i_j)}{t^i_j - t^i_{j-1}}, \quad t \in [t^i_{j-1}, t^i_j],$$

$j = 2, \ldots, n_i$. Moreover, $\tilde{Z}^{ij}_t = h_{\boldsymbol{\beta},\boldsymbol{b}^i}(X^{ij}_t)$, where X^{ij}, $j = 2, \ldots, n_i, i = 1, \ldots, N$ are independent $(t^i_{j-1}, u^i_{j-1} - \varepsilon^i_{j-1}, t^i_j, u^i_j - \varepsilon^i_j)$-bridges of the basic diffusion given by (2.1) conditionally on $\boldsymbol{U}^i_{\text{obs}}$, $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ and $\boldsymbol{\varepsilon}^i$. The likelihood function for the augmented data set is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau^2; \boldsymbol{U}_{\text{obs}}, \boldsymbol{U}_{\text{mis}}) = \prod_{i=1}^{N} \left[\tilde{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{U}^i_{\text{obs}} - \boldsymbol{\varepsilon}^i, \boldsymbol{X}^i_{\text{mis}}) p_{\boldsymbol{\gamma}}(\boldsymbol{a}^i, \boldsymbol{b}^i) \prod_{j=1}^{n_i} \varphi_{\tau^2}(\varepsilon^i_j) \right],$$

where φ_{τ^2} is the Gaussian density function with mean zero and variance τ^2 and $\tilde{L}_i$ is defined as L_i in Section 2 except that $\ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}$ is replaced by $\tilde{\ell}^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i,\boldsymbol{\varepsilon}^i}$.

Apart from the modified likelihood function and obvious minor changes, the only two changes to the Gibbs sampler are that in step 1 also τ^2 must be drawn, and that in each iteration an extra step must be added prior to step 3. In the extra step, $\boldsymbol{\varepsilon}^i$ is simulated conditionally on $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau^2, \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i)$ for $i = 1, \ldots, N$. For the EM-algorithm similar modifications are needed. In particular, $\boldsymbol{\varepsilon}^i$ must be added as an extra site to the Gibbs sampler used to create the MC-samples.

The algorithms could also have been formulated in terms of simulation of generalized diffusion bridges, i.e. diffusion bridges where the two endpoints are random, in this case $u^i_{j-1} - \varepsilon^i_{j-1}$ and $u^i_j - \varepsilon^i_j$.

5 Exponential family models

Considerable simplifications of the Gibb's sampler and the EM-algorithm can be obtained, when the drift is linear in the parameter $\boldsymbol{\alpha} \in \mathbb{R}^{p_1}$ and in the random effect $\boldsymbol{a} \in \mathbb{R}^{p_2}$, and the random effects $\boldsymbol{a}$ and $\boldsymbol{b}$ are independent with densities $p_{\boldsymbol{\gamma}_1}$ and $p_{\boldsymbol{\gamma}_2}$, respectively. Specifically,

we will in this section investigate stochastic differential equations with mixed effects, where the drift has the form

$$d_{\boldsymbol{\alpha},\boldsymbol{a}}(x) = \sum_{k=1}^{p_1} \alpha_k f_k(x) + \sum_{k=1}^{p_2} a_k g_k(x) = \boldsymbol{\alpha}\boldsymbol{f}(x)^\top + \boldsymbol{a}\boldsymbol{g}(x)^\top, \tag{5.1}$$

where $\boldsymbol{f} = (f_1, \ldots, f_{p_1})$ and $\boldsymbol{g} = (g_1, \ldots, g_{p_2})$. In this paper vectors are row vectors, and $\top$ denotes transposition. A large part of the diffusion models used in practice have this form.

When the drift is of the form (5.1), the continuous time full model is an exponential family of processes (in the sense of Küchler and Sørensen (1997)) in the parameters and random effects in the drift (conditionally on these random effects). More specifically, we find that

$$H_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(x,y) = \boldsymbol{\alpha}\int_x^y \frac{\boldsymbol{f}(y)^\top}{\sigma^2_{\boldsymbol{\beta},\boldsymbol{b}}(y)}dy + \boldsymbol{a}\int_x^y \frac{\boldsymbol{g}(y)^\top}{\sigma^2_{\boldsymbol{\beta},\boldsymbol{b}}(y)}dy - \tfrac{1}{2}\log\frac{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(y)}{\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x)}, \tag{5.2}$$

cf. (2.6), and

$$\begin{aligned}
\phi_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{b}}(x) &= \boldsymbol{\alpha}\bar{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))^\top + \boldsymbol{a}\bar{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))^\top + 2\boldsymbol{\alpha}\tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))^\top\tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))\boldsymbol{a}^\top \\
&\quad + \boldsymbol{\alpha}\tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))^\top\tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))\boldsymbol{\alpha}^\top + \boldsymbol{a}\tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))^\top\tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))\boldsymbol{a}^\top \\
&\quad + \tfrac{1}{4}\left(\sigma'_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))\right)^2 - \tfrac{1}{2}\sigma''_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x))\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}}(x)),
\end{aligned}$$

where $h_{\boldsymbol{\beta},\boldsymbol{b}}$ is given by (2.2), and the kth coordinate of the functions $\bar{\boldsymbol{f}}, \bar{\boldsymbol{g}}, \tilde{\boldsymbol{f}}$ and $\tilde{\boldsymbol{g}}$ are given by

$$\begin{aligned}
(\bar{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}})_k(x) &= f'_k(x) - 2f_k(x)(\log(\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x))' \\
(\bar{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}})_k(x) &= g'_k(x) - 2g_k(x)(\log(\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x))' \\
(\tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}})_k(x) &= f_k(x)/\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x) \\
(\tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}})_k(x) &= g_k(x)/\sigma_{\boldsymbol{\beta},\boldsymbol{b}}(x).
\end{aligned}$$

If we assume that the distribution of the random effects in the drift is Gaussian

$$\boldsymbol{a}^i \sim N_{p_2}(\boldsymbol{\xi}, \boldsymbol{\Gamma}^{-1}), \tag{5.3}$$

then the likelihood function for the augmented data has the form

$$L(\boldsymbol{\theta}; \boldsymbol{X}_{\text{obs}}, \boldsymbol{X}_{\text{mis}}) = \exp\left(\boldsymbol{\alpha}\boldsymbol{v}^\top_{\boldsymbol{\beta},\underline{\boldsymbol{a}},\underline{\boldsymbol{b}}} - \tfrac{1}{2}\boldsymbol{\alpha}\boldsymbol{D}_{\boldsymbol{\beta},\underline{\boldsymbol{b}}}\boldsymbol{\alpha}^\top + q(\boldsymbol{\beta},\underline{\boldsymbol{a}},\underline{\boldsymbol{b}})\right)\prod_{i=1}^N \left[p_{\boldsymbol{\gamma}_1}(\boldsymbol{a}^i)p_{\boldsymbol{\gamma}_2}(\boldsymbol{b}^i)\right] \tag{5.4}$$

$$= C(\boldsymbol{\theta},\underline{\boldsymbol{b}})\prod_{i=1}^N \exp\left(\boldsymbol{a}^i\left[(\boldsymbol{t}^i_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{b}^i})^\top + \boldsymbol{\Gamma}\boldsymbol{\xi}^\top\right] - \tfrac{1}{2}\boldsymbol{a}^i\left(\boldsymbol{B}^i_{\boldsymbol{\beta},\boldsymbol{b}^i} + \boldsymbol{\Gamma}\right)(\boldsymbol{a}^i)^\top\right). \tag{5.5}$$

Here we have used the notation $\underline{\boldsymbol{a}} = (\boldsymbol{a}^1, \ldots, \boldsymbol{a}^N)$ and $\underline{\boldsymbol{b}} = (\boldsymbol{b}^1, \ldots, \boldsymbol{b}^N)$. The likelihood function is Gaussian as a function of $\boldsymbol{\alpha}$ as well as of $\boldsymbol{a}^i$. The vectors and matrices in (5.4)

and (5.5) are defined as follows:

$$\begin{aligned}
\boldsymbol{t}^i_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{b}^i} &= \int_{x_1^i}^{x^i_{n_i}} \frac{\boldsymbol{g}(y)}{\sigma^2_{\boldsymbol{\beta},\boldsymbol{b}^i}(y)} dy - \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \boldsymbol{k}_{1i}(Y_s^{*ij} + \ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(s))ds \qquad (5.6)\\
\boldsymbol{v}_{\boldsymbol{\beta},\underline{\boldsymbol{a}},\underline{\boldsymbol{b}}} &= \sum_{i=1}^{N} \left\{ \int_{x_1^i}^{x^i_{n_i}} \frac{\boldsymbol{f}(y)}{\sigma^2_{\boldsymbol{\beta},\boldsymbol{b}^i}(y)} dy - \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \boldsymbol{k}_{2i}(Y_s^{*ij} + \ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(s))ds \right\} \qquad (5.7)\\
\boldsymbol{B}^i_{\boldsymbol{\beta},\boldsymbol{b}^i} &= \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \boldsymbol{K}_{3i}(Y_s^{*ij} + \ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(s))ds \qquad (5.8)\\
\boldsymbol{D}_{\boldsymbol{\beta},\underline{\boldsymbol{b}}} &= \sum_{i=1}^{N} \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \boldsymbol{K}_{4i}(Y_s^{*ij} + \ell^{ij}_{\boldsymbol{\beta},\boldsymbol{b}^i}(s))ds \qquad (5.9)
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{k}_{1i}(y) &= \tfrac{1}{2}\bar{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y)) + \boldsymbol{\alpha}\tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))^\top \tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))\\
\boldsymbol{k}_{2i}(y) &= \tfrac{1}{2}\bar{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y)) + \boldsymbol{a}^i\tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))^\top \tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))\\
\boldsymbol{K}_{3i}(y) &= \tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))^\top \tilde{\boldsymbol{g}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))\\
\boldsymbol{K}_{4i}(y) &= \tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))^\top \tilde{\boldsymbol{f}}_{\boldsymbol{\beta},\boldsymbol{b}^i}(h^{-1}_{\boldsymbol{\beta},\boldsymbol{b}^i}(y))
\end{aligned}$$

The real functions $q(\boldsymbol{\beta},\underline{\boldsymbol{a}},\underline{\boldsymbol{b}})$ and $C(\boldsymbol{\theta},\underline{\boldsymbol{b}})$ can be determined from the expressions above (and the multivariate normal density function).

5.1 The Gibbs sampler

For the Gibbs sampler the simplification is obtained if a conjugate prior is used for $\boldsymbol{\alpha}$ and for $\boldsymbol{\gamma}_1$. Specifically, we use the priors

$$\begin{aligned}
\boldsymbol{\alpha} &\sim N_{p_1}(\bar{\boldsymbol{\alpha}}, \boldsymbol{\Sigma})\\
\boldsymbol{\gamma}_1 &= (\boldsymbol{\xi}, \boldsymbol{\Gamma}) \sim NW_{p_2}(\boldsymbol{\xi}_0, \lambda, \boldsymbol{V}, \nu)\\
\boldsymbol{\beta} &\sim \pi_1 \quad \boldsymbol{\gamma}_2 \sim \pi_2
\end{aligned}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are independent. By $NW_{p_2}(\boldsymbol{\xi}_0, \lambda, \boldsymbol{V}, \nu)$ we denote the p_2-dimensional normal-Wishart distribution with parameters $\boldsymbol{\xi}_0 \in \mathbb{R}^{p_2}, \lambda > 0, \boldsymbol{V}, \nu > p_2 - 1$, where $\boldsymbol{V}$ is a positive definite $p_2 \times p_2-$matrix. The parameter $\boldsymbol{\gamma}_1 = (\boldsymbol{\xi}, \boldsymbol{\Gamma})$ can be simulated by first simulating $\boldsymbol{\Gamma}$ from a Wishard distribution with parameters $(\boldsymbol{V}, \nu)$, and then (conditionally on $\boldsymbol{\Gamma}$) simulating $\boldsymbol{\xi}$ from a multivariate normal distribution with mean $\boldsymbol{\xi}_0$ and covariance matrix $(\lambda\boldsymbol{\Gamma})^{-1}$.

If some of the parameters or random effects must necessarily be positive (for instance to ensure ergodicity), then these coordinates of the multivariate normal distributions must be restricted to the positive half-line. The same is then the case of the corresponding normal distributions in the following algorithm.

With these priors, it follows easily from (5.4) and (5.5) (and well-known results for the normal-Wishart prior) that the Gibbs sampler goes as follows.

0. To start the sampler, draw $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ from the prior distribution, and given the value of $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, draw $(\boldsymbol{a}^i, \boldsymbol{b}^i)$ from the distribution with parameter $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, independently for $i = 1, \ldots, N$. Complete the initialization by simulating independent sample paths Y^{*ij} conditionally on $\boldsymbol{\theta}, \boldsymbol{a}^i, \boldsymbol{b}^i$ and $\boldsymbol{X}^i_{\text{obs}}$ for $j = 2, \ldots, n_i,\ i = 1, \ldots, N$.

1. Draw $\boldsymbol{\alpha}$ from the $N_{p_1}((\boldsymbol{v}_{\boldsymbol{\beta},\underline{\boldsymbol{a}},\underline{\boldsymbol{b}}} + \bar{\boldsymbol{\alpha}}\boldsymbol{\Sigma}^{-1})(\boldsymbol{D}_{\boldsymbol{\beta},\underline{\boldsymbol{b}}} + \boldsymbol{\Sigma}^{-1})^{-1}, (\boldsymbol{D}_{\boldsymbol{\beta},\underline{\boldsymbol{b}}} + \boldsymbol{\Sigma}^{-1})^{-1})$ distribution (conditionally on $\boldsymbol{\beta}, \underline{\boldsymbol{a}}, \underline{\boldsymbol{b}}, \boldsymbol{X}_{\text{obs}}$ and $\boldsymbol{Y}^{*i}$, $i = 1, \ldots, N$)

2. Draw $\boldsymbol{\beta}$ from the distribution with density function proportional to

$$\pi_1(\boldsymbol{\beta}) \prod_{i=1}^{N} L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i)$$

   (all quantities other than $\boldsymbol{\beta}$ are fixed)

3. Draw $\boldsymbol{\gamma}_1 = (\boldsymbol{\xi}, \boldsymbol{\Gamma})$ (conditionally on $\boldsymbol{a}^1, \ldots, \boldsymbol{a}^N$) from the $NW_{p_2}(\lambda\boldsymbol{\xi}_0 + N\bar{\boldsymbol{a}})/(\lambda+N), \lambda+N, (\boldsymbol{V}^{-1} + N\hat{\boldsymbol{\Gamma}}^{-1} + (\bar{\boldsymbol{a}} - \boldsymbol{\xi}_0)(\bar{\boldsymbol{a}} - \boldsymbol{\xi}_0)^\top \lambda N/(\lambda+N))^{-1}, \nu+N)$ distribution, where $\bar{\boldsymbol{a}}$ and $\hat{\boldsymbol{\Gamma}}^{-1}$ are the mean and the sample covariance matrix of $\boldsymbol{a}^1, \ldots, \boldsymbol{a}^N$

4. Draw $\boldsymbol{\gamma}_2$ from the distribution with density function proportional to $\pi_2(\boldsymbol{\gamma}_2) \prod_{i=1}^N p_{\boldsymbol{\gamma}_2}(\boldsymbol{b}^i)$ (with $\boldsymbol{b}^1, \ldots, \boldsymbol{b}^N$ fixed)

5. Draw independent values of $\boldsymbol{a}^i$ from the $N_{p_2}((\boldsymbol{t}^i_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{b}^i} + \boldsymbol{\xi}\boldsymbol{\Gamma})(\boldsymbol{B}^i_{\boldsymbol{\beta},\boldsymbol{b}^i} + \boldsymbol{\Gamma})^{-1}, (\boldsymbol{B}^i_{\boldsymbol{\beta},\boldsymbol{b}^i} + \boldsymbol{\Gamma})^{-1})$ distribution, $i = 1, \ldots, N$ (conditionally on $\boldsymbol{\theta}, \boldsymbol{b}^i, \boldsymbol{X}^i_{\text{obs}}$ and $\boldsymbol{Y}^{*i}$)

6. Draw $\boldsymbol{b}^i$ from the distribution with density function proportional to

$$L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}^i_{\text{obs}}, \boldsymbol{Y}^{*i}, \boldsymbol{a}^i, \boldsymbol{b}^i) p_{\boldsymbol{\gamma}_2}(\boldsymbol{b}^i)$$

   (where all quantities other than $\boldsymbol{b}^i$ are fixed), independently for $i = 1, \ldots, N$

7. Simulate independent sample paths Y^{*ij} conditionally on $\boldsymbol{\theta}, \boldsymbol{a}^i, \boldsymbol{b}^i$ and $\boldsymbol{X}^i_{\text{obs}}$ for $j = 2, \ldots, n_i,\ i = 1, \ldots, N$

8. GO TO 1

In (5.3) we allowed all kinds of dependencies between the random effects in the drift. It might in some cases be reasonable to assume some structure in the matrix Γ, for instance that it is a diagonal matrix. This changes the posterior of $\boldsymbol{\gamma}_1$ in step 3. If we assume that the coordinates of $\boldsymbol{a}^i$ are independent, then the posterior is independent normal-gamma distributions.

If some of the parameters or random effects must necessarily be positive, the normal distributions, restricted to the positive half-line, can be replaced by exponential distributions independent of the other coordinates. For instance, if a^i is one-dimensional ($p_2 = 1$) and must be positive, then we can assume that a^i is exponential distributed with mean γ_1^{-1}, and that the prior of γ_1 is the $\Gamma(\nu, \lambda)$-distribution (with density proportional to $\eta^{\nu-1}e^{-\lambda\eta}$). Then step 5 is replaced by

5*. For $i = 1, \ldots, N$, draw independent values of a^i from the normal distribution with mean $(t^i_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{b}^i} - \gamma_1)/B^i_{\boldsymbol{\beta},\boldsymbol{b}^i}$ and variance $(B^i_{\boldsymbol{\beta},\boldsymbol{b}^i})^{-1}$ restricted to the positive half-axis (conditionally on $\boldsymbol{\theta}, \boldsymbol{b}^i, \boldsymbol{X}^i_{\text{obs}}$ and $\boldsymbol{Y}^{*i}$),

while step 3 is replaced by

3*. Draw γ_1 from the $\Gamma(\nu + N, \lambda + a^1 + \ldots + a^N)$-distribution (conditionally on $a^1, \ldots, a^N$).

A considerable simplification of step 2 can be obtained when

$$\sigma_{\beta,\boldsymbol{b}}(x) = \beta c_{\boldsymbol{b}}(x), \tag{5.10}$$

where $\beta > 0$ and $c_{\boldsymbol{b}}(x) > 0$. Then

$$H_{\boldsymbol{\alpha},\beta,\boldsymbol{a},\boldsymbol{b}}(x, y) = \beta^{-2} \int_x^y \frac{\boldsymbol{\alpha}\boldsymbol{f}(z)^\top + \boldsymbol{a}\boldsymbol{g}(z)^\top}{c^2_{\boldsymbol{b}}(z)} dz - \tfrac{1}{2} \log \frac{c_{\boldsymbol{b}}(y)}{c_{\boldsymbol{b}}(x)},$$

where the β in the last term in (5.2) has been omitted as it cancels in the likelihood function. Further,

$$\phi_{\boldsymbol{\alpha},\beta,\boldsymbol{a},\boldsymbol{b}}(x) = m^{(1)}_{\boldsymbol{\alpha},1,\boldsymbol{a},\boldsymbol{b}}(\beta x) + \beta^2 m^{(2)}_{1,\boldsymbol{b}}(\beta x) + \beta^{-2} m^{(3)}_{\boldsymbol{\alpha},1,\boldsymbol{a},\boldsymbol{b}}(\beta x)$$

where

$$\begin{aligned}
m^{(1)}_{\boldsymbol{\alpha},\beta,\boldsymbol{a},\boldsymbol{b}}(x) &= \boldsymbol{\alpha}\bar{\boldsymbol{f}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))^\top + \boldsymbol{a}\bar{\boldsymbol{g}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))^\top \\
m^{(2)}_{\beta,\boldsymbol{b}}(x) &= \tfrac{1}{4}(\sigma'_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x)))^2 - \tfrac{1}{2}\sigma''_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))\sigma_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x)) \\
m^{(3)}_{\boldsymbol{\alpha},\beta,\boldsymbol{a},\boldsymbol{b}}(x) &= 2\boldsymbol{\alpha}\tilde{\boldsymbol{f}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))^\top \tilde{\boldsymbol{g}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))\boldsymbol{a}^\top + \boldsymbol{\alpha}\tilde{\boldsymbol{f}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))^\top \tilde{\boldsymbol{f}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))\boldsymbol{\alpha}^\top \\
&\qquad + \boldsymbol{a}\tilde{\boldsymbol{g}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))^\top \tilde{\boldsymbol{g}}_{\beta,\boldsymbol{b}}(h^{-1}_{\beta,\boldsymbol{b}}(x))\boldsymbol{a}^\top
\end{aligned} \tag{5.11}$$

Thus, in step 2 of the Gibbs sampler, β must be drawn from a distribution with density proportional to

$$\pi_1(\beta)\beta^{-(n_\cdot - N)} \exp\left(-\beta^{-2}(G_1 + G_2) + F(\beta)\right),$$

where $n_\cdot = n_1 + \cdots + n_N$ and

$$G_1 = \sum_{i=1}^N \sum_{j=2}^{n_i} \frac{(h_{1,\boldsymbol{b}^i}(x^i_j) - h_{1,\boldsymbol{b}^i}(x^i_{j-1}))^2}{2(t^i_j - t^i_{j-1})} \qquad G_2 = -\sum_{i=1}^N \int_{x^i_1}^{x^i_{n_i}} \frac{\boldsymbol{\alpha}\boldsymbol{f}(z)^\top + \boldsymbol{a}^i\boldsymbol{g}(z)^\top}{c^2_{\boldsymbol{b}^i}(z)} dz$$

and

$$F(\beta) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \phi_{\boldsymbol{\alpha},\beta,\boldsymbol{a}^i,\boldsymbol{b}^i}(Y^{*ij}_s + \beta^{-1}\ell^{ij}_{1,\boldsymbol{b}^i}(s))ds,$$

where the function F in general depends on β in a complicated way.

If we introduce the parameter $\eta = \beta^{-2}$ and as the prior of η choose the $\Gamma(\kappa, \delta)$-distribution, then in step 2, η should be drawn from the weighted gamma-distribution with density function proportional to

$$\eta^{(n_\cdot - N)/2 + \kappa - 1} \exp\left(-\eta(\delta + G_1 + G_2) + F(\eta^{-1/2})\right). \tag{5.12}$$

This can be done in several ways.

a) We can replace step 2 by one iteration of a Metropolis-Hastings algorithm.

2*. If $\delta + G_1 + G_2 > 0$, then draw η^* from the $\Gamma((n_\cdot - N)/2 + \kappa, \delta + G_1 + G_2)$ distribution, and with probability

$$\min\left(1, \exp(F(\eta^{*-1/2}) - F(\eta^{-1/2}))\right)$$

accept the proposed value and set $\eta := \eta^*$.

If $\delta + G_1 + G_2 \leq 0$, then draw η^* from the $\Gamma((n_\cdot - N)/2 + \kappa, \delta + G_1)$ distribution, and with probability

$$\min\left(1, \exp(F(\eta^{*-1/2}) - F(\eta^{-1/2}) + (\eta - \eta^*)G_2)\right)$$

accept the proposed value and set $\eta := \eta^*$.

If the proposed value is not accepted, η is unchanged.

b) If the function F is bounded, step 2 can be replaced by a rejection sampler. Suppose, for instance, that the prior of η is truncated to the interval $[E_1, E_2]$, $0 < E_1 < E_2$, and that F is a continuous function of $\beta \in [E_2^{-1/2}, E_1^{-1/2}]$. Then there exists $M > 0$ such that $\sup_{E_1 \leq \eta \leq E_2} \exp(F(\eta^{-1/2})) \leq M$. If $\delta + G_1 + G_2 > 0$, we can replace step 2 by the following algorithm.

2**.

(1) Draw η^* from the $\Gamma((n_\cdot - N)/2 + \kappa, \delta + G_1 + G_2)$-distribution truncated to the interval $[E_1, E_2]$.

(2) With probability $\exp(F(\eta^{*-1/2}))/M$, accept the proposed value and set $\eta := \eta^*$. Otherwise go to (1).

If $\delta + G_1 + G_2 \leq 0$, a similar step can be used, where the scale parameter of the gamma distribution is modified as above, and the acceptance probability is $\exp(F(\eta^{*-1/2}) - \eta^* G_2)/M_2$ where $\exp(F(\eta^{-1/2}) - \eta G_2) \leq M_2$.

c) Finally, we can use *approximate direct sampling*. Again we consider the case $\delta + G_1 + G_2 > 0$. For numerical reasons we choose M such that $(G_1 + G_2)/M$ has a reasonable magnitude. Draw $Z_1, \ldots, Z_K$ independently from the $\Gamma((n_\cdot - N)/2 + \kappa, (\delta + G_1 + G_2)/M)$-distribution, and define

$$p_i := \frac{\exp(F((Z_i/M)^{-1/2}))}{\sum_{j=1}^K \exp(F((Z_j/M)^{-1/2}))}.$$

Let I be a random variable with $P(I = i) = p_i$. Then the distribution of $\eta = Z_I/M$ is approximately equal to the weighted gamma-distribution (5.12), where the approximation improves as K increases. Again there is a modification in case $\delta + G_1 + G_2 \leq 0$

For concrete models, it may be more efficient to take advantage of the particular structure of the function F, when drawing values of β in step 2 of the Gibbs sampler, see examples below.

Step 6 can be simplified in a similar way when

$$\sigma_{\boldsymbol{\beta},b}(x) = b c_{\boldsymbol{\beta}}(x), \tag{5.13}$$

where $b > 0$ and $c_{\boldsymbol{\beta}}(x) > 0$. For more details see the next subsection.

5.2 The EM algorithm

For the EM algorithm we obtain a further simplification, if we assume that the diffusion coefficient has the form (5.13) and that the random variable $E = b^{-2}$ is $\Gamma(\kappa, \delta)$ distributed. Thus $\boldsymbol{\gamma} = (\boldsymbol{\xi}, \boldsymbol{\Gamma}, \kappa, \delta)$, and the function $q(\boldsymbol{\beta})$, implicitly given by (5.4) (here we suppress the arguments $\underline{\boldsymbol{a}}$ and $\underline{\boldsymbol{b}}$), has the form

$$\begin{aligned} q(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \Bigg[\frac{\boldsymbol{a}^i}{(b^i)^2} \int_{x_1^i}^{x_{n_i}^i} \frac{\boldsymbol{g}(z)^\top}{c_{\boldsymbol{\beta}}^2(z)} dz - (n_i - 1)\log(b^i) - \frac{1}{2}\log(c_{\boldsymbol{\beta}}(x_{n_i}^i)/c_{\boldsymbol{\beta}}(x_1^i)) - \sum_{j=2}^{n_i} \log c_{\boldsymbol{\beta}}(x_j^i) \\ &\quad - \frac{1}{(b^i)^2} \sum_{j=2}^{n_i} \frac{(h_{\boldsymbol{\beta},1}(x_j^i) - h_{\boldsymbol{\beta},1}(x_{j-1}^i))^2}{2(t_j^i - t_{j-1}^i)} - \frac{1}{2} \sum_{j=2}^{n_i} \int_{t_{j-1}^i}^{t_j^i} r_{\boldsymbol{\beta},\boldsymbol{a}^i,b^i}(b^i Y_s^{*ij} + \ell_{\boldsymbol{\beta},1}^{ij}(s)) ds \Bigg], \quad (5.14) \end{aligned}$$

where

$$r_{\boldsymbol{\beta},\boldsymbol{a}^i,b^i}(x) = \boldsymbol{a}^i \bar{\boldsymbol{g}}_{\boldsymbol{\beta},1}(h_{\boldsymbol{\beta},1}^{-1}(x))^\top + (b^i)^{-2} \boldsymbol{a}^i \tilde{\boldsymbol{g}}_{\boldsymbol{\beta},1}(h_{\boldsymbol{\beta},1}^{-1}(x))^\top \tilde{\boldsymbol{g}}_{\boldsymbol{\beta},1}(h_{\boldsymbol{\beta},1}^{-1}(x))(\boldsymbol{a}^i)^\top + (b^i)^2 m_{\boldsymbol{\beta},1}^{(2)}(x)$$

with $m^{(2)}$ given by (5.11).

It follows from (5.4) and standard results on maximum likelihood estimation for the multivariate normal and the gamma distribution that the EM-algorithm works as follows. We use the vector and the matrix defined by (5.7) and (5.9), but we suppress the dependence on $\underline{\boldsymbol{a}}$ and $\underline{\boldsymbol{b}}$ in the notation and write $\boldsymbol{v}_{\boldsymbol{\beta}}$ and $\boldsymbol{D}_{\boldsymbol{\beta}}$.

0. Choose initial values $\hat{\boldsymbol{\alpha}}_0, \hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\gamma}}_0$, $k := 0$

1. For $i = 1, \ldots, N$, generate MC-samples $\boldsymbol{X}_{\text{mis}}^{m,i} = \{\boldsymbol{Y}^{*m,i}, \boldsymbol{a}^{m,i}, b^{m,i}\}$, $m = 1, \ldots, M$, conditionally on $\boldsymbol{X}_{\text{obs}}^i$ under the parameter values $\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\gamma}}_k$. Then for each m, calculate the averages

$$\bar{\boldsymbol{a}} = \tfrac{1}{NM} \textstyle\sum_{i,m} \boldsymbol{a}^{m,i}, \qquad \boldsymbol{S} = \tfrac{1}{NM} \textstyle\sum_{i,m} (\boldsymbol{a}^{m,i} - \bar{\boldsymbol{a}})^\top (\boldsymbol{a}^{m,i} - \bar{\boldsymbol{a}})$$

$$\bar{e} = \tfrac{1}{NM} \textstyle\sum_{i,m} (b^{m,i})^{-2}, \qquad \bar{l} = -\tfrac{2}{NM} \textstyle\sum_{i,m} \log(b^{m,i}).$$

Finally, for each m let $\boldsymbol{v}_{\boldsymbol{\beta}}^m$, $\boldsymbol{D}_{\boldsymbol{\beta}}^m$ and $q^m(\boldsymbol{\beta})$ denote the quantities given by (5.7), (5.9) and (5.14)

2. Set

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{k+1} &:= \text{argmax}_{\boldsymbol{\beta}} \left(\tfrac{1}{2} \hat{\boldsymbol{v}}_{\boldsymbol{\beta}} \hat{\boldsymbol{D}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{v}}_{\boldsymbol{\beta}}^\top + \hat{q}(\boldsymbol{\beta}) \right) \\ \hat{\boldsymbol{\alpha}}_{k+1} &:= \hat{\boldsymbol{v}}_{\hat{\boldsymbol{\beta}}_{k+1}} \hat{\boldsymbol{D}}_{\hat{\boldsymbol{\beta}}_{k+1}}^{-1} \\ \hat{\xi}_{k+1} &:= \bar{\boldsymbol{a}}, \quad \hat{\boldsymbol{\Gamma}}_{k+1} := \boldsymbol{S}^{-1} \quad \hat{\delta}_{k+1} := \hat{\kappa}_{k+1}/\bar{e}, \end{aligned}$$

where

$$\hat{\boldsymbol{v}}_{\boldsymbol{\beta}} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{v}_{\boldsymbol{\beta}}^m, \qquad \hat{\boldsymbol{D}}_{\boldsymbol{\beta}} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{D}_{\boldsymbol{\beta}}^m, \qquad \hat{q}(\boldsymbol{\beta}) = \frac{1}{M} \sum_{m=1}^{M} q^m(\boldsymbol{\beta}),$$

and where $\hat{\kappa}_{k+1}$ is the unique solution to $\log(\hat{\kappa}_{k+1}) - \psi(\hat{\kappa}_{k+1}) = \log(\bar{e}) - \bar{l}$ (ψ denotes the digamma function)

3. k:=k+1; GO TO 1.

The MC-samples in step 1 can be generated for each value of i by means of a simplified version of the previous Gibbs sampler in which the following definitions are used:

$$G_{i1} = \sum_{j=2}^{n_i} \frac{(h_{\hat{\boldsymbol{\beta}}_k,1}(x_j^i) - h_{\hat{\boldsymbol{\beta}}_k,1}(x_{j-1}^i))^2}{2(t_j^i - t_{j-1}^i)} \qquad G_{i2} = -\int_{x_1^i}^{x_{n_i}^i} \frac{\hat{\boldsymbol{\alpha}}_k \boldsymbol{f}(y)^\top + \boldsymbol{a}^i \boldsymbol{g}(y)^\top}{c^2_{\hat{\boldsymbol{\beta}}_k}(y)} dy$$

and

$$F_i(b) = -\frac{1}{2} \sum_{j=2}^{n_i} \int_{t_{j-1}^i}^{t_j^i} \phi_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, \boldsymbol{a}^i, b}(Y_s^{*ij} + b^{-1} \ell^{ij}_{\hat{\boldsymbol{\beta}}_k,1}(s)) ds,$$

where

$$\phi_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, \boldsymbol{a}^i, b}(y) = m^{(1)}_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, \boldsymbol{a}^i, 1}(by) + b^2 m^{(2)}_{\hat{\boldsymbol{\beta}}_k,1}(by) + b^{-2} m^{(3)}_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, \boldsymbol{a}^i, 1}(by)$$

with $m^{(1)}, m^{(2)}$ and $m^{(3)}$ given by (5.11).

The Gibbs sampler goes as follows, cf. (5.5).

0. Draw independent values of $\boldsymbol{a}^i$ and E from the $N_{p_2}(\hat{\boldsymbol{\xi}}_k, \hat{\boldsymbol{\Gamma}}_k^{-1})$-distribution and the $\Gamma(\hat{\kappa}_k, \hat{\delta}_k)$-distribution, respectively, and set $b^i := E^{-2}$

1. Simulate independent sample paths Y^{*ij}, $j = 2, \ldots, n_i$, conditionally on $\boldsymbol{a}^i, b^i, \boldsymbol{X}^i_{\text{obs}}$ (with the parameter values $\hat{\boldsymbol{\alpha}}_k$ and $\hat{\boldsymbol{\beta}}_k$), and use these to calculate $\boldsymbol{t}^i_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, b^i}$ and $\boldsymbol{B}^i_{\hat{\boldsymbol{\beta}}_k, b^i}$ given by (5.6) and (5.8).

2. Draw $\boldsymbol{a}^i$ from the $N_{p_2}((\boldsymbol{t}^i_{\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k, b^i} + \hat{\boldsymbol{\xi}}_k \hat{\boldsymbol{\Gamma}}_k)(\boldsymbol{B}^i_{\hat{\boldsymbol{\beta}}_k, b^i} + \hat{\boldsymbol{\Gamma}}_k)^{-1}, (\boldsymbol{B}^i_{\hat{\boldsymbol{\beta}}_k, b^i} + \hat{\boldsymbol{\Gamma}}_k)^{-1})$-distribution (conditionally on $b^i, \boldsymbol{X}^i_{\text{obs}}$ and $\boldsymbol{Y}^{*i}$)

3. Draw E from the weighted gamma distribution with density proportional to

$$x^{(n_i-1)/2+\hat{\kappa}_k-1} \exp\left(-x(\hat{\delta}_k + G_{i1} + G_{i2}) + F_i(x^{-1/2})\right), \quad x > 0 \tag{5.15}$$

(conditionally on $a^i, \boldsymbol{X}^i_{\text{obs}}$ and $\boldsymbol{Y}^{*i}$), and set $b^i := E^{-2}$

4. GO TO 1

Draws from the weighted gamma distribution (5.15) can be done in various ways, as indicated in the previous subsection. Here we just mention that if $\hat{\delta}_k + G_{i1} + G_{i2} > 0$, then step 3 in the Gibbs sampler can be replaced by

3*. Draw E from the $\Gamma((n_i - 1)/2 + \hat{\kappa}_k, \hat{\delta}_k + G_{i1} + G_{i2})$-distribution. With probability

$$\min\left(1, \exp(F_i(E^{-1/2}) - F_i(b^i))\right)$$

the proposed value is accepted and $b^i := E^{-2}$. Otherwise b^i is unchanged.

If $\hat{\delta}_k + G_{i1} + G_{i2} \leq 0$, the scale parameter of the gamma distribution must be changed to $(\hat{\delta}_k + G_{i1})^{-1}$, and F_i must be replaced by the function $\tilde{F}_i(b) = b^{-2}G_{i2} + F_i(b)$.

If some of the random effects in the drift must necessarily be positive, these coordinates of the multivariate normal distribution can restricted to the positive half-line. Perhaps a more satisfactory solution is to assume that these coordinates are exponential distributed and independent of the other coordinates. For instance, if a^i is one-dimensional ($p_2 = 1$) and must be positive, then we can assume that a^i is exponential distributed with mean λ^{-1}. Then in step 2 of the EM-algorithm $\hat{\xi}_{k+1}$ and $\hat{\mathbf{\Gamma}}_{k+1}$ are replaced by

$$\hat{\lambda}_{k+1} := 1/\bar{a}$$

(and $\boldsymbol{S}$ is not needed in step 1). In step 2 of the Gibbs sampler, a^i is drawn from the normal distribution with mean $(t^i_{\hat{\boldsymbol{\alpha}}_k,\hat{\boldsymbol{\beta}}_k,b^i} - \hat{\lambda}_k)/B^i_{\boldsymbol{\beta},b^i}$ and variance $(B^i_{\boldsymbol{\beta},b^i})^{-1}$ restricted to the positive half-axis (conditionally on $b^i, \boldsymbol{X}^i_{\text{obs}}$ and $\boldsymbol{Y}^{*i}$).

6 Simulation studies

6.1 The Ornstein-Uhlenbeck process

Consider the Ornstein-Uhlenbeck process

$$dX^i_t = -a^i X^i_t dt + \beta dW^i_t,$$

with random speed parameter a^i, where $\beta > 0$ and a^i is exponential distributed with mean γ^{-1}.

First we apply the *Gibbs sampler*. As the prior, we choose the $\Gamma(\nu, \lambda)$ distribution for γ and the $\Gamma(\kappa, \delta)$ distribution for $\eta = \beta^{-2}$, and we assume independence of γ and β.

Since $f(x) = 0, g(x) = -x$, $h_\beta(x) = x/\beta$ and $\phi_a(x) = -a + a^2x^2$, the quantities needed in the algorithm are

$$\begin{aligned}
t^i_\beta &= -\frac{1}{2\beta^2}\left((x^i_{n_i})^2 - (x^i_1)^2\right) + \frac{1}{2}(t^i_{n_i} - t^i_1) \\
B^i_\beta &= \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \left(Y^{*ij}_s + \beta^{-1}\ell^{ij}_1(s)\right)^2 ds \\
G_1 &= \sum_{i=1}^{N}\sum_{j=2}^{n_i} \frac{(x^i_j - x^i_{j-1})^2}{2(t^i_j - t^i_{j-1})} \\
G_2 &= \frac{1}{2}\sum_{i=1}^{N} a^i \left((x^i_{n_i})^2 - (x^i_1)^2\right) \\
E_1 &= \frac{1}{2}\sum_{i=1}^{N}(a^i)^2 \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} \ell^{ij}_1(s)^2 ds \\
E_2 &= -\sum_{i=1}^{N}(a^i)^2 \sum_{j=2}^{n_i} \int_{t^i_{j-1}}^{t^i_j} Y^{*ij}_s \ell^{ij}_1(s) ds,
\end{aligned}$$

where

$$\ell^{ij}_1(t) = \frac{(t^i_j - t)x^i_{j-1} + (t - t^i_{j-1})x^i_j}{t^i_j - t^i_{i-j}}, \quad t \in [t^i_{j-1}, t^i_j]. \tag{6.1}$$

In this example $F(\beta) = -\beta^{-2}E_1 + \beta^{-1}E_2$ (apart from an additive term independent of β), but here we can make E_1 a part of the scale parameter of the weighted gamma distribution, because it is positive.

The Gibbs sampler goes as follows, with N_+ denoting the normal distribution restricted to the positive half-line.

0. First draw β and γ independently from the prior distribution, and given γ, draw a^i from the exponential distribution with mean γ^{-1}, independently for $i = 1, \ldots, N$.

1. Simulate independent sample paths Y^{*ij} conditionally on a^i, β and X^i_{obs} for $j = 2, \ldots, n_i$, $i = 1, \ldots, N$

2. Draw a^i with distribution $N_+((t^i_\beta - \gamma)/B^i_\beta, (B^i_\beta)^{-1})$, independently for $i = 1, \ldots, N$

3. Draw η from the distribution with density function proportional to

$$\eta^{(n_\cdot - N)/2+\kappa-1} \exp\left(-\eta(\delta + G_1 + G_2 + E_1) + \sqrt{\eta}E_2\right),$$

   and set $\beta := \eta^{-1/2}$

4. Draw γ from the $\Gamma(\nu + n, \lambda + a^1 + \cdots + a^N)$-distribution

5. GO TO 1

If $\delta + G_1 + G_2 + E_1 > 0$, the distribution in step 3 is a weighted gamma distribution with scale parameter $(\delta + G_1 + G_2 + E_1)^{-1}$ and weight function $\exp(\sqrt{\eta}E_2)$. If $\delta + G_1 + G_2 + E_1 \leq 0$, the scale parameter is $(\delta + G_1 + E_1)^{-1}$ (which is always positive) and the weight function is $\exp(\sqrt{\eta}E_2 - \eta G_2)$.

As data we simulated 100 diffusions with $\beta = 1$ and $\gamma = 1$ at the time points $t^i_j = j, j = 1, \ldots, 100$. Then we ran 1000 iterations of the Gibbs sampler using prior distributions with $\kappa = 1, \delta = 0.5, \nu = 1, \lambda = 2$. In step 1 we simulated the approximate diffusion bridges proposed by Bladt and Sørensen (2014), and in step 3 we used the approximate direct sampling method described in Subsection 5.1. In all cases $\delta + G_1 + G_2 + E_1$ was positive.

Because the burn-in was almost immediate, we based the estimation on the last 900 draws of the parameters. The mean posterior estimators are $\hat{\beta} = 0.9800$ and $\hat{\gamma} = 1.017$, and the 95% credibility intervals are $[0.9368, 1.0200]$ for β and $[0.8264, 1.2291]$ for γ. The last 900 draws and histograms of the parameter values are plotted in Figure 6.1.

Next we apply the *EM algorithm*. In this example

$$q(\beta) = -\beta^{-2}(G_1 + E_1) + \beta^{-1}E_2 - (n_\cdot - N)\log(\beta)$$

(apart from additive terms that do not depend on β). The EM-algorithm goes as follows:

0. Choose initial values $\hat{\beta}_0$ and $\hat{\gamma}_0$, $k := 0$

1. For $i = 1, \ldots, N$, generate MC-samples $\boldsymbol{X}^{m,i}_{\text{mis}} = \{\boldsymbol{Y}^{*m,i}, a^{m,i}\}$, $m = 1, \ldots, M$, conditionally on $\boldsymbol{X}^i_{\text{obs}}$ under the parameter values $\hat{\beta}_k$ and $\hat{\gamma}_k$, and for each m calculate G^m_2, E^m_1 and E^m_2. Finally, calculate the averages

$$\hat{G}_2 = \frac{1}{M}\sum_{m=1}^{M} G^m_2, \qquad \hat{E}_1 = \frac{1}{M}\sum_{m=1}^{M} E^m_1, \qquad \hat{E}_2 = \frac{1}{M}\sum_{m=1}^{M} E^m_2, \qquad \bar{a} = \frac{1}{NM}\sum_{i,m} a^{m,i}$$

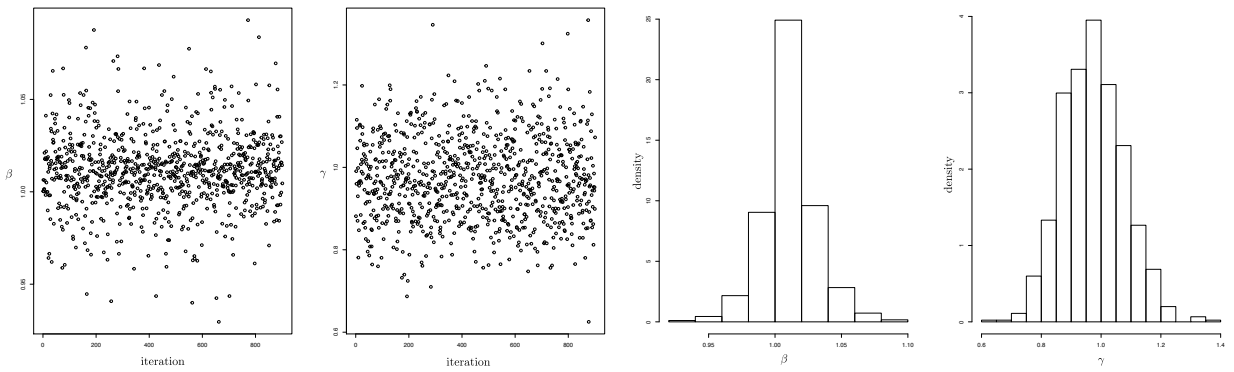

Figure 6.1: The Gibbs sampler for the Ornstein-Uhlenbeck process with random drift parameter: scatter plots of the draws of the parameters β and γ (after a burn-in of 100 draws) and histograms of the draws. The true parameter values are $\beta = 1$ and $\gamma = 1$.

2.

$$\hat{\beta}_{k+1} := \text{argmax}_{\boldsymbol{\beta}} \left(-\beta^{-2}(G_1 + \hat{G}_2 + \hat{E}_1) + \beta^{-1}\hat{E}_2 - (n. - N)\log(\beta) \right), \qquad \hat{\gamma}_{k+1} := 1/\bar{a}$$

3. k:=k+1; GO TO 1.

The maximization in step 2 is elementary. If $G_1 + \hat{G}_2 + \hat{E}_1 > 0$ (which is usually the case), then

$$\hat{\beta}_{k+1} = \frac{4(G_1 + \hat{G}_2 + \hat{E}_1)}{-\hat{E}_2 + \sqrt{\hat{E}_2^2 + 8(G_1 + \hat{G}_2 + \hat{E}_1)(n. - N)}}.$$

Moreover, $\hat{\beta}_{k+1} = -4(G_1 + \hat{G}_2 + \hat{E}_1)/(\hat{E}_2 + \sqrt{\hat{E}_2^2 + 8(G_1 + \hat{G}_2 + \hat{E}_1)(n. - N)})$ when $0 > G_1 + \hat{G}_2 + \hat{E}_1 \geq -\frac{1}{8}\hat{E}_2^2/(n. - N)$ and $\hat{E}_2 > 0$, and if $G_1 + \hat{G}_2 + \hat{E}_1 = 0$ and $\hat{E}_2 < 0$, then $\hat{\beta}_{k+1} = -\hat{E}_2/(n. - N)$. In other cases, a positive maximum does not exist.

The MC-samples in step 1 can be generated for each value of i by means of the following simplified version of the previous Gibbs sampler.

0. Draw a^i from the exponential distribution with mean $\hat{\gamma}_k^{-1}$

1. Simulate independent sample paths Y^{*ij}, $j = 2, \ldots, n_i$, conditionally on a^i and $\boldsymbol{X}^i_{\text{obs}}$ (with the parameter value $\hat{\beta}_k$), and use these to calculate $t^i_{\hat{\beta}_k}$ and $B^i_{\hat{\beta}_k}$

2. Draw a^i from the $N_+((t^i_{\hat{\beta}_k} - \hat{\gamma}_k)/B^i_{\hat{\beta}_k}, (B^i_{\hat{\beta}_k})^{-1})$-distribution

3. GO TO 1

We simulated 100 diffusions with $\beta = 1$ and $\gamma = 1$ at the time points $t^i_j = j, j = 1, \ldots, 100$ as our data. Then we ran 50 iterations of the EM algorithm with initial values $\hat{\beta}_0 = 3$ and $\hat{\gamma}_0 = 5$. The parameter values in the iterations are plotted in Figure 6.2. The convergence was fast. As estimators we use the final parameter values. The EM maximum likelihood estimators estimators are $\hat{\beta} = 1.017$ and $\hat{\gamma} = 1.003$.

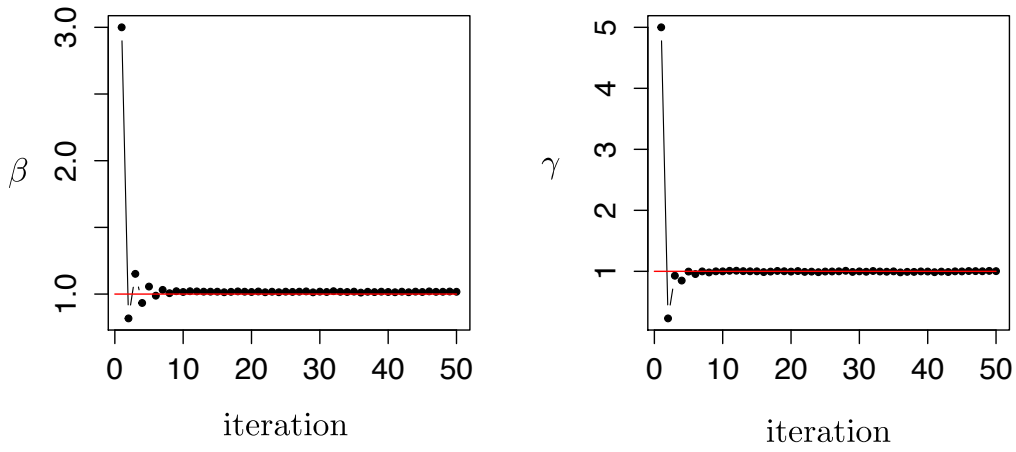

Figure 6.2: The EM algorithm for the Ornstein-Uhlenbeck process with random drift parameter: parameter values in 50 iterations. The true parameter values are $\beta = 1$ and $\gamma = 1$.

6.2 A t-diffusion

Consider the t-diffusion

$$dX_t^i = -a^i X_t^i dt + \beta\sqrt{1 + (X_t^i)^2} dW_t^i,$$

with random speed parameter a^i, where $\beta > 0$ and a^i is exponential distributed with mean γ^{-1}. The t-diffusion is one of the Pearson diffusions; see Forman and Sørensen (2008). For fixed a^i it is an ergodic process on $\mathbb{R}$ provided that $a^i > 0$. The invariant probability measure is a re-scaled t-distribution with $\nu = 2a^i/\beta^2 + 1$ degrees of freedom (and scale-parameter $\nu^{-\frac{1}{2}}$).

For this model

$$\begin{aligned}
h_\beta(x) &= \beta^{-1}\log(x + \sqrt{x^2+1}) \\
h_\beta^{-1}(y) &= \sinh(\beta y) \\
\mu_{a^i,\beta}(y) &= -(a^i/\beta + \beta/2)\tanh(\beta y) \\
\phi_{a^i,\beta}(y) &= (a^i/\beta + \beta/2)^2\tanh^2(\beta y) - \frac{a^i + \frac{1}{2}\beta^2}{\cosh^2(\beta y)} \\
s_{a^i,\beta}(x) &= -\frac{a^i}{2\beta^2}\log(1+x^2).
\end{aligned}$$

Note that $a^i/\beta + \beta/2 = \frac{1}{2}\beta\nu$.

First we consider the *Gibbs sampler*. As the prior, we choose again the $\Gamma(\nu,\lambda)$ distribution for γ and the $\Gamma(\kappa,\delta)$ distribution for $\eta = \beta^{-2}$, and we assume independence of γ and β. In the algorithm we use following quantities

$$t_\beta^i = -\frac{1}{2\beta^2}\log\left(\frac{1+(x_{n_i}^i)^2}{1+(x_1^i)^2}\right) + \frac{1}{2}(t_{n_i}^i - t_1^i) - \sum_{j=2}^{n_i}\int_{t_{j-1}^i}^{t_j^i}\tanh^2(\beta Y_s^{*ij} + \ell_1^{ij}(s))ds$$

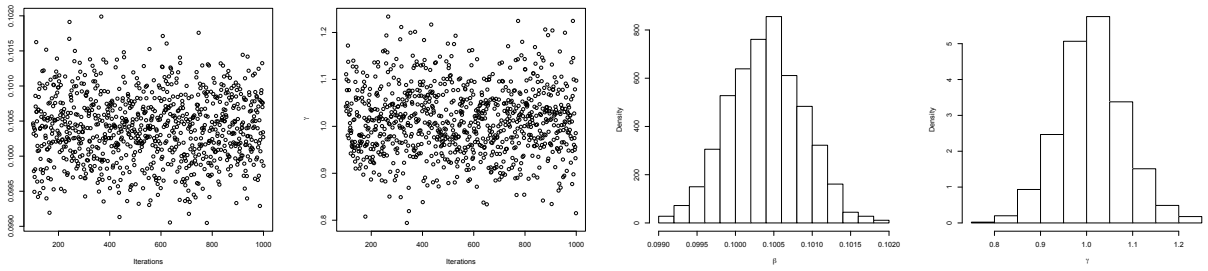

Figure 6.3: The Gibbs sampler for the t-diffusion with random drift parameter: scatter plots of the draws of the parameters β and γ (after a burn-in of 100 draws) and histograms of the draws. The true parameter values are $\beta = 0.1$ and $\gamma = 1$.

$$
\begin{aligned}
B^i_\beta &= \frac{1}{\beta^2}\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j} \tanh^2(\beta Y^{*ij}_s + \ell^{ij}_1(s))ds \\
G_1 &= \sum_{i=1}^{N}\sum_{j=2}^{n_i} \frac{\left(\log\left(\frac{x^i_j+\sqrt{(x^i_j)^2+1}}{x^i_{j-1}+\sqrt{(x^i_{j-1})^2+1}}\right)\right)^2}{2(t^i_{j-1}-t^i_j)} \\
G_2 &= \frac{1}{2}\sum_{i=1}^{N} a^i \log\left(\frac{1+(x^i_{n_i})^2}{1+(x^i_1)^2}\right) \\
F(\beta) &= -\frac{1}{2}\sum_{i=1}^{N}\left[\left(\frac{(a^i)^2}{\beta^2}+\frac{3}{4}\beta^2+a^i(\beta+1)\right)\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}\tanh^2(\beta Y^{*ij}_s+\ell^{ij}_1(s))ds \right. \\
&\qquad \left. -\left(a^i+\frac{1}{2}\beta^2\right)(t^i_{n_i}-t^i_1)\right]
\end{aligned}
$$

where

$$
\ell^{ij}_1(t) = \frac{(t^i_j - t)\log\left(x^i_{j-1}+\sqrt{(x^i_{j-1})^2+1}\right) + (t-t^i_{j-1})\log\left(x^i_j+\sqrt{(x^i_j)^2+1}\right)}{t^i_j - t^i_{i-j}}, \quad t\in[t^i_{j-1}, t^i_j].
$$

With these definitions the Gibbs sampler is as in the previous example except that step 3 is replaced by step 2* in Subsection 5.1.

As data we simulated 100 diffusions with $\beta = 0.1$ and $\gamma = 1$ at the time points $t^i_j = j, j = 0, \ldots, 100$. We then ran 1000 iterations of the Gibbs sampler using prior distributions with $\kappa = 1, \delta = 5, \nu = 1, \lambda = 0.75$. We used the approximate diffusion bridges of Bladt and Sørensen (2014). In all cases it turned out that $\delta + G_1 + G_2 > 0$.

The estimation is based on the last 900 draws of the parameters, after a burn-in of 100 iterations. The mean posterior estimators are $\hat{\beta} = 0.0998$ and $\hat{\gamma} = 1.0248$, and the 95% credibility intervals are $[0.0989, 0.1008]$ for β and $[0.8944, 1.1648]$ for γ. The last 900 draws and histograms of the parameter values are plotted in Figure 6.3.

Next we turn to the *EM algorithm*. For the t-diffusion

$$
q(\beta) = -\beta^{-2}(G_1+G_2) + \tfrac{1}{4}\beta^2\sum_{i=1}^{N}(t^i_{n_i}-t^i_1) - (n_\cdot - N)\log(\beta)
$$

$$-\tfrac{1}{2}\sum_{i=1}^{N}\left(2a_i + a_i^2\beta^{-2} + \tfrac{3}{4}\beta^2\right)\sum_{j=2}^{n_i}\int_{t_{j-1}^i}^{t_j^i}\tanh^2(\beta Y_s^{*ij} + \ell_1^{ij}(s))ds$$

(apart from additive terms that do not depend on β). The EM-algorithm goes as follows:

0. Choose initial values $\hat{\beta}_0$ and $\hat{\gamma}_0$, $k := 0$

1. For $i = 1, \ldots, N$, generate MC-samples $\boldsymbol{X}_{\text{mis}}^{m,i} = \{\boldsymbol{Y}^{*m,i}, a^{m,i}\}$, $m = 1, \ldots, M$ conditionally on $\boldsymbol{X}_{\text{obs}}^i$ under the parameter values $\hat{\beta}_k$ and $\hat{\gamma}_k$, and for each m calculate G_2^m. Finally, calculate
$$\hat{G}_2 = \frac{1}{M}\sum_{m=1}^{M} G_2^m, \qquad \bar{a} = \frac{1}{MN}\sum_{i,m} a^{m,i}$$

2.
$$\begin{aligned}\hat{\gamma}_{k+1} &:= 1/\bar{\boldsymbol{a}}\\ \hat{\beta}_{k+1} &:= \text{argmax}_{\boldsymbol{\beta}}\left(-\beta^{-2}(G_1 + \hat{G}_2) + \tfrac{1}{4}\beta^2\sum_{i=1}^{N}(t_{n_i}^i - t_1^i) - (n. - N)\log(\beta) - \tilde{q}(\beta)\right),\end{aligned}$$

where

$$\tilde{q}(\beta) = \tfrac{1}{2}\frac{1}{M}\sum_{i,m}\left(2a_i^m + (a_i^m)^2\beta^{-2} + \tfrac{3}{4}\beta^2\right)\sum_{j=2}^{n_i}\int_{t_{j-1}^i}^{t_j^i}\tanh^2(\beta Y_s^{*m,ij} + \ell_1^{ij}(s))ds$$

3. k:=k+1; GO TO 1.

As in the previous example, the MC-samples in step 1 can be generated for each value of i by means of a simplified version of the previous Gibbs sampler, which is identical to the previous sampler (except that $t_{\hat{\beta}_k}^i$ and $B_{\hat{\beta}_k}^i$ are defined as in this subsection).

We simulated 100 diffusions with $\beta = 0.1$ and $\gamma = 1$ at the time points $t_j^i = j, j = 1, \ldots, 100$. We ran 50 iterations of the EM algorithm with initial values $\hat{\beta}_0 = 0.2$ and $\hat{\gamma}_0 = 2$. The parameter values in the iterations are plotted in Figure 6.4. The convergence was fast. As estimators we use the final parameter values. The EM maximum likelihood estimators estimates are $\hat{\beta} = 0.1011$ and $\hat{\gamma} = 1.0098$.

7 Application to neuronal data

In this section we estimate the parameters of a leaky integrate-and-fire neuronal model (Ornstein-Uhlenbeck process) with random level on the basis of measurement of the membrane potential between firings.

The data consists of measurements every 0.15 ms of the membrane potential of a single auditory neuron of a guinea pig (for details on the experiment, see Yu *et al.* (2004)). When the potential crosses a certain threshold, the neuron fires (produces an electric signal), resets to an initial resting value and starts increasing towards a certain level around which it fluctuates in a stationary way until the neuron fires again. The data recorded between firings (in so-called inter-spike intervals) can be considered realisations of independent random processes with identical parameters except that the stationary level varies randomly from one

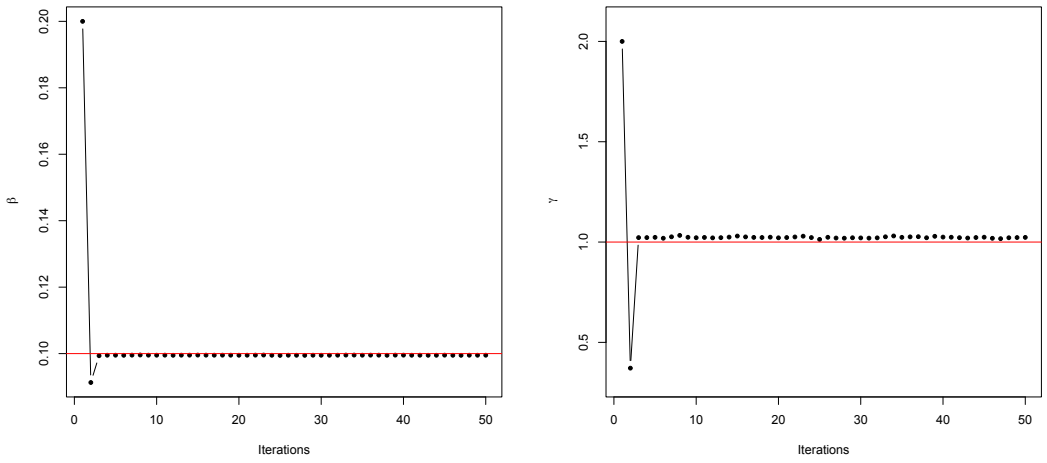

Figure 6.4: The EM algorithm for the t-diffusion with random drift parameter: parameter values in 50 iterations. The true parameter values are $\beta = 0.1$ and $\gamma = 1$.

interval to another. There are 312 inter-spike intervals, for which the number of observations per interval, n_i, varies from hundreds to several thousands. The data was previously analysed by Lansky *et al.* (2006), who fitted a fixed effect Ornstein-Uhlenbeck process individually to each inter-spike interval. Picchini *et al.* (2008), Picchini *et al.* (2010) and Wiqvist *et al.* (2021) fitted an Ornstein-Uhlenbeck process with a random level to the data using methods different from ours, see the introduction, while Dion (2016) obtained a non-parametric estimate of the density of the random level.

The model is the Ornstein-Uhlenbeck process

$$dX_t^i = (a^i - \alpha X_t^i)dt + \beta dW_t^i,$$

with random level a^i, where $\alpha > 0$, $\beta > 0$ and $a^i \sim N(\xi, \sigma^2)$. We apply the Gibbs sampler, and as the prior distribution we take $\alpha, \eta = \beta^{-2}, \xi$ and $\gamma = \sigma^{-2}$ to be independent and exponential distributed with mean λ_i^{-1}, $i = 1, 2, 3, 4$, respectively.

Since $f(x) = -x, g(x) = 1, h_\beta(x) = x/\beta$ and $\phi_{\alpha,\beta,a}(x) = -\alpha - 2\alpha a x/\beta + \alpha^2 x^2 + a^2/\beta^2$, the quantities needed in the algorithm are

$$\begin{aligned}
t^i_{\alpha,\beta} &= \beta^{-2}\left(x^i_{n_i} - x^i_1\right) + \frac{\alpha}{\beta}\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}\left(Y_s^{*ij} + \beta^{-1}\ell_1^{ij}(s)\right)^2 ds \\
B^i_\beta &= \beta^{-2}(t^i_{n_i} - t^i_1) \\
v_{\beta,\underline{\boldsymbol{a}}} &= \sum_{i=1}^{N}\left[\frac{1}{2}\beta^{-2}\left((x^i_1)^2 - (x^i_{n_i})^2\right) + \frac{1}{2}(t^i_{n_i} - t^i_1) + \beta^{-1}a^i\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}\left(Y_s^{*ij} + \beta^{-1}\ell_1^{ij}(s)\right)ds\right] \\
D_\beta &= \sum_{i=1}^{N}\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}\left(Y_s^{*ij} + \beta^{-1}\ell_1^{ij}(s)\right)^2 ds \\
G_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=2}^{n_i}\frac{(x^i_j - x^i_{j-1})^2}{(t^i_j - t^i_{j-1})}
\end{aligned}$$

$$
\begin{aligned}
G_2 &= \sum_{i=1}^{N}\left[\frac{\alpha}{2}\left((x^i_{n_i})^2-(x^i_1)^2\right)+a^i\left(x^i_{n_i}-x^i_1\right)\right]\\
E_1 &= \frac{1}{2}\sum_{i=1}^{N}\left[(a^i)^2(t^i_{n_i}-t^i_1)+\alpha^2\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}\ell_1^{ij}(s)^2ds\right]\\
E_2 &= \sum_{i=1}^{N}\left[\alpha a^i\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}Y_s^{*ij}ds-\alpha^2\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}Y_s^{*ij}\ell_1^{ij}(s)ds\right]\\
E_3 &= -\alpha\sum_{i=1}^{N}a^i\sum_{j=2}^{n_i}\int_{t^i_{j-1}}^{t^i_j}\ell_1^{ij}(s)ds
\end{aligned}
$$

where $\ell_1^{ij}(s)$ is given by (6.1). For this model $F(\beta)=-\beta^{-2}(E_1+E_3)+\beta^{-1}E_2$ (apart from an additive term independent of β), but here we can make E_1 a part of the scale parameter of the weighted gamma distribution in step 4 of the Gibbs sampler below, because E_1 is always positive. The Gibbs sampler goes as follows.

0. First draw α,β,ξ and γ independently from the prior distribution, and draw a_i from the normal distribution with mean ξ and variance γ^{-1}, independently for $i=1,\ldots,N$.

1. Simulate independent sample paths Y^{*ij} conditionally on a^i,α,β and X^i_{obs} for $j=2,\ldots,n_i,\ i=1,\ldots,N$

2. Draw a_i with distribution $N((t^i_{\alpha,\beta}+\xi\gamma)/(B^i_\beta+\gamma),(B^i_\beta+\gamma)^{-1})$, independently for $i=1,\ldots,N$

3. Draw α from the distribution $N_+((v_{\beta,\underline{\boldsymbol{a}}}-\lambda_1)/D_\beta,D_\beta^{-1})$

4. Draw η from the distribution with density function proportional to

$$\eta^{(n_\cdot-N)/2}\exp\left(-\eta(\lambda_2+G_1+E_1+G_2+E_3)+\sqrt{\eta}E_2\right),$$

   and set $\beta:=\eta^{-1/2}$

5. Draw ξ from the $N(\bar a-\lambda_3/(\gamma N),(\gamma N)^{-1})$-distribution, where $\bar a=(a^1+\cdots+a^N)/N$

6. Draw γ from the $\Gamma(N/2+1,\lambda_4+\frac{1}{2}\sum_{i=1}^N(a^i-\xi)^2)$-distribution.

7. GO TO 1

If $\lambda_2+G_1+E_1+G_2+E_3>0$, the distribution in step 4 is a weighted gamma distribution with scale parameter $(\lambda_2+G_1+E_1+G_2+E_3)^{-1}$ and weight function $\exp(\sqrt{\eta}E_2)$. If $\lambda_2+G_1+E_1+G_2+E_3\le 0$, the scale parameter is $(\lambda_2+G_1+E_1)^{-1}$ (which is always positive) and the weight function is $\exp(\sqrt{\eta}E_2-\eta(G_2+E_3))$. In step 1 we simulated the approximate diffusion bridges of Bladt and Sørensen (2014), and in step 4 we used approximate direct sampling, see Subsection 5.1.

The Gibbs sampler was run with 1000 iterations using prior distributions with parameters $\lambda_1=5,\lambda_2=0.02,\lambda_3=400,\lambda_4=0.35$, which are expected to be essentially non-informative. After a burn-in of 100 iterations, the estimates are based on the last 900 draws of the parameters. Histograms of the last 900 parameter draws are plotted in Figure 7.1. The

Parameter	mean	0.025-quantile	0.975-quantile
α	21.580	21.574	21.586
β	0.013438	0.013431	0.013446
ξ	0.2500	0.2498	0.2502
σ	0.0576	0.0536	0.0624

Table 7.1: Estimates obtained by 900 iterations of the Gibbs sampler after a burn-in of 100 iterations.

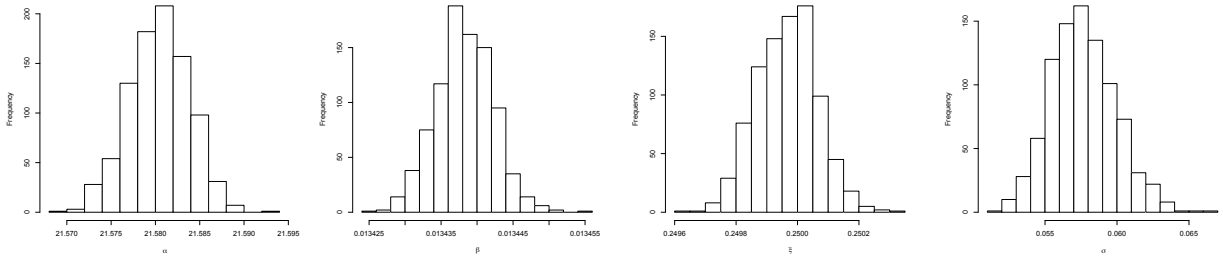

Figure 7.1: Histograms of the Gibbs sampler draws of the parameters α, β, ξ and γ of the neuron model (after a burn-in of 100 draws).

results, see Table 7.1, are in reasonable accordance with the estimates obtained in Lansky *et al.* (2006), Picchini *et al.* (2010), Wiqvist *et al.* (2021) and Dion (2016). The estimates of β are essentially the same, while the estimate of σ is close to the mean of the estimates in the other papers (which vary quite a bit between the papers). Picchini *et al.* (2010) are unhappy with their own estimate of α and prefer a value from the literature, namely 25.6, which is close to our estimate.

A Appendix: Simple diffusion bridge simulation

In this appendix we briefly present the simple method of simulating diffusion bridges introduced in Bladt and Sørensen (2014) and the corrigendum Bladt *et al.* (2021), which is key to the algorithms proposed in this paper.

The aim is to simulate a (t_1, x_1, t_2, x_2)-bridge for the diffusion process given by (2.1), which is assumed to be ergodic with invariant probability density function ν w.r.t. Lebesgue measure on the state space. The method is based on the following simple construction of a process that starts from x_1 at time zero and at time t_1 ends at x_2, which is a good *approximation to a* (t_1, x_1, t_2, x_2)*-bridge*. One diffusion process, X_t, that solves (2.1), is started from the point x_1, while another independent solution to (2.1), $\bar{X}_t$, is started from the point x_2. The time of the second diffusion is then reversed to obtain the process $X'_t = \bar{X}_{t_2-t}$ Suppose there is a time point $\tau \in [t_1, t_2]$ at which $X_\tau = X'_\tau$. Then the process that is equal to X_t for $t \in [t_1, \tau]$ and for $t \in [\tau, t_2]$ equals X'_t is a process that starts at x_1 and ends at x_2. Because the probability that X and X' meet in $[t_1, t_2]$ is considerable, provided $t_2 - t_1$ is not very small, such a process can be obtained by rejection sampling and can be shown

to be an approximation to a (t_1, x_1, t_2, x_2)-bridge in a sense explained in Bladt *et al.* (2021).

In practice, the approximate diffusion bridge is simulated as follows. Let $Y_{\delta i}$, $i = 0, 1, \ldots, N$ and $\bar{Y}_{\delta i}$, $i = 0, 1, \ldots, N$ be (independent) simulations, for instance using the Eule scheme, of X and $\bar{X}$ in $[t_1, t_2]$ with step size $\delta = (t_2 - t_1)/N$. Then a simulation of an approximation to a (t_1, x_1, t_2, x_2)-bridge is obtained by the following rejection sampling algorithm. Keep simulating Y and $\bar{Y}$ until there is an i such that either $Y_{\delta i} \geq \bar{Y}_{\delta(N-i)}$ and $Y_{\delta(i+1)} \leq \bar{Y}_{\delta(N-(i+1))}$ or $Y_{\delta i} \leq \bar{Y}_{\delta(N-i)}$ and $Y_{\delta(i+1)} \geq \bar{Y}_{\delta(N-(i+1))}$. Once this has been achieved, define

$$B_{\delta i} = \begin{cases} Y_{\delta i} & \text{for } \; i = 0, 1, \ldots, \mu - 1 \\ \\ \bar{Y}_{\delta(N-i)} & \text{for } \; i = \mu, \ldots N, \end{cases}$$

where $\mu = \min\{i \in \{1, \ldots N\} | Y_{\delta i} \leq \bar{Y}_{\delta(N-i)}\}$ if $Y_0 \geq \bar{Y}_\Delta$, and $\mu = \min\{i \in \{1, \ldots N\} | Y_{\delta i} \geq \bar{Y}_{\delta(N-i)}\}$ if $Y_0 \leq \bar{Y}_\Delta$. Then B approximates a (t_1, x_1, t_2, x_2)-bridge. Apart from the usual discretization error, the step size δ also controls the probability that a trajectory crossing is not detected. Therefore, it is advisable to choose δ smaller than usual.

In several cases the approximate diffusion bridge is a sufficiently good approximation to be used as it is, as we do in the simulation studies and data applications in this paper. However, an *exact diffusion bridge* (appart from the discretization error) can be obtained via a pseudo-marginal Metropolis-Hastings algorithm with target distribution equal to that of an exact diffusion bridge, see Andrieu and Roberts (2009). In the kth step of the algorithm, we simulate as proposal an approximate diffusion bridge $X^{(k)}$ as explained above supplemented by a geometric random variable $S^{(k)}$, which is obtained as follows. Simulate a sequence of independent solutions $Z^{(i)}$ to (2.1) in $[t_1, t_2]$ with $Z^{(i)}_{t_1} \sim \nu$ until a sample path is obtained that intersects $X^{(k)}$ in $[t_1, t_2]$. Then

$$S^{(k)} = \min\{i \, : \, Z^{(i)} \text{ intersects } X^{(k)}\}.$$

The proposed diffusion bridge $X^{(k)}$ is accepted with probability $\min\{1, S^{(k)}/S^{(k-1)}\}$. We call $S^{(k)}$ the *geometric variable associated with* $X^{(k)}$. In order to reduce the variance, we could alternatively replace $S^{(k)}$ by the average $T^{(k)}$ of a number of independent geometric random variables associated with $X^{(k)}$, but if the probability that $Z^{(i)}$ intersects $X^{(k)}$ is small, this might be time consuming.

Suppose the distribution of the diffusion process depends on a parameter vector $\boldsymbol{\theta}$ with prior distribution $\pi(\boldsymbol{\theta})$, and that we want to use the simple bridge simulation method in a Gibbs sampler that alternates between drawing $\boldsymbol{\theta}$ conditional on a (t_1, x_1, t_2, x_2)-bridge X and drawing X conditional on $\boldsymbol{\theta}$. If we use the approximate bridge, this is simple to do. If we want to simulate exact bridges, we can use a *Metropolis within Gibbs* algorithm. In the ith iteration we draw $\boldsymbol{\theta}_i$ and X_i. Conditional on $\boldsymbol{\theta}_i$, we draw an approximate bridge $X_{a,i}$ and its associated geometric variable S_i. With probability $\min\{1, S_i/S_{i-1}\}$ we accept the proposed value, i.e. $X_i := X_{a,i}$. Otherwise, $X_i := X_{i-1}$.

Everything described above can be generalized to multivariate diffusion processes, see Bladt *et al.* (2016, 2022). The construction of the approximate diffusion bridge and simulation of the associated geometric variables are more complicated in the multivariate case because coupling methods for diffusions must be applied to ensure that sample paths intersect with positive probability. The methods for multivariate diffusions can also be used to improve the computational efficiency of the simulation of one-dimensional diffusion bridges.

Acknowledgement

The work of F. Baltazar-Larios was supported by UNAM-DGAPA-PAPIIT-IN102224.

References

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, **36**, 906–937.

Andrieu, C. and Roberts, G. (2009). A pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, **37**, 697–725.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. Roy. Statist. Soc. B*, **68**, 333–382.

Bishwal, J. P. N. (2022). Mixingale estimation function for mixed fractional SPDEs with random effect and random sampling. *Eur. J. Math Appl.*, **2**, 12.

Bladt, M. and Sørensen, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, **20**, 645–675.

Bladt, M., Finch, S., and Sørensen, M. (2016). Simulation of multivariate diffusion bridges. *J.R. Statist. Soc. B*, **78**, 343–369.

Bladt, M., Mider, M., and Sørensen, M. (2021). Corrigendum to "Simple simulation of diffusion bridges with application to likelihood inference for diffusions". *Bernoulli*, **27**, 218 – 220.

Bladt, M., Finch, S., and Sørensen, M. (2022). Corrigendum to "Simulation of multivariate diffusion bridges". *J.R. Statist. Soc. B*, **84**, 1581–1585.

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

Dai, M., Duan, J., Liao, J., and Wang, X. (2021). Maximum likelihood estimation of stochastic differential equations with random effects driven by fractional brownian motion. *Applied Mathematics and Computation*, **397**, 125927.

Delattre, M. (2021). A review on asymptotic inference in stochastic differential equations with mixed effects. *Japanese Journal of Statistics and Data Science*, **4**, 543–575.

Delattre, M. and Lavielle, M. (2013). Coupling the saem algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effect diffusion models. *Statistics and Its Interface*, **6**, 519–532.

Delattre, M., Genon-Catalot, V., and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scand. J. Statist.*, **40**, 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. *Eur. Ser. Appl. Indust. Math. Probab. Statist.*, **19**, 671–688.

Delattre, M., Genon-Catalot, V., and Samson, A. (2016). Mixtures of stochastic differential equations with random effects: Application to data clustering. *J. Statist. Planning Inf.*, **173**, 109–124.

Delattre, M., Genon-Catalot, V., and Laredo, C. (2018a). Approximate maximum likelihood estimation for stochastic differential equations with random effects in the drift and the diffusion. *Metrika*, **81**, 953–983.

Delattre, M., Genon-Catalot, V., and Laredo, C. (2018b). Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic Process. Appl.*, **128**, 1929–1957.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic version of the em algorithm. *Annals of Statistics*, **27**, 94–128.

Diebolt, J. and Ip, E. H. S. (1996). Stochastic em: method and application. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 259–273. Chapman & Hall.

Dion, C. (2016). Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model. *Metrika*, **79**, 919–951.

Dion, C. and Genon-Catalot, V. (2016). Bidimensional random effect estimation in mixed stochastic differential models. *Stat. Inference Stoch. Process.*, **19**, 131–158.

Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *Eur. Ser. Appl. Indust. Math. Probab. Statist.*, **12**, 196–218.

Donnet, S. and Samson, A. (2014). Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. *Journal de la Société Française de Statistique*, **155**, 49–72.

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

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Business & Econom. Statist.*, **20**, 297–338.

El Maroufy, H., Ichi, S., El Omari, M., and Slaoui, Y. (2024). Nonparametric estimation for random effects models driven by fractional brownian motion using hermite polynomials. *Statistical Inference for Stochastic Proceses*, **53**.

El Omari, M., El Maroufy, H., and Fuchs, C. (2019). Non parametric estimation for fractional diffusion processes with random effects. *Statistics*, **53**, 753–769.

Forman, J. L. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scand. J. Statist.*, **35**, 438–465.

Jamba, N. T., Jacinto, G., Filipe, P. A., and Braumann, C. A. (2022). Likelihood function through the delta approximation in mixed sde models. *Mathematics*, **10**, 385.

Jamba, N. T., Filipe, P. A., Jacinto, G., and Braumann, C. A. (2024). Stochastic differential equations mixed model for individual growth with inclusion of genetic characteristics. *Statistics, Optimization and Information Computing*, **12**.

Küchler, U. and Sørensen, M. (1997). *Exponential Families of Stochastic Processes*. Springer, New York.

Lansky, P., Sanda, P., and He, J. F. (2006). Parameters of the stochastic leaky integrate-and-fire neuronal model. *J. Comput. Neurosci.*, **21**, 211–223.

Maitra, T. and Bhattacharya, S. (2020). On classical and Bayesian asymptotics in stochastic differential equations with random effects having mixture normal distributions. *J. Statist. Planning Inf.*, **208**, 36–57.

McLachlan, G. J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

Overgaard, R. V., Jonsson, N., e, C. W. T., and Madsen, H. (2005). Non-linear mixed-effect models with stochastic differential equations: implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn*, **32**, 85–107.

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

Picchini, U. and Forman, J. L. (2019). Bayesian inference for stochastic differential equation mixed effects models of a tumour xenography study. *Appl. Statist*, **68**, 887–913.

Picchini, U., Ditlevsen, S., Gaetano, A. D., and Lansky, P. (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neural Computation*, **20**, 2696–2714.

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

Prakasa Rao, B. L. S. (2023). Maximum likelihood estimation for stochastic differential equations driven by mixed fractional brownian motion with random effects. *Communications in Statistics: Theory and Methods*, **52**, 3816–3824.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using metropolis-hastings algorithms. *Biometrika*, **88**, 603–621.

Ruse, M. G., Samson, A., and Ditlevsen, S. (2020). Inference for biomedical data by using diffusion models with covariates and mixed effects. *Appl. Statist*, **69**, 167–193.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

Whitaker, G. A., Golightly, A., Boys, R. J., and Sherlock, C. (2017). Bayesian inference for diffusion-driven mixed-effect models. *Bayesian Analysis*, **12**, 435–463.

Wiqvist, S., Golightly, A., McLean, A. T., and Picchini, U. (2021). Efficient inference for stochastic differential equation mixed-effects models using correlated particle pseudo-marginal algorithms. *Computational Statistics and Data Analysis*, **157**, 107151.

Yu, Y. Q., Xiong, Y., Chan, Y. S., and He, J. F. (2004). Corticofugal gating of auditory information in the thalamus: An in vivo intracellular recording study. *J. Neurosci.*, **24**, 3060–3069.