Bayesian inference for linear stochastic differential equations with application to biological processes

Ashleigh Taylor McLean

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics, Statistics & Physics
Newcastle University
Newcastle upon Tyne
United Kingdom

December, 2020

Abstract

Stochastic differential equations (SDEs) provide a natural framework for describing the stochasticity inherent in physical processes that evolve continuously over time. In this thesis, we consider the problem of Bayesian inference for a specific class of SDE – one in which the drift and diffusion coefficients are linear functions of the state. Although a linear SDE admits an analytical solution, the inference problem remains challenging, due to the absence of a closed form expression for the posterior density of the parameter of interest and any unobserved components. This necessitates the use of sampling-based approaches such as Markov chain Monte Carlo (MCMC) and, in cases where observed data likelihood is intractable, particle MCMC (pMCMC). When data are available on multiple experimental units, a stochastic differential equation mixed effects model (SDEMEM) can be used to further account for between-unit variation. Integrating over this additional uncertainty is computationally demanding.

Motivated by two challenging biological applications arising from physiology studies of mice, the aim of this thesis is the development of efficient sampling-based inference schemes for linear SDEs. A key contribution is the development of a novel Bayesian inference scheme for SDEMEMs.

Acknowledgements

Here we are, finally at the end! I owe thanks to many people for helping me get to this point - on the top of that list by a long way, is my supervisor, Andrew Golightly; thank you so much for your guidance and support. You put up with a lot during our time (tears, confusion, never ending bugs in my code, not to mention the weekly gossip). You provided so much more than supervisory advice: you listened to my ever changing political opinions and worries, you supported my unusual lifestyle choices and recommended some great old school drum 'n' bass. I will dearly miss our sessions.

For the Djs, the MCs, the crew and the venues that provided the sound track that helped me will this thesis into being, I am forever grateful for the joy you bring. Thank you Mandidextrous. Thank you Inja. Thank you Lively Up, and thank you Worldies.

Also, I owe immense thanks to my Mum. Not only did you birth me, your continued belief in my abilities and vision for my future has fuelled my successes and achievements. I appreciate the patience and encouragement that you showed when I decided to go back to University to continue my academic journey, who would have thought that this is where we would be 10 years on? In addition, I must thank the rest of my family unit; your continual input ensures that I strive towards aspirational targets of personal excellence.

Furthermore, my deepest gratitude goes to Gabriel, whose unwavering support saw me through the darkest moments of this work, as well as celebrating its highlights. At times, when I struggled with confidence and my desire to continue, you guided me through my thinking process, aiding me to come to the conclusion to persist (not to mention your uncanny ability to put up with my various moods).

Finally, I am indebted to the best bunch of weirdos: Azza, Beth, Claire, Curry, Dan, Emily, Gary, Hind, Hjörvar, Jan, John, Jonny C, Jonny L, Kim, Laura, Lauren, Matt B, Matt Y, Mawds, Robyn and Sooz. Thank you so much for helping me unwind on those wild weekends and thank you for keeping me sane, or, more accurately, thanks for helping me return to sanity. I feel like the luckiest person alive to count you all as my friends and have you in my life.

Declaration

Parts of this thesis have been submitted for publication by the author:

- Chapter 5 and Chapter 7 have been accepted as: Samuel Wiqvista, Andrew Golightly, Ashleigh T. McLean, Umberto Picchini. 'Efficient inference for stochastic differential equation mixed-effects models using correlated particle pseudo-marginal algorithms', for *Computational Statistics & Data Analysis*.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Stochastic differential equations (SDEs) are arguably the most used and studied stochastic dynamic models. SDEs (Fuchs, 2013) allow the representation of stochastic time-dynamics, and are ubiquitous in applied research, most notably in finance (Steele, 2012), systems biology (Wilkinson, 2018), pharmacokinetic/pharmacodynamic modelling (Lavielle, 2014) and neuronal modelling (Saarinen *et al.*, 2006). SDEs extend the possibilities offered by ordinary differential equations (ODEs), by allowing random dynamics. As such, they can in principle replace ODEs in practical applications, to offer a richer mathematical representation for complex phenomena that are intrinsically non-deterministic.

However, in practice switching from ODEs to SDEs is usually far from trivial, due to the absence of closed form solutions to SDEs (except for the simplest toy problems), implying the need for numerical approximation procedures (Kloeden & Platen, 1992). Numerical approximation schemes, while useful for simulation purposes, considerably complicate statistical inference for model parameters. For reviews of inference strategies for SDE models, see e.g. Fuchs (2013) (including Bayesian approaches), Sørensen (2004) (classical approaches) and Wilkinson (2018) (for sampling based approaches in the biological context). Generally, in the non-Bayesian framework, the literature for parametric inference approaches for SDEs is vast, however there is no inference procedure that is applicable to general nonlinear SDEs that is also easy to implement on a computer. This is due to the lack of explicit transition densities for most SDE models.

In this thesis, we consider a linear class of SDE model, for which the governing transition densities are available in closed form. We assume that observations are available at discrete times, and the inferential goal is learning the SDE parameters and any unobserved dynamic states. By adopting a Bayesian approach to inference, our prior beliefs are encapsulated via a prior density, which is subsequently combined with the observed data likelihood to give a posterior density. Unfortunately, the latter is rarely tractable, and we turn to computationally intensive techniques to generate samples from the poste-

rior distribution. We place particular emphasis on Markov chain Monte Carlo (MCMC) methods (see e.g, book by Gamerman & Lopes (2006)) which will be used as the basis for inference.

In this thesis, we consider "repeated measurement experiments" (longitudinal data), modeled via mixed-effects, where the dynamics are Markov processes expressed via linear stochastic differential equations. These dynamics are assumed directly unobservable, i.e. are only observable up to measurement error. The practical goal is to fit observations pertaining to several individuals (i.e. independent experiments) simultaneously, by formulating state-space models having parameters randomly varying between those individuals. The resulting model is referred to as a *stochastic differential equation mixed-effects model* (SDEMEM). SDEMEMs are interesting because, in addition to explaining intrinsic stochasticity in the time-dynamics, they also take into account random variation between experimental units. The latter variation permits the understanding of between-subjects variability within a population. When considered in conjunction with an observation model, these two types of variability (population variation and intrinsic stochasticity) are separated from the third source of variation, namely residual variation (measurement error). Thanks to their generality, and the ability to separate the three levels of variation, SDEMEMs have attracted attention, see e.g. Donnet & Samson (2013*a*) for a review and Whitaker (2016) for a more recent account.

Here we review key papers on inference for SDEMEMs, and refer the reader to `https://umbertopicchini.github.io/sdemem/` for a comprehensive list of publications. Early attempts at inference for SDEMEMs use methodology borrowed from standard (deterministic) nonlinear mixed-effects literature such as FOCE (first order conditional estimation) combined with the extended Kalman filter, as in Overgaard *et al.* (2005). This approach could only deal with SDEMEMs having a constant diffusion term. The resulting inference is approximate maximum likelihood estimation, and no uncertainty quantification is given. Moreover, only Gaussian random effects are allowed and measurement error is also assumed Gaussian. Other maximum likelihood approaches are in Picchini *et al.* (2010) and Picchini & Ditlevsen (2011), where a closed-form series expansion for the unknown transition density is found using the method in Ait-Sahalia (2008), however the methodology could only be applied to reducible multivariate diffusions without measurement error. Donnet *et al.* (2010) discuss inference for SDEMEMs in a Bayesian framework. They implement a Gibbs sampler when the SDE (for each subject) has an explicit solution, and consider Gaussian random effects and Gaussian measurement error. When no explicit solution exists, they approximate the diffusion process using the Euler-Maruyama approximation. Donnet & Samson (2013*b*) construct an exact maximum likelihood strategy based on stochastic approximation Euler-Maruyama (SAEM), where latent trajectories are "proposed" via particle Markov chain Monte Carlo (Andrieu *et al.*,

2010). The major problem with using SAEM is the need for sufficient summary statistics for the "complete likelihood", which makes the methodology essentially impractical for arbitrarily complex models. Delattre & Lavielle (2013) also use SAEM, but they avoid the need for the (usually unavailable) summary statistics for the complete likelihood, and propose trajectories using the extended Kalman filter instead of particle MCMC. Unlike in Donnet & Samson (2013b), the inference in Delattre & Lavielle (2013) is approximate and measurement error and random effects are required to be Gaussian. Whitaker *et al.* (2017) work with the Euler-Maruyama approximation and adopt a data augmentation approach to integrate over the uncertainty associated with the latent diffusion process by employing carefully constructed bridge constructs inside a Gibbs sampler. A linear noise approximation (LNA) is also considered, see e.g. Golightly *et al.* (2015). However, the limitations are that the observation equation has to be a linear combination of the latent states and measurement error has to be Gaussian. In addition, constructing the bridge construct in the data augmentation approach or the LNA-based likelihood requires some careful analytical derivations. Consequently, neither approach can be regarded as a plug-and-play method (that is, a method that only requires forward simulation and evaluation of the measurement error density). In Picchini & Forman (2019), approximate and exact Bayesian approaches for a tumour growth study were considered: the approximate approach was based on synthetic likelihoods, where summary statistics of the data are used for the inference, while exact inference used pseudo-marginal methodology via an auxiliary particle filter, which is suited to target measurements observed with a small error. It was found that using a particle approach to integrate out the random effects was very time consuming. Even though the dataset was small (comprising 5–8 subjects to fit, depending on the experimental group, and around 10 observations per subject), the number of particles required in the procedure was in the order of thousands.

1.1 Contribution of this thesis

The aim of thesis is to investigate Bayesian inference for linear SDEs with focus on SDE-MEMs whose underlying dynamics are driven by linear SDEs. We consider separately, scenarios for which the observed data likelihood is tractable and intractable. The former arises, for example, when the linear SDE admits a (linear) Gaussian transition density and the observation process is both linear and Gaussian. In this case, the observed data likelihood can be calculated efficiently using a forward filter (Bucy & Joseph, 2005). Metropolis-within-Gibbs scheme is then used to alternate between draws of blocks consisting of parameters governing each experimental unit, the populations level parameters, and the parameters governing the observation process. An intractable observed data likelihood is encountered, for example, when a nonlinear observation model is assumed.

In this case, we derive a pseudo-marginal Metropolis-Hastings scheme, within which the intractable likelihood is replaced with an unbiased estimator thereof. A particle filter (see Andrieu *et al.*, 2009) is used to generate non-negative unbiased estimates (Del Moral *et al.*, 2006) giving an algorithm known as particle MCMC. We present a novel pMCMC scheme, applicable to SDEMEMs. Our contribution here is two-fold:

1. We consider a blocking strategy that reduces the variance of the acceptance probability of the move-step for the parameters common to all experimental units, and

2. we exploit recent advances based on correlated pMCMC (see e.g. Deligiannidis *et al.*, 2018).

We apply the resulting methodology in two challenging applications. The first application involves data consisting of hourly average temperature values in 20 mice, over a period of six months. Of the 20 mice, 10 were fed ad libitum (AL) and 10 were caloric restricted (CR), in which calorie intake is reduced but adequate nutrition is maintained. The goal of this study is to look at the ways in which mice compensate for the reduction in calories by studying their core body temperature. Caloric restriction has been shown to delay the onset of some cancers and other age related diseases in organisms such as yeast, worms, flies and mice (Weindruch & Walford, 1988). Most of the previous studies focus on lifespan (see Spindler, 2005) or cancer incidence (see Volk *et al.*, 1994) and few focus on the whole-animal physiological response to CR. One such study jointly models temperature and activity for late onset, short term caloric restriction (see Golightly *et al.*, 2019) and found that core body temperature was lower for CR mice, as found in other related work (see Weindruch & Walford, 1988; Duffy *et al.*, 1989; and Roth *et al.*, 2002). Our focus is to study how rodents may physiologically compensate for reduced food availability during long term late onset CR whilst attempting to capture the inherent stochasticity in the data, both within and between groups.

The second application considers synthetic data on tumour volume generated from the SDEMEM described in Picchini & Forman (2019). We fit the synthetic data using our novel pMCMC scheme. We seek to verify the accuracy of our approach and compare efficiency with standard pMCMC approaches and a linear noise approximation (LNA).

1.2   Organisation of the thesis

This thesis is organised as follows. In Chapter 2 we introduce Monte Carlo methods for estimating integrals including importance sampling and weighted resampling. This takes us to an outline of Markov chain Monte Carlo (MCMC) as a method for simulating from distributions whose densities are only available up to a normalising constant and we introduce some variations on this technique; the Metropolis-Hastings algorithm, the

Gibbs sampler, a pseudo-marginal MCMC scheme, and finally we introduce the correlated version of the pseudo-marginal MCMC scheme.

In Chapter 3 we build up an intuitive understanding of SDEs by first considering ODEs. We introduce Brownian motion and its properties, which form the building blocks of SDEs. Geometric Brownian motion is used as the motivation for requiring a method to integrate when Brownian motion is the integrator, introducing Itô integrals and the fundamental concepts of Itô calculus. We then generalise to multivariate processes. We focus on linear SDEs and review the different types by classifying them depending on the characteristics of the governing coefficients. We further describe general solutions to linear SDEs in the narrow sense and in the homogeneous case. Finally, we discuss how to create a linear SDE approximation of a non-linear SDE using the linear noise approximation (LNA).

In Chapter 4 we describe how to perform Bayesian inference for parameters that govern linear SDEs, given observations at discrete times that may be incomplete and subject to observational error. Inference in this case may be required for the joint distribution of the parameters and the latent process, but we focus on inference via the marginal parameter posterior given the observational data. We derive the forward filter which can be used to calculate the marginal likelihood when the observation model is linear, and we briefly discuss a backwards sweep to allow inference regarding the latent process. We describe an MCMC scheme that we can use to perform inference on the unknown parameters that makes use of the forward filter to calculate the marginal likelihood. We introduce having a nonlinear observation model so that the marginal likelihood is intractable and we look at how to linearise the model using the LNA from Chapter 3 allowing us to obtain a tractable (but approximate) marginal likelihood. By recalling Chapter 2 we adapt the theory of the pseudo-marginal MCMC scheme and the correlated version along with the use of a bootstrap particle filter to obtain estimates of the observed data likelihood. We conclude this chapter by performing Bayesian inference on an Ornstein-Uhlenbeck process comparing PMMH with CPMMH schemes.

In Chapter 5 we consider models for repeated measurement experiments where we allow for fixed effects and random effects. In the resulting mixed-effects framework, dynamics of each experimental unit are described by linear SDEs. We look at Bayesian inference in this framework considering linear and non-linear observation models. We utilise a Metropolis-within-Gibbs strategy and introduce auxiliary variables to allow pseudo-marginal Metropolis-Hastings updates followed by correlated pseudo-marginal Metropolis-Hastings updates. We again use a bootstrap particle filter to obtain estimates of the observed data likelihood for each experimental unit.

In Chapter 6 we work on a real data set comprising of minutely observations of mice temperatures over six months. The mice were subject to two different feeding regimes,

or treatments, and we seek to understand the physiological impact of these treatments via studying their core temperatures. We build an SDEMEM to jointly model core temperature over multiple experimental units (mice). This allows for the incorporation of intrinsic stochasticity inherent in observed temperature traces for time varying amplitude. We start by building an SDE for individual units which are linear in the narrow sense that we are then able to solve. Proceeding via Bayesian inference we utilise a forward filter to find the marginal likelihood within a Metropolis-Hastings scheme. Following this we extend the SDE to an SDEMEM allowing us to consider both fixed and mixed effects. The results show a clear distinction between the two treatments.

In Chapter 7 we use a stochastic differential mixed effects model to describe the tumour volume dynamics in mice receiving treatment for tumours. This is based on the work of Picchini & Forman (2019). We use their model to generate synthetic data and compare four approaches to perform Bayesian inference. The approaches we consider are a naive PMMH (where the auxiliary variables are updated with both the subject specific and common parameters), PMMH (where the auxiliary variables are only updated with the subject specific parameters), CPMMH and the LNA based approach. The CPMMH scheme shows a clear advantage in efficiency compared to the other schemes.

Finally, in Chapter 8 we discuss the thesis and results that we found and also present some ideas for future work.

Chapter 2

Monte Carlo methods

Monte Carlo methods use the law of large numbers and repeated sampling of random variables to approximate an expected value. Probabilities, integrals and summations can all be expressed as expectations, therefore Monte Carlo methods are widely used in Bayesian statistics, particularly when the posterior distribution is intractable. We begin by reviewing key concepts such as Monte Carlo integration, importance sampling and weighted resampling before giving an introduction to Markov chain Monte Carlo (MCMC).

2.1 Monte Carlo integration

Monte Carlo integration is a particular Monte Carlo method that uses random sampling to numerically compute an estimate of an integral. Suppose we want to evaluate an integral of the form

$$I = \int_D \phi(\theta)d\theta$$

over some domain D, for which there is no closed-form analytical solution. If the integrand can be written as

$$\phi(\theta) = \widetilde{\phi}(\theta)\pi(\theta)$$

for some density function π with support D, then the integral has the form

$$\int_D \phi(\theta)d\theta = \int_D \widetilde{\phi}(\theta)\pi(\theta)d\theta = E[\widetilde{\phi}(\Theta)]$$

where Θ is a random variable with PDF $\pi(\cdot)$. If we know how to generate independent realisations of Θ, say $\theta^{(1)}, \ldots, \theta^{(N)}$ then we may construct the estimate

$$\int_D \phi(x)dx = E[\widetilde{\phi}(\Theta)] \approx \frac{1}{N}\sum_{i=1}^{N}\widetilde{\phi}(\theta^{(i)}) = \hat{I}.$$

Abusing notation, we will also denote the corresponding estimator by $\hat{I}$. This method of approximating integrals is known as Monte Carlo integration. In its simplest form, the domain D is just an interval $[a, b]$ and we take the density $\pi(\theta)$ to be the uniform density $\frac{1}{(b-a)}$, $a < \theta < b$.

Note that $E(\hat{I}) = I$ and $Var(\hat{I}) \propto \frac{1}{N}$ assuming that $Var(\hat{\phi}(\Theta))$ is finite. Hence $\hat{I}$ is an unbiased and consistent estimator of I, and the estimate will converge to the true value of the expectation with large enough N. Now, consider again the variance of the estimator. That is

$$
\begin{aligned}
Var(\hat{I}) &= Var\left[\frac{1}{N}\sum\widetilde{\phi}(\Theta^{(i)})\right] \\
&= \frac{1}{N}Var\left[\widetilde{\phi}(\Theta)\right] \\
&= \frac{1}{N}\int_D \pi(\theta)\left(\widetilde{\phi}(\theta) - E\left[\widetilde{\phi}(\Theta)\right]\right)^2 d\theta \\
&= \frac{1}{N}\int_D \pi(\theta)\left(\widetilde{\phi}(\theta) - \int_D \widetilde{\phi}(y)\pi(y)dy\right)^2 d\theta.
\end{aligned}
$$

This integral is a measure of the roughness of the function $\widetilde{\phi}$. However, we usually cannot evaluate this integral either. Instead we work out the standard error of the estimator $\hat{I}$ which is an estimator of the variance. Explicitly

$$
\text{standard error}^2 = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\phi}(\theta^{(i)}) - \hat{I})^2.
$$

2.2   Importance sampling

Importance sampling is another way of estimating integrals and may give an estimator with smaller variance than the procedure described above. Suppose, as before we have

$$
I = \int_D \phi(\theta)d\theta = \int_D \widetilde{\phi}(\theta)\pi(\theta)d\theta
$$

but that we cannot easily simulate from $\pi(\cdot)$. Suppose that we can simulate from a density $g(\cdot)$ that has the same support as $\pi(\cdot)$. Then

$$
\begin{aligned}
\int_D \phi(\theta)d\theta &= \int_D \frac{\widetilde{\phi}(\theta)\pi(\theta)}{g(\theta)}g(\theta)d\theta \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\frac{\widetilde{\phi}(\theta^{(i)})\pi(\theta^{(i)})}{g(\theta^{(i)})} = \hat{I}_{is}.
\end{aligned}
$$

This is very similar to simple Monte Carlo integration, but we weight each contribution to the sum by the ratio $\pi(\theta^{(i)})/g(\theta^{(i)})$. It is easy to show that $\hat{I}_{is}$ gives an unbiased and consistent estimator of I, assuming $Var(\frac{\hat{\phi}(\Theta)\pi(\Theta)}{g(\Theta)})$ is finite. Note that the variance of the estimator can be reduced by finding an importance density $g(\cdot)$ that is a good approximation of $\pi(\cdot)\hat{\phi}(\cdot)$.

2.3 Weighted resampling

Consider now the problem of generating draws from $\pi(\cdot)$. The idea behind resampling methods is to simulate from one distribution that is easy to simulate from and then correct that value according to some criterion ensuring we get samples from the right distribution.

At the first step of the algorithm, N points $\{\theta^{(1)}, \ldots, \theta^{(N)}\}$ are sampled from some proposal density $g(\cdot)$. For each $\theta^{(j)}$, a (normalised) weight $w^{(j)}$ is constructed as

$$w^{(j)} = \frac{\pi(\theta^{(j)})/g(\theta^{(j)})}{\sum_{i=1}^{N} \pi(\theta^{(i)})/g(\theta^{(i)})}, \qquad j = 1, \ldots, N. \tag{2.1}$$

Finally, the second sample of size M (where $M = N$ in practice) is drawn from the discrete distribution on $\{\theta^{(1)}, \ldots, \theta^{(N)}\}$ with probabilities $\{w^{(1)}, \ldots, w^{(N)}\}$. The resulting sample $\{\theta^{(1)}, \ldots, \theta^{(M)}\}$ has approximate distribution $\pi(\cdot)$. Note that if only $\tilde{\pi}(\cdot) = k\pi(\cdot)$ can be evaluated (so that only the unnormalised target is available), then weighted resampling can still be applied. The weights are

$$w^{(j)} = \frac{\frac{1}{k}\tilde{\pi}(\theta^{(j)})/g(\theta^{(j)})}{\frac{1}{k}\sum_{i=1}^{N} \tilde{\pi}(\theta^{(i)})/g(\theta^{(i)})}, \qquad j = 1, \ldots, N.$$

and as the ks cancel we are left with the earlier form of (2.1). Hence weighted resampling is particularly useful in Bayesian statistics, where only the posterior density up to proportionality is known. The method can be justified as follows.

For simplicity, consider a univariate Θ. The distribution function of a univariate Θ generated by the algorithm is

$$\begin{aligned}
\widetilde{F}_{\Theta}(a) &= \sum_{j:\theta^{(j)} \leq a} w^{(j)} \\
&= \frac{\sum_{j=1}^{N} \pi(\theta^{(j)})/g(\theta^{(j)})\mathcal{I}(\theta^{(j)} \leq a)}{\sum_{i=1}^{N} \pi(\theta^{(i)})/g(\theta^{(i)})}
\end{aligned}$$

where $\mathcal{I}(\theta^{(j)} \leq a)$ takes the value 1 if $\theta^{(j)} \leq a$ and 0 otherwise. Taking $N \to \infty$ yields

$$\widetilde{F}_\Theta(a) \to \frac{\int_\Theta \left[\pi(\theta)/g(\theta)\right] \mathcal{I}(\theta \leq a) g(\theta) d\theta}{\int_\Theta \left[\pi(\theta)/g(\theta)\right] g(\theta) d\theta}$$
$$= \frac{\int_\Theta \pi(\theta) \mathcal{I}(\theta \leq a) d\theta}{\int_\Theta \pi(\theta) d\theta}$$
$$= P(\Theta \leq a)$$

Example

If $\pi(\theta)$ is only known up to proportionality, that is we have $\pi(\cdot) = \frac{\widetilde{\pi}(\cdot)}{k}$, then $k = \int \widetilde{\pi}(\cdot) d\theta$. It is clear that $\frac{1}{N} \Sigma_{j=1}^N \frac{\widetilde{\pi}(\theta^{(j)})}{g(\theta^{(j)})}$ is an unbiased (and consistent) estimator of k as

$$k = \int \frac{\widetilde{\pi}(\theta) g(\theta)}{g(\theta)} d\theta = E_g \left[\frac{\widetilde{\pi}(\Theta)}{g(\Theta)}\right].$$

We see that the average unnormalised weight gives an unbiased and consistent estimator of k.

Choosing a suitable proposal density can be far from straightforward in practice. We therefore consider another simulation based technique, Markov chain Monte Carlo, which has been ubiquitously applied in Bayesian statistics. In what follows we provide an intuitive introduction and refer the reader to Gamerman & Lopes (2006) for further details.

2.4 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a generic tool for simulating from distributions whose density may be known only up to proportionality, which is particularly useful for Bayesian inference. Suppose we have a generic target density $\pi(\theta)$ where $\theta = (\theta_1, \ldots, \theta_p)^T \in S$, from which we wish to generate samples. The MCMC strategy takes advantage of the fact that it is easy to simulate from a Markov chain. The basic strategy is as follows.

- Construct a Markov chain with stationary distribution $\pi(\theta)$.

- Simulate realisations of this chain.

- When the chain is in equilibrium, take the realisations as a (dependent) sample from $\pi(\theta)$.

- Use this sample to evaluate integrals/perform inference.

Thus, providing that the chain has converged, any value sampled will be from the density of interest $\pi(\cdot)$.

2.4.1 Metropolis-Hastings algorithm

Metropolis *et al.* (1953) introduced this algorithm which was generalised by Hastings (1970), hence the name Metropolis-Hastings. Central to the Metropolis-Hastings algorithm is the idea of a proposal density, denoted $q(\cdot|\cdot)$. It can be advantageous to have a proposal density which is easy to simulate from, however it need not (necessarily) have $\pi(\theta)$ as its stationary distribution. The Metropolis-Hastings algorithm is as follows:

1. Initialise the iteration counter to $j = 1$, and initialise the chain to $\theta^{(0)}$ chosen from somewhere in the support of $\pi(\theta)$.

2. Generate a proposed value θ^* using the proposal density $q(\theta^*|\theta^{(j-1)})$.

3. Evaluate the acceptance probability $\alpha(\theta^*|\theta^{(j-1)})$ of the proposed move, defined by

$$\alpha(\theta^*|\theta^{(j-1)}) = \min\left\{1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)}\right\}.$$

4. Put $\theta^{(j)} = \theta^*$ with probability $\alpha(\theta^*|\theta^{(j-1)})$; otherwise put $\theta^{(j)} = \theta^{(j-1)}$.

5. Put j to $j + 1$ and go to step 2.

At each stage a new value is generated from the proposal distribution. This is either accepted, in which case the chain moves, or is rejected, in which case the chain stays at the same point. Note that the target density $\pi(\cdot)$ only enters into the acceptance probability as a ratio, and so the method can be used when the target density is only known up to a multiplicative constant. Plainly, the algorithm defines a first order Markov chain. To see that the Markov chain has $\pi(\cdot)$ as an invariant distribution, we can check that the detailed balance equation (see e.g. Chapter 4 Gamerman & Lopes, 2006) is satisfied. Detailed balance is shown to hold for the chain $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots$ in the following way. First we need to obtain the transition kernel of the Markov chain. The transition kernel is given by

$$p(\phi|\theta) = \alpha(\phi|\theta)q(\phi|\theta) \qquad \text{when } \theta \neq \phi.$$

There is also a finite probability that the chain stays at θ, which depends on θ, and which we denote as $w(\theta)$. The full transition kernel is

$$p(\phi|\theta) = \alpha(\phi|\theta)q(\phi|\theta) + w(\theta)\delta(\theta - \phi)$$

where $\delta(\cdot)$ is the Dirac delta function ($\delta(\cdot) = 1$ when $\theta = \phi$ and $\delta(\cdot) = 0$ when $\theta \neq \phi$). To find $w(\theta)$, note that the probability of the chain staying at θ is 1 minus the probability

that the chain moves. Therefore

$$w(\theta) = 1 - \int_S q(\phi|\theta)\alpha(\phi|\theta)d\phi.$$

So the final expression for the transition kernel is

$$p(\phi|\theta) = q(\phi|\theta)\alpha(\phi|\theta) + \delta(\theta - \phi)\left[1 - \int_S q(\phi|\theta)\alpha(\phi|\theta)d\phi\right].$$

We can then check whether detailed balance holds:

$$\pi(\theta)p(\phi|\theta) = \pi(\theta)q(\phi|\theta)\min\left\{1, \frac{\pi(\phi)q(\theta|\phi)}{\pi(\theta)q(\phi|\theta)}\right\} + \pi(\theta)w(\theta)\delta(\theta - \phi)$$

$$= \min\left\{\pi(\theta)q(\phi|\theta), \pi(\phi)q(\theta|\phi)\right\} + \pi(\theta)w(\theta)\delta(\theta - \phi)$$

The first term is clearly symmetric in θ and ϕ. Also, the second term is symmetric because it is non-zero only when $\theta = \phi$. Detailed balance is satisfied since

$$\pi(\theta)p(\phi|\theta) = \pi(\phi)p(\theta|\phi)$$

and an easy consequence of this is that the Metropolis-Hastings algorithm defines a reversible Markov chain with stationary distribution $\pi(\cdot)$. It remains that we choose a suitable proposal density $q(\cdot|\cdot)$. In particular we want a chain that

- converges rapidly, and

- mixes well. That is, the chain

  - moves often and

  - moves around the support of $\pi(\cdot)$.

Some commonly used special cases are now considered.

2.4.2 Special cases of the Metropolis-Hastings algorithm

We now review some commonly used proposal mechanisms, and where appropriate give the form of the acceptance probability for the resulting scheme.

Symmetric chains

If the proposal distribution is symmetric

$$q(\theta^*|\theta) = q(\theta|\theta^*) \qquad \forall \theta, \theta^* \in S$$

then the acceptance probability simplifies to

$$\alpha(\theta^*|\theta) = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta)}\right\}$$

and hence does not involve the proposal density at all.

Random Walk chains

We can define the proposed move at iteration j to be $\theta^* = \theta^{(j-1)} + w^j$ where w^j is a $p \times 1$ random vector (completely independent of the state of the chain). Suppose that the w^j have density $g(\cdot)$ which is easy to simulate from. We can then simulate an innovation w^j and set the candidate point to $\theta^* = \theta^{(j-1)} + w^j$. The proposal distribution is then $q(\theta^*|\theta) = g(\theta - \theta^*)$ and this can be used to calculate the acceptance probability. Of course, if $g(\cdot)$ is symmetric about zero, then we have a symmetric random walk chain, and the acceptance probability does not depend on $g(\cdot)$.

Example: Normal innovations Suppose we take $w^j \sim N_p(0, \Sigma)$ and so the proposal distribution is

$$\theta^*|\theta \sim N_p(\theta, \Sigma).$$

We should therefore choose (or tune) Σ to maximise the efficiency of the algorithm (in terms of mixing). Note that choices of Σ that lead to large innovations will be capable of moving the sampler further around the parameter space but is likely to lead to a large number of rejected proposals. Conversely, choices of Σ that lead to small innovations will lead to lots of small steps around the parameter space. This suggests an optimal value of Σ.

Under certain constraints of the target distribution (see Roberts & Rosenthal, 2001), it has been shown that the optimal choice of Σ (for large p) is

$$\Sigma = \frac{2.38^2}{p} Var(\Theta)$$

and this leads to an optimal acceptance rate of 0.234; see e.g. Sherlock *et al.* (2013) for further details. Of course, we typically don't know the variance $Var(\Theta)$. However, we could first run the MCMC algorithm (for example by using the prior variance $Var(\Theta)$ in place of Σ) to obtain an estimate of $Var(\Theta)$. We should also note that in practice, and especially for small p, the above formula for Σ should just be used as a guide – an acceptance rate anywhere between 0.1 and 0.4 could be close to optimal.

Finally, note that for large p, sampling a new θ^* from a multivariate Normal can be expensive. An alternative approach is to take the components of $w^j = (w_1^j, \ldots, w_p^j)^T$

as iid (univariate) Normal random variables. That is, for a component $i, w_i^j \sim N(0, s_i^2)$ where, for example, $s_i^2 = \frac{2.38^2}{p} Var(\Theta_i)$.

Independence chains

In this case the proposed transition is completely independent of the current position of the chain and so $q(\theta^*|\theta) = g(\theta^*)$ for some density $g(\cdot)$. The acceptance probability becomes

$$\alpha(\theta^*|\theta) = \min \left\{ 1, \frac{\pi(\theta^*)}{g(\theta^*)} \times \frac{g(\theta)}{\pi(\theta)} \right\}$$

In this case, $g(\cdot)$ should ideally be as close to $\pi(\cdot)$ as possible, to give an acceptance rate close to 1.

Componentwise transitions

Constructing a suitable $q(\cdot|\cdot)$ may be difficult. Moreover, for many problems of interest, full conditional distributions (FCDs) may be available for directly sampling from a subset of components of θ.

Denote the FCD for the ith component of θ by $\pi(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p)$. The algorithm is as follows:

1. Initialise the iteration counter to $j = 1$.

2. Obtain a new value $\theta^{(j)}$ from $\theta^{(j-1)}$ by successive generation of values.

   - $\theta_1^{(j)} \sim \pi(\theta_1|\theta_2^{(j-1)}, \ldots, \theta_p^{(j-1)})$ using a Metropolis-Hastings step with proposal distribution $q_1(\theta_1|\theta_1^{(j-1)})$
   - $\theta_2^{(j)} \sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_p^{(j-1)})$ using a Metropolis-Hastings step with proposal distribution $q_2(\theta_2|\theta_2^{(j-1)})$.
   $\vdots$
   - $\theta_p^{(j)} \sim \pi(\theta_p|\theta_1^{(j)}, \ldots, \theta_{p-1}^{(j)})$ using a Metropolis-Hastings step with proposal distribution $q_p(\theta_p|\theta_p^{(j-1)})$.

3. Change counter j to $j + 1$ and return to step 2.

This is in fact the original form of the Metropolis algorithm. Proving that $\pi(\theta)$ is the stationary distribution of a Markov chain defined in this way can be achieved by induction (see Gamerman & Lopes, 2006). Note that the Metropolis-Hastings algorithm as presented in Section 2.4.1 can be seen as a special case of the above algorithm. If the FCD is available for sampling from directly, for a particular component θ_i, it is easy to show that the resulting acceptance probability is 1. When all FCDs are available for sampling from, we obtain an algorithm known as the Gibbs sampler.

2.4.3 Gibbs sampler

The Gibbs sampler (or generically Gibbs sampling) originated in the field of image processing. It was introduced by Gemen & Gemen (1984) before being brought to the attention of the larger statistical community by Gelfand & Smith (1990). In essence the Gibbs sampler is an MCMC scheme in which the full conditional distributions are used to form the transition kernel. Assume that, for all components of θ, the full conditional distributions are available and can easily be sampled from. The algorithm has the following form:

1. Initialise the iteration counter to $j = 1$
   Initialise the state of the chain to $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})^T$

2. Obtain a new value $\theta^{(j)}$ from $\theta^{(j-1)}$ by successive generation of values

   - $\theta_1^{(j)} \sim \pi(\theta_1 | \theta_2^{(j-1)}, \ldots, \theta_p^{(j-1)})$
   - $\theta_2^{(j)} \sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_p^{(j-1)})$
   - $\vdots$
   - $\theta_p^{(j)} \sim \pi(\theta_p | \theta_1^{(j)} \ldots, \theta_{p-1}^{(j)})$

3. Change counter j to $j + 1$ and return to step 2.

This clearly defines a homogeneous Markov chain, as each simulated value depends only on the previous simulated value and not on any previous value or the iteration counter j. However, we need to show that $\pi(\theta)$ is a stationary distribution of the chain. The transition kernel is

$$p(\phi|\theta) = \prod_{i=1}^{p} \pi(\phi_i | \phi_1, \ldots, \phi_{i-1}, \theta_{i+1}, \ldots, \theta_p).$$

Therefore, we need to check that $\pi(\theta)$ is the stationary distribution of the chain with this transition kernel. Unfortunately, the form of the Gibbs sampler specified here is not reversible, so we cannot use detailed balance. We need to check stationarity directly, ie prove that

$$\pi(\phi) = \int p(\phi|\theta)\pi(\theta)d\theta$$

For the bivariate case we have:

$$\int p(\phi|\theta)\pi(\theta)d\theta = \int \pi(\phi_1|\theta_2)\pi(\phi_2|\phi_1)\pi(\theta_1,\theta_2)d\theta_1 d\theta_2$$

$$= \pi(\phi_2|\phi_1)\int\int \pi(\phi_1|\theta_2)\pi(\theta_1,\theta_2)d\theta_1 d\theta_2$$

$$= \pi(\phi_2|\phi_1)\int \pi(\phi_1|\theta_2)\left[\int \pi(\theta_1,\theta_2)d\theta_1\right]d\theta_2$$

$$= \pi(\phi_2|\phi_1)\int \pi(\phi_1|\theta_2)\pi(\theta_2)d\theta_2$$

$$= \pi(\phi_2|\phi_1)\pi(\phi_1)$$

$$= \pi(\phi_1,\phi_2)$$

$$= \pi(\phi).$$

The general case is similar, see e.g. Gamerman & Lopes (2006) for further details.

Example   Consider a bivariate normal target with zero mean and unit variance for the marginals, and a correlation of ρ between the two components. The target density is

$$f(\theta_1,\theta_2) \propto \exp\left\{-\frac{1}{2\rho^2}(\theta_1^2 + \theta_2^2 - 2\rho\theta_1\theta_2)\right\}, \qquad -\infty < \theta_1,\theta_2 < \infty.$$

We construct a Gibbs sampler for this target. The full conditional density for Θ_1 is

$$f(\theta_1|\theta_2) = \frac{f(\theta_1,\theta_2)}{f(\theta_2)}$$

$$\propto f(\theta_1,\theta_2)$$

$$\propto \exp\left\{-\frac{1}{2(1-\rho^2)}(\theta_1^2 - 2\rho\theta_1\theta_2)\right\}$$

$$\propto \exp\left\{-\frac{1}{2(1-\rho^2)}(\theta_1 - \rho\theta_2)^2\right\}.$$

Hence, we recognise the full conditional distribution for Θ_1 as

$$\Theta_1|\Theta_2 = \theta_2 \sim N(\rho\theta_2, 1-\rho^2).$$

By symmetry, the full conditional distribution for Θ_2 is

$$\Theta_2|\Theta_1 = \theta_1 \sim N(\rho\theta_1, 1-\rho^2).$$

The Gibbs sampler for the target above then has the following form

1. Initialise the iteration counter to $j = 1$

Initialise the state of the chain to $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)})^T$

2. Obtain a new value $\theta^{(j)}$ from $\theta^{(j-1)}$ by successive generation of values

   - $\theta_1^{(j)} \sim N(\rho\theta_2^{(j-1)}, 1 - \rho^2)$
   - $\theta_2^{(j)} \sim N(\rho\theta_1^{(j)}, 1 - \rho^2)$

3. Change counter j to $j + 1$ and return to step 2.
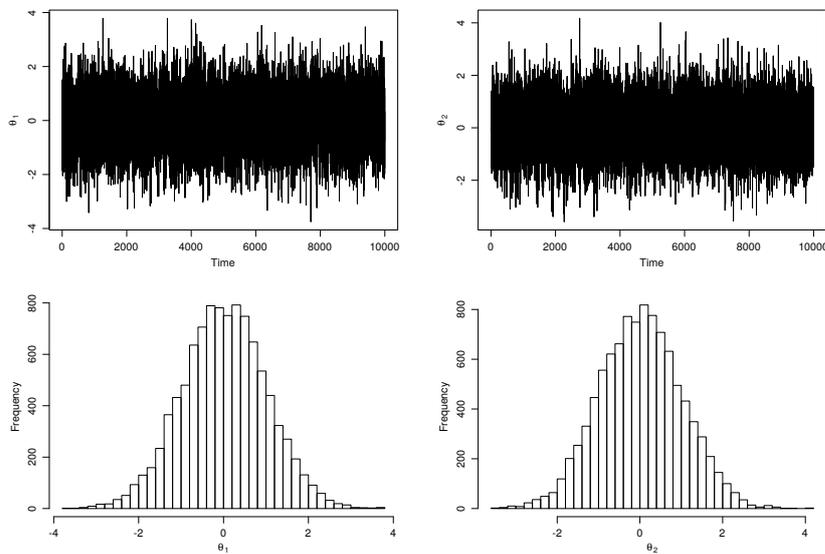
Figure 2.1: Trace plot (upper panel) and histogram (lower panel) from Gibbs sampler targeting a bivariate normal, with zero mean and unit variance with a correlation of $\rho = 0.1$.

Figures 2.1 – 2.4 summarise the output of this Gibbs sampler for $\rho \in \{0.1.0.5, 0.9, 0.999\}$. Trace plots can be used as a visual aid to assess mixing of the Markov chain. Trace plots are constructed by plotting each Metropolis-Hastings sample against the iteration number. Ideally, the trace plot should look like a "thick pen". The trace plot can also be used to diagnose poor mixing (for example, as a result of a small acceptance probability showing periods of sticking at the same value), assess burn-in and convergence. Although formal convergence tests are possible (Gamerman & Lopes, 2006), we typically eschew these in favour of a simple graphical approach. As ρ approaches 1, we see the mixing deteriorate, as the sample is slow to explore the parameter space, due to the high correlation between Θ_1 and Θ_2 that can be seen in Figure 2.5.
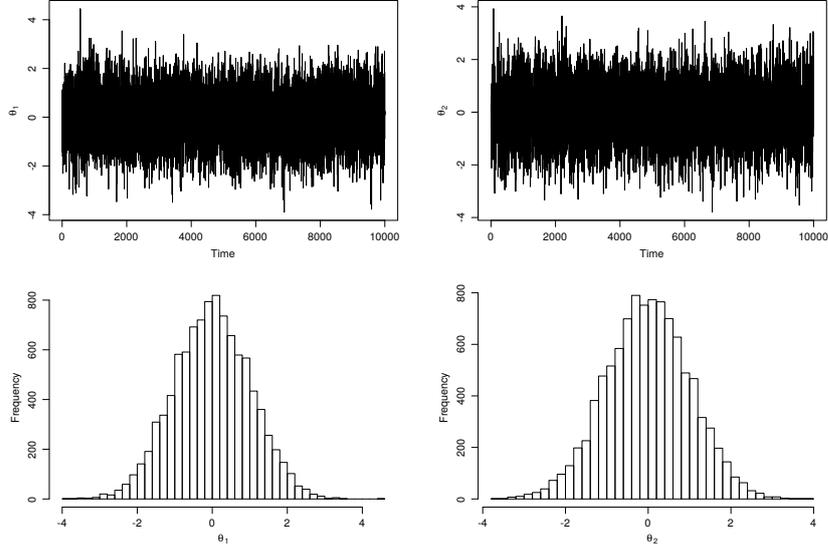
Figure 2.2: Trace plot (upper panel) and histogram (lower panel) from Gibbs sampler targeting a bivariate normal, with zero mean and unit variance with a correlation of $\rho = 0.5$.

2.5 Pseudo-marginal MCMC

Suppose that only an estimate of $\pi(\theta)$ is available. This may occur, for example, if $\pi(\cdot)$ is obtained through an integration (by marginalising out nuisance variables). We will denote the estimator by $\hat{\pi}_U(\theta)$ where U denotes all the random variables used to construct the estimator. Given $U \sim g(u)$, the corresponding estimate is denoted $\hat{\pi}_u(\theta)$.

Suppose that $E_U(\hat{\pi}_U(\theta)) = \pi(\theta)$ where $U \sim g(\cdot)$ so that the estimator is unbiased. Consider a joint density over U and θ of the form

$$\hat{\pi}(\theta, u) \propto \hat{\pi}_u(\theta)g(u) \tag{2.2}$$

Running a Metropolis-Hastings scheme with joint proposal density $q(\theta^*|\theta)g(u^*)$ yields an acceptance probability of $\min\{1, A\}$ where

$$
\begin{aligned}
A &= \frac{\hat{\pi}(\theta^*, u^*)}{\hat{\pi}(\theta, u)} \times \frac{q(\theta|\theta^*)g(u)}{q(\theta^*|\theta)g(u^*)} \\
&= \frac{\hat{\pi}_{u^*}(\theta^*)g(u^*)}{\hat{\pi}_u(\theta)g(u)} \times \frac{q(\theta|\theta^*)g(u)}{q(\theta^*|\theta)g(u^*)} \\
&= \frac{\hat{\pi}_{u^*}(\theta^*)}{\hat{\pi}_u(\theta)} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}
\end{aligned}
$$
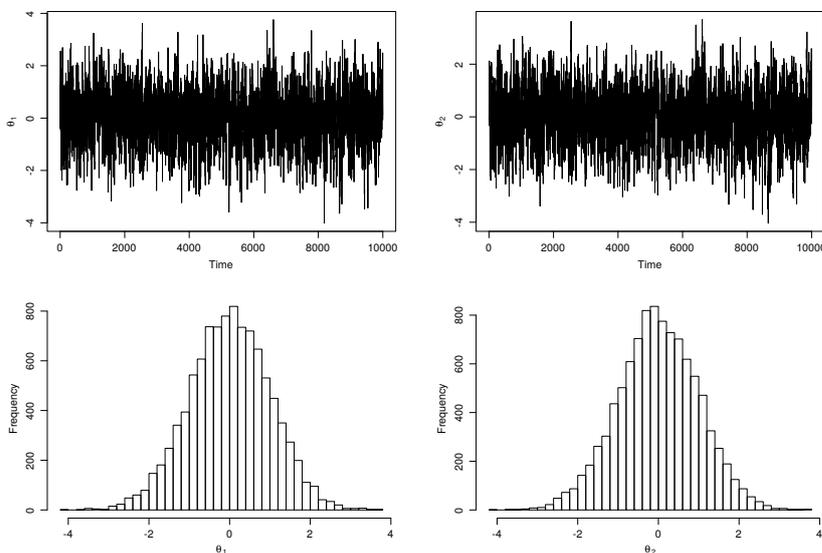
Figure 2.3: Trace plot (upper panel) and histogram (lower panel) from Gibbs sampler targeting a bivariate normal, with zero mean and unit variance with a correlation of $\rho = 0.9$.

We recognise this as the acceptance probability of an idealised Metropolis-Hastings scheme targeting $\pi(\theta)$, with the estimates $\hat{\pi}_u(\theta)$ and $\hat{\pi}_{u^*}(\theta^*)$ used in place of $\pi(\theta)$ and $\pi(\theta^*)$. Nevertheless, this scheme can be shown to exactly target $\pi(\cdot)$ since

$$\int \hat{\pi}(\theta, u) du = \int \hat{\pi}_u(\theta) g(u) du$$
$$= E_U[\hat{\pi}_U(\theta)]$$
$$\propto \pi(\theta).$$

2.5.1 Illustration

Consider a $N(0,1)$ target with density proportional to $\exp\left\{-\frac{1}{2}\theta^2\right\}$. To illustrate the pseudo-marginal approach, we will take a joint density $\hat{\pi}(\theta, u) \propto \pi(\theta) u g(u)$ where $U \sim LN\left(-\frac{\tau^2}{2}, \tau^2\right)$.
Note that

$$\int \hat{\pi}(\theta, u) du \propto \pi(\theta) \int u g(u) du$$
$$\propto \pi(\theta) E(U)$$
$$\propto \pi(\theta),$$
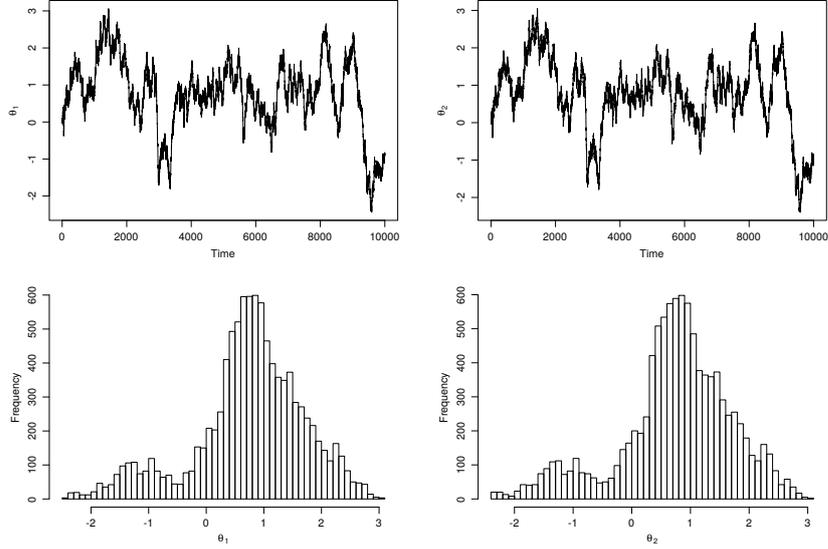
Figure 2.4: Trace plot (upper panel) and histogram (lower panel) from Gibbs sampler targeting a bivariate normal, with zero mean and unit variance with a correlation of $\rho = 0.999$.

since it is easy to show that $E(U) = 1$. Equivalently, let $\hat{\pi}_u(\theta) = \pi(\theta)u$, which is unbiased for $\pi(\theta)$ since

$$E_u(\hat{\pi}_u(\theta)) = \pi(\theta)E(U) = \pi(\theta).$$

We run a standard Metropolis-Hastings scheme targeting $\hat{\pi}(\theta, u)$ with proposal density $q(\theta^*, u^*|\theta, u) = q(\theta^*|\theta)g(u^*)$ and we calculate the acceptance probability as

$$\min \left\{ 1, \frac{\hat{\pi}(\theta^*, u^*)}{\hat{\pi}(\theta, u)} \times \frac{q(\theta, u|\theta^*, u^*)}{q(\theta^*, u^*|\theta, u)} \right\}$$

$$\min \left\{ 1, \frac{u^*\pi(\theta^*)g(u^*)}{u\pi(\theta)g(u)} \times \frac{q(\theta|\theta^*)g(u)}{q(\theta^*|\theta)g(u^*)} \right\}$$

$$\min \left\{ 1, \frac{\pi(\theta^*)u^*}{\pi(\theta)u} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \right\}$$

$$\min \left\{ 1, \frac{\hat{\pi}_{u^*}(\theta^*)}{\hat{\pi}_u(\theta)} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \right\}.$$

Running this method for 10,000 iterations with a Gaussian random walk proposal mechanism and an initial value of θ to be zero, we can see the effect of varying τ. Figure 2.6 shows that increasing τ makes the exploration of the space 'sticky', meaning the chain gets stuck in one place for a while as the proposals are all being rejected. As τ is increased
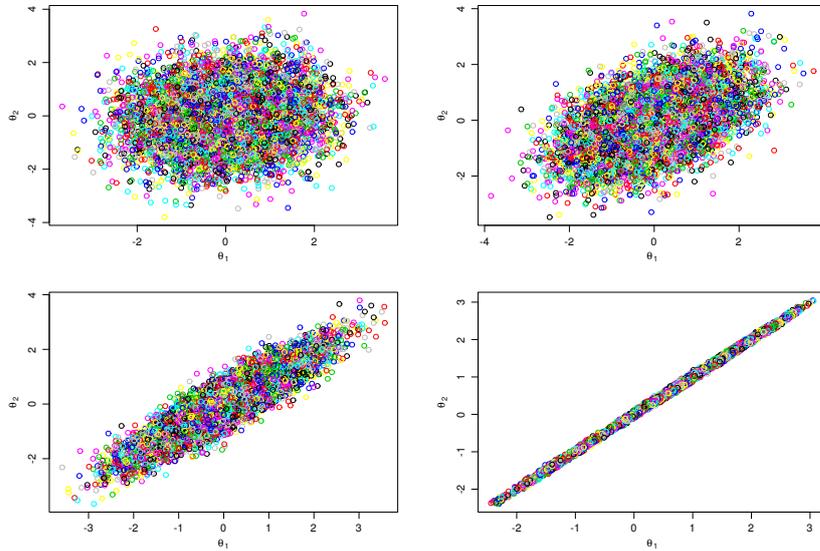
Figure 2.5: Plots of Θ_1 vs Θ_2 from top left to bottom right with $\rho = 0.1$, $\rho = 0.5$, $\rho = 0.9$ and $\rho = 0.999$ respectively.

the variance of U increases as

$$Var(U) = e^{-\tau^2 + \tau^2}(e^{\tau^2} - 1)$$
$$= e^{\tau^2} - 1.$$

Hence, for large τ, large values of $\hat{\pi}_{u^*}(\theta^*)$ are possible, leading to acceptance of θ^* followed by long periods of rejection, until a suitably large value of $\hat{\pi}_{u^*}(\theta^*)$ is generated, allowing the chain to move on.

As a further diagnostic check, we may compute the effective sample size (ESS) for each chain corresponding to $\tau \in \{0.1, 2, 5\}$. ESS is the equivalent number of independent samples, obtained as

$$ESS = \frac{n_{iters}}{1 + \sum_{k=1}^{\infty} \psi(k)}$$

where $\psi(k)$ denotes the lag-k auto-correlation. Using the coda package in R (Plummer *et al.*, 2006) we obtain effective sample sizes for $\tau = 0.1, 2$ and 5 as 1167, 242 and 100 respectively.

In order to reduce the variance of the Metropolis-Hastings acceptance ratio (and alleviate the sticky behaviour of the chain) we consider a simple modification of the pseudo-marginal Metropolis-Hastings scheme, whereby positive correlation is introduced between successive values of u and u^*.
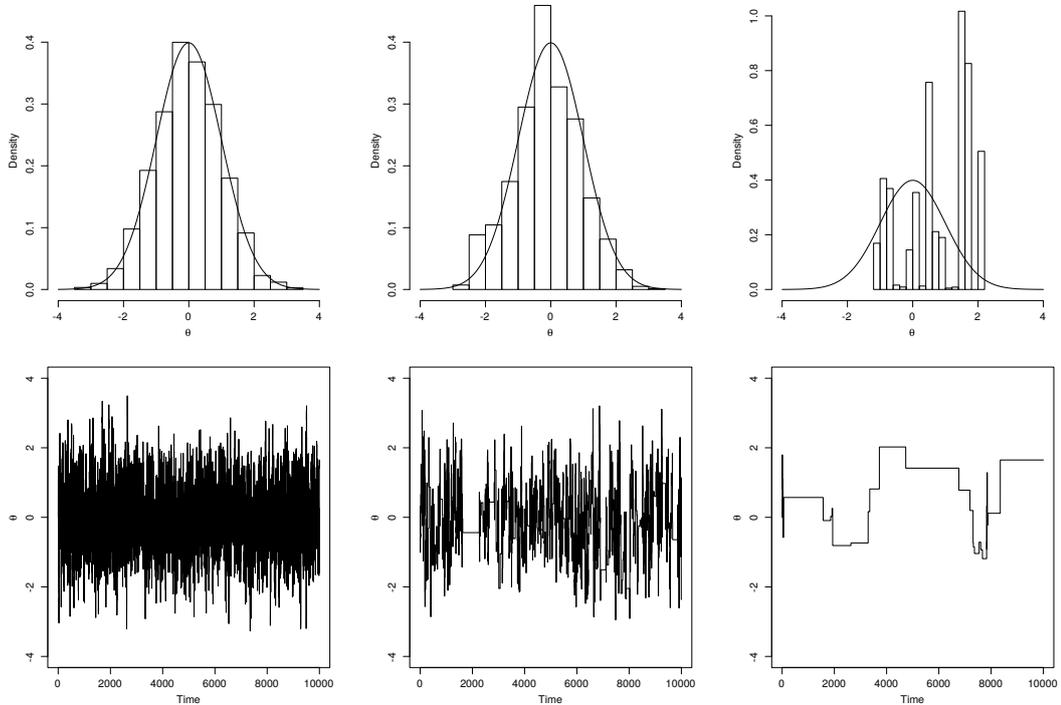
Figure 2.6: Histogram and trace plot of samples of θ using pseudo-marginal Metropolis-Hastings with 10,000 iterations and a starting value of zero. The $N(0,1)$ target density is overlaid. From left to right $\tau = 0.1, 2$, and 5

2.6   Correlated pseudo-marginal MCMC

We see in PMCMC that a highly variable $\hat{\pi}_u(\theta)$ results in 'sticky' MCMC chains. Correlated PMCMC (CPMCMC) allows us to induce positive correlation between successive values of $\hat{\pi}_u(\theta)$, which in turn reduces the variance of the pseudo marginal acceptance ratio.

Consider again the joint density given in (2.2). Suppose that $g(u) = N(u; 0, I_d)$ where I_d is the $d \times d$ identity matrix and d is the dimension of the auxiliary variables u. The correlated pseudo-marginal scheme (Deligiannidis *et al.*, 2018, Dahlin *et al.*, 2015) recognises that rather than propose u^* from $g(\cdot)$, we may use a kernel $k(u^*|u)$, carefully chosen to allow correlation between u and u^* (and therefore $\hat{\pi}_u(\theta)$ and $\hat{\pi}_{u^*}(\theta^*)$). In particular $k(u^*|u)$ should be g-invariant, that is the detailed balance equation

$$k(u^*|u)g(u) = k(u|u^*)g(u^*)$$

should be satisfied. A proposal kernel that meets these requirements is the Crank-Nicolson

proposal (Cotter *et al.*, 2013) which has density

$$k(u^*|u) = N(u^*; \rho u, (1 - \rho^2)I_d) \tag{2.3}$$

and choosing ρ to be close to 1 will induce the required positive correlation between $\hat{\pi}_u(\theta))$ and $\hat{\pi}_{u^*}(\theta^*)$. In the case where $\rho = 0$ we recover $k(u^*|u) = g(u^*)$ which corresponds to the PMCMC scheme. The role of correlation here should not be confused with the effect of correlation between components in a Gibbs sampler. We emphasise that correlation is used in CPMMH as a variance reduction technique.

The acceptance probability used in the CPMCMC scheme is $\min\{1, A\}$ where $A = \frac{\hat{\pi}_{u^*}(\theta^*)}{\hat{\pi}_u(\theta)} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}$, and remains unchanged from the PMCMC scheme. The Crank-Nicolson proposal in equation (2.3) is g-invariant where $g(\cdot)$ is a standard normal density. In practice, the innovations u required to generate $\hat{\pi}_u(\theta)$ may not be Gaussian. Nevertheless, innovations from commonly used densities can be obtained by applying the inverse Gaussian CDF to u to obtain standard uniform quantities, followed by the inverse CDF of the desired distribution. The technique is known as the inversion method (see e.g. Gamerman & Lopes, 2006).

2.6.1   Illustration

Consider again the scenario in Section 2.5.1 with a $N(0,1)$ target and joint density $\hat{\pi}(\theta, u) \propto \pi(\theta) u g(u)$ where $u \sim LN(-\frac{\tau^2}{2}, \tau^2)$ and therefore $g(u) = LN(u; -\frac{\tau^2}{2}, \tau^2)$. To illustrate the CPMCMC scheme, let us rewrite $\hat{\pi}(\theta, u)$ as

$$\hat{\pi}(\theta, u) \propto \pi(\theta) g(u) \exp\{-\frac{\tau}{2} + \tau u\}$$

where $g(u) = N(u; 0, 1)$. It is easily checked that

$$\int \hat{\pi}(\theta, u) du \propto \pi(\theta) \int_{-\infty}^{\infty} \exp\{-\frac{\tau}{2} + \tau u\} g(u) du$$

$$\propto \pi(\theta) E_g(\exp\{-\frac{\tau}{2} + \tau U\})$$

$$\propto \pi(\theta)$$

since $\exp\{-\frac{\tau}{2} + \tau U\} \sim LN(-\frac{\tau}{2}, \tau^2)$ with expectation 1.

We ran CPMCMC with a Gaussian random walk proposal for 10,000 iterations with $\tau \in 0.1, 2, 5$ and $\rho = 0.99$. Each MCMC scheme ran for 10,000 iterations and the effective sample sizes for $\tau = 0.1, 2$ and 5 were 1766, 1705, and 1029 respectively, this is considerably better than the effective sample size when the innovations for u were not correlated.

As discussed in Deligiannidis *et al.* (2018), care must be taken on choosing ρ. For $\rho \approx 1$, the sampler will fail to adequately mix u, resulting in long range dependence in
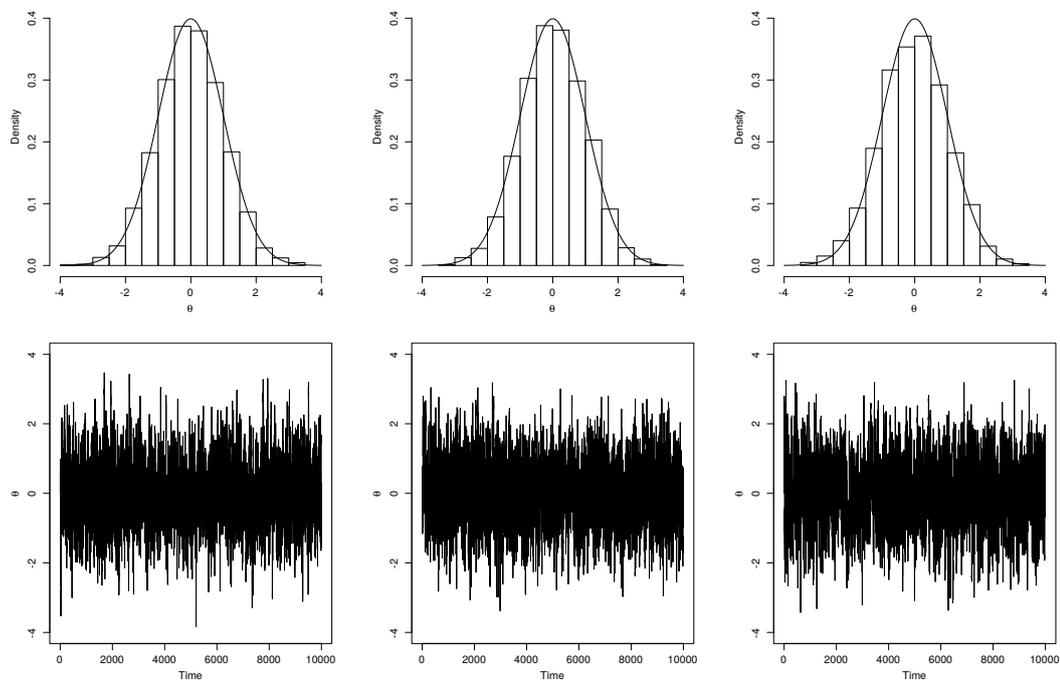
Figure 2.7: Histogram and trace plot of samples of θ using pseudo-marginal Metropolis Hastings with 10,000 iterations and a starting value of zero. From left to right $\tau = 0.1, 2$, and 5

the θ chain. We consider the effect of ρ in the context of an empirical application in Chapter 7.

Chapter 3

Stochastic Differential Equations

Stochastic differential equations are used as opposed to ordinary differential equations (ODEs) in order to capture the inherent stochasticity that may be present in a given process. These processes could be temperatures over time, see e.g. Figure 3.1, stock exchange prices, or the population size of a species, to name a few. In what follows
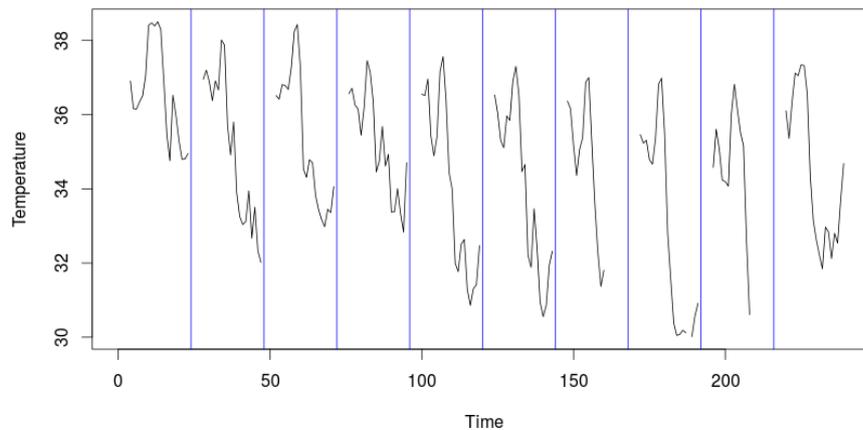
Figure 3.1: Hourly average temperature of a mouse.

we give an intuitive introduction to SDEs, by first considering ODEs. For a rigorous introduction to SDEs, we refer the reader to Øksendal (1995).

3.1 ODEs to SDEs

Let us take the evolution of the number of cells x_t in an organism infected by a virus. Suppose that over an infinitesimal time period the growth rate is θx_t. The ODE system

characterising this process is therefore

$$\frac{dx_t}{dt} = \theta x_t, \qquad X_0 = x_0 \qquad \Rightarrow \qquad X_t = x_0 e^{\theta t}.$$

Although the solution to this simple ODE is available, analytic solutions of generic ODE system are rarely tractable. In the absence of an analytic solution, an approximate numerical solution can be used. We therefore compare the analytic solution $x_{t+\Delta t} = x_t e^{\theta \Delta t}$ with a simple numerical solution – the Euler approximation. The Euler approximation has that

$$\frac{dx_t}{dt} \approx \frac{x_{t+\Delta t} - x_t}{\Delta t}$$

which we rearrange to give $x_{t+\Delta t} = x_t + \theta x_t \Delta t$. We can see several numerical solutions plotted in Figure 3.2 for decreasing Δt. As $\Delta t \to 0$, the analytic solution is obtained. Assume now that the growth rate is on average θ, but that there are fluctuations due to
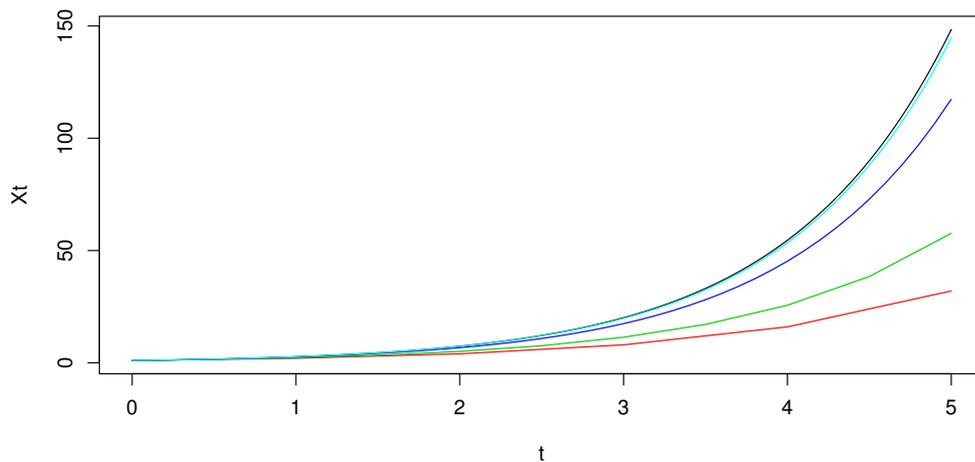
Figure 3.2: Growth of a virus with $\theta = 1$ and $x_0 = 1$. Analytic solution (black) and Euler approximations (coloured) with $\Delta t = 1$ (red), $\Delta t = 0.5$ (green), $\Delta t = 0.1$ (dark blue), and $\Delta t = 0.01$ (light blue).

changing unpredictable biological conditions. So, at any given time the growth rate is θ + 'noise'. Using uppercase to denote the resulting stochastic process, we obtain

$$\frac{dX_t}{dt} = (\theta + \text{'noise'})X_t, \qquad X_0 = x_0$$

and we specify the random noise terms so that their mean is 0. Figure 3.3 shows a curve described by a deterministic ODE and corresponding curves that take into account stochasticity by including the noise term. We denote the noise term as W_t. To formalise
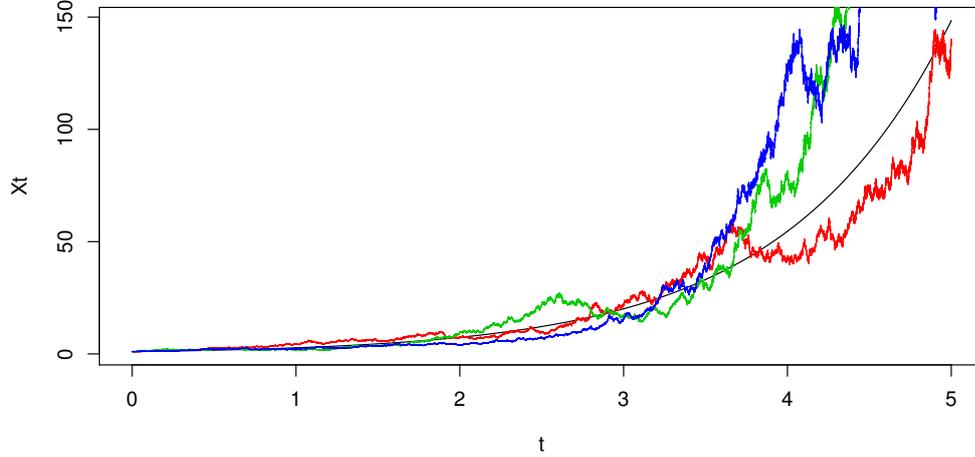
Figure 3.3: Deterministic growth curve of a virus (black) where $\theta = 1$ and $x_0 = 1$ and realisations of growth of the virus taking unpredictable fluctuations into account (coloured).

the role of W_t, take an ordinary integral equation of the form

$$X_t = x_0 + \int_0^t \alpha(X_s)ds + \sigma W_t, \qquad t \geq 0 \tag{3.1}$$

where α is some differentiable function and σ is a positive scaling constant used to tune the effect of W on X. For $\sigma = 0$ we obtain a deterministic integral equation with equivalent differential form given by $dX_t = \alpha(X_t)dt$. We wish for W_t to capture the discrepancy between the smooth and rough path. At any time t, W_t has to be a random variable, hence $\{W_t, t \geq 0\}$ will be a stochastic process and we need $W_0 = 0$ to recover x_0 at time 0. Working with increments of X we have

$$X_{t+\Delta t} - X_t = \int_t^{t+\Delta t} \alpha(X)ds + \sigma(W_{t+\Delta t} - W_t)$$
$$\simeq \alpha(X_t)\Delta t + \sigma(W_{t+\Delta t} - W_t).$$

We would then like the process W_t to have the following four properties:

1. The expected value of the noise increment should be zero

$$E[W_{t+\Delta t} - W_t] = 0 \qquad \forall t, \Delta t.$$

2. The increment $W_{t_1} - W_{t_0}$ is independent of the increment $W_{t_2} - W_{t_1}$ for all times $0 < t_0 < t_1 < t_2 < \infty$ and the increments are identically distributed.

3. The distribution of $W_{t+\Delta t} - W_t$ depends only on $|\Delta t|$ and not on t.

4. W has differentiable sample paths.

However, such a stochastic process does not exist, but there is a process satisfying 1-3 with continuous sample paths. This process is known as (standard) Brownian motion (BM) also known as the Wiener process and is described below. Note that the integral equation in (3.1) can be equivalently written in differential form as

$$dX_t = \alpha(X_t)dt + \sigma dW_t$$

which is known as a stochastic differential equation (SDE). Since W and X are not differentiable the above SDE only makes sense when integrated with respect to t, in which case the integral equation is recovered. Therefore, we obtain

$$\int_0^t dW_s = W_t - W_0 = W_t.$$

Formally, $\{W_t, t \geq 0\}$ is a standard Brownian motion (BM) if W_t depends continuously on t and the following three assumptions hold

1. $W_0 = 0$.

2. For all times $0 \leq t_0 < t_1 < t_2 < \infty$, the increment $W_{t_2} - W_{t_1}$ is independent of the increment $W_{t_1} - W_{t_0}$.

3. For all times $0 \leq t_0 < t_1 < \infty$, $W_{t_1} - W_{t_0} \sim N(0, t_1 - t_0)$.

Several important properties can be deduced from the definition. For example the distribution of $W_{t_i}|W_{t_{i-1}}$ for $t_{i-1} < t_i$ are deduced by writing $W_{t_i} = W_{t_i} - W_{t_{i-1}} + W_{t_{i-1}}$ which gives

$$W_{t_i}|W_{t_{i-1}} = w_{t_{i-1}} \sim N(w_{t_{i-1}}, t_i - t_{i-1}). \tag{3.2}$$

Hence, for an interval $[0,T]$ partitioned as $0 = t_0 < t_1 < \ldots < t_n = T$ the process can be simulated at these discrete times by recursively drawing from (3.2). A realisation of a continuous-time process at discrete times is known as a skeleton path. Figure 3.4 shows four realisations of BM giving random walk-like behaviour.

3.1.1 SDE examples

We now provide some illustrative examples of SDEs and motivate the need for a stochastic integral.
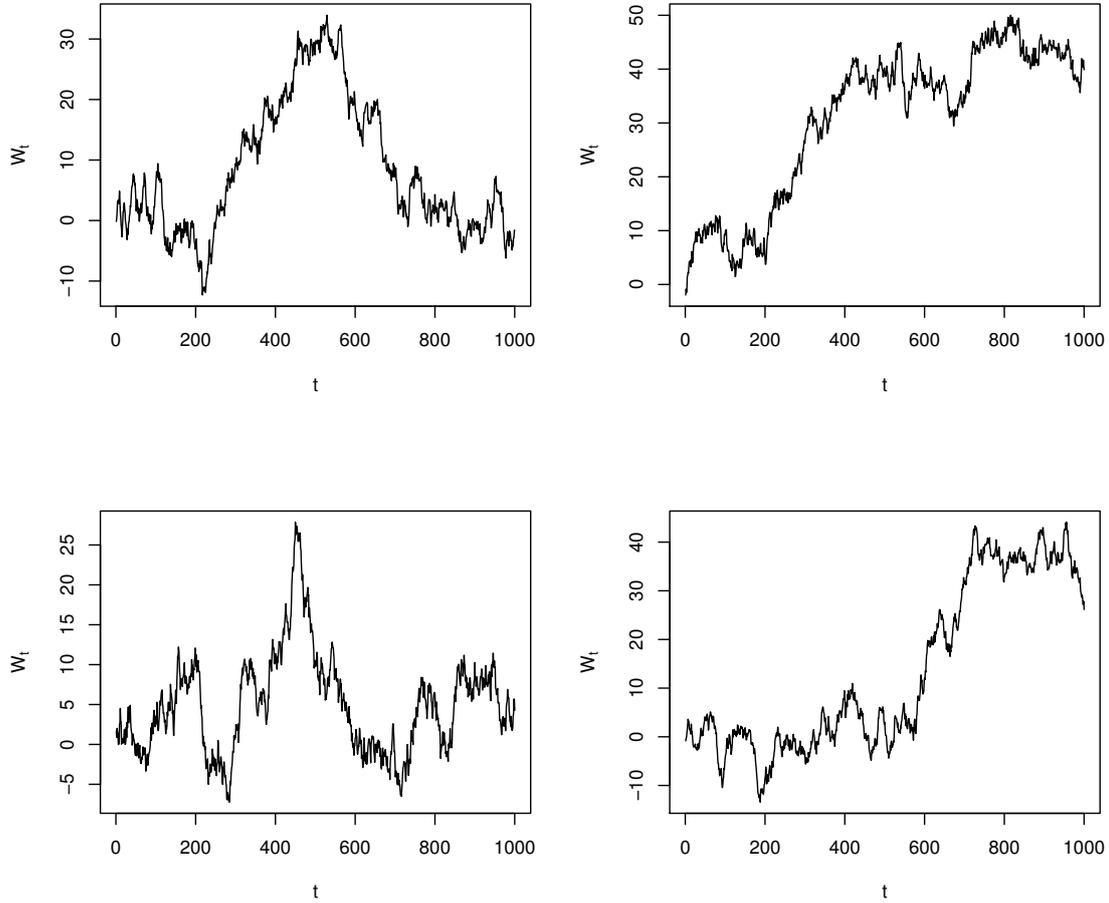
Figure 3.4: Brownian motion sample paths.

Generalised Brownian motion

Consider a process $\{X_t, t \geq 0\}$ satisfying an SDE of the form

$$dX_t = \mu dt + \sigma dW_t, \qquad X_0 = x_0, \qquad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$$

$$\Rightarrow X_t = x_0 + \mu t + \sigma W_t$$

where the last line is found by integrating both sides of the SDE between 0 and t. Note that μ is known as the drift coefficient, and σ is known as the diffusion coefficient. Using $W_t \sim N(0, t)$ gives $X_t | X_0 = x_0 \sim N(x_0 + \mu t, \sigma^2 t)$. Moreover, for times $t_i < t_{i-1}$ and

$\Delta t = t_i - t_{i-1}$, we obtain

$$
\begin{aligned}
X_{t_i} &= X_{t_i} - X_{t_{i-1}} + X_{t_{i-1}} \\
&= X_{t_{i-1}} + \mu(t_i - t_{i-1}) + \sigma(W_{t_i} - W_{t_{i-1}}) \\
&\Rightarrow X_{t_i}|X_{t_{i-1}} = x_{t_{i-1}} \sim N(x_{t_{i-1}} + \mu\Delta t, \sigma^2 \Delta t).
\end{aligned}
$$

Hence, standard BM is generalised via a linear drift controlled by μ and a scaling parameter σ. Recursive simulation from the above gives a skeleton path. Figure 3.5 shows four
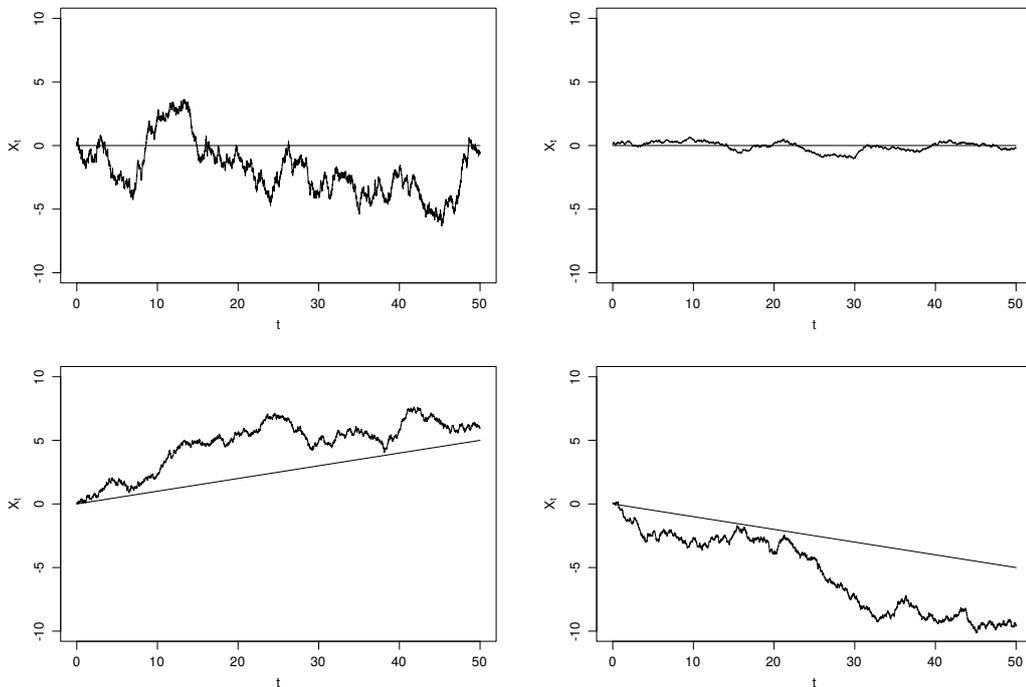
Figure 3.5: Generalised Brownian motion for different choices of μ and σ.

skeleton paths of generalised Brownian motion for different choices of μ and σ. The straight line in each plot indicates the drift and the effect of σ can clearly be seen.

Geometric Brownian motion

Geometric Brownian motion was first introduced in 1973 as a way of modelling stock prices (Merton, 1973). Consider an SDE of the form

$$
dX_t = \mu X_t dt + \sigma X_t dW_t, \qquad X_0 = x_0, \qquad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ \tag{3.3}
$$

so that the drift and diffusion functions are proportional to the value of the process. Note here that the integral equation is

$$\int_0^t dX_s = \mu \int_0^t X_s ds + \sigma \int_0^t X_s dW_s,$$

but we do not yet know how to deal with the last term, where Brownian motion is the integrator. We require a method of dealing with integrals of the form,

$$\int_0^t \beta(X_s) dW_s$$

known as Itô integrals.

3.1.2   Itô integrals

We cannot interpret

$$\int_a^b \beta(X_t) \frac{dW_t}{dt} dt$$

in the Riemann sense as although sample paths of standard Brownian motion are continuous, W_t is nowhere differentiable. We consider a stochastic integral of the form

$$\int_a^b \beta(X_t) dW_t = \lim_{n \to \infty} \sum_{i=0}^{n-1} \beta(X_{t_i}) \Delta W_{t_i} \tag{3.4}$$

for a partition of $[a, b]$ as $a = t_0 < t_1 < \ldots < t_n = b$, $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$ and the limit is in the mean square sense. The integral is obtained by considering appropriate limits of the approximating sum as $n \to \infty$. The choice of t_i is important as different choices give different limits. As written we obtain the Itô integral. Let us consider a simple Itô integral where $\beta(X_t) = 1$. Using (3.4) directly gives

$$
\begin{aligned}
\int_0^t W_s &= \lim_{n \to \infty} \sum_{i=0}^{n-1} \Delta W_{t_i} \\
&= \lim_{n \to \infty} \left\{ (W_{t_1} - W_{t_0}) + (W_{t_2} - W_{t_1}) + \cdots + (W_{t_n} - W_{t_{n-1}}) \right\} \\
&= \lim_{n \to \infty} (W_{t_n} - W_{t_0}) \\
&= \lim_{n \to \infty} (W_t - W_0) \\
&= W_t - W_0 \\
&= W_t.
\end{aligned}
$$

Throughout this thesis, we will typically be interested in integrals of the form

$$\int_0^t g(s)dW_s$$

for some real valued, square-integrable function $g(\cdot)$. It can be shown that

$$\int_0^t g(t)dW_t \sim N\left(0, \int_0^t g^2(t)dt\right),\tag{3.5}$$

where the result for the variance is typically referred to as the Itô isometry (Øksendal, 1995). The left hand side of (3.5) is an Itô integral, therefore we may write

$$\int_0^t g(t)dW_t = \lim_{n\to\infty} \sum_{i=0}^{n-1} g(t_i)\Delta W_{t_i}$$

for a partition $0 = t_0 < \ldots < t_n = t$ with $t_{i+1} - t_i = \Delta t$. Using that $\Delta W_{t_i} \sim N(0, \Delta t)$ we notice that within the summation we have a sum of constants multiplied by normal random variables. Therefore

$$\sum_{i=0}^{n-1} g(t_i)\Delta W_{t_i} \sim N\left(0, \sum_{i=0}^{n-1} g^2(t_i)\Delta t\right).$$

As we take the limit to infinity, the partition gets finer and $\Delta t \to 0$, we get a Riemann integral, as given in (3.5). Now suppose that we may write $X_t = \int_0^t g(s)dW_s$. The corresponding differential form is $dX_t = g(t)dW_t$. A generic (Itô) stochastic differential equation satisfied by a process $\{X_t, t \geq 0\}$ is

$$dX_t = \alpha(t, X_t)dt + \sqrt{\beta(t, X_t)}dW_t \tag{3.6}$$

where $\alpha(\cdot, \cdot)$ is the drift function (characterising the infinitesimal mean) and $\beta(\cdot, \cdot)$ is the diffusion coefficient (characterising the infinitesimal variance). The corresponding integral representation of X_t is

$$X_t = X_0 + \int_0^t \alpha(s, X_s)ds + \int_0^t \sqrt{\beta(s, X_s)}dW_s$$

where the first integral is of (deterministic) Riemann type and the second is of (stochastic) Itô type. For a rigorous discussion of uniqueness and existence of solutions to SDEs we refer the reader to Øksendal (1995). Of particular use to us is the chain rule for Itô processes that we give in the following section.

3.1.3 Itô formula

Consider an Itô process $\{X_t, t \geq 0\}$ satisfying the Itô SDE

$$dX_t = \alpha(t, X_t)dt + \sqrt{\beta(t, X_t)}dW_t \tag{3.7}$$

where $\alpha(t, X_t)$ is the drift and $\beta(t, X_t)$ is the diffusion coefficient. Let $g(t, x)$ be a real valued function, once differentiable in t and twice differentiable in x. Let

$$g_t = \frac{\partial g}{\partial t}, \qquad g_x = \frac{\partial g}{\partial x}, \qquad g_{xx} = \frac{\partial^2 g}{\partial x^2}$$

denote the first partial derivatives of g with respect to t, and the first two partial derivatives with respect to x. Itô's formula then gives the SDE satisfied by the process $Y_t, t \geq 0$, where $Y_t = g(t, X_t)$ as

$$\begin{aligned} dY_t = &\left(g_t(t, X_t) + \alpha(t, X_t)g_x(t, X_t) + \frac{1}{2}\beta(t, X_t)g_{xx}(t, X_t) \right) dt \\ &+ \sqrt{\beta(t, X_t)}g_x(t, X_t)dW_t. \end{aligned} \tag{3.8}$$

Itô's formula is a method to obtain SDEs satisfied by transformation of Itô processes. Equation (3.8) is the Itô calculus equivalent of the chain rule in classical calculus. When applying Itô's formula the following identities are used

$$dt^2 = dt\,dW_t = dW_t\,dt = 0 \qquad \text{and} \qquad dW_t^2 = dt.$$

To give an indication of why (3.8) is as it appears above, take a Taylor series expansion of $g(t + \Delta t, x + \Delta x)$ about (t, x) to give

$$dg(t, x) \quad \approx \quad \Delta t\, g_t + \Delta x\, g_x + \frac{1}{2}(\Delta t)^2\, g_{tt} + \frac{1}{2}\Delta t \Delta x\, g_{tx} + \frac{1}{2}(\Delta x)^2\, g_{xx} + \dots$$

where $g_{tt} = \partial^2 g/\partial t^2$ and $g_{tx} = \partial^2 g/\partial t\partial x$. Using shorthand notation for $\alpha(\cdot, \cdot)$ as α and $\beta(\cdot, \cdot)$ as β, replace Δx with $\alpha\Delta t + \sqrt{\beta}\Delta W$ and $(\Delta x)^2$ by $\alpha^2(\Delta t)^2 + 2\alpha\sqrt{\beta}\Delta t\Delta W + \beta(\Delta W)^2$ to obtain

$$\begin{aligned} dg(t, x) \quad \approx \quad &\Delta t\, g_t + (\alpha\Delta t + \sqrt{\beta}\Delta W)\, g_x + \frac{1}{2}(\Delta t)^2\, g_{tt} + \frac{1}{2}(\alpha(\Delta t)^2 + \sqrt{\beta}\Delta t\Delta W)\, g_{tx} + \\ &+ \frac{1}{2}(\alpha^2(\Delta t)^2 + 2\alpha\sqrt{\beta}\Delta t\Delta W + \beta(\Delta W)^2)\, g_{xx} + \dots \end{aligned}$$

Now approximate $(\Delta W)^2$ by Δt (and note in fact that $\mathbb{E}[(\Delta W^2)] = \Delta t$) to write the preceding expression as

$$dg(t,x) \quad \approx \quad \left(g_t + \alpha g_x + \frac{\beta}{2}g_{xx}\right)\Delta t + \sqrt{\beta}g_x \Delta W + o(\Delta t).$$

Itô's formula then follows by letting $\Delta t \to 0$.

Geometric Brownian motion revisited

Let us again consider the SDE in (3.3). Let $g(t,x) = \log(X_t)$ and apply Itô's formula. We have

$$g_t = 0, \qquad g_x = \frac{1}{X_t}, \qquad g_{xx} = -\frac{1}{X_t^2}$$

$$d\log(X_t) = 0 + \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}(dX_t)^2.$$

Substituting in dX_t and $(dX_t)^2 = \sigma^2 X_t^2 dt$ gives

$$d\log(X_t) = \frac{1}{X_t}X_t(\mu dt + \sigma dW_t) - \frac{1}{2}\frac{1}{X_t^2}\sigma^2 X_t^2 dt$$

$$= \mu dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt$$

which we recognise as a generalised Brownian motion process. Now, integrating between 0 and t gives

$$\log(X_t) - \log(X_0) = \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t$$

and exponentiating both sides and applying our initial condition we obtain

$$X_t = x_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\}.$$

Now, using $W_t \sim N(0,t)$ gives

$$\log\frac{X_t}{X_0} \sim N\left((\mu - \frac{\sigma^2}{2})t, \sigma^2 t\right)$$

so that

$$X_t|X_0 = x_0 \sim LN(\log x_0 + (\mu - \frac{\sigma^2}{2})t, \sigma^2 t)$$

where $LN(\cdot, \cdot)$ denotes the log normal distribution. More generally, for times $0 < s < t < \infty$,

$$X_t|X_s = x_s \sim LN\left(\log x_s + (\mu - \frac{\sigma^2}{2})(t-s), \sigma^2(t-s)\right)$$

and sampling from this distribution recursively gives a skeleton path. The geometric Brownian motion is used as a model for stock indexes by Black & Scholes (1973). In Figure 3.6 we can see four sample paths of geometric Brownian motion, all with the same initial value, drift coefficient, and diffusion coefficient.
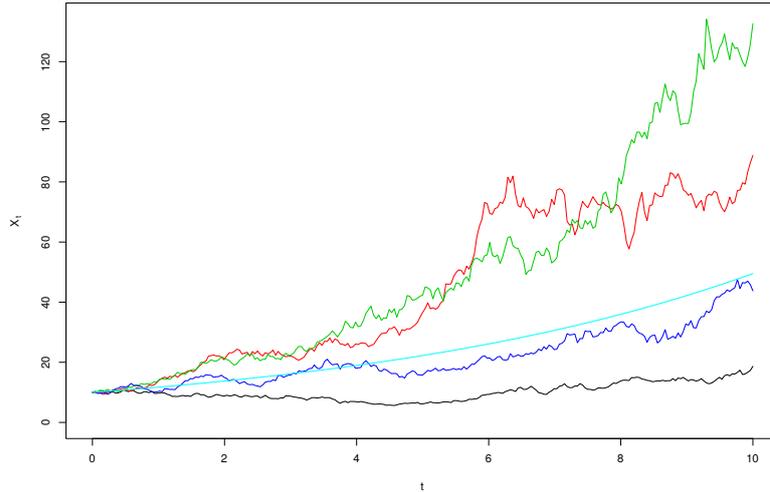
Figure 3.6: Realisations of four geometric Brownian motion paths with $\mu = 0.16$ and $\sigma = 0.2$. Mean trajectory overlaid (light blue).

3.2 Multivariate processes

We have so far only looked at univariate processes, however it is often the case that systems of interest involve many components. Here we shall extend the previously written theory to the case of multivariate scenarios.

Let us now consider a continuous-time d-dimensional Itô process $\{X_t, t \geq 0\}$ with $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})^T$ (where superscript T denotes the transpose) and initial condition $X_0 = x_0$, governed by the SDE

$$dX_t = \alpha(t, X_t)dt + \sqrt{\beta(t, X_t)}dW_t. \tag{3.9}$$

Here, the drift α is a d-vector, the diffusion coefficient β is a $d \times d$ positive definite matrix with a square root representation $\sqrt{\beta}$ such that $\sqrt{\beta}\sqrt{\beta}^T = \beta$ and W_t is a d-vector of (uncorrelated) standard Brownian motion. The equation in (3.9) is the natural extension to the univariate case in (3.7).

A non-linear transformation can be applied to (3.9) through the use of a multivariate Itô formula. Again, we take $Y_t = g(t, X_t)$ where $g(t, x)$ is a real-valued function, once

differentiable in t and x_i, $i = 1, \ldots, d$. Let

$$g_{t,k} = \frac{\partial g_k}{\partial t}, \qquad g_{x_i,k} = \frac{\partial g_k}{\partial x_i} \qquad \text{and} \qquad g_{x_i,x_j,k} = \frac{\partial^2 g_k}{\partial x_i x_j}$$

denote the first partial derivative of the kth element of g with respect to t, the first with respect to x_i and the mixed derivative with respect to x_i and x_j. Thus the kth component of $Y_t, t \geq 0$ will satisfy the SDE given by

$$dY_{t,k} = g_{t,k}dt + \sum_{i=1}^{d} g_{x_i,k}dX_{t,i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} g_{x_i,x_j,k}dX_{t,i}dX_{t,j}. \tag{3.10}$$

The following identities are useful in the above calculations:

$$dt^2 = dtdW_{t,i} = dW_{t,i}dt = 0 \qquad \text{and} \qquad dW_{t,i}dW_{t,j} = \delta_{ij}dt$$

where δ_{ij} is the Kronecker delta.

3.3 Linear SDEs

Consider an Itô process $\{X_t, t \leq 0\}$ satisfying the Itô SDE

$$dX_t = \alpha(t, X_t)dt + \sqrt{\beta(t, X_t)}dW_t \tag{3.11}$$

$$\alpha(t, X_t) = a_1(t)X_t + a_2(t), \qquad \left[\sqrt{\beta(t, X_t)}\right]_j = b_{1,j}(t)X_t + b_{2,j}(t)$$

where $X_t = d$-vector, $a_1 = d \times d$ matrix, $a_2 = d$-vector, $\left[\sqrt{\beta(t, X_t)}\right]_j$ denotes the jth column of $\sqrt{\beta(t, X_t)}$ with $b_{1,j}(t) = d \times d$ matrix, $b_{2,j}(t) = d$-vector, $j = 1, \ldots \ell$ and $dW_t = \ell$-vector of standard Brownian motion. We note the following:

- The linear SDE is autonomous if all coefficients are constant.

- The linear SDE is homogeneous if $a_2(t) = 0$ and $b_{2,j}(t) = 0$.

- The linear SDE is linear in the 'narrow' sense (additive noise) if $b_{1,j}(t) = 0$.

- The noise is multiplicative if $b_{2,j}(t) = 0$.

3.3.1 General solution to a linear SDE in the narrow sense

The linear SDE in the narrow sense means that $b_1 = 0$, so the SDE becomes

$$dX_t = (a_1(t)X_t + a_2(t))dt + b_2(t)dW_t.$$

We now consider the solution to the SDE. Let the fundamental matrix P_t satisfy

$$\frac{dP_t}{dt} = a_1(t)P_t, \qquad\qquad P_0 = I_d \qquad\qquad (3.12)$$

where I_d is the $d \times d$ identity matrix. Note that

$$\frac{d}{dt}P_t P_t^{-1} = P_t \frac{dP_t^{-1}}{dt} + \frac{dP_t}{dt}P_t^{-1} = 0$$

$$\Rightarrow \frac{dP_t^{-1}}{dt} = -P_t^{-1}a_1(t).$$

Now let $U_t = P_t^{-1}X_t$ with initial condition $U_0 = P_0^{-1}X_0 = X_0$ using equation 3.12. The SDE satisfied by U_t is given by

$$dU_t = d(P_t^{-1}X_t)$$
$$= (dP_t^{-1}X_t) + P_t^{-1}dX_t.$$

Hence

$$dU_t = -P_t^{-1}a_1(t)X_t dt + P_t^{-1}(a_1(t)X_t + a_2(t))dt + P_t^{-1}b_2(t)dW_t$$
$$\Rightarrow dU_t = P_t^{-1}a_2(t)dt + P_t^{-1}b_2(t)dW_t.$$

Integrating both sides of dU_t between 0 and t gives

$$U_t = U_0 + \int_0^t P_s^{-1}a_2(s)ds + \int_0^t P_s^{-1}b_2(s)dW_s.$$

Recall the Itô isometry (3.5)

$$\int_0^t g(s)dB_s \sim N\left(0, \int_0^t g(s)g(s)^T ds\right)$$

for a real-valued, square-integrable function $g(\cdot)$. Hence, the distribution of $U_t|U_0$ is therefore

$$U_t|U_0 \sim N\left(U_0 + \int_0^t P_s^{-1}a_2(s)ds, \int_0^t P_s^{-1}b_2(s)b_2(s)^T(P_s^{-1})^T ds\right).$$

Now, using $U_0 = x_0$ and $U_t = P_t^{-1}X_t \Rightarrow X_t = P_t U_t$ to give

$$X_t|X_0 = x_0 \sim N\left(P_t x_0 + P_t \int_0^t P_s^{-1}a_2(s)ds, P_t \int_0^t P_s^{-1}b_2(t)b_2(t)^T(P_s^{-1})^T ds P_t^T\right).$$

An equivalent representation of the mean and variance can be found as follows. Let $m_t = P_t x_0 + P_t \int_0^t P_s^{-1} a_2(s) ds$ denote the expectation of X_t. We have that

$$\frac{dm_t}{dt} = \frac{dP_t}{dt} x_0 + P_t P_t^{-1} a_2(t) + \frac{d}{dt} P_t \int_0^t P_s^{-1} a_2(s) ds.$$

Hence, using (3.12)

$$\frac{dm_t}{dt} = a_1(t) P_t x_0 + a_2(t) + a_1(t) P_t \int_0^t P_s^{-1} a_2(s) ds$$

$$\Rightarrow \frac{dm_t}{dt} = a_1(t) m_t + a_2(t), \qquad m_0 = X_0. \tag{3.13}$$

Similarly, let

$$V_t = P_t \int_0^t P_s^{-1} b_2(t) b_2(t)^T (P_s^{-1})^T ds P_t^T$$

then by the product rule

$$\frac{dV_t}{dt} = P_t \frac{d}{dt} \left(\int_0^t P_s^{-1} b_2(t) b_2(t)^T (P_s^{-1})^T ds P_t^T \right) + \frac{d}{dt} P_t \left(\int_0^t P_s^{-1} b_2(t) b_2(t)^T (P_s^{-1})^T ds P_t^T \right)$$

and applying the product rule again

$$\frac{dV_t}{dt} = P_t \left(\int_0^t P_s^{-1} b_2(t) b_2(t)^T (P_s^{-1})^T ds \frac{d}{dt} P_t^T + P_t^{-1} b_2(t) b_2(t)^T (P_t^{-1})^T P_t^T \right)$$

$$+ a_1(t) \underbrace{\left(P_t \int_0^t P_s^{-1} b_2(t) b_2(t)^T (P_s^{-1})^T ds P_t^T \right)}_{V_t} \qquad \text{[using (3.12)]}$$

$$= \underbrace{P_t \int_0^t P_s^{-1} b_2(t) b_2(t)^T (P_s^{-1})^T ds [a_1(t) P_t]^T}_{V_t a_1(t)^T} + b_2(t) b_2(t)^T + a_1(t) V_t$$

$$\Rightarrow \frac{dV_t}{dt} = V_t a_1(t)^T + b_2(t) b_2(t)^T + a_1(t) V_t, \qquad V_0 = 0. \tag{3.14}$$

To summarise, $X_t | X_0 = x_0 \sim N(m_t, V_t)$

$$X_t | X_0 = x_0 \sim N(m_t, V_t) \tag{3.15}$$

where m_t and V_t satisfy the ODE system given by equations (3.13) and (3.14). It will be helpful later to note that for an interval $[t_j, t_{j_{j+1}}]$

$$X_{t_{j+1}} | X_{t_j} = x_{t_j} \sim N(m_{t_{j+1}}, V_{t_{j+1}}) \tag{3.16}$$

where $m_{t_{j+1}}$ is the solution to (3.13), integrated over $[t_j, t_{j+1}]$ with initial condition $m_{t_j} = x_{j+1}$ and $V_{t_{j+1}}$ is the solution to (3.14), integrated over $[t_j, t_{j+1}]$ with initial condition $V_{t_j} = 0$. That is

$$m_{t_{j+1}} = m_{t_j} + \int_{t_j}^{t_{j+1}} \left(a_1(s)m_t + a_2(s) \right) dt$$

and

$$V_{t_{j+1}} = \int_{t_j}^{t_{j+1}} \left(V_t a_1(s)^T + b_2(s)b_2(s)^T + a_1(s)V_t \right) dt.$$

Example: Ornstein-Uhlenbeck SDE

Take $a_1(t) = \theta_1$, $a_2(t) = \theta_2$ and $b_2(t) = \theta_3$

$$\Rightarrow dX_t = (\theta_1 X_t + \theta_2)dt + \theta_3 dW_t.$$

The process satisfying the above SDE is known as an Ornstein-Uhlenbeck process. We apply (3.13) and (3.14) to generate the following ODE system

$$\frac{dm_t}{dt} = \theta_1 m_t + \theta_2, \qquad m_0 = X_0 \tag{3.17}$$

$$\frac{dV_t}{dt} = V_t \theta_1 + \theta_3^2 + \theta_1 V_t = 2\theta_1 V_t + \theta_3^2, \qquad V_0 = 0 \tag{3.18}$$

Using the method in Appendix A.1.2 we obtain the integrating factor $e^{-\theta_1 t}$ and therefore

$$m_t = m_0 e^{\theta_1 t} + e^{\theta_1 t} \times \int_0^t e^{-\theta_1 t dt} \theta_2$$

$$= m_0 e^{\theta_1 t} + e^{\theta_1 t} \times \left[-\frac{\theta_2}{\theta_1} e^{-\theta_1 t} \right]_0^t$$

$$= m_0 e^{\theta_1 t} + e^{\theta_1 t} \times \frac{\theta_2}{\theta_1} \left(1 - e^{-\theta_1 t} \right)$$

$$= m_0 e^{\theta_1 t} + \frac{\theta_2}{\theta_1} e^{\theta_1 t} \left(1 - e^{-\theta_1 t} \right).$$

To solve (3.18) we again use the integrating factor method with $e^{-2\theta_1 t}$ as the integrating factor and hence

$$V_t = e^{2\theta_1 t} \times \int_0^t e^{-2\theta_1 t dt} \theta_3^2$$

$$= e^{2\theta_1 t} \times \left[-\frac{\theta_3^2}{2\theta_1} e^{-2\theta_1 t} \right]_0^t$$

$$= \frac{\theta_3^2}{2\theta_1} \left(e^{2\theta_1 t} - 1 \right).$$

Thus the distribution of X_t is

$$X_t | X_0 = x_0 \sim N\left(x_0 e^{\theta_1 t} + \frac{\theta_2}{\theta_1} e^{\theta_1 t} \left(1 - e^{-\theta_1 t}\right), \frac{\theta_3^2}{2\theta_1}\left(e^{2\theta_1 t} - 1\right)\right).$$

For θ_1 negative, the expectation tends to $-\frac{\theta_2}{\theta_1}$ and $Var(X_t) \to -\frac{\theta_3^2}{2\theta_1}$ as $t \to \infty$. Figure 3.7 shows 10 skeleton paths of the Ornstein-Uhlenbeck processes, with $\theta_1 = -0.75$, $\theta_2 = 3$, $\theta_3 = 0.5$ on the interval $[0, 10]$ with a time step of 0.1. We took X_0 to be a random draw from a $N(0, 25)$ distribution.
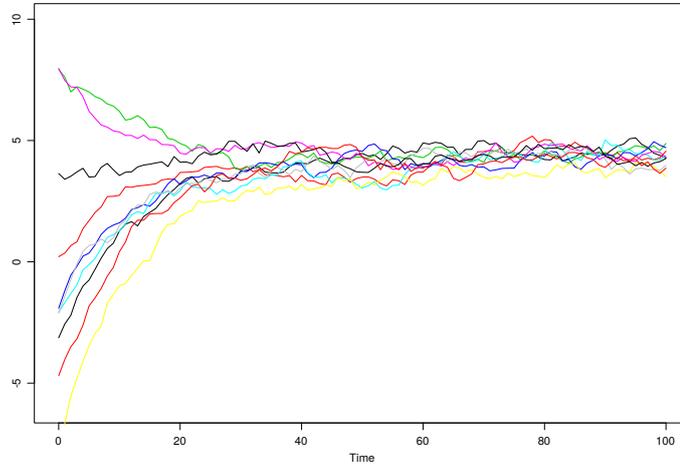
Figure 3.7: Ten skeleton paths of the Ornstein-Uhlenbeck process.

3.3.2 Homogeneous case

Consider a univariate linear homogeneous SDE

$$dX_t = a_1(t)X_t dt + b_1(t)X_t dW_t.$$

To solve the SDE, we apply the Itô formula (3.8) with $g(t, x) = \log x$. The three partial derivatives we require are

$$g_t = \frac{\partial g}{\partial t} = 0 \Rightarrow g_t(t, X_t) = 0$$

$$g_x = \frac{\partial g}{\partial x} = \frac{1}{x} \Rightarrow g_x(t, X_t) = \frac{1}{X_t}$$

$$g_{xx} = \frac{\partial^2 g}{\partial x^2} = -\frac{1}{x^2} \Rightarrow g_{xx}(t, X_t) = -\frac{1}{X_t^2}.$$

Hence

$$d \log X_t = \left(0 + \frac{1}{\cancel{X_t}} a_1(t) \cancel{X_t} + \frac{1}{2} \left(-\frac{1}{\cancel{X_t^2}} \right) b_1^2(t) \cancel{X_t^2} \right) dt + \frac{1}{\cancel{X_t}} b_1(t) \cancel{X_t} dW_t$$

$$\Rightarrow d \log X_t = \left(a_1(t) - \frac{b_1^2(t)}{2} \right) dt + b_1(t) dW_t$$

$$\Rightarrow \int_0^t d \log X_s = \int_0^t \left(a_1(s) - \frac{b_1^2(s)}{2} \right) ds + \int_0^t b_1(s) dW_s$$

$$\Rightarrow X_t = X_0 \exp \left\{ \int_0^t \left(a_1(s) - \frac{b_1^2(s)}{2} \right) dt + \int_0^t b_1(s) dW_t \right\}$$

and using Itô isometry (3.5) the distribution of X_t is

$$X_t | X_0 = x_0 \sim LN \left(\log x_0 + \int_0^t \left(a_1(s) - \frac{b_1^2(s)}{2} \right) dt, \int_0^t b_1^2(s) dt \right) \qquad (3.19)$$

Example: geometric Brownian motion

Take $a_1(t) = \theta_1$ and $b_1(t) = \theta_2$ to give

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dW_t$$

Using equation (3.19) we obtain

$$X_t | X_0 = x_0 \sim LN \left(\log x_0 + \left(\theta_1 - \frac{\theta_2^2}{2} \right) t, \theta_2^2 t \right)$$

and we recognise $\{X_t, t \geq 0\}$ as a geometric Brownian motion process.

3.4 Linear SDEs from non-linear SDEs

Consider a generic SDE of the form

$$dX_t = \alpha(t, X_t) dt + \epsilon \sqrt{\beta(t, X_t)} dW_t \qquad (3.20)$$

where the roll of ϵ is to indicate that the stochastic term is small. For generic $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ this SDE cannot be solved analytically (to give a closed form solution for X_t) so we examine a computationally efficient approximation to the solution via a linear noise approximation (LNA). The LNA can be derived by approximating the forward Fokker-Planck equation through a Taylor series expansion (see e.g. Komorowski *et al.*, 2009 and Elf & Ehrenberg, 2003). In what follows we conduct a more intuitive derivation.

Partition X_t into a deterministic path z_t and a residual stochastic process, M_t such that $X_t = z_t + \epsilon M_t$. Let z_t be the solution to

$$\frac{dz_t}{dt} = \alpha(t, z_t), \tag{3.21}$$

an ODE for the deterministic process. We assume that $\|X_t - z_t\|$ is $\mathcal{O}(\epsilon)$ over a time interval of interest. As X_t satisfies (3.20) the residual stochastic process M_t satisfies

$$dM_t = \frac{1}{\epsilon}\{\alpha(t, X_t) - \alpha(t, z_t)\}dt + \sqrt{\beta(t, X_t)}dW_t \tag{3.22}$$

This SDE is typically intractable. A tractable approximation can be obtained via Taylor expanding $\alpha(t, X_t)$ and $\beta(t, X_t)$ about z_t. We then obtain

$$\alpha(t, z_t + \epsilon M_t) = \alpha(t, z_t) + \epsilon H_t M_t + \cdots$$

and

$$\beta(t, z_t + \epsilon M_t) = \beta(t, z_t) + \cdots,$$

where H_t is the Jacobian matrix with (i, j)th element

$$(H_t)_{i,j} = \frac{\partial \alpha_i(t, z_t)}{\partial z_{j,t}}. \tag{3.23}$$

Collecting terms of $\mathcal{O}(\epsilon)$ gives an SDE satisfied by an approximate residual process $\{\hat{M}_t, t \leq 0\}$ of the form

$$d\hat{M}_t = H_t \hat{M}_t dt + \sqrt{\beta(t, z_t)}dW_t. \tag{3.24}$$

To indicate that the stochastic term in (3.20) is small we have used ϵ, so the drift term $\alpha(t, X_t)$ dominates the diffusion coefficient $\beta(t, X_t)$, equivalently, diffusion $<<$ drift. However, ϵ does not feature in (3.21) or (3.24). From here, therefore, we assume $\epsilon = 1$. In (3.24) for z_t in equilibrium and if α and β are time homogeneous, we get an Ornstein-Uhlenbeck process for $\hat{M}_t$.

3.4.1 Solving the linear noise approximation

Let us assume that the initial condition for (3.24) is $\hat{M}_0 = \hat{m}_0$ and follows a Gaussian distribution for all $t > 0$, so $\hat{M}_0 \sim N(\hat{m}_0, \hat{V}_0)$. Furthermore, let P_t be the $d \times d$ fundamental matrix for the deterministic ODE

$$\frac{d\hat{m}_t}{dt} = H_t \hat{m}_t$$

which satisfies

$$\frac{dP_t}{dt} = H_t P_t, \qquad P_0 = I_d \tag{3.25}$$

where I_d is the $d \times d$ identity matrix. Now,

$$\frac{d}{dt} P_t P_t^{-1} = P_t \frac{dP_t^{-1}}{dt} + \frac{dP_t}{dt} P_t^{-1} = 0.$$

Therefore using (3.25) it follows that

$$\frac{dP_t^{-1}}{dt} = -P_t^{-1} H_t. \tag{3.26}$$

Set $U_t = P_t^{-1} \hat{M}_t$. Then $U_0 = \hat{M}_0$. We write

$$\begin{aligned}
dU_t &= d(P_t^{-1} \hat{M}_t) \\
&= (dP_t^{-1}) \hat{M}_t + P_t^{-1}(d\hat{M}_t).
\end{aligned}$$

Then using (3.24) and (3.26) gives

$$\begin{aligned}
dU_t &= (-P_t^{-1} H_t dt) \hat{M}_t + P_t^{-1}(H_t \hat{M}_t dt + \sqrt{\beta(t, z_t)} dW_t) \\
&= -P_t^{-1} H_t \hat{M}_t dt + P_t^{-1} H_t \hat{M}_t dt + P_t^{-1} \sqrt{\beta(t, z_t)} dW_t \\
&= P_t^{-1} \sqrt{\beta(t, z_t)} dW_t.
\end{aligned}$$

Hence we can write

$$U_t = U_0 + \int_0^t P_s^{-1} \sqrt{\beta(s, z_s)} dW_s.$$

Appealing to linearity and Itô isometry, we obtain

$$U_t | U_0 \sim N\left(U_0, \int_0^t P_s^{-1} \beta(s, z_s)(P_s^{-1})^T ds\right). \tag{3.27}$$

Therefore, for the initial assumption above, that is $\hat{M}_0 = U_0 \sim N(\hat{m}_0, \hat{V}_0)$, we have that

$$\hat{M}_t \sim N\left(P_t \hat{m}_0, P_t \psi_t P_t^T\right), \tag{3.28}$$

where

$$\psi_t = \hat{V}_0 + \int_0^t P_s^{-1} \beta(s, z_s)(P_s^{-1})^T ds.$$

Thus, the SDE (3.24) satisfied by $\hat{M}_t$ can be solved analytically, where P_t and ψ_t satisfy the ODE system

$$\frac{dP_t}{dt} = H_t P_t, \qquad P_0 = I_d \tag{3.29}$$

$$\frac{d\psi_t}{dt} = P_t^{-1} \beta(t, z_t)(P_t^{-1})^T, \qquad \psi_0 = \hat{V}_0. \tag{3.30}$$

Hence, the approximating distribution of $X_t | X_0$ is given by

$$X_t | X_0 \sim N\left(z_t + P_t \hat{m}_0, P_t \psi_t P_t^T\right). \tag{3.31}$$

Note that (3.28) can be written as

$$\hat{M}_t | \hat{M}_0 = \hat{m}_0 \sim N(m_t, V_t)$$

where it is clear from (3.29) that

$$\frac{dm_t}{dt} = H_t m_t, \qquad m_0 = \hat{m}_0 \tag{3.32}$$

and the ODE for $V_t = P_t \psi_t P_t^T$ can be obtained as

$$\frac{dV_t}{dt} = \frac{d\left(P_t \psi_t P_t^T\right)}{dt},$$

to which we apply the product rule and obtain

$$\begin{aligned}
\frac{dV_t}{dt} &= P_t \frac{d}{dt}\left(\psi P_t^T\right) + \left(\frac{dP_t}{dt}\right)\psi_t P_t^T \\
&= P_t \left\{ \psi_t \frac{dP_t^T}{dt} + \left(\frac{d\psi_t}{dt}\right)P_t^T \right\} + H_t P_t \psi_t P_t^T \\
&= P_t \left\{ \psi_t P_t^T H_t^T + P_t^{-1}\beta(z_t, \theta)\left(P_t^{-1}\right)^T P_t^T \right\} + H_t P_t \psi_t P_t^T \\
&= P_t \psi_t P_t^T H_t + \beta(z_t, \theta) + H_t P_t \psi_t P_t^T \\
&= V_t H_t^T + \beta(z_t, \theta) + H_t V_t, \qquad V_0 = 0. \tag{3.33}
\end{aligned}$$

Now we can obtain a less computationally intensive solution by solving (3.21), (3.32) and (3.33) instead of (3.21), (3.29) and (3.30), where now the approximating distribution of X_t is

$$X_t \sim N(z_t + m_t, V_t).$$

3.4.2 LNA birth-death example

Consider a birth-death process satisfied by the following SDE

$$dX_t = (\theta_1 - \theta_2)X_t dt + \sqrt{(\theta_1 + \theta_2)X_t} dW_t \qquad (3.34)$$

where X_t denotes a population of some species at time t, θ_1 and θ_2 denote the birth rate and death rate respectively, and $W_t, t \geq 0$ is standard Brownian motion. Here, $H_t = (\theta_1 - \theta_2)$ and the ODE system ((3.21), (3.29), (3.30)) governing the linear noise approximation of (3.34) is given by

$$\frac{dz_t}{dt} = (\theta_1 - \theta_2)z_t, \qquad z_0 = x_0$$

$$\frac{dP_t}{dt} = (\theta_1 - \theta_2)P_t, \qquad P_0 = 1$$

$$\frac{d\psi_t}{dt} = P_t^{-2}(\theta_1 + \theta_2)z_t$$

The above system of ODEs can be solved explicitly to give

$$z_t = x_0 \exp\{(\theta_1 - \theta_2)t\},$$

$$P_t = \exp\{(\theta_1 - \theta_2)t\},$$

$$\psi_t = \frac{(\theta_1 + \theta_2)x_0}{(\theta_1 - \theta_2)}\left[1 - \exp\{-(\theta_1 - \theta_2)t\}\right].$$

To generate a skeleton path of the birth-death process, we can iteratively draw from the transition density under the LNA. We partition an interval $[0, T]$ as $0 = t_0 < t_i < \ldots < t_n = T$. Hence, given x_{t_i} at a time t_i, the ODEs for z_t, P_t and ψ_t are initialised at $x_{t_i}, 1$ and 0 respectively, and integrated over $(t_i, t_{i_1}]$. We then draw $X_{t_{i+1}} \sim N(z_{t_{i+1}}, P_{t_{i+1}}^2 \psi_{t_{i+1}})$ to obtain $x_{t_{i+1}}$. Note that since we are initialising z_t at the simulated value of x_t in each interval, we have that $m_t = 0$ for $t \in [0, T]$. Figure 3.8 shows a single realisation of the process, and Figure 3.9 shows the mean and 95% credible region for X_t from 10, 000 simulations, where $x_0 = 10$, $\theta = (0.2, 0.18)$ and a time-step $\Delta t = 0.05$. In the next section we will look at different methods of applying Bayesian inference for linear SDEs.
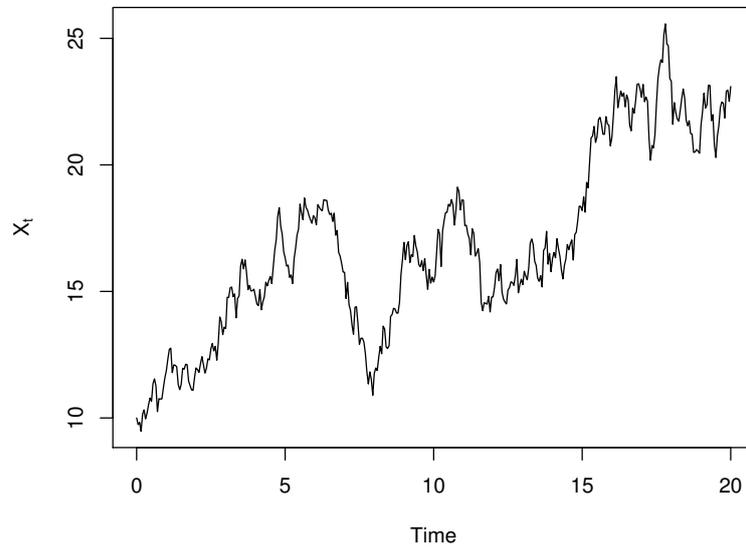
Figure 3.8: A single realisation of a species X_t in the birth-death model, $x_0 = 10$ and $\theta = (0.2, 0.18)$ with time-step $\Delta t = 0.05$
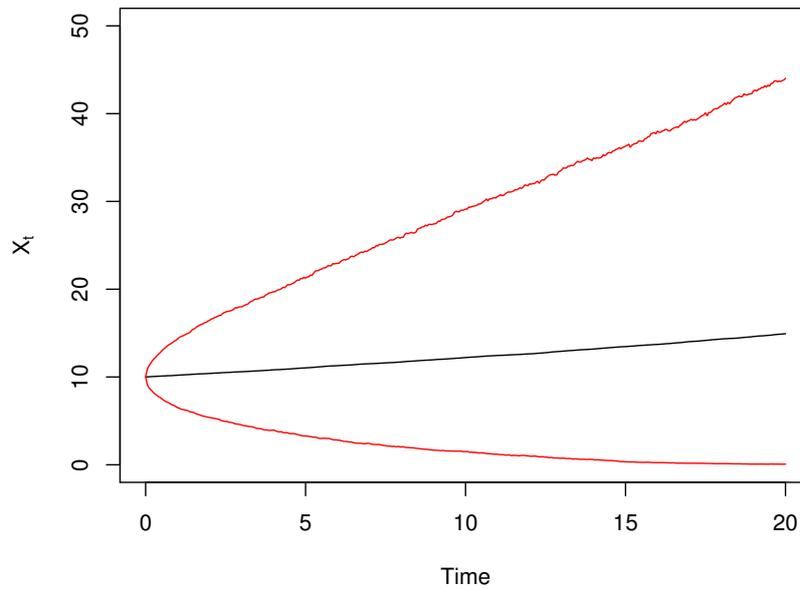
Figure 3.9: Birth-death model. 95% credible region (red line) and mean (black line) for X_t, with $x_0 = 10$ and $\theta = (0.2, 0.18)$ with time-step $\Delta t = 0.05$ from 10,000 simulations.

Chapter 4

Bayesian inference for linear stochastic differential equations

We now consider the problem of performing Bayesian inference for the parameters θ governing an SDE of linear form given by (3.11). That is

$$dX_t = \alpha(t, X_t, \theta)dt + \sqrt{\beta(t, X_t, \theta)}dW_t$$

where we now allow explicit dependence of $\alpha(\cdot)$ and $\beta(\cdot)$ on an unknown parameter vector $\theta = (\theta_1, \ldots, \theta_p)^T$. Hence the drift and jth column of the diffusion coefficient take the form

$$\alpha(t, X_t, \theta) = a_1(t, \theta)X_t + a_2(t, \theta)$$
$$\left[\sqrt{\beta(t, X_t, \theta)}\right]_j = b_{1,j}(t, \theta)X_t + b_{2,j}(t, \theta).$$

Given data at discrete times $x = (x_{t_0}, x_{t_1}, \ldots, x_{t_n})^T$, the likelihood is given by

$$\pi(x|\theta) = \prod_{i=1}^{n} \pi(x_{t_i}|x_{t_{i-1}}, \theta)$$

where $\pi(x_{t_i}|x_{t_{i-1}}, \theta)$ is the transition density, obtained by solving the SDE. Hence upon ascribing a prior density $\pi(\theta)$ to θ, Bayesian inference may proceed via the posterior

$$\pi(\theta|x) \propto \pi(\theta)\pi(x|\theta).$$

In practice, we anticipate that $\pi(\theta|x)$ will be intractable and we resort to the sampling based approaches from Chapter 2 to generate draws from $\pi(\theta|x)$. In what follows, we assume that $\{X_t, t \geq 0\}$ cannot be observed directly, and adopt an observation model of

the form

$$Y_t = h(X_t) + \epsilon_t, \qquad \epsilon_t \overset{indep}{\sim} N(0, \Sigma) \tag{4.1}$$

allowing for observation of a transformation of (a subset of) components of X_t, subject to additive Gaussian noise. We note that the observations are conditionally independent given $\{X_t, t \geq 0\}$ and if inference is required for Σ, we augment θ to include the components of Σ. Hence, the density linking the observations and latent process is

$$\pi(y|x, \theta) = \prod_{i=0}^{n} \pi(y_{t_i} | x_{t_i}, \theta).$$

If interest lies in the joint posterior for x and θ, we may construct this joint posterior

$$\pi(x, \theta|y) \propto \pi(\theta)\pi(x_{t_0})\pi(x|\theta, x_{t_0})\pi(y|x, \theta) \tag{4.2}$$

where $\pi(x_{t_0})$ is the prior density ascribed to x_{t_0}. Note that the marginal parameter posterior $\pi(\theta|y)$ is given by

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta) \tag{4.3}$$

where the marginal likelihood term $\pi(y|\theta)$ is given by

$$\pi(y|\theta) = \int \pi(x_{t_0})\pi(x|\theta)\pi(y|x, \theta)dx. \tag{4.4}$$

In what follows, we assume that interest lies in the marginal parameter posterior $\pi(\theta|y)$. We now consider two cases: in the first case, we assume a linear observation model so that the marginal likelihood is tractable. In the second case, we assume a non-linear observation model and consider a linear approximation approach, and a pseudo-marginal sampling-based approach.

4.1 Linear SDE in the narrow sense and linear observation model

In this instance we assume a linear observation model so in (4.1) we take $h(X_t) = F^T X_t$ so we have

$$Y_t = F^T X_t + \epsilon_t, \qquad \epsilon_t \overset{indep}{\sim} N(0, \Sigma), \tag{4.5}$$

where F is a $d \times d_0$ constant matrix allowing for partial observation of the components of X_t. In order to calculate the marginal likelihood $\pi(y|\theta)$ we make use of a forward filter, often referred to in this context as a hybrid Kalman filter due to the latent process being in continuous time and the observations being in discrete time. First note that the marginal

likelihood can be factorised as

$$\pi(y|\theta) = \pi(y_{t_0}|\theta) \prod_{j=1}^{n} \pi(y_{t_j}|y_{t_0:t_{j-1}}, \theta) \tag{4.6}$$

We use a forward filter to compute each term in (4.6) recursively. To simplify the notation, we suppress the parameters θ where possible. Suppose at time t_j we have $X_{t_j}|Y_{t_0:t_j} \sim N(a_j, C_j)$. This is the filtering distribution at time t_j. Using (3.16)

$$X_{t_{j+1}}|Y_{t_0:t_j} \sim N(m_{t_{j+1}}, V_{t_{j+1}})$$

where $m_{t_{j+1}}$ is the solution to (3.13), integrated over $[t_j, t_{j+1}]$ with initial condition a_j, and $V_{t_{j+1}}$ is the solution to (3.14) integrated over $[t_j, t_{j+1}]$ with initial condition $V_{t_j} = C_j$. Hence, using

$$Y_{t_{j+1}}|X_{t_{j+1}} \sim N(F^T X_{t_{j+1}}, \Sigma)$$

and

$$Y_{t_{j+1}}|Y_{t_0:t_j} \sim N(F^T m_{t_{j+1}}, F^T V_{t_{j+1}} F + \Sigma)$$

we see that the marginal likelihood contribution is

$$\pi(y_{t_{j+1}}|y_{t_0:t_j}) = N(y_{t_{j+1}}; F^T m_{t_{j+1}}, F^T V_{t_{j+1}} F + \Sigma).$$

Now, to update the filtering distribution, we note that

$$\begin{pmatrix} X_{t_{j+1}} \\ Y_{t_{j+1}} \end{pmatrix} \bigg| y_{t_0:t_j} \sim N\left(\begin{pmatrix} m_{t_{j+1}} \\ F^T m_{t_{j+1}} \end{pmatrix}, \begin{pmatrix} V_{t_{j+1}} & V_{t_{j+1}} F \\ F^T V_{t_{j+1}} & F^T V_{t_{j+1}} F + \Sigma \end{pmatrix} \right).$$

Hence conditioning on $Y_{t_{j+1}} = y_{t_j}$ gives

$$X_{t_{j+1}}|Y_{t_0:t_j} = y_{t_0:t_j} \sim N(a_{j+1}, C_{j+1})$$

where

$$a_{j+1} = m_{t_{j+1}} + V_{t_{j+1}} F \left(F^T V_{t_{j+1}} F + \Sigma \right)^{-1} \left(y_{t_{j+1}} - F^T m_{t_{j+1}} \right)$$

$$C_{j+1} = V_{t_{j+1}} + V_{t_{j+1}} F \left(F^T V_{t_{j+1}} F + \Sigma \right)^{-1} F V_{t_{j+1}}.$$

The forward filter steps are as follows in Algorithm 1.

Note that for the forward filter there is no need to store values of a_j or C_j. However, this algorithm can be extended should we wish to find the smoothing distribution of $X_{t_j}|y_{t_0:t_j}$ enabling us to make inferences about the latent process.

Consider the marginal posterior for the latent process, $\pi(x|y)$. We note the factorisa-

---

Algorithm 1 Forward filter

---

1. Initialisation: Compute $\pi(y_{t_0}) = N\left(y_{t_0}; F^T a, F^T C F + \Sigma\right)$. The posterior at time $t_0 = 0$ is therefore $X_{t_0}|y_{t_0} \sim N(a_0, C_0)$ where

$$a_0 = a + CF\left(F^T C F + \Sigma\right)^{-1}\left(y_{t_0} - F^T a\right)$$
$$C_0 = C - CF\left(F^T C F + \Sigma\right)^{-1} F^T C.$$

2. For iteration $j = 0, 1, \ldots, n-1$:

   (a) Prior at t_{j+1}. We have $X_{t_j}|Y_{t_0:t_j} \sim N\left(a_{t_j}, C_{t_j}\right)$
   $\Rightarrow X_{t_{j+1}}|Y_{t_0:t_j} \sim N\left(m_{t_{j+1}}, V_{t_{j+1}}\right)$ where $m_{t_{j+1}}$ and $V_{t_{j+1}}$ are the solutions to the ODEs (3.13) and (3.14) initialised with $m_{t_j} = a_{t_j}$ and $V_{t_j} = C_{t_j}$.

   (b) One step forecast:

   $$Y_{t_{j+1}}|Y_{t_0:t_j} \sim N\left(F^T m_{t_{j+1}}, F^T V_{t_{j+1}} F + \Sigma\right)$$

   Compute the updated marginal likelihood

   $$\pi\left(y_{t_0:t_{j+1}}\right) = \pi\left(y_{t_0:t_j}\right) \pi\left(y_{t_{j+1}}|y_{t_0:t_j}\right)$$
   $$= \pi\left(y_{t_0:t_j}\right) N\left(y_{t_{j+1}}; F^T m_{t_{j+1}}, F^T V_{t_{j+1}} F + \Sigma\right)$$

   (c) Posterior at t_{j+1}. Combining the distributions in (a) and (b) gives the joint distribution of $X_{t_{j+1}}$ and $Y_{t_{j+1}}$ conditional on $y_{t_0:t_j}$

   $$\begin{pmatrix} X_{t_{j+1}} \\ Y_{t_{j+1}} \end{pmatrix}\Bigg|y_{t_0:t_j} \sim N\left(\begin{pmatrix} m_{t_{j+1}} \\ F^T m_{t_{j+1}} \end{pmatrix}, \begin{pmatrix} V_{t_{j+1}} & V_{t_{j+1}} F \\ F^T V_{t_{j+1}} & F^T V_{t_{j+1}} F + \Sigma \end{pmatrix}\right)$$

   and therefore $X_{t_{j+1}}|y_{t_0:t_{j+1}} \sim N(a_{t_{j+1}}, C_{t_{j+1}})$, where

   $$a_{t_{j+1}} = m_{t_{j+1}} + V_{t_{j+1}} F\left(F^T V_{t_{j+1}} F + \Sigma\right)^{-1}\left(y_{t_{j+1}} - F^T m_{t_{j+1}}\right),$$
   $$C_{t_{j+1}} = V_{t_{j+1}} + V_{t_{j+1}} F\left(F^T V_{t_{j+1}} F + \Sigma\right)^{-1} F V_{t_{j+1}}.$$

---

tion

$$\pi(x|y) = \pi(x_{t_n}|y_{t_0:t_n}) \prod_{j=0}^{n-1} \pi(x_{t_j}|x_{t_{j+1}}, y_{t_0:t_j})$$

where $\pi(x_{t_n}|y_{t_0:t_n})$ is available at the end of the forward sweep of the filter and the subsequent terms can be obtained using a backwards sweep. In order to sample the latter densities we require storage of $P_{t_{j+1}}$ and can be obtained during step 2 of Algorithm 1 by additionally solving (3.12). We also store each value of $a_{t_{j+1}}$ and $C_{t_{j+1}}$. Then we may proceed with backwards sampling as in Algorithm 2.

Returning to the problem of inference for the parameters governing an SDE of the form

---

Algorithm 2 Backwards sampler

---

1. Draw x_{t_n} from $X_{t_n}|y \sim N(a_{t_j}, C_{t_j})$

2. For $j = n-1, n-2, \ldots, 0$,

   (a) Joint distribution of X_{t_j} and $X_{t_{j+1}}$. Note that $X_{t_j}|y_{t_0:t_j} \sim N(a_{t_j}, C_{t_j})$. The joint distribution of X_{t_j} and $X_{t_{j+1}}$ conditional on $y_{t_0:t_j}$ is

$$\begin{pmatrix} X_{t_j} \\ X_{t_{j+1}} \end{pmatrix} \Bigg| y_{t_0:t_j} \sim N\left(\begin{pmatrix} a_{t_j} \\ m_{t_{j+1}} \end{pmatrix}, \begin{pmatrix} C_{t_j} & C_{t_j}P_{t_{j+1}}^T \\ P_{t_{j+1}}C_{t_j} & V_{t_{j+1}}. \end{pmatrix} \right)$$

   (b) Backwards distribution. The distribution of $X_{t_j}|x_{t_{j+1}}, y_{t_0:t_j}$ is $N(\hat{a}_{t_j}, \hat{C}_{t_j})$ where

$$\hat{a}_{t_j} = a_{t_j} + C_{t_j}P_{t_{j+1}}^T V_{t_{j+1}}^{-1}(x_{t_{j+1}} - m_{t_{j+1}}),$$
$$\hat{C}_{t_j} = C_{t_j} - C_{t_j}P_{t_{j+1}}^T V_{t_{j+1}}^{-1} P_{t_{j+1}}C_{t_j}.$$

   Draw x_{t_j} from $X_{t_j}|x_{t_{j+1}}, y_{t_0:t_j} \sim N(\hat{a}_{t_j}, \hat{C}_{j_j})$.

(3.11) for an observation model of the form (4.1), we note that the marginal posterior

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta)$$

can be evaluated (up to a multiplicative constant) using the forward filter above. This leads to the MCMC scheme in Algorithm 3 which assumes a Gaussian random walk proposal mechanism.

4.2 Linear SDE and non-linear observation model

In this instance we assume a non-linear observation model. In this case, the marginal likelihood is intractable. In what follows, we therefore consider two approaches to inference: an approximate inferential model based on a linearisation of the observed components and exact (simulation-based) inference using pseudo-marginal methods.

4.2.1 Approximation via linearisation

Recall the general observation model (4.1) and take $h(X_t)$ to be some non-linear function. The marginal likelihood $\pi(y|\theta)$ here is intractable. However we may construct an appropriate linear noise approximation enabling us to obtain a tractable marginal likelihood.

For the case of $d = 1$, so that X_t is univariate, we apply the Itô formula to obtain the stochastic differential equation satisfied by $\tilde{X}_t = h(X_t)$. The LNA (see Section 3.4) can

---

Algorithm 3 MCMC with Gaussian random walk proposal

---

1. Initialise chain. Set $\theta^{(0)}$ at some appropriate initial value. Set $i = 0$.

2. Propose $\theta^* \sim N(\theta^{(i)}, \Sigma_\theta)$.

3. Evaluate the acceptance probability:

$$\alpha(\theta^*|\theta^{(i)}) = \min\left\{1, \frac{\pi(\theta^*|y)}{\pi(\theta^{(i)}|y)} \times \frac{q(\theta^{(i)}|\theta^*)}{q(\theta^*|\theta^{(i)})}\right\}$$

   where

$$q(\theta^*|\theta^{(i)}) = N(\theta^*; \theta^{(i)}, \Sigma_\theta)$$

   hence:

$$\alpha(\theta^*|\theta^{(i)}) = \min\left\{1, \frac{\pi(\theta^*)p(y|\theta^*)}{\pi(\theta^{(i)})p(y|\theta^{(i)})}\right\}.$$

4. With probability $\alpha(\theta^*|\theta^{(i)})$ set $\theta^{(i+1)} = \theta^*$, otherwise set $\theta^{(i+1)} = \theta^{(i)}$.

5. Set $i = i + 1$ and return to step 2.

---

then be applied to $d\tilde{X}_t$. For $d > 1$, we note that in general the SDE satisfied by $h(X_t)$ may explicitly depend on (a subset of) components of X_t. Hence, in this case, we define the $(d_0 + d)$ vector $\tilde{X}_t = (h(X_t)^T, X_t^T)^T$ and apply the multivariate Itô formula (see Section 3.2) to obtain

$$d\tilde{X}_t = \tilde{\alpha}(t, \tilde{X}_t, \theta)dt + \sqrt{\tilde{\beta}(t, \tilde{X}_t, \theta)}dW_t.$$

Applying the LNA with the assumption that $z_0 = \tilde{x}_0$, then $m_0 = 0$, gives

$$\tilde{X}_t|\tilde{X}_0 \sim N(z_t, P_t\psi_t P_t^T)$$

where

$$\frac{dz_t}{dt} = \tilde{\alpha}(t, z_t, \theta), \qquad z_0 = \tilde{x}_0$$

$$\frac{dP_t}{dt} = \tilde{H}_t P_t, \qquad P_0 = I_d$$

$$\frac{d\psi_t}{dt} = P_t^{-1}\tilde{\beta}(t, z_t, \theta)(P_t^{-1})^T, \qquad \psi_0 = \hat{V}_0$$

and

$$(\tilde{H}_t)_{ij} = \frac{\partial \tilde{\alpha}_i(t, z_t, \theta)}{\partial z_{j,t}}.$$

Using that $\tilde{X}_t = (h(X_t)^T, X_t^T)^T$ we see that the observation model can be written as

$$Y_t = F^T \tilde{X}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \Sigma)$$

where F is a $(d_0 + d) \times d_0$ matrix, specifically

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Hence, F is block partitioned with upper $d_0 \times d_0$ elements as the $d_0 \times d_0$ identity matrix and the lower $d \times d_0$ elements as 0. Thus, we may compute the marginal likelihood under the LNA,

$$\pi(y|\theta) = \int \pi(\tilde{x}, y|\theta) d\tilde{x}$$

using a forward filter, as described in Section 4.1.

4.2.2 Pseudo-marginal Metropolis-Hastings

Recall that we have an intractable observed data likelihood (4.4) due to our observation model in (4.1) being non-linear with $h(X_t)$ some non-linear function. Given the ability to unbiasedly estimate the observed data likelihood $\pi(y|\theta)$, the pseudo-marginal Metropolis-Hastings (PMMH) scheme of Section 2.5 is applicable. Denote the unbiased estimator by $\hat{\pi}(y|\theta) = \hat{\pi}_U(y|\theta)$ where U denotes all the random variables used to construct the estimator. Hence, the unbiasedness property gives $E_U(\hat{\pi}_U(y|\theta)) = \pi(y|\theta)$. Now, consider a joint target

$$\pi(\theta, u|y) \propto \pi(\theta)\hat{\pi}_u(y|\theta)g(u) \tag{4.7}$$

where $g(u)$ denotes the density associated with U. It is easily checked that

$$\int \pi(\theta, u|y)du \propto \pi(\theta) \int \hat{\pi}_u(y|\theta)g(u)du$$
$$\propto \pi(\theta)\pi(y|\theta)$$
$$\propto \pi(\theta|y).$$

and so marginalising out u gives the marginal parameter posterior in (4.3).

The pseudo-marginal Metropolis-Hastings approach can be seen as a standard Metropolis-

---

Algorithm 4 Pseudo-marginal Metropolis-Hastings

1. For iteration ($i = 0$).

   (a) Set $\theta^{(0)}$ in the support of $\pi(\theta)$ and draw $u^{(0)} \sim g(\cdot)$.

   (b) Compute $\hat{\pi}_{u^{(0)}}(y|\theta^{(0)})$

2. For iteration $i = 1, \ldots, n_{\text{iters}}$:

   (a) Draw $\theta^* \sim q(|\theta^{(i-1)})$ and $u^* \sim g(\cdot)$

   (b) Compute $\hat{\pi}_{u^*}(y|\theta^*)$.

   (c) With probability $\alpha(\theta^*, u^*|\theta^{(i-1)}, u^{(i-1)})$ given by (4.8), put $(\theta^{(i)}, u^{(i)}) = (\theta^*, u^*)$ otherwise store the current values $(\theta^{(i)}, u^{(i)}) = (\theta^{(i-1)}, u^{(i-1)}.)$

Hastings scheme targeting (4.7) with a proposal density $q(\theta^*, u^*|\theta, u) = q(\theta^*|\theta)g(u^*)$ for which the acceptance probability is

$$\alpha(\theta^*, u^*|\theta, u) = \min\left\{1, \frac{\hat{\pi}_{u^*}(y|\theta^*)\pi(\theta^*)}{\hat{\pi}_u(y|\theta)\pi(\theta)} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}\right\}. \tag{4.8}$$

The PMMH scheme is given by Algorithm 4. It remains that we can obtain estimates $\hat{\pi}_u(y|\theta)$. As shown by (Del Moral, 2004) (also see Pitt *et al.*, 2012) a particle filter can be used to give realisations of an unbiased estimator $\hat{\pi}_U(y|\theta)$. The bootstrap particle filter recursively draws from the filtering distribution $\pi(x_{t_j}|y_{t_0:t_j}, \theta)$ for each $j = 0, \ldots, n$ (Gordon *et al.*, 1993) (see also Künsch, 2013). Essentially, a sequence of importance sampling and resampling steps are used to propagate a weighted sample $\{(x_{t_j,k}, w(u_{t_j,k})), k = 1, \ldots, N\}$ from the filtering distribution. Note that we let the weight depend explicitly on the corresponding component of the auxiliary variable $u = (u_0, \ldots, u_n)$. At time t, the particle filter uses the approximation

$$\hat{\pi}(x_{t_j}|y_{t_0:t_j}, \theta) \propto \pi(y_{t_j}|x_{t_j}, \theta) \sum_{k=1}^{N} \pi(x_{t_j}|x_{t_{j-1},k}, \theta)w(u_{t_{j-1},k}). \tag{4.9}$$

A simple importance sampling/resampling strategy follows, where particles are resampled (with replacement) in proportion to their weights, propagated via $x_{t_j,k} = f_j(u_{t_j,k}) \sim \pi(\cdot|x_{t_{j-1}}, \theta)$ and reweighted by $\pi(y_{t_j}|x_{t_j,k}, \theta)$. Here, $f_j(\cdot)$ is a deterministic function of $u_{t_j,k}$ (as well as the parameters and previous latent state, suppressed for simplicity) that gives an explicit connection between the particles and auxiliary variables. Algorithm 5 provides a complete description of the bootstrap particle filter with input given by parameters θ, auxiliary variables u, data y and the number of particles N. However notice the addition of a non-standard and optional sorting step (a), which is particularly useful when

---

Algorithm 5 Bootstrap particle filter

---

1. Initialisation $(t = 0)$.

   (a) **Sample** the prior. Put $x_{t_0,k} = f_0(u_{t_0,k}) \sim \pi(\cdot)$, $k = 1, \ldots, N$.

   (b) **Compute** the weights. For $k = 1, \ldots, N$ set

$$\tilde{w}(u_{t_0,k}) = \pi(y_{t_0}|x_{t_0,k}, \theta), \qquad w(u_{t_0,k}) = \frac{\tilde{w}(u_{t_0,k})}{\sum_{m=1}^{N} \tilde{w}(u_{t_0,m})}.$$

   (c) **Update** observed data likelihood estimate. Compute $\hat{\pi}_{u_{t_0}}(y_{t_0}|\theta) = \sum_{k=1}^{N} \tilde{w}(u_{1,k})/N$.

2. For $j = 1, \ldots, n$:

   (a) **(optional) Sorting.** Use Euclidean sorting on particles $\{x_{t_{j-1},1}, \ldots, x_{t_{j-1},N}\}$ and their weights if using CPMMH.

   (b) **Resample.** Obtain ancestor indices $a_{t_{j-1}}^k$, $k = 1, \ldots, N$ using systematic resampling on the collection of weights $\{w(u_{t_{j-1},1}), \ldots, w(u_{t_{j-1},N})\}$.

   (c) **Propagate.** Put $x_{t_j,k} = f_{t_j}(u_{t_j,k}) \sim \pi\left(\cdot | x_{t_{j-1},a_{t_{j-1}}^k}, \theta\right)$, $k = 1, \ldots, N$.

   (d) **Compute** the weights. For $k = 1, \ldots, N$ set

$$\tilde{w}(u_{t_j,k}) = \pi(y_{t_j}|x_{t_j,k}, \theta), \qquad w(u_{t_j,k}) = \frac{\tilde{w}(u_{t_j,k})}{\sum_{m=1}^{N} \tilde{w}(u_{t_j,m})}.$$

   (e) **Update** observed data likelihood estimate. Compute

$$\hat{\pi}_{u_{t_0:t_j}}(y_{t_0:t_j}|\theta) = \hat{\pi}_{u_{t_0:t_{j-1}}}(y_{t_0:t_{j-1}}|\theta)\hat{\pi}_{u_{t_j}}(y_{t_j}|y_{t_0:t_j}, \theta)$$

   where $\hat{\pi}_{u_{t_j}}(y_{t_j}|y_{t_0:t_{j-1}}, \theta) = \sum_{k=1}^{N} \tilde{w}(u_{t_j,k})/N$.

---

implementing a correlated pseudo-marginal approach, as described in Section 4.2.3. For the resampling step we follow Deligiannidis *et al.* (2018) among others and use systematic resampling (see e.g. Murray *et al.*, 2016), which only requires simulating a single uniform random variable at each time point.

It is straightforward to augment the auxiliary variable u to include the random variables used in the resampling step. As a by-product of the particle filter, the observed data likelihood $\pi(y|\theta)$ can be estimated via the quantity

$$\hat{\pi}_u(y|\theta) = N^{-n} \prod_{j=0}^{n} \sum_{k=1}^{N} \tilde{w}(u_{t_j,k}). \tag{4.10}$$

4.2.3 Correlated pseudo-marginal Metropolis-Hastings

Using the PMMH scheme may lead to highly variable realisations of $\hat{\pi}_u(y|\theta)$ as we saw in the example in Section 2.5. So we turn to correlated PMMH (CPMMH) to induce positive correlation between successive values of $\hat{\pi}_u(y|\theta)$ as in Section 2.6. Consider the joint density in (4.7). In step $(2a)$ of Algorithm 4, pseudo-marginal Metropolis-Hastings, the auxiliary variables u are proposed from the associated density $g(\cdot)$. The correlated PMMH scheme generates a new u^* from $K(u^*|u)$ where $K(\cdot|\cdot)$ satisfies detailed balance equation

$$g(u)K(u^*|u) = g(u^*)K(u|u^*).$$

It is then straightforward to show that a MH scheme with proposal kernel $q(\theta^*|\theta)K(u^*|u)$ and acceptance probability (4.8) satisfies detailed balance with respect to the target $\pi(\theta, u|y)$. Upon negating the trivial scenario that the chain does not move, we have that

$$
\begin{aligned}
\pi(\theta, u)&q(\theta^*|\theta)K(u^*|u)\alpha(\theta^*, u^*|\theta, u)\\
&= \min\left\{\pi(\theta)g(u)\hat{\pi}_u(y|\theta)q(\theta^*|\theta)K(u^*|u), \hat{\pi}_{u^*}(y|\theta^*)\pi(\theta^*)q(\theta|\theta^*)g(u)k(u^*|u)\right\}\\
&= \min\left\{\pi(\theta)g(u)\hat{\pi}_u(y|\theta)q(\theta^*|\theta)K(u^*|u), \hat{\pi}_{u^*}(y|\theta^*)\pi(\theta^*)q(\theta|\theta^*)g(u^*)k(u|u^*)\right\}\\
&= \pi(\theta^*, u^*)q(\theta|\theta^*)K(u|u^*)\alpha(\theta, u|\theta^*, u^*)
\end{aligned}
$$

We take $g(u)$ as a standard Gaussian density and $K(u^*|u)$ as the kernel associated with a Crank–Nicolson proposal (Deligiannidis *et al.*, 2018). Hence

$$g(u) = \mathrm{N}\left(u; 0, I_d\right) \qquad \text{and} \qquad K(u^*|u) = \mathrm{N}\left(u^*; \rho u, \left(1 - \rho^2\right)I_d\right)$$

where I_d is the identity matrix whose dimension d is determined by the number of elements in u. The parameter ρ is chosen to be close to 1, to induce strong positive correlation between $\hat{\pi}_u(y|\theta)$ and $\hat{\pi}_{u^*}(y|\theta)$, thus reducing the variance of the acceptance probability in (4.8), which is beneficial because it reduces the chance of accepting an overestimation of the likelihood function. Taking $\rho = 0$ gives the special case that $K(u^*|u) = g(u^*)$, which corresponds to the standard PMMH. Iteration i of step 2 of Algorithm 4 then becomes

2. For $i = 1, \dots, n_{\text{iters}}$:

   (a) Draw $\theta^* \sim q(\cdot|\theta^{(i-1)})$. Draw $\omega \sim \mathrm{N}(0, I_d)$ and put $u^* = \rho u^{(i-1)} + \sqrt{1 - \rho^2}\omega$.

   (b) Compute $\hat{\pi}_{u^*}(y|\theta^{(i-1)})$ by running Algorithm 5 with u^*, θ^*, and y.

   (c) With probability given by (4.8) put $\theta^{(i)} = \theta^*$ and $u^{(i)} = u^*$. Otherwise, store the current values $\theta^{(i)} = \theta^{(i-1)}$ and $u^{(i)} = u^{(i-1)}$.

Care must be taken here when executing Algorithm 5 in Step 2(b). Upon changing θ

and u, the effect of the resampling step is likely to prune out different particles, thus breaking the correlation between successive estimates of observed data likelihood. Sorting the particles before resampling can alleviate this problem (Deligiannidis *et al.*, 2018). We follow Choppala *et al.* (2016) (see also Golightly *et al.*, 2019) by using a simple Euclidean sorting procedure which, for the case of a 1-dimensional latent state (e.g. when $\dim(X_t^i) = 1$ for every t) implies, prior to resampling the particles, that the particles are sorted from the smallest to the largest. Deligiannidis *et al.* (2018) sort the particles before resampling via the Hilbert sort procedure of Gerber & Chopin (2015). This is step (2b) in Algorithm 5, denoted "optional" as it only applies to CPMMH, not PMMH.

Tuning advice

It remains that we can choose the number of particles N to be used to obtain estimates of the observed data likelihood contributions $\hat{\pi}_u(y|\theta)$. In the case of PMMH, a simple strategy is to fix θ, and Σ at some central posterior value (obtained from a pilot run), and choose N so that the variance $\log(y|\theta)$ (denoted σ_N^2) is around 2 (Doucet *et al.*, 2015; Sherlock *et al.*, 2015)). When using a CPMMH kernel, we follow Tran *et al.* (2017) by choosing N so that $\sigma_N^2 = 2.16^2/(1 - \rho_l^2)$ where ρ_l is the estimated correlation between $\hat{\pi}_u(y|\theta)$ and $\hat{\pi}_{u^*}(y|\theta)$. Hence, an initial pilot run (with the number of particles set at some conservative value) is required to determine plausible values of the parameters. This pilot run can also be used to give estimates of $\mathrm{var}(\theta|y)$, which can subsequently be used as the innovation variance in a Gaussian random walk proposal for θ.

Example: Ornstein-Uhlenbeck SDE

We consider the following Ornstein-Uhlenbeck (OU) process

$$\begin{cases} Y_t & = X_t + \epsilon_t, \quad \epsilon_t \overset{indep}{\sim} \mathrm{N}(0, \sigma_\epsilon^2) \\ dX_t & = \theta_1(\theta_2 - X_t)dt + \theta_3 dW_t. \end{cases} \tag{4.11}$$

We adopt a parameterisation where $\theta_2 \in \mathbb{R}$ is the stationary mean for the $\{X_t\}$ process, $\theta_1 > 0$ is a growth rate (expressing how rapidly the system reacts to perturbations) and θ_3 is the diffusion coefficient. Let $\theta = (\theta_1, \theta_2, \theta_3)$. The SDE satisfied by X_t can be solved analytically (following the procedure in 3.3.1) to give

$$X_t|X_0 = x_0 \sim N\left(x_0 e^{-\theta_1 t} + \theta_2(1 - e^{-\theta_1 t}), \frac{\theta_3^2}{2\theta_1}(1 - e^{-2\theta_1 t})\right).$$

Although the linear Gaussian solution and the additive Gaussian observation model permit a tractable marginal likelihood, we apply the pseudo-marginal schemes discussed above. We compare "PMMH", which is Algorithm 4 with Algorithm 5 to find the data likelihood

without step 2a, "CPMMH-099", which is Algorithm 4 with a Crank-Nicolson proposal for the u using a correlation of $\rho = 0.99$ in step 2a, and "CPMMH-0999" where we use a correlation of $\rho = 0.999$ in step 2a. The number of particles used for each method was selected using the methods described previously.

We simulate data with the following settings, $n = 100$ observations using an inter-observation time of 1, $\theta = (1, 20, 1)$ and $\sigma_\epsilon \in \{0.1, 0.5, 1\}$. We assumed an initial condition of $x_0 = 5$ so that the process has a reasonable burn in period before it approaches the reversion level of 20. We further assume x_0 and σ_ϵ to be known, for simplicity. The priors assigned to θ were

$$\log(\theta_1) \sim N(3, 10^2),$$
$$\log(\theta_2) \sim N(3, 10^2),$$
$$\log(\theta_3) \sim N(3, 0.5^2).$$

We ran all three methods for the same simulated data set with $\sigma_\epsilon \in \{0.1, 0.5, 1\}$ for 50k iterations, considering the first 10k iterations to be the burn-in period. We set the initial value for $\sigma_\epsilon = 0.7$ and the starting values for θ were set to their actual values.

Results are in Tables 4.1 – 4.2. As a reference for the efficiency of the considered samplers, we take the minimum ESS per minute (mESS/m in Table 4.1 – 4.2) as measured on PMMH as "base/default" value and set it to 1 in the rightmost column of Table 4.1 – 4.2. The minimum ESS per minute for the other samplers are relative to the PMMH value. We can see the effect of varying σ_e, as the observation error decreases, the number of particles required increases. However, the benefit in minimum effective sample size per second due to increasing ρ from 0.99 to 0.999 is smaller as σ_e decreases. From Table 4.1 and Table 4.2 we conclude that CPMMH is about three times more efficient than PMMH in terms of mESS/m when $\sigma_e \in 0.5, 1$, and from Table 4.3 we can conclude that CPMMH is 15 – 25 times more efficient than PMMH in terms of mESS/m when $\sigma_e = 0.1$.

The marginal posteriors in Figures 4.1–4.3 show that the three methods generate very similar posterior inferences when $\sigma_e = 1$ and the true values do not fall outside of the tails, this was the case for each value of σ_e but plots for these cases are omitted.

Algorithm	ρ	N	CPU (s)	mESS	mESS/s	Rel.
PMMH	0	23	929	200	0.215	1
CPMMH-099	0.99	3	319	197	0.618	2.9
CPMMH-0999	0.999	3	362	264	0.729	3.4

Table 4.1: OU SDEMEM with $\sigma_e = 1$. Correlation ρ, number of particles N, CPU time (in minutes m), minimum ESS (mESS), minimum ESS per minute (mESS/m) and relative minimum ESS per minute (Rel.) as compared to PMMH. All results are based on 50k iterations of each scheme.

Algorithm	ρ	N	CPU (s)	mESS	mESS/s	Rel.
PMMH	0	87	1982	3214	1.621	1
CPMMH-099	0.99	22	795	3348	4.213	2.6
CPMMH-0999	0.999	22	744	3189	4.285	2.6

Table 4.2: OU SDEMEM with $\sigma_e = 0.5$. Correlation ρ, number of particles N, CPU time (in minutes m), minimum ESS (mESS), minimum ESS per minute (mESS/m) and relative minimum ESS per minute (Rel.) as compared to PMMH. All results are based on 50k iterations of each scheme.

Algorithm	ρ	N	CPU (s)	mESS	mESS/s	Rel.
PMMH	0	445	9221	2318	0.251	1
CPMMH-099	0.99	20	702	2529	3.803	15.1
CPMMH-0999	0.999	7	419	2644	6.309	25.1

Table 4.3: OU SDEMEM with $\sigma_e = 0.1$. Correlation ρ, number of particles N, CPU time (in minutes m), minimum ESS (mESS), minimum ESS per minute (mESS/m) and relative minimum ESS per minute (Rel.) as compared to PMMH. All results are based on 50k iterations of each scheme.

4.2.4 Computational considerations

Although the approximate inferential approach based on the LNA gives a tractable marginal likelihood, an order $(d_0 + d)^2$ system of coupled ODEs must be solved per marginal likelihood evaluation. Consequently, for systems with a large number of components (e.g. $d > 10$), the LNA approach is likely to be computationally infeasible. Nevertheless, it requires minimal tuning beyond a suitable innovation variance (if using a normal random walk proposal mechanism) that can be obtained from a short pilot run. (C)PMMH requires $N \times n$ one-step simulations per evaluation of an estimate of marginal likelihood. For PMMH N should be order n (Bérard *et al.*, 2013) and for CPMMH, N should be order $n^{1/2}$ for $d = 1$ and order $n^{2/3}$ for $d = 2$ (see Deligiannidis *et al.* (2018) for further discussion). Consequently, in scenarios with relatively short time series, CPMMH may outperform the LNA approach computationally. However, we note that CPMMH requires additional tuning. For example, choosing N is likely to require several short pilot runs of the scheme. Moreover, and as discussed in Owen *et al.* (2015), pseudo-marginal schemes can suffer from long burn-in times if poorly initialised.
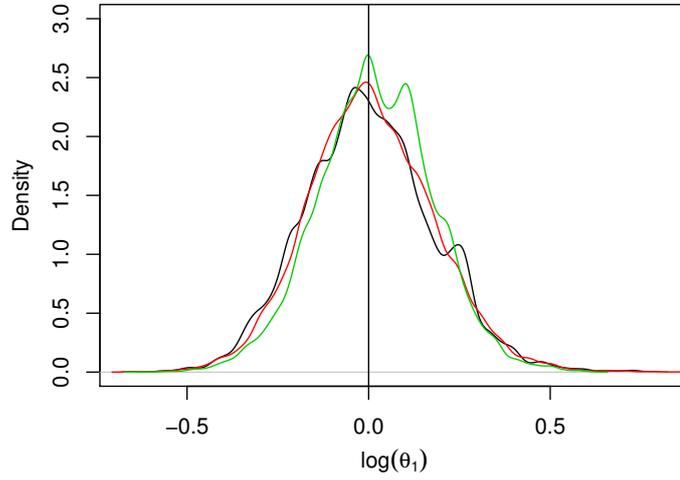
Figure 4.1: Marginal posterior densities for θ_1 when $\sigma_e = 1$. Solid line PMMH, red line CPMMH-099, green line CPMMH-0999, vertical line truth.
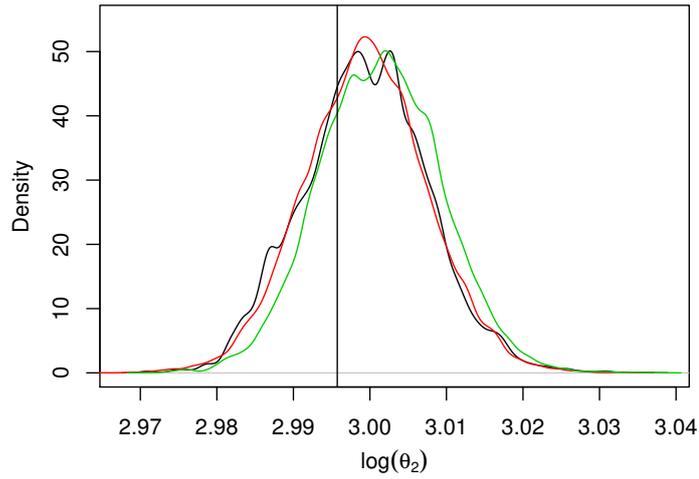
Figure 4.2: Marginal posterior densities for θ_2 when $\sigma_e = 1$. Solid line PMMH, red line CPMMH-099, green line CPMMH-0999, vertical line truth.
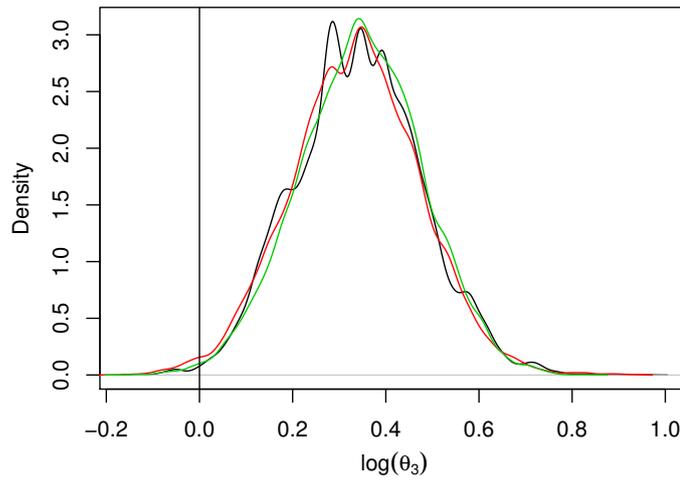
Figure 4.3: Marginal posterior densities for θ_3 when $\sigma_e = 1$. Solid line PMMH, red line CPMMH-099, green line CPMMH-0999, vertical line truth.

Chapter 5

Bayesian inference for mixed effects stochastic differential equations

In this chapter we consider appropriate models for repeated measurement experiments. We consider a mixed-effects framework, where dynamics of each experimental unit are described by linear SDEs. We describe the modelling framework before considering inference in two settings; one in which a linear observation model is assumed and one in which observations are assumed to arise from a nonlinear model.

5.1 Stochastic differential mixed-effects models

Consider the case where we have M experimental units randomly chosen from a theoretical population, and associated with each unit i is a continuous-time d-dimensional Itô process $\{X_t^i, t \geq 0\}$ governed by the SDE

$$dX_t^i = \alpha(X_t^i, \kappa, \phi^i)\, dt + \sqrt{\beta(X_t^i, \kappa, \phi^i)}\, dW_t^i, \quad X_0^i = x_0^i, \quad i = 1, \ldots, M. \qquad (5.1)$$

Here, α is a d-vector of drift functions, the diffusion coefficient β is a $d \times d$ positive definite matrix with a square root representation $\sqrt{\beta}$ such that $\sqrt{\beta}\sqrt{\beta}^T = \beta$ and W_t^i is a d-vector of (uncorrelated) standard Brownian motion processes. Note that we surpress dependence of α and β on time t, for notational simplicity. Allowing explicit time dependence is straightforward in what follows. The p-vector parameter $\kappa = (\kappa_1, \ldots, \kappa_p)^T$ is common to all units whereas the q-vectors $\phi^i = (\phi_1^i, \ldots, \phi_q^i)^T$, $i = 1, \ldots, M$, are unit-specific effects, which may be fixed or random. In the most general random effects scenario we let $\pi(\phi^i|\eta)$ denote the joint distribution of ϕ^i, parameterised by the r-vector $\eta = (\eta_1, \ldots, \eta_r)^T$. The

model defined by (5.1) allows for differences between experimental units through different realisations of the Brownian motion paths W_t^i and the random effects ϕ^i, accounting for inherent stochasticity within a unit, and variation between experimental units respectively. We assume that each experimental unit $\{X_t^i, t \geq 0\}$ cannot be observed exactly, but observations $y^i = (y_1^i, \ldots, y_n^i)^T$ are available. We assume the same number of observations per experimental unit for notational simplicity and note that allowing n to depend on i is easily accommodated. The observations are assumed conditionally independent (given the latent process) and we link them to the latent process via

$$Y_t^i = h(X_t^i, D^i, \epsilon_t^i), \qquad \epsilon_t^i | \Sigma \overset{indep}{\sim} p_\epsilon(\Sigma), \qquad i = 1, ..., M \qquad (5.2)$$

where Y_t^i is a d_o-vector, ϵ_t is a random d_o-vector, $d_o \leq d$, D^i is a unit-specific static or time-dependent deterministic input (e.g. covariates, forcing functions etc.), ϵ_t^i is the measurement noise and $h(\cdot)$ is a possibly nonlinear function of its arguments. We denote the density linking Y_t^i and X_t^i by $\pi(y_t^i | x_t^i, \Sigma)$. A special case that arises from our flexible observation model is when $h(X_t^i, \epsilon_t^i) = F^T X_t^i + \epsilon_t^i$ for a constant matrix F and $\epsilon_t^i | \Sigma \overset{indep}{\sim} N(0, \Sigma)$, allowing for observation of a linear combination of components of X_t^i, subject to additive Gaussian noise. Note that our methodology in Sections 5.2–5.4.4 can be applied to an arbitrary $h(\cdot)$, provided this can be evaluated pointwisely for any value of its arguments.

We refer to the model constituted by the system (5.1)–(5.2) as a stochastic differential equation mixed effects model (SDEMEM). This is a state-space model, due to the Markov property of the Itô processes $\{X_t^i, t \geq 0\}$, and the assumption of conditional independence of the observations given the latent process. The model is flexible: equation (5.1) explains the intrinsic stochasticity in the dynamics (via β) and the variation between-units (via the random effects ϕ^i), while (5.2) explains residual variation (measurement error, via Σ).

5.2 Linear stochastic differential equations with linear observation model

Suppose that the latent process associated with experimental unit i follows (5.1) where the drift and diffusion functions are given by

$$\alpha(X_t^i, \kappa, \phi^i) = a_1(\kappa, \phi^i)X_t^i + a_2(\kappa, \phi^i)$$
$$\beta(X_t^i, \kappa, \phi^i) = b(\kappa, \phi^i)$$

where a_1, a_2 and b may depend on t but we suppress this dependence for notational simplicity. Hence, we assume a linear SDE (in the narrow sense). Suppose further that the observation model is

$$Y_t^i = F^T X_t^i + \epsilon_t^i, \qquad \epsilon_t \overset{indep}{\sim} N(0, \Sigma).$$

Given data $y = (y^1, \ldots, y^M)^T$, the joint posterior for the common parameters κ, fixed/random effects $\phi = (\phi^1, \ldots, \phi^M)^T$, hyperparameters η, measurement error parameters Σ and latent values x is

$$\pi(\kappa, \eta, \Sigma, \phi, x | y) \propto \pi(\kappa)\pi(\eta)\pi(\Sigma)\pi(\phi|\eta)\pi(x|\kappa, \phi)\pi(y|x, \Sigma) \tag{5.3}$$

where $\pi(\kappa)\pi(\eta)\pi(\Sigma)$ is the joint prior density ascribed to κ, η and Σ. In addition we have that

$$\pi(\phi|\eta) = \prod_{i=1}^{M} \pi(\phi^i|\eta), \tag{5.4}$$

$$\pi(y|x, \Sigma) = \prod_{i=1}^{M} \prod_{j=1}^{n} \pi(y_j^i | x_j^i, \Sigma) \tag{5.5}$$

and

$$\pi(x|\kappa, \phi) = \prod_{i=1}^{M} \pi(x_1^i) \prod_{j=2}^{n} \pi(x_j^i | x_{j-1}^i, \kappa, \phi^i). \tag{5.6}$$

In what follows, we assume that interest lies in the marginal posterior for all parameters, which is given by

$$\pi(\kappa, \eta, \Sigma, \phi | y) \propto \pi(\kappa)\pi(\eta)\pi(\Sigma)\pi(\phi|\eta)\pi(y|\kappa, \Sigma, \phi) \tag{5.7}$$

$$\propto \pi(\kappa)\pi(\eta)\pi(\Sigma) \prod_{i=1}^{M} \pi(\phi^i|\eta)\pi(y^i|\kappa, \Sigma, \phi^i). \tag{5.8}$$

This factorization suggests a Gibbs sampler with separate blocks for each parameter vector that sequentially takes draws from the full conditionals

1. $\pi(\phi|\kappa, \eta, \Sigma, y) \propto \prod_{i=1}^{M} \pi(\phi^i|\eta)\pi(y^i|\kappa, \Sigma, \phi^i)$,

2. $\pi(\kappa|\eta, \Sigma, \phi, y) = \pi(\kappa|\phi, \Sigma, y) \propto \pi(\kappa) \prod_{i=1}^{M} \pi(y^i|\kappa, \Sigma, \phi^i)$,

3. $\pi(\Sigma|\kappa, \eta, \phi, y) = \pi(\Sigma|\kappa, \phi, y) \propto \pi(\Sigma) \prod_{i=1}^{M} \pi(y^i|\kappa, \Sigma, \phi^i)$,

4. $\pi(\eta|\kappa, \Sigma, \phi, y) = \pi(\eta|\phi) \propto \pi(\eta) \prod_{i=1}^{M} \pi(\phi^i|\eta)$.

Although the likelihood terms $\pi(y^i|\kappa, \Sigma, \phi^i)$ are tractable, and can be computed using a forward filter of Algorithm 1 the full conditionals in steps 1, 2 and 3 will typically be intractable necessitating Metropolis-Hastings steps. In scenarios where the (log of the) random effects parameters ϕ^i follow normal distributions (as is the case in Chapters 3 and 4) a semi-conjugate prior specification for η is possible, admitting tractable Gibbs steps for these parameters. The full inference scheme is given by Algorithm 6. Note that for ease of presentation Σ and κ have been treated as a single block.

Algorithm 6 Metropolis within Gibbs (tractable marginal likelihood)

---

1. Initialise $\eta^{(0)}, \kappa^{(0)}, \Sigma^{(0)}, \phi^{(0)}$. Run M forward filters (conditional on $\kappa^{(0)}, \Sigma^{(0)}, \phi^{(0)}$) to obtain

$$\pi(y|\kappa^{(0)}, \Sigma^{(0)}, \phi^{(0)}).$$

   Set the iteration counter to $j = 1$

2. (a) Propose $(\Sigma^*, \kappa^*) \sim q(\cdot|\Sigma^{(j-1)}, \kappa^{(j-1)})$

   (b) Run M forward filters (conditional on (Σ^*, κ^*)) to obtain $\pi(y|\kappa^*, \Sigma^*, \phi^{(j-1)})$

   (c) With probability

$$\alpha\big([\Sigma^*, \kappa^*]|[\Sigma^{(j-1)}, \kappa^{(j-1)}]\big) =$$
$$\min\left\{1, \frac{\pi(\Sigma^*)\pi(\kappa^*)}{\pi(\Sigma_e^{(j-1)})\pi(\kappa^{(j-1)})} \times \frac{\pi(y|\kappa^*, \Sigma^*, \phi^{(j-1)})}{\pi(y|\kappa^{(j-1)}, \Sigma^{(j-1)}, \phi^{(j-1)})}\right\}$$

   Put $(\Sigma^{(j)}, \kappa^{(j)}) = (\Sigma^*, \kappa^*)$ otherwise put $(\Sigma^{(j)}, \kappa^{(j)}) = (\Sigma^{(j-1)}, \kappa^{(j-1)})$.

3. For $i = 1, \ldots, M$:

   (a) Propose $\phi^{i*} \sim q(\cdot|\phi^{i,(j-1)})$

   (b) Run a forward filter (conditional on $\phi^{i*}, \Sigma^{(j)}, \kappa^{(j)}$) to obtain

$$\hat{\pi}(y^i|\kappa^{(j)}, \Sigma^{(j)}, \phi^{i*})$$

   (c) With probability

$$\alpha(\phi^{i*}|\phi^{i,(j-1)}) = \min\left\{1, \frac{\pi(\phi^{i*}|\eta)}{\pi(\phi^{i,(j-1)}|\eta^{(j-1)})} \times \frac{\pi(y^i|\kappa^{(j)}, \Sigma^{(j)}, \phi^{i*})}{\pi(y^i|\kappa^{(j)}, \Sigma^{(j)}, \phi^{i,(j-1)})}\right\}$$

   Put $\phi^{i,(j)} = \phi^{i*}$ otherwise put $\phi^{i,(j)} = \phi^{i,(j-1)}$.

4. Draw $\eta^{(j)} \sim \pi(\cdot|\phi^{(j)})$ [Gibbs or Metropolis-Hastings]

5. Set $j := j + 1$. Go to step 2.

5.3 Linear stochastic differential equations with non-linear observation model

Consider now an SDEMEM where the dynamics of each experimental unit is described by a linear SDE and the observation model is of the form (5.2), where $h(\cdot)$ is a non-linear function of X_t^i. We assume interest lies in the marginal posterior

$$\pi(\kappa, \eta, \Sigma, \phi|y) \propto \pi(\kappa)\pi(\eta)\pi(\Sigma)\pi(\phi|\eta)\pi(y|\kappa, \Sigma, \phi) \qquad (5.9)$$

$$\propto \pi(\kappa)\pi(\eta)\pi(\Sigma)\prod_{i=1}^{M}\pi(\phi^i|\eta)\pi(y^i|\kappa, \Sigma, \phi^i) \qquad (5.10)$$

which suggests a Gibbs sampler that alternates between draws of 1, 2, 3, 4 as in Section 5.2. Of course, in practice, the observed data likelihood $\pi(y^i|\kappa, \Sigma, \phi^i)$ will be intractable. In what follows, therefore, we consider a Metropolis within Gibbs strategy, and in particular introduce auxiliary variables u to allow pseudo-marginal Metropolis-Hastings updates.

5.4 A pseudo-marginal approach

Consider again the intractable target in (5.9) and suppose that we can unbiasedly estimate the intractable observed data likelihood $\pi(y|\kappa, \Sigma, \phi)$. To this end let

$$\hat{\pi}_U(y|\kappa, \Sigma, \phi) = \prod_{i=1}^{M}\hat{\pi}_{U^i}(y^i|\kappa, \Sigma, \phi^i)$$

denote a (non-negative) unbiased estimator of $\pi(y|\kappa, \Sigma, \phi)$, where $u = (u^1, \ldots, u^M)^T$ is the collection of auxiliary variables used to produce the corresponding estimate, with density $\pi(u) = \prod_{i=1}^{M} g(u^i)$. Now, the pseudo-marginal Metropolis-Hastings (PMMH) scheme targets

$$\pi(\kappa, \eta, \Sigma, \phi, u|y) \propto \pi(\kappa)\pi(\eta)\pi(\Sigma)\pi(\phi|\eta)\hat{\pi}_u(y|\kappa, \Sigma, \phi)\pi(u) \qquad (5.11)$$

for which it is easily checked that

$$\int \pi(\kappa, \eta, \Sigma, \phi, u|y)du \propto \pi(\kappa)\pi(\eta)\pi(\Sigma)\pi(\phi|\eta)\int \hat{\pi}_u(y|\kappa, \Sigma, \phi)\pi(u)du$$

$$\propto \pi(\kappa, \eta, \Sigma, \phi|y).$$

Hence, marginalising out U gives the marginal parameter posterior in (5.9). Directly targeting the high dimensional posterior $\pi(\kappa, \eta, \Sigma, \phi, u|y)$ with PMMH is likely to give very small acceptance rates. The structure of the SDEMEM naturally admits a Gibbs

sampling strategy that we outline in the next section.

5.4.1 Gibbs sampling and blocking strategies

The form of (5.11) immediately suggests a Gibbs sampler that sequentially takes draws from the full conditionals

1. $\pi(\phi^i, u^i | \kappa, \eta, \Sigma, y^i) \propto \pi(\phi^i | \eta) \hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i) g(u^i)$, $i = 1, \ldots, M$,

2. $\pi(\kappa, u | \eta, \Sigma, \phi, y, u) = \pi(\kappa, u | \phi, \Sigma, y) \propto \pi(\kappa) \prod_{i=1}^{M} \hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i) g(u^i)$,

3. $\pi(\Sigma, u | \kappa, \eta, \phi, y, u) = \pi(\Sigma, u | \kappa, \phi, y) \propto \pi(\Sigma) \prod_{i=1}^{M} \hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i) g(u^i)$,

4. $\pi(\eta | \kappa, \Sigma, \phi, y, u) = \pi(\eta | \phi) \propto \pi(\eta) \prod_{i=1}^{M} \pi(\phi^i | \eta)$.

Note that step 1 consists of a set of draws of M conditionally independent random variables since

$$\pi(\phi, u | \kappa, \eta, \Sigma, y) = \prod_{i=1}^{M} \pi(\phi^i, u^i | \kappa, \eta, \Sigma, y^i).$$

Hence, step 1 gives a sample from $\pi(\phi, u | \kappa, \eta, \Sigma, y)$. Draws from the full conditionals in 1-3 can be obtained by using Metropolis-Hastings within Gibbs. Taking the $[\phi^i, u^i]$ block as an example, we use a proposal density of the form $q(\phi^{i*} | \phi^i) g(u^{i*})$ and accept a move from $[\phi^i, u^i]$ to $[\phi^{i*}, u^{i*}]$ with probability

$$\min\left\{1, \frac{\pi(\phi^{i*} | \cdot)}{\pi(\phi^i | \cdot)} \times \frac{\hat{\pi}_{u^{i*}}(y^i | \phi^{i*}, \cdot)}{\hat{\pi}_{u^i}(y^i | \phi^i, \cdot)} \times \frac{q(\phi^i | \phi^{i*})}{q(\phi^{i*} | \phi^i)}\right\}.$$

Effectively, samples from the full conditionals in 1–3 are obtained via draws from pseudo-marginal MH kernels. However, the above strategy is somewhat naive, since the auxiliary variables U need only be updated once per Gibbs iteration. We therefore propose to update the blocks $[\phi^i, u^i]$, $i = 1, \ldots, M$ in step 1, and condition on the most recent value of u in the remaining steps. Explicitly, we take draws from

1. $\pi(\phi^i, u^i | \kappa, \eta, \Sigma, y^i) \propto \pi(\phi^i | \eta) \hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i) g(u^i)$, $i = 1, \ldots, M$,

2. $\pi(\kappa | \eta, \Sigma, \phi, y, u) = \pi(\kappa | \phi, \Sigma, y, u) \propto \pi(\kappa) \prod_{i=1}^{M} \hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i)$,

3. $\pi(\Sigma | \kappa, \eta, \phi, y, u) = \pi(\Sigma | \kappa, \phi, y, u) \propto \pi(\Sigma) \prod_{i=1}^{M} \hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i)$,

4. $\pi(\eta | \kappa, \Sigma, \phi, y, u) = \pi(\eta | \phi) \propto \pi(\eta) \prod_{i=1}^{M} \pi(\phi^i | \eta)$.

The aim of blocking in this way is to reduce the variance of the acceptance probability associated with steps 2 and 3, which involve the product of M estimates as opposed to a single estimate in each constituent part of step 1. The effect of blocking in this way is explored empirically in Chapter 7.

5.4.2 Estimating the likelihood

It remains that we can generate non-negative unbiased estimates $\hat{\pi}_u(y|\kappa, \Sigma, \phi)$. This can be achieved by running a (bootstrap) particle filter that, for a single experimental unit, recursively draws from the filtering distribution $\pi(x_t^i|y_{1:t}^i, \kappa, \Sigma, \phi^i)$ for each $t = 1, \ldots, n$, as described in Section 4.2.2. Essentially, a sequence of importance sampling and resampling steps is used to propagate a weighted sample $\{(x_{t,k}^i, w(u_{t,k}^i)), k = 1, \ldots, N_i\}$ from the filtering distribution. Note that we let the weight depend explicitly on the tth component of the auxiliary variable $u^i = (u_1^i, \ldots, u_n^i)$, associated with experimental unit i. At time t, the particle filter uses the approximation

$$\hat{\pi}(x_t^i|y_{1:t}^i, \kappa, \Sigma, \phi^i) \propto \pi(y_t^i|x_t^i, \Sigma) \sum_{k=1}^{N_i} \pi(x_t^i|x_{t-1,k}^i, \kappa, \phi^i) w(u_{t-1,k}^i). \tag{5.12}$$

A simple importance sampling/resampling strategy follows, where particles are resampled (with replacement) in proportion to their weights, propagated via $x_{t,k}^i = f_t(u_{t,k}^i) \sim \pi(\cdot|x_{t-1,k}^i, \kappa, \phi^i)$ and reweighted by $p(y_t^i|x_{t,k}^i, \Sigma)$. Recall that $f_t(\cdot)$ is a deterministic function of $u_{t,k}^i$ (as well as the parameters and previous latent state, suppressed for simplicity) that gives an explicit connection between the particles and auxiliary variables. An example of $f_t(\cdot)$ is to take the Euler-Maruyama approximation

$$f_t(u_{t,k}^i) = x_{t-1,k}^i + \alpha(x_{t-1,k}^i, \kappa, \phi^i)\,\Delta t + \sqrt{\beta(x_{t-1,k}^i, \kappa, \phi^i)\Delta t}\,u_{t,k}^i$$

where $u_{t,k}^i \sim N(0, I_d)$ and Δt is a suitably chosen time-step. In practice, unless Δt is sufficiently small to allow an accurate Euler-Maruyama approximation, $f_t(u_{t,k}^i)$ will describe recursive application of the numerical approximation.

For a linear SDE in the narrow sense, we have that

$$f_t(U_{t,k}^i) = x_{t-1,k}^i + m_{t|t-1}^i + \sqrt{V_{t|t-1}^i}\,u_{t,k}^i$$

where $m_{t|t-1}^i$ is given by (3.13) integrated over $(t, t-1]$ with initial condition $x_{t-1,k}^i$ and $V_{t|t-1}^i$ is (3.14) integrated over $(t, t-1]$ with initial condition 0. Algorithm 7 provides a complete description of the particle filter. For the resampling step, we again use systematic resampling (see e.g. Murray *et al.*, 2016), which only requires simulating a single uniform random variable at each time point. It is straightforward to augment the auxiliary variable u^i to include the random variables used in the resampling step. As a by-product of the particle filter, the observed data likelihood $\pi(y^i|\kappa, \Sigma, \phi^i)$ can be estimated via the quantity

$$\hat{\pi}(y^i|\kappa, \Sigma, \phi^i) = N_i^{-n} \sum_{k=1}^{N_i} \tilde{w}(u_{1,k}^i) \prod_{t=2}^{n} \sum_{k=1}^{N_i} \tilde{w}(u_{t,k}^i). \tag{5.13}$$

Algorithm 7 Bootstrap particle filter for experimental unit i

---

1. Initialisation $(t = 0)$.

   (a) **Sample** the prior. Put $x^i_{t_0,k} = f_0(u^i_{t_0,k}) \sim \pi(\cdot)$, $k = 1, \ldots, N_i$.

   (b) **Compute** the weights. For $k = 1, \ldots, N_i$ set

   $$\tilde{w}(u^i_{t_0,k}) = \pi(y^i_{t_0}|x^i_{t_0,k}, \Sigma), \qquad w(u^i_{t_0,k}) = \frac{\tilde{w}(u^i_{t_0,k})}{\sum_{j=1}^{N_i} \tilde{w}(u^i_{t_0,j})}.$$

   (c) **Update** observed data likelihood estimate. Compute $\hat{\pi}_{u^i_{t_0}}(y^i_{t_0}|\kappa, \Sigma, \phi^i) = \sum_{k=1}^{N_i} \tilde{w}(u^i_{t_0,k})/N_i$.

2. For times $t = 2, 3, \ldots, n$:

   (a) **Resample.** Obtain ancestor indices $a^k_{t_{j-1}}$, $k = 1, \ldots, N_i$ using systematic resampling on the collection of weights $\{w(u^i_{t_{j-1},1}), \ldots, w(u^i_{t_{j-1},N_i})\}$.

   (b) **Propagate.** Put $x^i_{t_j,k} = f_t(u^i_{t_j,k}) \sim \pi\big(\cdot|x^i_{t_{j-1},a^k_{t_{j-1}}}, y^i_{t_j}, \kappa, \Sigma, \phi^i\big)$, $k = 1, \ldots, N_i$.

   (c) **Compute** the weights. For $k = 1, \ldots, N_i$ set

   $$\tilde{w}(u^i_{t_j,k}) = \pi(y^i_{t_j}|x^i_{t_j,k}, \Sigma), \qquad w(u^i_{t_j,k}) = \frac{\tilde{w}(u^i_{t_j,k})}{\sum_{m=1}^{N_i} \tilde{w}(u^i_{t_j,m})}.$$

   (d) **Update** observed data likelihood estimate. Compute

   $$\hat{\pi}_{u^i_{t_0:t_j}}(y^i_{t_0:t_j}|\kappa, \Sigma, \phi^i) = \hat{\pi}_{u^i_{t_0:t_{j-1}}}(y^i_{t_0:t_{j-1}}|\kappa, \Sigma, \phi^i)\hat{\pi}_{u^i_{t_j}}(y^i_{t_j}|y^i_{t_0:t_{j-1}}, \kappa, \Sigma, \phi^i)$$

   where $\hat{\pi}_{u^i_{t_j}}(y^i_{t_j}|y^i_{t_0:t_{j-1}}, \kappa, \Sigma, \phi^i) = \sum_{k=1}^{N_i} \tilde{w}(u^i_{t_j,k})/N_i$.

We remind the reader that the corresponding estimator can be shown to be unbiased (Del Moral, 2004; Pitt *et al.*, 2012).

The full Gibbs sampler for generating draws from the joint posterior (5.11) is given by Algorithm 8. For ease of exposition, we have blocked the updates for κ and Σ, but note that the use of separate updates for these parameters is straightforward. The precise implementation of step 4 is likely to be example specific, and we anticipate that a direct draw of $\eta^{(j)} \sim \pi(\cdot|\phi^{(j)})$ will often be possible, for example when the components of ϕ are assumed to be normally distributed and η consists of the corresponding means and precisions, for which a semi-conjugate prior specification is possible. Executing Algorithm 8 requires order $n \sum_{i=1}^{M} N_i$ draws from the transition density governing the SDE in (5.1) *per iteration*. Although not considered in this thesis, we note that in scenarios

---

Algorithm 8 Gibbs sampler

---

1. Initialise $\phi^{(0)} = (\phi^{1,(0)}, \ldots, \phi^{M,(0)})$, $\kappa^{(0)}$, $\Sigma^{(0)}$. Draw $u^{i,(0)} \sim g(\cdot)$ and run Algorithm 7 for $i = 1, \ldots, M$ with $u^{i,(0)}$, $\phi^{i,(0)}, \kappa^{(0)}, \Sigma^{(0)}$ and y^i to obtain $\hat{\pi}_{u^{i,(0)}}(y^i|\kappa^{(0)}, \Sigma^{(0)}, \phi^{i,(0)})$. Set the iteration counter $j = 1$.

2. Update subject specific parameters. For $i = 1, \ldots, M$:

   (a) Propose $u^{i*} \sim g(\cdot)$ and $\phi^{i*} \sim q(\cdot|\phi^{i,(j-1)})$.

   (b) Compute $\hat{\pi}_{u^{i*}}(y^i|\kappa^{(j-1)}, \Sigma^{(j-1)}, \phi^{i*})$ by running Algorithm 7 with u^{i*}, ϕ^{i*}, $\kappa^{(j-1)}$, $\Sigma^{(j-1)}$ and y^i.

   (c) With probability

   $$\min\left\{1, \frac{\pi(\phi^{i*}|\eta)}{\pi(\phi^{i,(j-1)}|\eta)} \times \frac{\hat{\pi}_{u^{i*}}(y^i|\kappa^{(j-1)}, \Sigma^{(j-1)}, \phi^{i*})}{\hat{\pi}_{u^{i,(j-1)}}(y^i|\kappa^{(j-1)}, \Sigma^{(j-1)}, \phi^{i,(j-1)})} \times \frac{q(\phi^{i,(j-1)}|\phi^{i*})}{q(\phi^{i*}|\phi^{i,(j-1)})}\right\}$$
   (5.14)

   put $\phi^{i,(j)} = \phi^{i*}$ and $u^{i,(j)} = u^{i*}$. Otherwise, store the current values $\phi^{i,(j)} = \phi^{i,(j-1)}$ and $u^{i,(j)} = u^{i,(j-1)}$.

3. Update common parameters.

   (a) Propose $(\kappa^*, \Sigma^*) \sim q(\cdot|\kappa^{(j-1)}, \Sigma^{(j-1)})$.

   (b) Compute $\hat{\pi}_{u^{(j)}}(y|\kappa^*, \Sigma^*, \phi^{(j)}) = \prod_{i=1}^{M} \hat{\pi}_{u^{i,(j)}}(y^i|\kappa^*, \Sigma^*, \phi^{i,(j)})$ by running Algorithm 7 for $i = 1, \ldots, M$ with $u^{i,(j)}$, $\phi^{i,(j)}$, κ^*, Σ^* and y^i.

   (c) With probability

   $$\min\left\{1, \frac{\pi(\kappa^*)\pi(\Sigma^*)}{\pi(\kappa^{(j-1)})\pi(\Sigma^{(j-1)})} \times \frac{\hat{\pi}_{u^{(j)}}(y|\kappa^*, \Sigma^*, \phi^{(j)})}{\hat{\pi}_{u^{(j)}}(y|\kappa^{(j-1)}, \Sigma^{(j-1)}, \phi^{(j)})} \times \frac{q(\kappa^{(j-1)}, \Sigma^{(j-1)}|\kappa^*, \Sigma^*)}{q(\kappa^*, \Sigma^*|\kappa^{(j-1)}, \Sigma^{(j-1)})}\right\}$$
   (5.15)

   put $(\kappa^{(j)}, \Sigma^{(j)}) = (\kappa^*, \Sigma^*)$. Otherwise, store the current values $(\kappa^{(j)}, \Sigma^{(j)}) = (\kappa^{(j-1)}, \Sigma^{(j-1)})$.

4. Update random effect population parameters. Draw $\eta^{(j)} \sim \pi(\cdot|\phi^{(j)})$.

5. If $j = n_{\text{iters}}$, stop. Otherwise, set $j := j + 1$ and go to step 2.

---

where the transition density is intractable, draws of a suitable numerical approximation are required. For example, we may use the Euler-Maruyama discretisation with time step $\Delta t = 1/m$, where m is chosen to limit the associated discretisation bias. As discussed by Andrieu *et al.* (2010), the number of particles per experimental unit, N_i, should be scaled in proportion to the number of data points n. Consequently, the use of PMMH kernels is likely to be computationally prohibitive in practice. We therefore consider the adaptation of the correlated PMMH scheme (discussed in Section 4.2.2) to our problem.

5.4.3 A correlated pseudo-marginal approach

Consider again the task of sampling the full conditional $\pi(\phi^i, u^i | \kappa, \eta, \Sigma, y^i)$ associated with the ith experimental unit. In step 2(a) of Algorithm 8, a (pseudo-marginal) Metropolis-Hastings step is used whereby the auxiliary variables u^i are proposed from the associated pdf $g(\cdot)$. Recall that as discussed by Deligiannidis *et al.* (2018) (see also Dahlin *et al.*, 2015), the proposal kernel need not be restricted to the use of $g(u^i)$. The correlated PMMH (CPMMH) scheme generates a new u^{i*} from $K(u^{i*}|u^i)$ where $K(\cdot|\cdot)$ satisfies the detailed balance equation

$$g(u^i)K(u^{i*}|u^i) = g(u^{i*})K(u^i|u^{i*}). \tag{5.16}$$

It is then straightforward to show that a MH scheme with proposal kernel $q(\phi^{i*}|\phi^i)K(u^{i*}|u^i)$ and acceptance probability (5.14) satisfies detailed balance with respect to the target $\pi(\phi^i, u^i | \kappa, \eta, \Sigma, y^i)$ (see the discussion in 4.2.2).

We take $g(u^i)$ as a standard Gaussian density and $K(u^{i*}|u^i)$ as the kernel associated with a Crank–Nicolson proposal (Deligiannidis *et al.*, 2018). Hence

$$g(u^i) = \mathrm{N}\left(u^i; 0, I_d\right) \qquad \text{and} \qquad K(u^{i*}|u^i) = \mathrm{N}\left(u^{i*}; \rho u^i, \left(1 - \rho^2\right) I_d\right)$$

where I_d is the identity matrix whose dimension d is determined by the number of elements in u^i. The parameter ρ is chosen to be close to 1, to induce positive correlation between $\hat{\pi}_{u^i}(y^i | \kappa, \Sigma, \phi^i)$ and $\hat{\pi}_{u^{i*}}(y^i | \kappa, \Sigma, \phi^{i*})$, thus reducing the variance of the acceptance probability in (5.14). Taking $\rho = 0$ gives the special case that $K(u^{i*}|u^i) = g(u^{i*})$, which corresponds using PMMH. Iteration j of step 2 of Algorithm 8 becomes

2. For $i = 1, \ldots, M$:

   (a) Propose $\phi^{i*} \sim q(\cdot|\phi^{i,(j-1)})$. Draw $\omega \sim \mathrm{N}(0, I_d)$ and put $u^{i*} = \rho u^{i,(j-1)} + \sqrt{1 - \rho^2}\omega$.

   (b) Compute $\hat{\pi}_{u^{i*}}(y^i | \kappa^{(j-1)}, \Sigma^{(j-1)}, \phi^{i*})$ by running Algorithm 7 with $u^{i*}, \phi^{i*}, \kappa^{(j-1)}$, $\Sigma^{(j-1)}$ and y^i.

   (c) With probability given by (5.14) put $\phi^{i,(j)} = \phi^{i*}$ and $u^{i,(j)} = u^{i*}$. Otherwise, store the current values $\phi^{i,(j)} = \phi^{i,(j-1)}$ and $u^{i,(j)} = u^{i,(j-1)}$.

Care must be taken here when executing Algorithm 7 in Step 2(b). Upon changing ϕ^i and u^i, the effect of the resampling step is likely to prune out different particles, thus breaking the correlation between successive estimates of observed data likelihood. Sorting the particles before resampling can alleviate this problem (Deligiannidis *et al.*, 2018). We follow the simple Euclidean sorting procedure of Section 4.2.2 to sort the particles before resampling. We find that this works well for the example considered in Chapter 7.

5.4.4 Tuning advice

It remains that we can choose the number of particles N_i to be used to obtain estimates of the observed data likelihood contributions $\hat{\pi}_{u^i}(y^i|\kappa, \Sigma, \phi^i)$. Note that we allow a different number of particles per experimental unit to accommodate differing lengths of the y^i and potential model misspecification at the level of an individual unit. In the case of PMMH, a simple strategy is to fix ϕ^i, κ and Σ at some central posterior value (obtained from a pilot run), and choose N_i so that the variance of the log-posterior (denoted $\sigma^2_{N_i}$) is around 2 (Doucet *et al.*, 2015; Sherlock *et al.*, 2015). When using a CPMMH kernel, we follow Tran *et al.* (2017) by choosing N_i so that $\sigma^2_{N_i} = 2.16/(1 - \rho_l^2)$ where ρ_l is the estimated correlation between $\hat{\pi}_{u^i}(y^i|\kappa, \Sigma, \phi^i)$ and $\hat{\pi}_{u^{i*}}(y^i|\kappa, \Sigma, \phi^i)$. Hence, an initial pilot run (with the number of particles set at some conservative value) is required to determine plausible values of the parameters. This pilot run can also be used to give estimates of $\text{var}(\phi^i|y^i)$, $i = 1, \ldots, M$, each of which can subsequently be used as the innovation variance in a Gaussian random walk proposal for ϕ^i.

Chapter 6

Application I

6.1 Background

Caloric restriction (CR) is when adequate nutrition is maintained but total caloric intake is reduced. This has been shown to delay the onset of cancer and other age related diseases, for example in organisms such as yeast, worms, flies and mice (Weindruch & Walford, 1988). It has been shown that the most dramatic effect on lifespan is when CR is early onset, but there is also some evidence to suggest that late onset CR can result in beneficial effects (Weindruch & Walford, 1988; Yu *et al.*, 1985; Spindler, 2005). Most of the studies focus on lifespan (Spindler, 2005) or cancer incidence (Weindruch & Walford, 1988; Volk *et al.*, 1994; Pugh *et al.*, 1999; Spindler, 2005) and very few focus on the whole-animal physiological response to late onset CR. In this work we look at the ways in which mice compensate for the reduction in calories by studying their core body temperature. We have data on 10 mice that were fed *ad libitum* (AL) and 10 mice that were subject to the late onset CR diet. It has been widely shown that a reduction in core body temperature is a key factor in increasing lifespan with CR, this result can vary with strain (Liao *et al.*, 2010). In this research we seek to find out if mice exposed to late onset CR also show a reduction in body temperature.

Based on data arising from a 70 day study into the effect of late onset short term caloric restriction on mice, Golightly *et al.* (2012) developed a joint model for physical activity and core body temperature. They describe a sinusoidal pattern due to the circadian cycle in both the activity and temperature along with a linear relationship between temperature and (transformed) activity. A dynamic linear model was used to allow a time varying amplitude and phase. Their initial analysis suggested that amplitude and phase were plausibly constant and a simple hierarchical model was then used to quantify differences between mice fed ad libitum and those that were caloric restricted. It was found that core body temperature generally showed a decrease during caloric restriction and they found

little difference in the linear coefficient characterising the relationship between temperature and (transformed) activity between AL and CR mice. They also found that basal core body temperature appeared to be fixed at a lower point in CR mice.

Unlike the approach of Golightly *et al.* (2012), we will use a SDEMEM to jointly model core temperature over multiple experimental units (mice). This allows for the incorporation of intrinsic stochasticity inherent in observed temperature traces. Moreover, our hierarchical model allows for time varying amplitude.

6.2 The experimental data

Mice were taken from a long established colony of the C57/BL (ICRFa) strain and placed into one of 20 cages (10 for males and 10 for females), with between four and six mice in each cage. Mice were fed *ad libitum* (AL) until they were fourteen months old and cages were then divided into two groups, with 10 cages in each group; five male cages and five female cages. One group was allocated to AL feeding, and the other to caloric restriction (CR) for 36 months. CR mice were provided with a 40% food restriction relative to the AL group, delivered as one daily ration at around 9:30am. One randomly chosen mouse in each cage was implanted with a wireless E-mitter which continuously monitored body temperature. The data consist of hourly average temperature measurements, but due to some power cuts there is some missingness in the data. Data were also removed for the period between 9am and 12 noon every day due to the mice being disturbed then for feeding. The first two weeks and the last two weeks worth of data were also removed. The full data set corresponds to hourly temperature averages over a period of six months, for each of the 20 mice.

Exploratory plots of the data are always useful for building good statistical models. From Figure 6.1 it can be seen that all mice have a strong circadian pattern but there is a clear difference in this pattern between CR and AL mice, along with an obvious difference in overall temperature with CR mice appearing hold a lower temperature than the AL mice. In what follows we construct and fit an appropriate SDEMEM for these data.

6.3 Modelling a single experimental unit

6.3.1 SDE model and solution

Consider a bivariate diffusion process $\{X_t, t \geq 0\}$ where

$$X_t = (Z_t, A_t)^T.$$
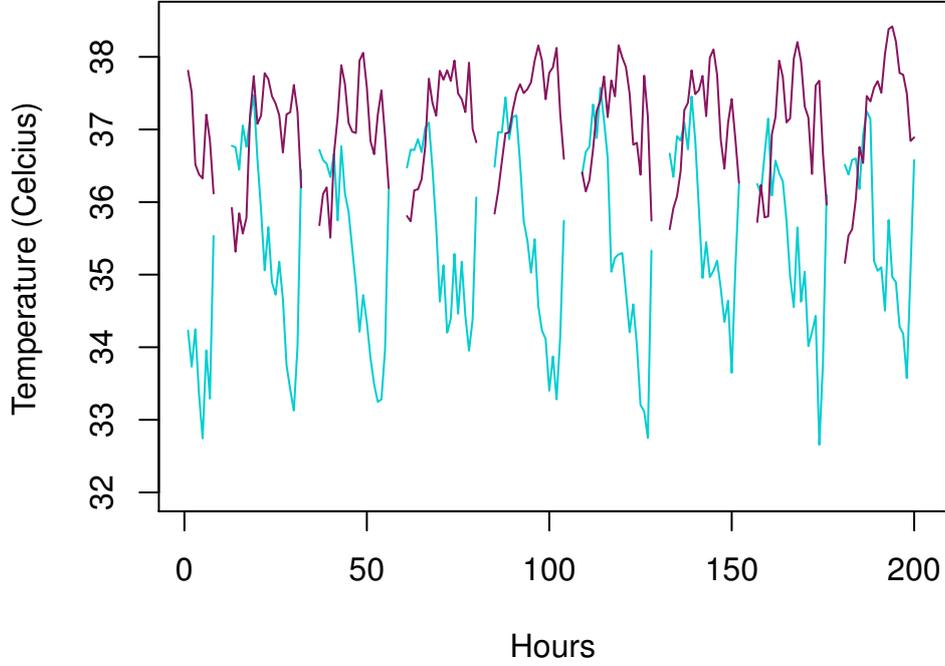
Figure 6.1: The average hourly temperatures of a single mouse randomly selected from each group over eight days. The pink line denotes mice fed *ad libitum* and the blue denotes mice fed on the caloric restricted diet.

Here Z_t denotes temperature and A_t denotes amplitude. The dynamics of X_t are described by a coupled SDE of the form

$$dZ_t = \theta_1 \left(\theta_2 - Z_t\right) dt + \frac{\pi}{12} A_t \cos\left(\frac{\pi t}{12} + B\right) dt + \sigma_1 dW_{1,t} \tag{6.1}$$

$$dA_t = \theta_3 \left(\theta_4 - A_t\right) dt + \sigma_2 dW_{2,t} \qquad W_{1,t} \perp W_{2,t}$$

Hence, we allow both amplitude and temperature to vary stochastically over time. We anticipate that both processes will not deviate too far from some overall mean. We incorporate this belief via the form of the drift governing dA_t and dZ_t, which ensures that amplitude mean-reverts around a value θ_4 at rate θ_3 and the temperature process mean reverts around θ_2 at rate θ_1. We have that σ_1 and σ_2 control the intrinsic noise for the temperature and amplitude respectively, and B denotes the phase shift for the sinusoidal behaviour of the temperature. See Figure 6.2 for how each parameter impacts the trajectories of the temperature paths. Note that the coupled SDE above can be written
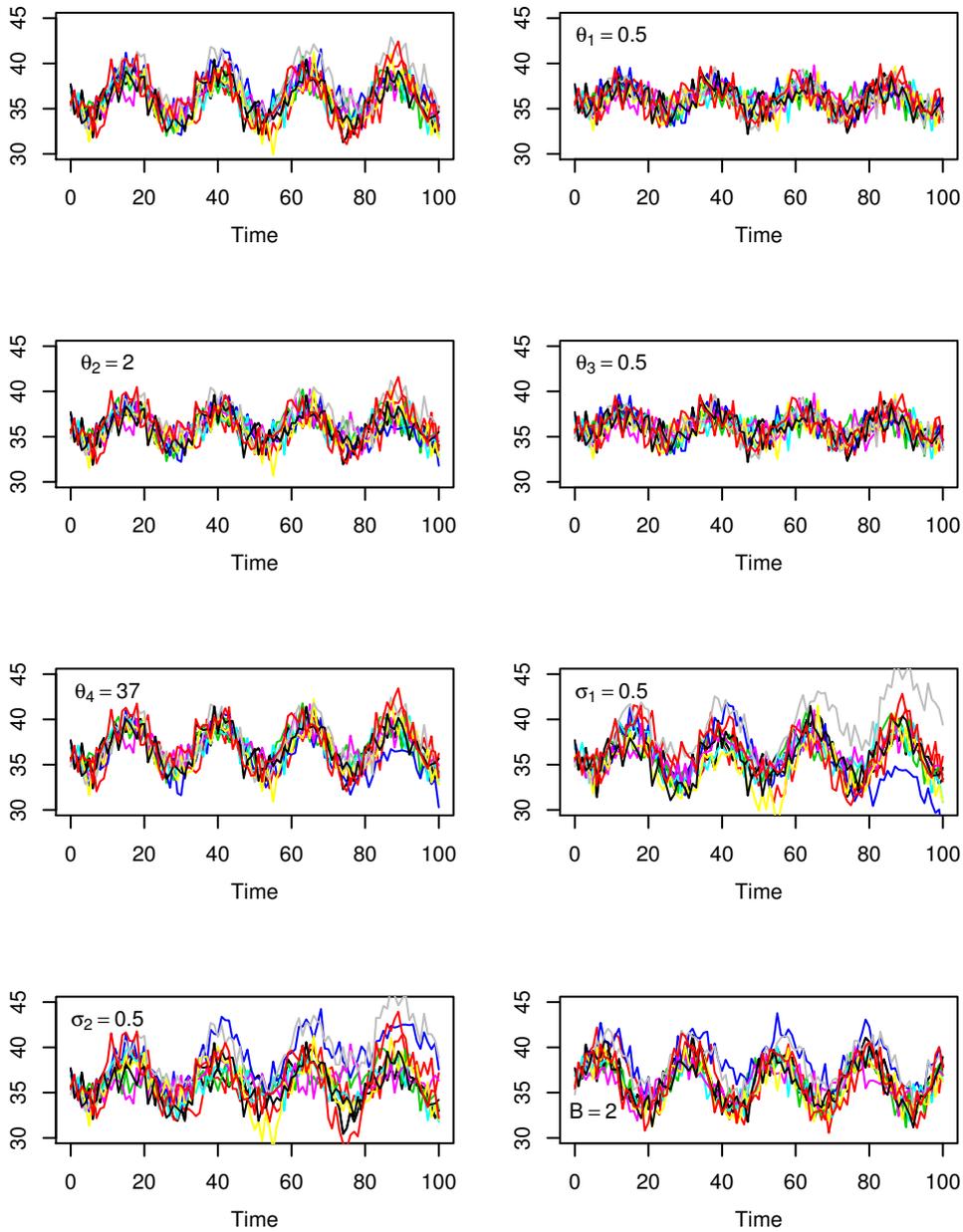
Figure 6.2: Each plot shows 10 simulated paths of mice temperatures using the model described by (6.1) with an added observation error $\sigma_\epsilon \sim N(0,1)$. Using $(\theta_1 = 0.1, \theta_2 = 36, \theta_3 = 0.1, \theta_4 = 3, \sigma_1 = 0.1, \sigma_2 = 0.1, B = 0)^T$ as the starting variables, we change each variable value separately and in turn to see the effect this has on typical trajectories.

as

$$dX_t = [\alpha_1(t)X_t + \alpha_2(t)]\, dt + \sqrt{\beta(t)}dW_t \tag{6.2}$$

where

$$\alpha_1(t) = \begin{pmatrix} -\theta_1 & \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right) \\ 0 & -\theta_3 \end{pmatrix},$$

$$\alpha_2(t) = \begin{pmatrix} \theta_1\theta_2 \\ \theta_3\theta_4 \end{pmatrix},$$

$$\beta(t) = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix},$$

which is the form of a linear SDE in the narrow sense. This particular SDE can be solved as follows; see also Section 3.3.1. Let $U_t = P_t^{-1}X_t$. Recall that P_t is known as a fundamental matrix and satisfies the ODE:

$$\frac{dP_t}{dt} = \alpha_1(t)P_t, \qquad P_0 = I_{2\times 2}$$

Note that $U_0 = P_0^{-1}X_0 = X_0$. Now

$$dU_t = d(P_t^{-1}X_t)$$

and, since U_t is linear in X_t, Itô's formula becomes

$$dU_t = (dP_t^{-1})X_t + P_t^{-1}dX_t. \tag{6.3}$$

Note that

$$\frac{d}{dt}P_tP_t^{-1} = P_t\frac{dP_t^{-1}}{dt} + \frac{dP_t}{dt}P_t^{-1} = 0$$

$$\Rightarrow P_t\frac{dP_t^{-1}}{dt} = -\frac{dP_t}{dt}P_t^{-1}$$

$$\Rightarrow P_t^{-1}P_t\frac{dP_t^{-1}}{dt} = -P_t^{-1}\alpha_1(t)P_tP_t^{-1}$$

$$\Rightarrow dP_t^{-1} = -P_t^{-1}\alpha_1(t)dt.$$

Substituting the above form into equation 6.3 and additionally using the form of dX_t gives

$$dU_t = -P_t^{-1}\alpha_1(t)X_t dt + P_t^{-1}\left(\alpha_2(t) + \alpha_1(t)X_t\right)dt + P_t^{-1}\sqrt{\beta(t)}dW_t$$

$$= P_t^{-1}\alpha_2(t)dt + P_t^{-1}\sqrt{\beta(t)}dW_t$$

and hence

$$U_t = U_0 + \int_0^t P_s^{-1}\alpha_2(s)ds + \int_0^t P_s^{-1}\sqrt{\beta(s)}dWs.$$

Using the Itô Isometry (3.5) we get

$$U_t|U_0 \sim N\left(U_0 + \int_0^t P_s^{-1}\alpha_2(s)ds, \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds\right).$$

Since $X_t = P_t U_t$ we obtain

$$X_t|X_0 \sim N\left(P_t X_0 + P_t \int_0^t P_s^{-1}\alpha_2(s)ds, P_t \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds P_t^T\right).$$

To induce a compact notation we will write

$$X_t|X_0 \sim N(m_t, V_t),$$

and consider m_t and V_t as the solutions of a coupled ODE system which we derive as follows.

We have $m_t = P_t X_0 + P_t \int_0^t P_s^{-1}\alpha_2(s)ds$, so we apply the product rule to give

$$\frac{dm_t}{dt} = \frac{dP_t}{dt}X_0 + \frac{dP_t}{dt}\int_0^t P_s^{-1}\alpha_2(s)ds + P_t P_t^{-1}\alpha_2(t)$$

$$= \alpha_1(t)P_t X_0 + \alpha_1(t)P_t \int_0^t P_s^{-1}\alpha_2(s)ds + \alpha_2(t)$$

$$= \alpha_1(t)\left(P_t X_0 + P_t \int_0^t P_s^{-1}\alpha_2(s)ds\right) + \alpha_2(t)$$

$$\Rightarrow \frac{dm_t}{dt} = \alpha_1(t)m_t + \alpha_2(t).$$

We have $V_t = P_t \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds P_t^T$, so we apply the product rule twice to obtain

$$\frac{dV_t}{dt} = P_t \left\{ P_t^{-1}\beta(t)(P_t^{-1})^T P_t^T + \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds \times \frac{dP_t^T}{dt} \right\}$$
$$+ \frac{dP_t}{dt} \left\{ \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds P_t^T \right\}$$
$$= \beta(t) + P_t \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds P_t^T \alpha_2(t)^T$$
$$+ \alpha_2(t) P_t \int_0^t P_s^{-1}\beta(s)(P_s^{-1})^T ds P_t^T$$
$$\Rightarrow \frac{dV_t}{dt} = V_t \alpha_1(t)^T + \beta(t) + \alpha_1(t)V_t.$$

This yields the coupled ODE system

$$\begin{cases} \dfrac{dV_t}{dt} = V_t \alpha_1(t)^T + \beta(t) + \alpha_1(t)V_t \\ \\ \dfrac{dm_t}{dt} = \alpha_2(t) + \alpha_1(t)m_t \end{cases} \tag{6.4}$$

with initial conditions $m_0 = x_0$ and $V_0 = 0_{2\times2}$, that is, the 2×2 matrix of zeros. The solutions to the above ODE system can be found in Appendix A.1.

6.4   Bayesian inference

Assume that we only observe the temperature component at discrete times, subject to additive Gaussian error. That is

$$Y_t = F^T X_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$$

and

$$F = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Let $y = (y_{t_0}, ..., y_{t_n})^T$ denote the observations at $n + 1$ discrete times $t_0, \ldots, t_n$. Upon ascribing a prior density $\pi(\lambda)$ to $\lambda = (\theta_1, \theta_2, \theta_3, \theta_4, \sigma_1, \sigma_2, B, \sigma_\epsilon)^T$, Bayesian inference may proceed via the posterior density

$$\pi(\lambda|y) \propto \pi(\lambda)\pi(y|\lambda) \tag{6.5}$$

where the marginal likelihood is

$$\pi(y|\lambda) = \pi(y_{t_0}) \prod_{i=0}^{n-1} \pi(y_{t_{i+1}}|y_{t_0:t_i}, \lambda)$$

and the constituent terms $\pi(y_{t_{i+1}}|y_{t_0:t_i}, \lambda)$ are

$$\pi(y_{t_{i+1}}|y_{t_0:t_i}, \lambda) = \int \pi(y_{t_{i+1}}|x_{t_{i+1}}, \lambda)\pi(x_{t_{i+1}}|x_{t_i}, \lambda)\pi(x_{t_i}|y_{t_0:t_i}, \lambda)dx_{t_i:t_{i+1}}.$$

Note that this integral is analytically tractable due to the linear Gaussian structure of the model. Recall from Section 4.1 that these terms can be computed efficiently using a forward filter. For completeness, we give the forward filter in full, for this particular model see Algorithm 9.

6.4.1 Forward filter

Assume that $X_{t_0} \sim N(a, C)$ such that if $X_{t_0} = x_{t_0}$ is known, $a = x_{t_0}$ and C is a 2×2 matrix of zeros.

Recall that $\lambda = (\theta_1, \theta_2, \theta_3, \theta_4, \sigma_1, \sigma_2, B, \sigma_\epsilon)^T$ is the object of inference. Note that in what follows, where appropriate, we suppress dependence on λ for notational simplicity. The forward filter steps for this model are shown in Algorithm 9. Note that if interest is also in the marginal posterior for the latent process, $\pi(x|y)$, a backward sampler can be used to obtain draws of x given y (and λ). The backward sampler is given in full in Algorithm 10.

6.4.2 Metropolis-Hastings

Note that the posterior in (6.5) is intractable. We therefore generate draws of λ using a Metropolis-Hastings scheme. We update λ using a random walk proposal with normal innovations. That is, if the current value of the chain is λ, the proposal mechanism takes the form

$$q(\lambda^*|\lambda) = N(\lambda^*; \lambda, \Sigma)$$

where Σ is a tuning matrix. For example Roberts & Rosenthal (2001) suggest taking $\Sigma = \frac{2.38^2}{2}\hat{Var}(\lambda|y)$ where $\hat{Var}(\lambda|y)$ can be estimated from a pilot run.

The acceptance probability is then

$$\alpha(\lambda^*|\lambda) = \min\left\{1, \frac{\pi(\lambda^*)\pi(y|\lambda^*)}{\pi(\lambda)\pi(y|\lambda)}\right\}$$

where $\pi(y|\lambda^*)$ can be obtained from a single run of the forward filter in Algorithm 9 (and $\pi(y|\lambda)$ is obtained from the last accepted value of λ).

Algorithm 9 Forward filter for temperature model

---

1. Initialise.

   (a) Marginal likelihood update. Note that $Y_{t_0}|X_{t_0} \sim N(F'X_{t_0}, \sigma_\epsilon^2)$, therefore

   $$Y_{t_0} \sim N\left(F'a, F'CF + \sigma_\epsilon^2\right)$$

   $$\Rightarrow \pi\left(y_{t_0}\right) = N\left(y_{t_0}; F'a, F'CF + \sigma_\epsilon^2\right)$$

   (b) Posterior at t_0. Compute $X_{t_0}|y_{t_0} \sim N(a_0, C_0)$ where

   $$a_0 = a + CF\left(F'CF + \sigma_\epsilon^2\right)^{-1}\left(y_{t_0} - F'a\right)$$
   $$C_0 = C - CF\left(F'CF + \sigma_\epsilon^2\right)^{-1}F'C$$

2. Perform the following for $j = 0, 1, \ldots, n - 1$.

   (a) Prior at t_{j+1}. We have $X_{t_j}|Y_{t_0:t_j} \sim N(a_j, C_j)$ which gives

   $$X_{t_{j+1}}|Y_{t_0:t_j} \sim N\left(m(t_{j+1}), V(t_{j+1})\right)$$

   where $m(t_{j+1})$ and $V(t_{j+1})$ are the solutions to the ODE system in (6.4) initialised with $m(t_j) = a_j$ and $V(t_j) = C_j$.

   (b) Marginal likelihood update. Using the observation equation we have

   $$Y_{t_{j+1}}|Y_{t_0:t_j} \sim N\left(F'm(t_{j+1}), F'V(t_{j+1})F + \sigma_\epsilon^2\right)$$

   and therefore $\pi\left(y_{t_{j+1}}|y_{t_0:t_j}\right) = N\left(y_{t_{j+1}}; F'm(t_{j+1}), F'V(t_{j+1})F + \sigma_\epsilon^2\right)$. Hence the updated marginal likelihood is

   $$\pi\left(y_{t_0:t_{j+1}}\right) = \pi\left(y_{t_0} : y_{t_j}\right)\pi\left(y_{t_{j+1}}|y_{t_0:t_j}\right).$$

   (c) Posterior at t_{j+1}.

   $$\begin{pmatrix} X_{t_{j+1}} \\ Y_{t_{j+1}} \end{pmatrix}\Bigg|y_{t_0:t_j} \sim N\left(\begin{pmatrix} m(t_{j+1}) \\ F'm(t_{j+1}) \end{pmatrix}, \begin{pmatrix} V(t_{j+1}) & V(t_{j+1})F \\ F'V(t_{j+1}) & F'V(t_{j+1})F + \sigma_\epsilon^2 \end{pmatrix}\right)$$

   $$\Rightarrow X_{t_{j+1}}|y_{t_0:t_{j+1}} \sim N(a_{j+1}, C_{j+1})$$

   where

   $$a_{j+1} = m(t_{j+1}) + V(t_{j+1})F\left(F'V(t_{j+1})F + \sigma_\epsilon^2\right)^{-1}\left(y_{t_{j+1}} - F'm(t_{j+1})\right),$$

   $$C_{j+1} = V(t_{j+1}) + V(t_{j+1})F\left(F'V(t_{j+1})F + \sigma_\epsilon^2\right)^{-1}FV(t_{j+1}).$$

Algorithm 10 Backwards sampler for temperature model

---

1. Draw x_{t_n} from $X_{t_n}|y \sim N(a_{t_j}, C_{t_j})$

2. For $j = n-1, n-2, \ldots, 0$,

    (a) Joint distribution of X_{t_j} and $X_{t_{j+1}}$. Note that $X_{t_j}|y_{t_0:t_j} \sim N(a_{t_j}, C_{t_j})$. The joint distribution of X_{t_j} and $X_{t_{j+1}}$ conditional on $y_{t_0:t_j}$ is

$$\begin{pmatrix} X_{t_j} \\ X_{t_{j+1}} \end{pmatrix} \Bigg| y_{t_0:t_j} \sim N\left(\begin{pmatrix} a_{t_j} \\ m_{t_{j+1}} \end{pmatrix}, \begin{pmatrix} C_{t_j} & C_{t_j} P_{t_{j+1}}^T \\ P t_{j+1} C_{t_j} & V_{t_{j+1}}. \end{pmatrix} \right)$$

    (b) Backwards distribution. The distribution of $X_{t_j}|x_{t_{j+1}}, y_{t_0:t_j}$ is $N(\hat{a}_{t_j}, \hat{C}_{t_j})$ where

$$\hat{a}_{t_j} = a_{t_j} + C_{t_j} P_{t_{j+1}}^T V_{t_{j+1}}^{-1} (x_{t_{j+1}} - m_{t_{j+1}}),$$
$$\hat{C}_{t_j} = C_{t_j} - C_{t_j} P_{t_{j+1}}^T V_{t_{j+1}}^{-1} P_{t_{j+1}} C_{t_j}.$$

    Draw x_{t_j} from $X_{t_j}|x_{t_{j+1}}, y_{t_0:t_j} \sim N(\hat{a}_{t_j}, \hat{C}_{j_j})$.

This algorithm can be applied to data for each experimental unit independently, at the expensive cost of ignoring intra-subject variability. Therefore, in what follows, we develop a joint model over all experimental units, formulated as an SDEMEM.

6.5 SDEMEM

To better quantify the differences between AL and CR mice we develop an SDEMEM. We allow each mouse to have different parameters, but these parameters are drawn from distributions that are dependent on feeding regime, with regime-specific parameters. Hence, we assess differences between groups by performing inference for the treatment-specific parameters.

6.5.1 Bottom level

Consider experimental unit i in treatment group j with $i = 1, \ldots, n_j$ and $j = 1, 2$, with AL mice designated as group 1 and CR mice designated as group 2. Let

$$X_t^{ij} = \left(Z_t^{ij}, A_t^{ij} \right)^T$$

and assume that X_t^{ij} follows the SDE in (6.1) so that

$$dZ_t^{ij} = \theta_1^{ij}\left(\theta_2^{ij} - Z_t^{ij}\right)dt + \frac{\pi}{12}A_t^{ij}\cos\left(\frac{\pi t}{12} + B^{ij}\right)dt + \sigma_1^{ij}dW_{1,t}^{ij}$$

$$dA_t^{ij} = \theta_3^{ij}\left(\theta_4^{ij} - A_t^{ij}\right)dt + \sigma_2^{ij}dW_{2,t}^{ij} \qquad W_{1,t}^{ij} \perp W_{2,t}^{ij}.$$

6.5.2   Top level

We complete the specification of the hierarchical model via the following.

$$\lambda_1^{ij} = \theta_1^{ij} \sim N\left(\mu_1^j, \left[\tau_1^j\right]^{-1}\right), \qquad \lambda_2^{ij} = \theta_2^{ij} \sim N\left(\mu_2^j, \left[\tau_2^j\right]^{-1}\right),$$

$$\lambda_3^{ij} = \theta_3^{ij} \sim N\left(\mu_3^j, \left[\tau_3^j\right]^{-1}\right), \qquad \lambda_4^{ij} = \theta_4^{ij} \sim N\left(\mu_4^j, \left[\tau_4^j\right]^{-1}\right),$$

$$\lambda_5^{ij} = \sigma_1^{ij} \sim N\left(\mu_5^j, \left[\tau_5^j\right]^{-1}\right), \qquad \lambda_6^{ij} = \sigma_2^{ij} \sim N\left(\mu_6^j, \left[\tau_6^j\right]^{-1}\right),$$

$$\lambda_7^{ij} = logit\left(\frac{B^{ij}}{2\pi}\right) \sim N\left(\mu_7^j, \left[\tau_7^j\right]^{-1}\right), \quad i = 1, \ldots, n_j, \quad j = 1, 2.$$

Note that we avoided using any log normal priors to enable better interpretability and comparability. It remains that we specify prior distributions for the μ_k^j and τ_k^j. We assume that these parameters are independent *a priori* and specify

$$\mu_k^j \sim N(b_k, d_k^{-1}), \qquad \tau_k^j \sim Ga(g_k, h_k), \qquad j = 1, 2, \quad k = 1, \ldots, 7.$$

Note that, for ease of interpretation of the bottom level parameters and to facilitate natural comparison between the two treatment groups, we specify normal distributions for the unit specific parameters (rather than their natural logarithms).

6.5.3   Bayesian inference

We assume that observations are subject to additive Gaussian error

$$Y_t^{ij} = Z_t^{ij} + \epsilon_t^{ij} \qquad \epsilon_t^{ij} \overset{iid}{\sim} N\left(0, \left[\sigma_e^2\right]^j\right),$$

allowing a different measurement error variance in each treatment group. We assume the same observation regime within each treatment group so that mouse i in treatment group j has measurements $y^{ij} = (y_{t_0}^{ij}, \ldots, y_{t_{n_j}}^{ij})^T$. The complete data set is denoted by

$$y = \left(y^{11}, y^{21}, \ldots, y^{n_1 1}, y^{12}, y^{22}, \ldots, y^{n_2 2}\right).$$

The bottom level parameters for mouse i in treatment group j are denoted by $\lambda^{ij} = \left(\lambda_1^{ij}, \ldots, \lambda_7^{ij}\right)^T$ with

$$\lambda = \left(\lambda^{11}, \lambda^{21}, \ldots, \lambda^{n_1 1}, \lambda^{12}, \lambda^{22}, \ldots, \lambda^{n_2 2}\right)$$

denoting all bottom level parameters. For the top level parameters we write $\mu^j = \left(\mu_1^j, \ldots, \mu_7^j\right)^T$ and $\tau^j = \left(\tau_1^j, \ldots, \tau_7^j\right)^T$. To induce a compact notation, we further write $\psi^j = \left((\mu^j)^T, (\tau^j)^T\right)^T$ so that

$$\psi = \left(\psi^1, \psi^2\right)$$

denotes all bottom level parameters. Finally, we have $\sigma_e = \left([\sigma_e]^1, [\sigma_e]^2\right)^T$ and drop the use of the inner parentheses when referring to treatment groups. By Bayes theorem

$$\pi(\lambda, \psi, \sigma_e | y) \propto \prod_{j=1}^{2} \prod_{i=1}^{n_j} \pi(\sigma_e^j) \pi(\psi^j) \pi(\lambda^{ij} | \psi^j) \pi(y^{ij} | \lambda^{ij}, \sigma_e^j) \tag{6.6}$$

where we have used that the y^{ij} are independent, given λ^{ij} and σ_e^j. Given the assumption that the components of ψ^j ($j = 1, 2$) are independent a priori we have

$$\pi(\psi^j) = \prod_{k=1}^{7} \pi(\mu_k^j) \pi(\tau_k^j).$$

Since (6.6) is intractable we use a Metropolis-Hastings scheme to generate draws of λ, ψ and σ_e.

6.5.4 Metropolis-Hastings

We update λ, ψ and σ_e by iterating over the following draws.

1. $\lambda^{ij} \sim \pi\left(\lambda^{ij} | \mu^j, \tau^j, \sigma_e^j, y^{ij}\right), \quad i = 1, \ldots, n_j, \quad j = 1, 2.$

2. $\mu_k^j \sim \pi\left(\mu_k^j | \tau_k^j, \lambda_k^{1j}, \ldots, \lambda_k^{n_j j}\right), \quad j = 1, 2, \quad k = 1, \ldots, 7.$

3. $\tau_k^j \sim \pi\left(\tau_k^j | \mu_k^j, \lambda_k^{1j}, \ldots, \lambda_k^{n_j j}\right), \quad j = 1, 2, \quad k = 1, \ldots, 7.$

4. $\sigma_e^j \sim \pi\left(\sigma_e^j | \lambda^{1j}, \ldots, \lambda^{n_j j}, y^{1j}, \ldots, y^{n_j j}\right), \quad j = 1, 2.$

Note that we have exploited the conditional dependencies between parameter blocks where possible. Steps 2 and 3 admit tractable updates. Metropolis-within-Gibbs steps are required for the remaining updates. We provide details as follows.

λ^{ij} **update**

We have that

$$\pi\left(\lambda^{ij}|\mu^j,\tau^j,\sigma_e^j,y^{ij}\right) \propto \pi(\lambda^{ij}|\psi^j)\pi(y^{ij}|\lambda^{ij},\sigma_e^j).$$

Since this full conditional density (FCD) is intractable we update λ^{ij} using a Metropolis-Hastings step. For a symmetric proposal (e.g. random walk with normal innovations) with density $q\left(\tilde{\lambda}^{ij}|\lambda^{ij}\right) = q\left(\lambda^{ij}|\tilde{\lambda}^{ij}\right)$ the acceptance probability is

$$\alpha\left(\tilde{\lambda}^{ij}|\lambda^{ij}\right) = \min\left\{1,\frac{\pi(\tilde{\lambda}^{ij})\pi(y^{ij}|\tilde{\lambda}^{ij},\sigma_e^j)}{\pi(\lambda^{ij})\pi(y^{ij}|\lambda^{ij},\sigma_e^j)}\right\}.$$

μ_k^j **update**

We have that

$$\pi\left(\mu_k^j|\tau_k^j,\lambda_k^{1j},\ldots,\lambda_k^{n_jj}\right)$$

$$\propto \pi\left(\mu_k^j\right)\prod_{i=1}^{n_j}\pi\left(\lambda_k^{ij}|\mu_k^j,\tau_k^j\right)$$

$$\propto \exp\left\{-\frac{d_k}{2}\left(\mu_k^j-b_k\right)^2\right\}\exp\left\{-\frac{\tau_k^j}{2}\sum_{i=1}^{n_j}\left(\lambda_k^{ij}-\mu_k^j\right)^2\right\}$$

$$\propto \exp\left\{-\frac{d_k}{2}\left[(\mu_k^j)^2-2b_k\mu_k^j\right]-\frac{\tau_k^j}{2}\left[\sum_{i=1}^{n_j}\left(\lambda_k^{ij}\right)^2-2\mu_k^j\sum_{i=1}^{n_j}\lambda_k^{ij}+n_j\left(\mu_k^j\right)^2\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\mu_k^j\right)^2\left(d_k+\tau_k^jn_j\right)-2\mu_k^j\left(d_kb_k+\tau_k^j\sum_{i=1}^{n_j}\lambda_k^{ij}\right)\right]\right\}.$$

So the full conditional distribution for μ_k^j is

$$\mu_k^j|\cdot \sim N\left(\frac{d_kb_k+\tau_k^j\sum_{i=1}^{n_j}\lambda_k^{ij}}{d_k+n_j\tau_k^j},\frac{1}{d_k+n_j\tau_k^j}\right).$$

τ_k^j update

We have that

$$\pi\left(\tau_k^j|\mu_k^j, \lambda_k^{1j}, \ldots, \lambda_k^{n_j j}\right)$$

$$\propto \pi\left(\tau_k^j\right) \prod_{i=1}^{n_j} \pi\left(\lambda_k^{ij}|\mu_k^j, \tau_k^j\right)$$

$$\propto \left(\tau_k^j\right)^{g_k-1} \exp\{-\tau_k^j h_k\} \left(\tau_k^j\right)^{n_j/2} \exp\left\{-\frac{\tau_k^j}{2} \sum_{i=1}^{n_j} \left(\lambda_k^{ij} - \mu_k^j\right)^2\right\}$$

$$\propto \left(\tau_k^j\right)^{g_k+n_j/2-1} \exp\left\{-\tau_k^j\left[h_k + \frac{1}{2}\sum_{i=1}^{n_j}\left(\lambda_k^{ij} - \mu_k^j\right)^2\right]\right\}.$$

So the full conditional distribution for τ_k^j is

$$\tau_k^j|\cdot \sim \Gamma\left(g_k + n_j/2, h_k + \frac{1}{2}\sum_{i=1}^{n_j}\left(\lambda_k^{ij} - \mu_k^j\right)^2\right).$$

σ_e^j update

We have that

$$\pi\left(\sigma_e^j|\lambda^{1j}, \ldots, \lambda^{n_j j}, y^{1j}, \ldots, y^{n_j j}\right) \propto \pi(\sigma_e^j)\prod_{i=1}^{n_j} \pi(y^{ij}|\lambda^{ij}, \sigma_e^j).$$

We sample this intractable FCD using a Metropolis-Hastings step. Since σ_e^j must be strictly positive we work with $\phi_e^j = \log \sigma_e^j$ and propose $\tilde{\phi}_e^j$ using a normal random walk. Hence the acceptance probability is

$$\alpha\left(\tilde{\phi}_e^j|\phi_e^j\right) = \min\left\{1, \frac{\pi\left(\tilde{\phi}_e^j\right)\prod_{i=1}^{n_j}\pi\left(y^{ij}|\lambda^{ij}, \tilde{\phi}_e^j\right)}{\pi\left(\phi_e^j\right)\prod_{i=1}^{n_j}\pi\left(y^{ij}|\lambda^{ij}, \phi_e^j\right)}\right\}$$

where $\pi\left(\phi_e^j\right) = N\left(\phi_e^j; b_\phi, d_\phi^{-1}\right)$ is the prior density ascribed to ϕ_e^j.

6.6 Application to real data

The model was fit to the complete data set. Recall that this consists of around 4000 hourly temperature averages per mouse. After completing a pilot run to obtain appropriate tuning values (e.g. suitable starting values and innovation variances used in the MH updates for the bottom level parameters and observation error parameters) the Metropolis-Hastings

scheme was carried out with 100k iterations. The priors for the hyper parameters were

$$\mu_1 \sim N\,(1, 0.1)\,, \qquad \tau_1 \sim \Gamma\,(10, 0.1)\,,$$
$$\mu_2 \sim N\,(36, 0.1)\,, \qquad \tau_2 \sim \Gamma\,(10, 0.1)\,,$$
$$\mu_3 \sim N\,(1, 0.1)\,, \qquad \tau_3 \sim \Gamma\,(10, 0.1)\,,$$
$$\mu_4 \sim N\,(3, 0.1)\,, \quad \tau_4 \sim \Gamma\,(10, 0.1)\,,$$
$$\mu_5 \sim N\,(0.5, 0.1)\,, \qquad \tau_5 \sim \Gamma\,(10, 0.1)\,,$$
$$\mu_6 \sim N\,(0.5, 0.1)\,, \qquad \tau_6 \sim \Gamma\,(10, 0.1)\,,$$
$$\mu_7 \sim N\,(0, 0.1)\,, \qquad \tau_7 \sim \Gamma\,(10, 0.1)\,,$$
$$\log(\sigma_e) \sim N\,(0, 1)\,.$$

This specification incorporates relatively strong prior beliefs regarding the average reversion levels of Z_t and A_t. In particular, we expect average core temperature level to be around 36, and the prior for average amplitude level reflects our belief of a strong diurnal pattern between 33 and 39. We have relatively weak prior beliefs regarding the precision parameters. Our prior beliefs regarding average reversion rates and intrinsic noise reflects beliefs that the process should be strongly mean reverting and that the level of noise is small. Note also that the probability of assigning a negative value is small under these priors.

Running the Metropolis-within-Gibbs scheme with 100k iterations including 10k for burn-in (without thinning) gave output with reasonable mixing. Figures 6.3 – 6.5 give the marginal posterior distributions for each hyper parameter for both treatment groups. In Figure 6.3, the top centre plot shows the marginal posterior distributions of average (across experimental units) of the overall temperature level (μ_2) for the two treatment groups. We see that the AL group exhibits a distinctly higher average temperature than the CR group and the tails of these distributions do not considerably overlap. This result is consistent with other works showing that mice compensate for the lack of calories with a reduced core body temperature (Weindruch & Walford, 1988; Duffy *et al.*, 1989; Roth *et al.*, 2002). The top left plot in Figure 6.3 shows the posterior distribution for the mean reversion rate for the temperature, and the top right shows the posterior distribution for the mean reversion rate for the amplitude, although the two groups of mice seem to revert to the amplitude at the same rate, their rates of reversion to the temperature appears a lot slower for those on the caloric restricted diet than for those fed ad libitum. We can see the posterior distribution for the mean of the intrinsic stochasticity parameters for the temperature process (μ_5) in Figure 6.4 (top centre plot). It appears to be small and similar for both groups. Also in Figure 6.4, the top left plot shows the posterior distribution of the mean of the overall amplitude level (μ_4) for the two treatment groups. We see that
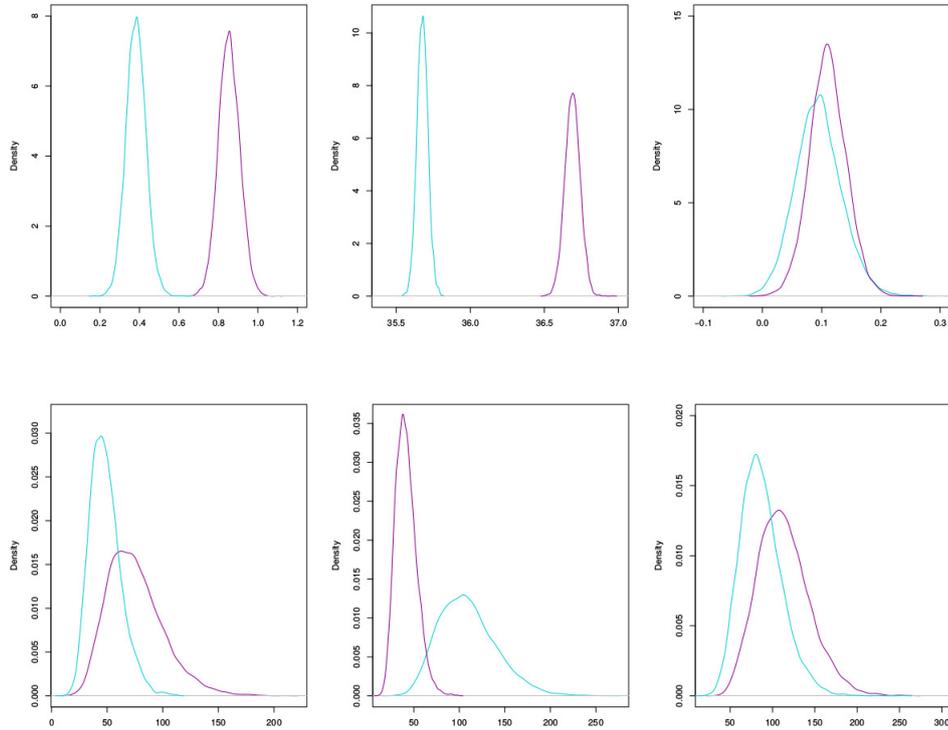
Figure 6.3: Marginal posterior distributions with top row showing the means μ_1, μ_2 and μ_3 from left to right and and the bottom row showing the precisions τ_1, τ_2 and τ_3 from left to right. Pink line shows AL mice, and blue line shows CR mice.

the AL group has a considerably higher amplitude than the CR group with the posteriors not overlapping significantly. This suggests that the diurnal variation is plausibly lower in the CR treatment group, as found previously (Golightly *et al.*, 2012). The intrinsic stochasticity for the amplitude is shown in the top right plot of Figure 6.4 and we can see that the CR treatment group has a smaller underlying random variation. Whilst we haven't overlaid the prior distributions it is clear that the analysis has been informative. We conclude that our findings support the hypothesis that mice have a lower temperature with a reduced calorie diet and that their diurnal variation is also smaller. In Figure 6.5, the left hand plot shows μ_7, the mean of B, the phase shift. The posterior variance of this parameter is much smaller for the AL group compared to CR, perhaps suggesting that CR mice may enter different phases of the sinusoid during the diurnal period.
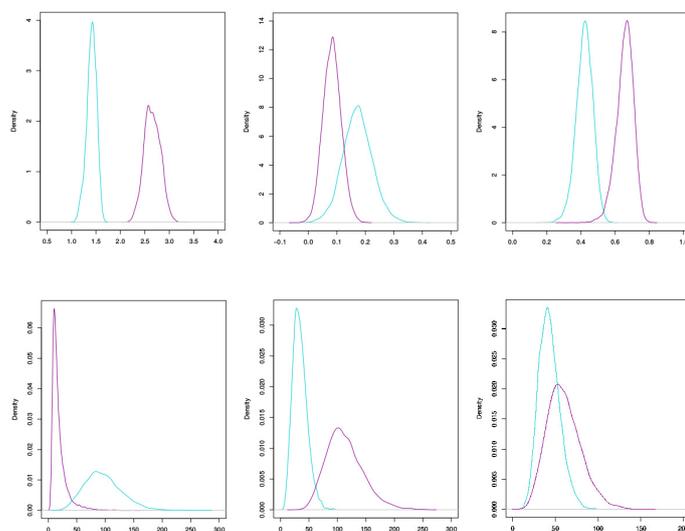
Figure 6.4: Marginal posterior distributions with top row showing the means μ_4, μ_5 and μ_6 from left to right and the bottom row showing the precisions τ_4, τ_5 and τ_6. Pink line shows AL mice, and blue line shows CR mice.
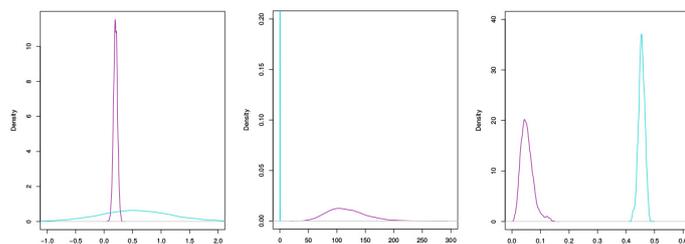
Figure 6.5: Marginal posterior distributions for mean μ_7, precision τ_7 and mean of σ_e. Pink line shows AL mice, and blue line shows CR mice.

6.7 Assessing model fit

Model fit can be assessed by drawing from the posterior predictive density (for mouse i in treatment j)

$$\pi(\tilde{y}^{ij}|y) = \int \pi(\tilde{y}^{ij}|\lambda^{ij})\pi(\lambda^{ij}|y)d\lambda^{ij}.$$

Although this predictive density will typically be intractable, samples can be obtained by taking a thinned set of posterior parameter samples $\{\lambda^{ij}(1), \ldots, \lambda^{ij}(N)\}$ and executing the following steps for each $m = 1, \ldots, N$:

1. Run the forward filter, backward sampler (Algorithms 9 and 10) conditional on $\lambda^{ij}(m)$ to obtain $x^{ij}(m)$

2. Set $\tilde{y}_t^{(ij)}(m) = x_t^{ij}(m) + \epsilon_t^{ij}, \qquad \epsilon_t^{ij} \sim N(0, [\sigma_\epsilon^2]^j(m)), \quad t = t_0, \dots, t_{n_j}.$

We assess model fit by comparing the data at each time point with their corresponding (marginal) predictive distributions. Figure 6.6 shows the marginal predictive mean and 95% credible intervals for two randomly chosen mice fed ad libitum, and two randomly chosen mice on the caloric restricted diet. We used 96 data points from the middle of the study and overlayed for comparison, with the disruptive periods removed. The model appears to fit the data reasonably well, with nearly all observations falling within the 95% credible intervals in each case. Similar results across the entire data set suggest a reasonable fit overall.
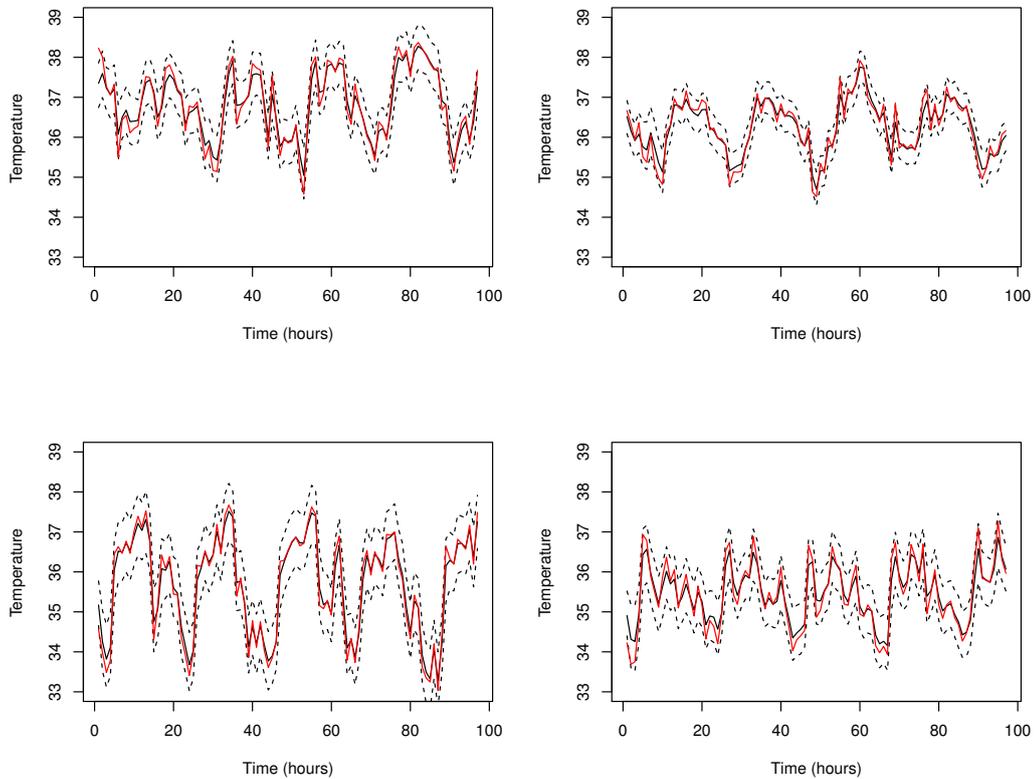
Figure 6.6: Predictive plots using the forward filter and backwards sampler. Top two plots show results for two mice fed ad libitum. Bottom two plots show results for mice on the caloric restricted diet. The black line shows the mean of predictive samples, the black dotted line shows the 95% credible interval, and the red line shows the real data.

Chapter 7

Application II

This application is based on the work of Picchini & Forman (2019), that used a stochastic differential mixed effects model to describe the tumour volume dynamics in mice receiving a treatment for tumours. The original study involved four treatment groups and one control group, and each group comprised seven or eight mice. Measurements of the tumours were taken every Monday, Wednesday and Friday for six weeks; however the majority of the mice were euthanized before the end of the study, once their tumour volumes exceeded 1000 cubic mm. We use the model derived in this study to generate a synthetic data set, in order to facilitate comparison between CPMMH accuracy and efficiency with PMMH. Although a transformation of the tumour growth process satisfies a linear SDE, a nonlinear observation model precludes tractability of the observed data likelihood, necessitating the use of (C)PMMH. We additionally derive a linear noise approximation (LNA) of the SDEMEM, and compare inferences made under the LNA with the output of (C)PMMH.

7.1 Tumor growth SDEMEM

Following Picchini & Forman (2019), we consider a stochastic differential mixed effects model with

$$dX_{1,t}^i = \left(\beta^i + (\gamma^i)^2/2\right) X_{1,t}^i dt + \gamma^i X_{1,t}^i dW_{1,t}^i$$
$$dX_{2,t}^i = \left(-\delta^i + (\psi^i)^2/2\right) X_{2,t}^i dt + \psi^i X_{2,t}^i dW_{2,t}^i \tag{7.1}$$

for experimental units $i = 1, \ldots, M$. Here, $W_{1,t}$ and $W_{2,t}$ are uncorrelated Brownian motion processes, $X_{1,t}^i$ and $X_{2,t}^i$ are respectively the volume of surviving tumor cells and volume of cells killed by a treatment for mouse i. We have that γ^2 controls the within-subject growth rate variance which therefore means that the instantaneous growth rate is not exactly β^i but deviates from this by a random normal perturbation. Similarly, ψ^2 controls the within-subject kill rate variance and so the instantaneous kill rate is not

exactly δ^i. We adopt the parameterisation in (7.1) (with $(\gamma^i)^2/2$ and $(\psi^i)^2/2$ included in the drift of the SDE) so that the individual growth/death processes are given by

$$X_{1,t}^i = x_{1,0}^i \exp(\beta_t^i + \gamma W_{1,t}^i)$$
$$X_{2,t}^i = x_{2,0}^i \exp(-\delta_t^i + \psi W_{2,t}^i) \tag{7.2}$$

which are log-normally distributed stochastic processes. Note that in the absence of intrinsic stochasticity (with $\gamma = \psi = 0$) we obtain $x_{1,0}^i \exp(\beta^i t)$ and $x_{2,0}^i \exp(-\delta^i t)$, which coincides with the ODE mixed effects model described later in Section 7.4.

Let $V_t^i = X_{1,t}^i + X_{2,t}^i$ denote the total tumor volume at time t in mouse i. The observation model is given by

$$Y_t^i = \log V_t^i + \epsilon_t^i, \qquad \epsilon_t^i \overset{indep}{\sim} N(0, \sigma_e^2). \tag{7.3}$$

That is, the logarithm of total tumour volume is observed subject to additive Gaussian noise. Let $\phi^i = (\log \beta^i, \log \gamma^i, \log \delta^i, \log \psi^i)$. We complete the SDEMEM specification via the assumption that

$$\phi_j^i | \eta \overset{indep}{\sim} N(\mu_j, \tau_j^{-1}), \qquad j = 1, \ldots, 4 \tag{7.4}$$

so that $\eta = (\mu_1, \ldots, \mu_4, \tau_1, \ldots, \tau_4)$.

We recognise that $X_{1,t}^i$ and $X_{2,t}^i$ are geometric Brownian motion processes and (7.1) can be solved analytically (by applying the Itô formula to the logarithm of both components) to give (7.2). Hence,

$$X_{1,t}^i | X_{1,0} = x_{1,0} \sim LN\left(\log(x_{1,0}) + \beta^i t, (\gamma^i)^2 t\right)$$
$$X_{2,t}^i | X_{2,0} = x_{2,0} \sim LN\left(\log(x_{2,0}) - \delta^i t, (\psi^i)^2 t\right) \tag{7.5}$$

where $LN(\cdot, \cdot)$ denotes the lognormal distribution. We emphasise that despite the availability of a closed form solution to the underlying SDE model, the observed data likelihood is intractable, due to the nonlinear form of (7.3).

We mimicked the real data application in Picchini & Forman (2019) by generating 21 observations at integer times for $M = 10$. We took

$$\eta = (\log 0.29, \log 0.25, \log 0.09, \log 0.34, 10, 10, 10, 10)$$

and sampled $\phi_j^i | \eta$ using (7.4). The latent SDE process was then generated using (7.5) with an initial condition of $x_0 = (75, 75)^T$ (assumed known), and each observation was corrupted according to (7.3) with $\sigma_e^2 = 0.2$. The resulting data traces are consistent with the observations on total tumor volume of those subjects receiving chemotherapy in Picchini & Forman (2019) and can be seen in Figure 7.1. Note that for simplicity we

only consider this treatment group and therefore do not include a treatment effect. We have also assumed data traces of equal length and on a regular time grid. We analyse this
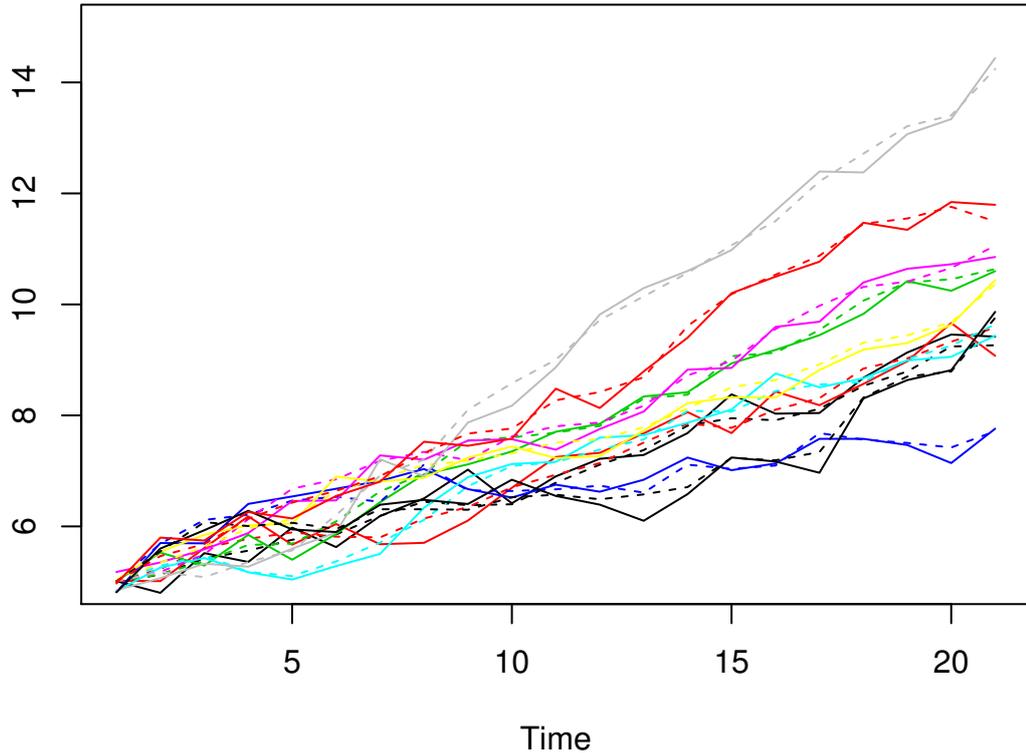
Figure 7.1: Simulated data for 10 units of tumour volume growth in mice. Dotted lines denote the tumour volume without observation error, solid lines have had observation error added.

synthetic data set in Section 7.3. In what follows, we note that a tractable approximation of the SDEMEM can be found by seeking a (linear) Gaussian approximation of $\log V_t^i$. The resulting linear noise approximation (LNA) is derived next in Section 7.2. We then compare inference under the gold standard SDEMEM to that obtained under the LNA.

7.2   LNA

Consider the tumor growth model in (7.1), (7.3) and (7.4) and a single experimental unit so that the superscript i can be dropped from the notation. To obtain a tractable observed data likelihood, we construct the linear noise approximation of $\log V_t = \log(X_{1,t} + X_{2,t})$.

Let $Z_t = (Z_{1,t}, Z_{2,t}, Z_{3,t})^T = (\log V_t, \log X_{1,t}, \log X_{2,t})^T$. The SDE satisfied by Z_t can

be found using the Itô formula, for which we obtain

$$dZ_t = \alpha(Z_t, \phi)dt + \sqrt{\beta(Z_t, \phi)}dW_t$$

where

$$\alpha(Z_t, \phi) =$$
$$\begin{pmatrix} \left\{\beta + 0.5\gamma^2\right\} e^{Z_{2,t}-Z_{1,t}} + \left\{-\delta + 0.5\tau^2\right\} e^{Z_{3,t}-Z_{1,t}} - 0.5 \left\{\gamma^2 e^{2(Z_{2,t}-Z_{1,t})} + \psi^2 e^{2(Z_{3,t}-Z_{1,t})}\right\} \\ \beta \\ -\delta \end{pmatrix}$$

$$\beta(Z_t, \phi) = \begin{pmatrix} \gamma^2 e^{2(Z_{2,t}-Z_{1,t})} + \tau^2 e^{2(Z_{3,t}-X_{1,t})} & \gamma^2 e^{2(Z_{2,t}-Z_{1,t})} & \psi^2 e^{2(Z_{3,t}-Z_{1,t})} \\ \gamma^2 e^{2(Z_{2,t}-Z_{1,t})} & \gamma^2 & 0 \\ \psi^2 e^{2(Z_{3,t}-Z_{1,t})} & 0 & \psi^2 \end{pmatrix}.$$

We apply the linear noise approximation (LNA) by partitioning Z_t as $Z_t = m_t + R_t$ where m_t is a deterministic process satisfying

$$\frac{dm_t}{dt} = \alpha(m_t, \phi) \tag{7.6}$$

and $\{R_t, t \geq 0\}$ is a residual stochastic process satisfying

$$dR_t = \left\{\alpha(Z_t, \phi) - \alpha(m_t, \phi)\right\} dt + \sqrt{\beta(Z_t, \phi)}dW_t.$$

By Taylor expanding α and β about the deterministic process m_t and retaining the first two terms in the expansion of α, and the first term in the expansion of β, we obtain an approximate residual stochastic process $\{\tilde{R}_t, t \geq 0\}$ satisfying

$$d\tilde{R}_t = J_t\tilde{R}_t dt + \sqrt{\beta(m_t, \phi)}dW_t$$

where J_t is the Jacobian matrix with (i, j)th element $(J_t)_{i,j} = \partial\alpha_i(m_t, \phi)/\partial m_{j,t}$. Assuming initial values $m_0 = z_0$ and $\tilde{R}_0 = 0$, the approximating distribution of Z_t is given by

$$Z_t|Z_0 = z_0 \approx \mathrm{N}(m_t, H_t) \tag{7.7}$$

where m_t satisfies (7.6) and, after several calculations which we omit for brevity, H_t is the solution to

$$\frac{dH_t}{dt} = H_t J_t^T + \beta(m_t, \phi) + J_t H_t. \tag{7.8}$$

7.2.1 Inference using the LNA

Note that the observation model in (7.3) can be written as

$$Y_t = P^T Z_t + \epsilon_t, \qquad \epsilon_t \overset{indep}{\sim} \mathrm{N}(0, \sigma_e^2). \tag{7.9}$$

where P is a 3×1 'observation vector' with first entry 1 and zeros elsewhere. The linearity of (7.7) and (7.9) yields a tractable approximation to the marginal likelihood $\pi(y|\phi, \sigma_e)$, which we denote by $\pi_{\mathrm{LNA}}(y|\phi, \sigma_e)$. The approximate marginal likelihood $\pi_{\mathrm{LNA}}(y|\phi, \sigma_e)$ can be factorised as

$$\pi_{\mathrm{LNA}}(y|\phi, \sigma_e) = \pi_{\mathrm{LNA}}(y_1|\phi, \sigma_e) \prod_{i=2}^{n} \pi_{\mathrm{LNA}}(y_i|y_{1:i-1}, \phi, \sigma_e) \tag{7.10}$$

where $y_{1:i-1} = (y_1, \ldots, y_{i-1})^T$. Suppose that $Z_1 \sim \mathrm{N}(a, C)$ *a priori*, for some constants a and C. The marginal likelihood under the LNA, $\pi_{\mathrm{LNA}}(y_{1:n}|\phi, \sigma_e) := \pi_{\mathrm{LNA}}(y|\phi, \sigma_e)$ can be obtained via a forward filter, which is given in Algorithm 1. Inference for the SDEMEM defined by (7.1), (7.3) and (7.4) may be performed via a Gibbs sampler that draws from the following full conditionals

1. $\pi_{\mathrm{LNA}}(\phi|\eta, \sigma_e, y) \propto \prod_{i=1}^{M} \pi(\phi^i|\eta) \pi_{\mathrm{LNA}}(y^i|\sigma_e, \phi^i)$,

2. $\pi_{\mathrm{LNA}}(\sigma_e|\eta, \phi, y) \propto \pi(\sigma_e) \prod_{i=1}^{M} \pi_{\mathrm{LNA}}(y^i|\sigma_e, \phi^i)$,

3. $\pi(\eta|\sigma_e, \phi, y) \propto \pi(\eta) \prod_{i=1}^{M} \pi(\phi^i|\eta)$.

The results for this scheme can be seen in the following section, in Table 7.1 and Figure 7.2 where we will also compare with the SDEMEM model using the various PMMH schemes.

7.3 SDEMEM vs LNA

We adopted semi conjugate, independent $N(-2, 1)$ and $Ga(2, 0.2)$ priors for the μ_j and τ_j respectively. We took $\log \sigma_e \sim N(0, 1)$ to complete the prior specification. Given the use of synthetic data of equal length for each experimental unit, we pragmatically took the number of particles as $N_i = N$, $i = 1, \ldots, 10$. Our choice of N was guided by the tuning advice of Section 5.4.4. We compare four approaches: naive PMMH (where the u^i are updated with both the subject specific and common parameters), PMMH (where the u^i are only updated with the subject specific parameters – Algorithm 8), CPMMH (Algorithm 8 with a Crank-Nicolson proposal on the u^i) and the LNA-based approach. We ran each scheme for $500k$ iterations.

The results are summarised in Table 7.1 and Figure 7.2.

Algorithm	ρ	N	CPU (m)	mESS	mESS/m	Rel.
LNA	-	-	1286	3676	2.858	16
PMMH - naive	0	40	8844	1598	0.181	1
PMMH	0	30	3842	2559	0.666	4
CPMMH	0.999	10	957	2311	2.415	13

Table 7.1: Tumour model. Correlation ρ, number of particles N, CPU time (in minutes m), minimum ESS, minimum ESS per minute and relative minimum ESS per minute. All results are based on 500k iterations of each scheme.

Figure 7.2 shows marginal posterior densities of the components of η, based on the ground truth SDEMEM and LNA. We see that inferences for these parameters are consistent with the true values that generated the data (with similar results obtained for the other parameters). Figure 7.3 shows the trace plots for μ_1 and τ_1, which we can see show that the CPMMH scheme converges well. Convergence is similar for all other parameters. We also note that the LNA based approach provides an accurate alternative to the SDEMEM.

Table 7.1 indicates that CPMMH with $\rho = 0.999$ yields results that are 13 times more efficient that naive PMMH and three times more so than standard PMMH. Although we can see that LNA is slightly more efficient than CPMMH there is a significant saving in CPU time making this method a reasonable competitor to using the LNA with the Gibbs sampler.

7.4 Comparison with ODEMEM

To highlight the potential issues that arise by ignoring inherent stochasticity, we consider inference for an ordinary differential equation mixed effects model (ODEMEM) of tumour growth. We take the SDEMEM in (7.1) and set $\gamma^i = \psi^i = 0$ to give

$$
\begin{aligned}
dx_{1,t}^i &= \beta^i x_{1,t}^i dt, \\
dx_{2,t}^i &= -\delta^i x_{2,t}^i dt
\end{aligned}
\tag{7.11}
$$

for $i = 1, \ldots, M$. The observation model and random effects distributions remain unchanged from (7.3) and (7.4) upon omitting $\log \gamma^i$ and $\log \psi^i$ from ϕ^i. The ODE system in (7.11) can be solved to give

$$
x_{1,t}^i = x_{1,0}^i \exp\{\beta^i t\}, \qquad x_{2,t}^i = x_{2,0}^i \exp\{-\delta^i t\}.
$$

The likelihood associated with each experimental unit is then obtained simply as

$$\pi(y^i|\phi^i, \sigma_e) = \prod_{t=1}^{21} \mathrm{N}\left(y_t^i; \log(x_{1,t}^i + x_{2,t}^i), \sigma_e^2\right).$$

Fitting the ODEMEM to the synthetic data set from Section 7.1 is straightforward, via a Metropolis-within-Gibbs scheme. Figures 7.4 and 7.5 summarise our findings.

Unsurprisingly, since the ODEMEM is unable to account for intrinsic stochasticity, the observation standard deviation is massively over-estimated. Figure 7.4 shows little agreement between the marginal posteriors under the ODEMEM and SDEMEM for this parameter. In terms of model fit, both the observation (Y_t^1) and latent process $(X_t^1 = \log V_t^1)$ predictive distributions for unit 1 are over-concentrated for the ODEMEM. Similar results (not shown) are obtained for the other experimental units. Notably, from Figure 7.5, around half of the actual simulated X_t values lie outside of the 95% credible interval under the ODEMEM.
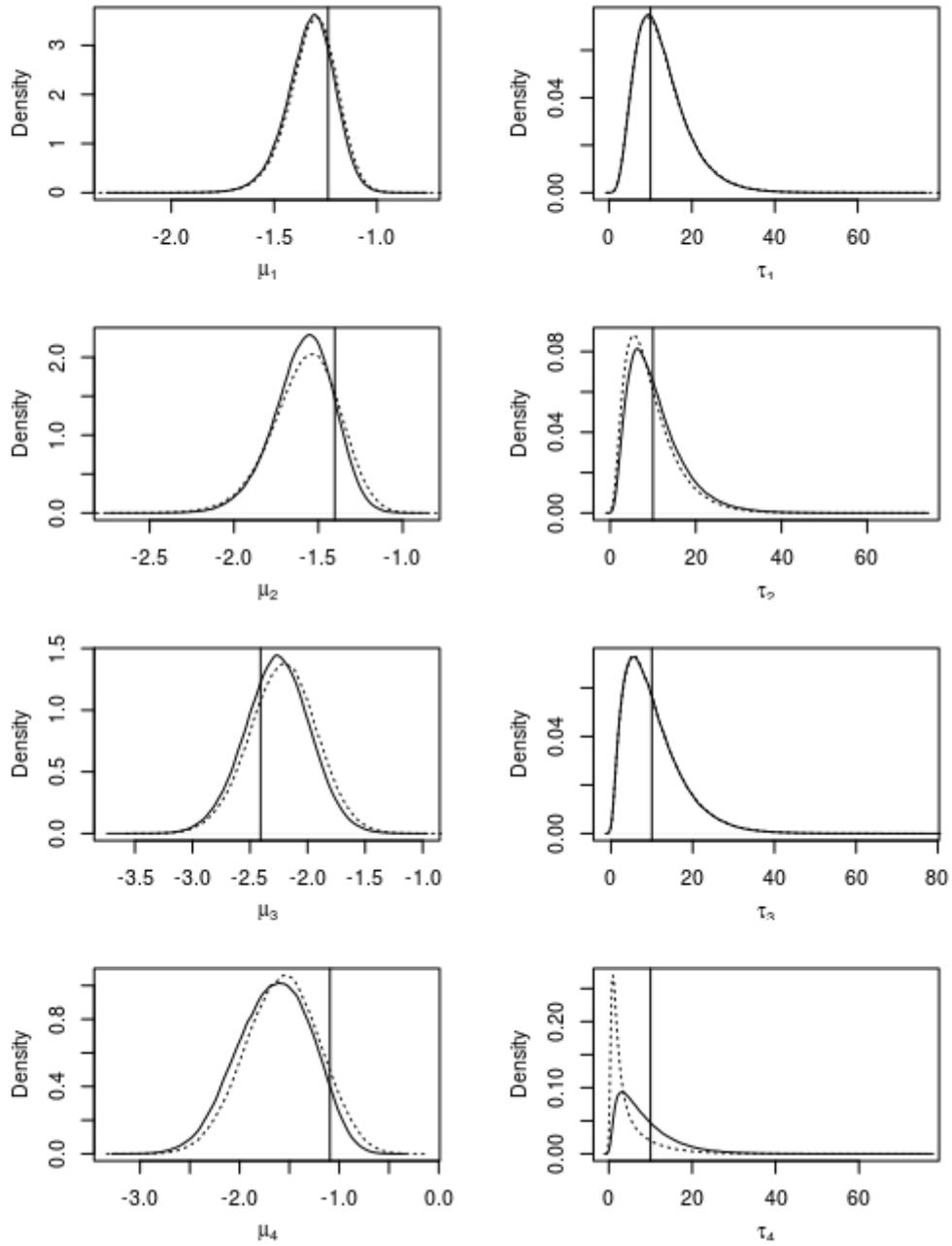
Figure 7.2: Marginal posterior distributions for μ_i and τ_i, $i = 1, \ldots, 4$. dotted line is from the LNA scheme, solid line is from the CPMMH scheme.
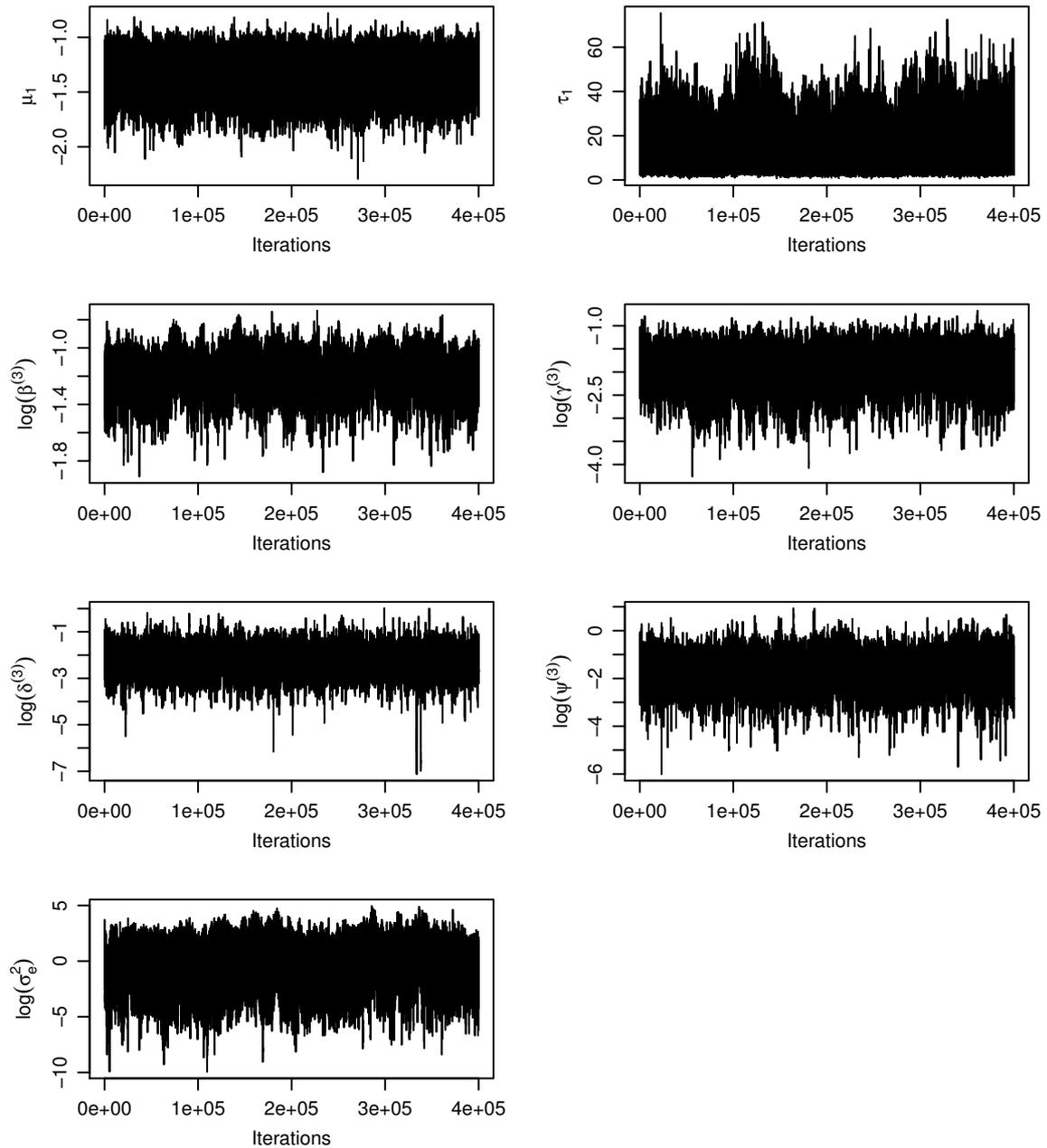
Figure 7.3: From top left to bottom right, trace plots showing the convergence of chains for μ_1, τ_1, $\log(\beta^3)$, $\log(\gamma^3)$, $\log(\delta^3)$, $\log(\psi^3)$ and at the bottom for $\log(\sigma_e^2)$.
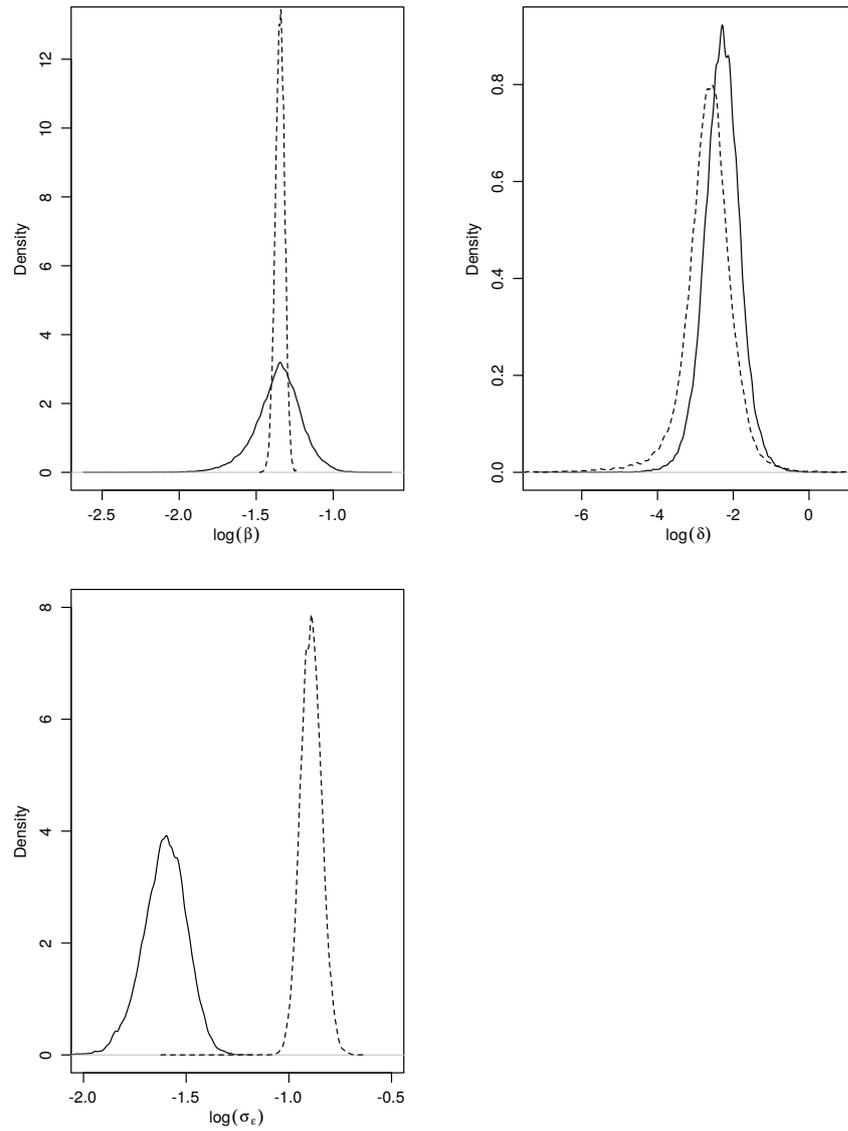
Figure 7.4: Marginal posterior distributions for the (logged) subject specific parameters $\log \beta^1$, $\log \delta^1$, and the observation standard deviation $\log \sigma_e$. Dashed line shows results from ODEMEM, solid line is from SDEMEM.
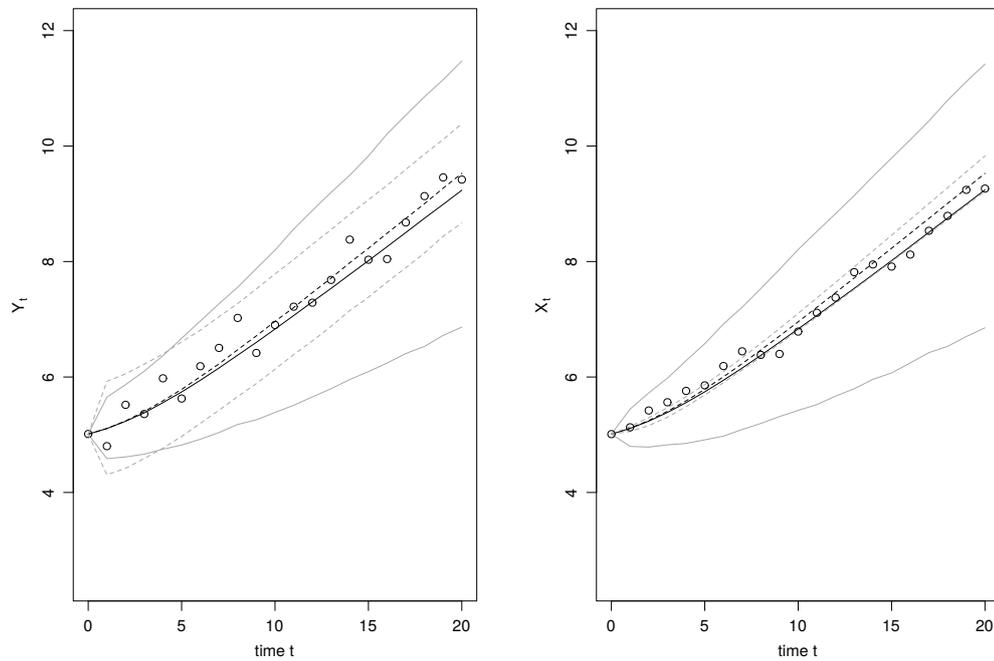
Figure 7.5: Posterior predictive mean (black) and 95% credible intervals (grey) for the observed process Y_t^1 (circles, left panel) and the latent process $X_t^1 = \log V_t^1$ (circles, right panel). Dashed line shows results from ODEMEM, solid line is from SDEMEM.

Chapter 8

Discussion and further work

The aim of this thesis was to efficiently infer model parameters of biological processes that evolve over time. These processes are often complicated by inherent stochasticity and the observations are usually subject to measurement error. When observations on multiple experimental units/subjects are available, it is important that the modelling approach additionally accounts for between subject variation. We sought to create Bayesian inference schemes that could capture these variations whilst inferring the model parameters. We focused particularly on SDEMEMs whose underlying dynamics were driven by linear SDEs. While SDEMEMs are a flexible class of model that capture both between- and within-subject variation, their widespread use has been limited by technical difficulties that make the execution of inference algorithms (both classic and Bayesian) computationally intensive. In the SDEMEMs we considered, the random effect parameters could have any distribution (not restricted to the Gaussian family) and the observation model does not have to be a linear combination of the latent states. Our contribution, therefore, is an inferential framework that applies to a large class of SDEMEMs, albeit under the assumption that the underlying SDE is linear in the state process.

We considered both instances of the observation process being linear and non-linear. These scenarios typically lead to the observed data likelihood being tractable and intractable respectively. For example, the former may arise if the SDE admits a Gaussian solution and the observation model is linear and Gaussian. In the scenario where the observed data likelihood was tractable it was calculated analytically using a forward filter. It is worth emphasising that even in this case, the marginal (in the sense of integrating out the latent dynamic process for each experimental unit) posterior distribution of all quantities of interest remains intractable. We therefore used a Metropolis-within-Gibbs scheme to generate draws from the posterior. We proposed alternating between draws of blocks consisting of parameters governing each experimental unit, population level parameters, and parameters governing the observation process.

For a non-linear (and/or non Gaussian) observation process, the observed data likelihood became intractable. Nevertheless, a particle filter (targeting the distribution of the latent process given the observations) can be used to unbiasedly estimate the observed data likelihood. This technique admits a class of Metropolis-Hastings (MH) scheme known as pseudo-marginal MH (PMMH). In essence, a carefully constructed MH scheme is used to target a joint density over an extended state space (that includes the variables used to generate the likelihood estimates) that admits the posterior density of interest as a marginal. We proposed a pseudo-marginal-within-Gibbs scheme applicable to SDEMEMs, and proposed to block together updates of the population level parameters and the innovations to improve mixing. It is worth noting again here that the efficiency of pseudo-marginal schemes will be sensitive to the number of particles N used in the particle filter. A likelihood estimator with large variance is likely to result in parameter chains that exhibit sticky behaviour. This problem can be alleviated by increasing N, but also at an increased computational cost. We therefore further adapted a recently proposed correlated pseudo-marginal MH method to the case of SDEMEMs. The basic idea is to induce strong and positive correlation between successive likelihood estimates, thereby reducing the variance of the MH acceptance ratio. This correlation is induced by proposing new innovations using a Crank-Nicolson kernel, which in turn requires specification of a correlation parameter ρ by the practitioner. We investigated the performance of the proposed methodology in the second of two applications, and additionally compared against an approximate approach, that constructs a linear noise approximation (LNA) of the SDEMEM.

In our first data application we used an SDEMEM to model core temperature over multiple experimental units (mice) that were involved in a feeding regime experiment. This allowed for the incorporation of intrinsic stochasticity inherent in observed temperature traces. Our hierarchical model allowed for time varying amplitude, was governed by underlying linear SDEs and had a linear observation model. We fitted the model to 4000 observations on 20 experimental units. We used the Metropolis-within-Gibbs algorithm described above to infer the model parameters. We found that the mice have a lower temperature when subjected to a reduced calorie diet and that their diurnal variation is also smaller. This is consistent with previous studies (e.g Weindruch & Walford, 1988; Duffy *et al.*, 1989; Roth *et al.*, 2002; Golightly *et al.*, 2012).

In our second data application we used a stochastic differential mixed effects model to describe the tumour volume dynamics in mice receiving treatment for tumours. We implemented the pseudo-marginal-Metropolis-within-Gibbs algorithm described above, and compared against a PMMH approach, and an MH scheme targeting the posterior under the linear noise approximation, where the latter was constructed by an appropriate linearisation of the observation model. We found that a correlated PMMH (CPMMH) approach permits fewer particles compared to PMMH, while still obtaining a comparable

effective sample size (ESS). Consequently, overall efficiency, as measured by ESS per second is much reduced. We found that an increase in overall efficiency of up to a factor of 25 was possible. The LNA gave improved performance, as expected (since in this case the observed data likelihood can be calculated without recourse to sampling), although there was relatively little difference between the CPMMH and LNA approaches. It is worth noting that some care must be taken when choosing ρ, which governs the level of correlation between successive likelihood estimates. Taking $\rho \approx 1$ can result in the sampler failing to adequately mix over the auxiliary variables. We found that this problem was exacerbated when using relatively few particles (such as $N = 1$), but can be overcome by reducing ρ. It was also necessary to set the number of particles N and, when correlated particle filters are used, the correlation parameter ρ (however this one is easily set within the interval $[0.90, 0.999]$). Finally, the usual settings for the MCMC proposal distribution should be decided (covariance matrix of the proposal density, $q(\cdot|\cdot)$). Nevertheless, it is clear that the CPMMH approach requires minimal tuning, over and above PMMH.

There are a number of ways in which our approach could be extended and improved upon. The proposed methodology relies on the use of the bootstrap particle filter, within which particles are propagated according to the SDE solution, myopically of the next observation. In scenarios where the density of process conditional on the next observation is available, or can be accurately approximated, its use inside the particle filter (for propagating particles) is likely to lead to gains in overall efficiency. Examples of this approach for generic SDEs can be found in Golightly & Wilkinson (2011). We have focused on SDEMEMs driven by a linear class of SDE. The extension of our inference approach to nonlinear SDEMEMs also remains of interest. However, we note that several difficulties become apparent. For example, nonlinear SDEs rarely admit analytic solutions necessitating a tractable approximation (such as the LNA) or the use of numerical approximations. If the latter approach is adopted, it is usually the case that intermediate times are added (between observation instants) to make the numerical approximation achieve a desired level of accuracy. Integrating over the uncertainty at these intermediate times is particularly challenging, and typically requires the use of a proposal construct that takes into account the observations, between which the latent process is required. This scenario is considered in Botha *et al.* (2019); see for example their component-wise pseudo-marginal (CWPM) method, which is similar to the naive Gibbs strategy we also propose. In order to correlate the particles, Botha *et al.* (2019) advocate the use of the blockwise pseudo-marginal strategy of Tran *et al.* (2017): this way, at each iteration of a CPMMH algorithm they randomly pick a unit in the set $\{1, .., M\}$, and only for that unit they update the corresponding auxiliary variates, whereas for the remaining $M - 1$ units they reuse the same auxiliary variates u_i as employed in the last accepted likelihood approximation. This approach implies an estimated correlation between loglikelihoods of around $1 - 1/M$, which

also implies that the correlation level is completely guided by the number of units. This means that for a small M (e.g. $M = 5$ or 10, implying a correlation of 0.80 and 0.90 respectively) a blockwise pseudo-marginal strategy might not be as effective as it could be. Nevertheless a comparison of our approach with that described in Botha *et al.* (2019) remains an interesting avenue for future research.

Appendix A

A.1 Solving a linear SDE model of temperature dynamics

A.1.1 SDE model

Recall that

$$X_t = (Z_t, A_t)^T$$

where

$$dZ_t = \theta_1(\theta_2 - Z_t)dt + \frac{\pi}{12}A_t \cos\left(\frac{\pi t}{12} + B\right)dt + \sigma dW_t$$

$$dA_t = \theta_3\left(\theta_4 - A_t\right)dt + \sigma_A dW_t^A$$

Therefore our SDE is

$$dX_t = \left[\begin{pmatrix} -\theta_1 & \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right) \\ 0 & -\theta_3 \end{pmatrix}\begin{pmatrix} Z_t \\ A_t \end{pmatrix} + \begin{pmatrix} \theta_1\theta_2 \\ \theta_3\theta_4 \end{pmatrix}\right]dt + \begin{pmatrix} \sigma & 0 \\ 0 & \sigma_A \end{pmatrix}d\begin{pmatrix} W_t \\ W_t^A \end{pmatrix}$$

$$\Rightarrow dX_t = [\alpha_1(t)X_t + \alpha_2(t)]\,dt + \sqrt{\beta(t)}dW_t$$

and the solution will be of the form

$$X_t | X_0 \sim N\left(m(t), V(t)\right).$$

Note that in what follows, for notational clarity, we denote deterministic functions of t as $m(t)$, $V(t)$ etc.

A.1.2 Solving first-order linear ODEs with integrating factors

Consider a first order linear differential equation of the form

$$\frac{dy(t)}{dt} = q(t) - p(t)y(t)$$

for known functions $q(\cdot)$ and $p(\cdot)$. It is straightforward to show that

$$y(t) = \frac{u(t_{i-1})y(t_{i-1})}{u(t)} + \frac{1}{u(t)} \int_{t_{i-1}}^{t} u(t)q(t)dt$$

where

$$u(t) = e^{\int p(t)dt}$$

is an integrating factor.

A.1.3 Mean

The mean of the process with varying amplitude satisfies

$$\frac{dm(t)}{dt} = \alpha_2(t) + \alpha_1(t)m(t)$$

$$= \begin{pmatrix} \theta_1\theta_2 \\ \\ \theta_3\theta_4 \end{pmatrix} + \begin{pmatrix} -\theta_1 & \frac{\pi}{12}\cos\left(\frac{2\pi t}{24} + B\right) \\ \\ 0 & -\theta_3 \end{pmatrix} \begin{pmatrix} m_1(t) \\ \\ m_2(t) \end{pmatrix}.$$

This yields a system of linear ODEs.

$$\begin{cases} \dfrac{dm_1(t)}{dt} = \theta_1\theta_2 - \theta_1 m_1(t) + \dfrac{\pi}{12}\cos\left(\dfrac{2\pi t}{24} + B\right)m_2(t) & (1) \\ \\ \dfrac{dm_2(t)}{dt} = \theta_3\theta_4 - \theta_3 m_2(t) & (2) \end{cases}$$

Equation (2) can be solved using the integrating factor with $p(t) = \theta_3$, and $q(t) = \theta_3\theta_4$, so that $u(t) = e^{\theta_3 t}$. Hence we obtain

$$
\begin{aligned}
m_2(t) &= \frac{e^{\theta_3(t_{i-1})}m_{2,0}}{e^{\theta_3 t}} + e^{-\theta_3 t}\int_{t_{i-1}}^{t}\theta_3\theta_4 e^{\theta_3 s}ds \\
&= m_2(t_{i-1})e^{-\theta_3(t-t_{i-1})} + e^{-\theta_3 t}\left[\theta_4 e^{\theta_3 s}\right]_{t_{i-1}}^{t} \\
&= m_2(t_{i-1})e^{-\theta_3(t-t_{i-1})} + e^{-\theta_3 t}\left[\theta_4 e^{\theta_3 t} - \theta_4 e^{\theta_3 t_{i-1}}\right] \\
&= m_2(t_{i-1})e^{-\theta_3(t-t_{i-1})} + \theta_4\left(1 - e^{-\theta_3(t-t_{i-1})}\right).
\end{aligned}
$$

Now, to solve equation (1), substitute in the form of $m_2(t)$ to give

$$
\begin{aligned}
\frac{dm_1(t)}{dt} &= \theta_1\theta_2 - \theta_1 m_1(t) + \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right)m_{2,t_{i-1}}e^{-\theta_3 t - t_{i-1}} \\
&\quad + \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right)\theta_4 - \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right)\theta_4 e^{-\theta_3(t-t_{i-1})}.
\end{aligned}
$$

The above can be solved using the integrating factor with

$$
\begin{aligned}
p(t) &= \theta_3 \\
q(t) &= \theta_1\theta_2 + \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right)m_{2,t_{i-1}}e^{-\theta_3 t - t_{i-1}} \\
&\quad + \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right)\theta_4 - \frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right)\theta_4 e^{-\theta_3(t-t_{i-1})} \\
u(t) &= e^{\theta_1 t}
\end{aligned}
$$

Hence we obtain

$$
m_1(t) = e^{-\theta_1(t-t_{i-1})}m_1(t_{i-1}) + e^{-\theta_1 t}\int_{t_{i-1}}^{t}u(s)q(s)ds
$$

Now

$$
\begin{aligned}
\int_{t_{i-1}}^{t}u(s)q(s)ds &= \theta_1\theta_2\int_{t_{i-1}}^{t}e^{\theta_1 s}ds \\
&\quad + \frac{\pi}{12}m_2(t_{i-1})e^{\theta_3(t_{i-1})}\int_{t_{i-1}}^{t}e^{s(\theta_1-\theta_3)}\cos\left(\frac{\pi s}{12} + B\right)ds \\
&\quad + \frac{\pi}{12}\theta_4\int_{t_{i-1}}^{t}e^{\theta_1 s}\cos\left(\frac{\pi s}{12} + B\right)ds \\
&\quad - \frac{\pi}{12}\theta_4 e^{\theta_4 t_{i-1}}\int_{t_{i-1}}^{t}e^{s(\theta_1-\theta_3)}\cos\left(\frac{\pi s}{12} + B\right)ds.
\end{aligned}
$$

Using $f_1(t, \nu)$ for $\int e^{s(\nu)} \cos\left(\frac{\pi s}{12} + B\right) ds$ which is derived in Appendix A.2.1 we get

$$
\int_{t_{i-1}}^{t} u(s)q(s)ds = \theta_2 \left[e^{\theta_1^t} - e^{\theta_1 t_{i-1}} \right]
$$
$$
+ \frac{\pi}{12} m_2(t_{i-1}) e^{\theta_3(t_{i-1})} \left[f_1(t, \nu_1) - f_1(t_{i-1}, \nu_1) \right]
$$
$$
+ \frac{\pi}{12} \theta_4 \left[f_1(t, \theta_1) - f_1(t_{i-1}, \theta_1) \right]
$$
$$
- \frac{\pi}{12} \theta_4 e^{\theta_4 t_{i-1}} \left[f_1(t, \nu_1) - f_1(t_{i-1}, \nu_1) \right]
$$

where $\nu_1 = (\theta_1 - \theta_3)$. Bringing this all together for $m_1(t)$ we have

$$
m_1(t) = e^{-\theta_1(t-t_{i-1})} m_1(t_{i-1})
$$
$$
+ e^{-\theta_1 t} \theta_2 \left[e^{\theta_1^t} - e^{\theta_1 t_{i-1}} \right]
$$
$$
+ e^{-\theta_1 t} \frac{\pi}{12} m_2(t_{i-1}) e^{\theta_3(t_{i-1})} \left[f_1(t, \nu_1) - f_1(t_{i-1}, \nu_1) \right]
$$
$$
+ e^{-\theta_1 t} \frac{\pi}{12} \theta_4 \left[f_1(t, \theta_1) - f_1(t_{i-1}, \theta_1) \right]
$$
$$
- e^{-\theta_1 t} \frac{\pi}{12} \theta_4 e^{\theta_4 t_{i-1}} \left[f_1(t, \nu_1) - f_1(t_{i-1}, \nu_1) \right].
$$

A.1.4 Variance

The variance of the process with varying amplitude satisfies

$$
\frac{dV(t)}{dt} = V(t)\alpha_1(t)^T + \beta(t)^2 + \alpha_1(t)V(t)
$$

$$
\Rightarrow \frac{d}{dt} \begin{pmatrix} V_{11}(t) & V_{12}(t) \\ V_{21}(t) & V_{22}(t) \end{pmatrix} = \begin{pmatrix} -\theta_1 V_{11}(t) + V_{12}(t)\frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right) & -\theta_3 V_{12}(t) \\ -\theta_1 V_{21}(t) + V_{22}(t)\frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right) & -\theta_3 V_{22}(t) \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma_A^2 \end{pmatrix}
$$

$$
+ \begin{pmatrix} -\theta_1 V_{11}(t) + V_{21}(t)\frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right) & -\theta_1 V_{12}(t) + V_{22}(t)\frac{\pi}{12}\cos\left(\frac{\pi t}{12} + B\right) \\ -\theta_3 V_{21}(t) & -\theta_3 V_{22}(t) \end{pmatrix}
$$

This yields a system of linear ODEs.

$$
\begin{cases}
\frac{dV_{11}(t)}{dt} = 2\left(V_{12}(t)\frac{\pi}{12}\cos\left(\frac{\pi t}{12}+B\right)-\theta_1 V_{11}(t)\right)+\sigma^2, & V_{11}(t_{i-1})=v_{11}(t_{i-1})\\[4mm]
\frac{dV_{12}(t)}{dt} = \frac{dV_{21}(t)}{dt} = -V_{12}(t)\left(\theta_1+\theta_3\right)+V_{22}(t)\frac{\pi}{12}\cos\left(\frac{\pi t}{12}+B\right), & V_{12}(t_{i-1})=v_{12}(t_{i-1})\\[4mm]
\frac{dV_{22}(t)}{dt} = -2\theta_3 V_{22}(t)+\sigma_A^2, & V_{22}(t_{i-1})=v_{22}(t_{i-1})
\end{cases}
$$
$$\tag{A.1}$$

The hierarchical nature of this system suggests starting with the ODE satisfied by $V_{22}(t)$.

Solution of $V_{22}(t)$

Using the integrating factor method as in A.1.2, let $p(t)=2\theta_3$, $q(t)=\sigma_A^2$, hence $u(t)=e^{2\theta_3 t}$. Then

$$
\begin{aligned}
V_{22}(t) &= V_{22}(0)e^{-2\theta_3 t_{i-1}}e^{-2\theta_3 t}+e^{-2\theta_3 t}\int_{t_{i-1}}^{t}\sigma_A^2 e^{2\theta_3 s}ds\\[2mm]
&= V_{22}(0)e^{-2\theta_3(t-t_{i-1})}+e^{-2\theta_3 t}\int_{t_{i-1}}^{t}\sigma_A^2 e^{2\theta_3 s}ds\\[2mm]
&= V_{22}(0)e^{-2\theta_3(t-t_{i-1})}+\left[\frac{\sigma_A^2}{2\theta_3}e^{2\theta_3 s}\right]_{t_{i-1}}^{t}\\[2mm]
&= V_{22}(0)e^{-2\theta_3(t-t_{i-1})}+\frac{\sigma_A^2}{2\theta_3}\left(1-e^{-2\theta_3(t-t_{i-1})}\right)
\end{aligned}
$$

Solution of $V_{12}(t)$

Substituting in the form of $V_{22}(t)$ we obtain

$$
\begin{aligned}
\frac{dV_{12}(t)}{dt} =& -V_{12}(t)\left(\theta_1+\theta_3\right)\\
&+\left[V_{22}(t_{i-1})e^{-2\theta_3(t-t_{i-1})}+\frac{\sigma_A^2}{2\theta_3}\left(1-e^{-2\theta_3(t-t_{i-1})}\right)\right]\frac{\pi}{12}\cos\left(\frac{\pi t}{12}+B\right)
\end{aligned}
$$

Again, using the integrating factor method, let

$$
\begin{aligned}
p(t) &= \theta_1+\theta_3\\
q(t) &= \frac{\pi}{12}\cos\left(\frac{\pi t}{12}+B\right)\left[V_{22}(t_{i-1})e^{-2\theta_3(t-t_{i-1})}+\frac{\sigma_A^2}{2\theta_3}\left(1-e^{-2\theta_3(t-t_{i-1})}\right)\right]\\
u(t) &= e^{t(\theta_1+\theta_3)}
\end{aligned}
$$

Then

$$V_{12}(t) = V_{12}(t_{i-1})e^{t_{i-1}(\theta_1+\theta_3)-t(\theta_1+\theta_3)} + e^{-t(\theta_1+\theta_3)} \int_{t_{i-1}}^{t} u(s)q(s)ds$$

Now

$$\int_{t_{i-1}}^{t} u(s)q(s)ds = \frac{\pi}{12}V_{22}(t_{i-1}) \int_{t_{i-1}}^{t} e^{s(\theta_1+\theta_3)}e^{-2\theta_3(s-s_{i-1})} \cos\left(\frac{\pi s}{12} + B\right) ds$$
$$+ \frac{\pi}{12}\frac{\sigma_A^2}{2\theta_3} \int_{t_{i-1}}^{t} e^{s(\theta_1+\theta_3)}\left(1 - e^{-2\theta_3(s-s_{i-1})}\right) \cos\left(\frac{\pi s}{12} + B\right) ds$$

$$\Rightarrow V_{12}(t) = V_{12}(t_{i-1})e^{(t_{i-1}-t)(\theta_1+\theta_3)}$$
$$+ \frac{\pi}{12}V_{22}(t_{i-1})e^{-t(\theta_1+\theta_3)}e^{2\theta_3(s_{i-1})} \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)} \cos\left(\frac{\pi s}{12} + B\right) ds}_{*}$$
$$+ \frac{\pi}{12}\frac{\sigma_A^2}{2\theta_3}e^{-t(\theta_1+\theta_3)} \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1+\theta_3)}\left(1 - e^{-2\theta_3(s-s_{i-1})}\right) \cos\left(\frac{\pi s}{12} + B\right) ds}_{**}$$

Looking at $*$ we have

$$* = \int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)} \cos\left(\frac{\pi s}{12} + B\right) ds$$
$$= f_1(t, \nu_1) - f_1(t_{i-1}, \nu_1)$$

Where $\nu_1 = \theta_1 - \theta_3$ and $f_1(t, \nu_1)$ is given in Appendix A.2.1.

Looking at $**$ we have

$$** = \int_{t_{i-1}}^{t} e^{s(\theta_1+\theta_3)}\left(1 - e^{-2\theta_3(s-s_{i-1})}\right) \cos\left(\frac{\pi s}{12} + B\right) ds$$
$$= \int_{t_{i-1}}^{t} e^{s(\theta_1+\theta_3)} \cos\left(\frac{\pi s}{12} + B\right) ds - e^{2\theta_3(t_{i-1})} \int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)} \cos\left(\frac{\pi s}{12} + B\right) ds$$
$$= f_1(t, \nu_2) - f_1(t_{i-1}, \nu_2) - e^{2\theta_3(t_{i-1})}\big(f_1(t, \nu_1) - f_1(t_{i-1}, \nu_1)\big)$$

Where $\nu_1 = \theta_1 - \theta_3$, $\nu_2 = \theta_1 + \theta_3$ and $f_1(t)$ is defined in Appendix A.2.1.

Pulling this all together we get an expression for $V_{12}(t)$

$$V_{12}(t) = V_{12}(t_{i-1})e^{-(t-t_{i-1})(\theta_1+\theta_3)}$$
$$+ \frac{\pi}{12}V_{22}(t_{i-1})e^{-t(\theta_1+\theta_3)}e^{2\theta_3 t_{i-1}}\big(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\big)$$
$$+ \frac{\pi\sigma_A^2}{24\theta_3}e^{-t(\theta_1+\theta_3)}\Big[f_1(t,\nu_2) - f_1(t_{i-1},\nu_2) - e^{2\theta_3 t_{i-1}}\big(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\big)\Big]$$

Solution to $V_{11}(t)$

Substituting in the form of $V_{12}(t)$, we get:

$$\frac{dV_{11}(t)}{dt} = 2\frac{\pi}{12}V_{12}(t_{i-1})e^{-(\theta_1+\theta_3)(t-t_{i-1})}\cos\left(\frac{\pi t}{12} + B\right)$$
$$+ 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{-t(\theta_1+\theta_3)}e^{2\theta_3 t_{i-1}}\big(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\big)\cos\left(\frac{\pi t}{12} + B\right)$$
$$+ \frac{\pi\sigma_A^2}{12\theta_3}\frac{\pi}{24}e^{-t(\theta_1+\theta_3)}\Big[f_1(t,\nu_2) - f_1(t_{i-1},\nu_2) - e^{2\theta_3 t_{i-1}}\big(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\big)\Big]\cos\left(\frac{\pi t}{12} + B\right)$$
$$- 2\theta_1 V_{11}(t) + \sigma^2.$$

Using the integrating factor method, let

$$p(t) = 2\theta_3$$
$$q(t) = 2\frac{\pi}{12}V_{12}(t_{i-1})e^{-(\theta_1+\theta_3)(t-t_{i-1})}\cos\left(\frac{\pi t}{12} + B\right)$$
$$+ 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{-t(\theta_1+\theta_3)}e^{2\theta_3 t_{i-1}}\big(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\big)\cos\left(\frac{\pi t}{12} + B\right)$$
$$+ \frac{\pi\sigma_A^2}{12\theta_3}\frac{\pi}{24}e^{-t(\theta_1+\theta_3)}\Big[f_1(t,\nu_2) - f_1(t_{i-1},\nu_2) - e^{2\theta_3 t_{i-1}}\big(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\big)\Big]\cos\left(\frac{\pi t}{12} + B\right)$$
$$+ \sigma^2.$$
$$u(t) = e^{2\theta_1 t}$$

Then

$$V_{11}(t) = V_{11}(t_{i-1})e^{-2\theta_1(t-t_{i-1})} + e^{-2\theta_1 t}\int_{t_{i-1}}^{t} u(s)q(s)ds$$

Now

$$\int_{t_{i-1}}^{t} u(s)q(s)ds = 2\frac{\pi}{12}V_{12}(t_{i-1})e^{(t_{i-1}(\theta_1-\theta_3))} \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1+\theta_3)}\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star}$$

$$+ 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{2\theta_3 t_{i-1}} \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)}f_1(s,\nu_1)\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star\star}$$

$$- 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{2\theta_3 t_{i-1}}f_1(t_{i-1},\nu_1) \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)}\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star}$$

$$+ \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right) \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)}f_1(s,\nu_2)\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star\star\star}$$

$$- \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)f_1(t_{i-1},\nu_2) \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)}\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star}$$

$$- \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)e^{2\theta_3 t_{i-1}} \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)}f_1(s,\nu_1)\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star\star}$$

$$+ \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)e^{2\theta_3 t_{i-1}}f_1(t_{i-1},\nu_1) \times \underbrace{\int_{t_{i-1}}^{t} e^{s(\theta_1-\theta_3)}\cos\left(\frac{\pi s}{12}+B\right)ds}_{\star}$$

$$+ \sigma^2 \int_{t_{i-1}}^{t} e^{2\theta_1 t}ds.$$

The solution to the integral in $\star$ is defined in A.2.1 to be $f_1(t,\nu)$ with $\nu = \theta_1 \pm \theta_3$, hence:

$$\star = f_1(t,\nu) - f_1(t_{i-1},\nu)$$

The solution to the integral in $\star\star$ can be found using integration by parts which states that:

$$\int u\,dv = uv - \int v\,du$$

We shall define $u = f_1(s,\nu_1)$ and $dv = e^{\nu_1 s}\cos\left(\frac{\pi t}{12}+B\right)$, hence:

$$du = e^{\nu_1 s}\cos\left(\frac{\pi t}{12}+B\right)$$

$$v = f_1(s,\nu_1)$$

so then we have that

$$\int_{t_{i-1}}^{t} f_1(s,\nu_1)e^{s\nu_1}\cos\left(\frac{\pi s}{12}+B\right)ds = f_1(s,\nu_1)f_1(s,\nu_1) - \int_{t_{i-1}}^{t} f_1(s,\nu_1)e^{\nu_1 s}\cos\left(\frac{\pi s}{12}+B\right)ds$$

$$\Rightarrow 2\int_{t_{i-1}}^{t} f_1(s,\nu_1)e^{s\nu_1}\cos\left(\frac{\pi s}{12}+B\right)ds = f_1(s,\nu_1)^2$$

$$\Rightarrow \int_{t_{i-1}}^{t} f_1(s,\nu_1)e^{s\nu_1}\cos\left(\frac{\pi s}{12}+B\right)ds = \frac{1}{2}\left(f_1(t,\nu_1)^2 - f_1(t_{i-1},\nu_1)^2\right)$$

then

$$\star\star = \frac{1}{2}\left(f_1(t,\nu_1)^2 - f_1(t_{i-1},\nu_1)^2\right)$$

Now, substituting the form for $f_1(t,\nu_2)$ into $\star\star\star$ we obtain

$$\star\star\star = \int_{t_{i-1}}^{t} e^{\nu_1 s} f_1(s,\nu_2)\cos\left(\frac{\pi s}{12}+B\right)ds$$

$$= \left(1+\left(\frac{12}{\pi}\right)^2\nu_2^2\right)^{-1}\frac{12}{\pi}\times$$

$$\int_{t_{i-1}}^{t} e^{\nu_1 s}\cos\left(\frac{\pi s}{12}+B\right)\left(e^{\nu_2 s}\sin\left(\frac{\pi s}{12}+B\right) - \frac{12}{\pi}\nu_2 e^{\nu_2 s}\cos\left(\frac{\pi s}{12}+B\right)\right)ds$$

$$= \left(1+\left(\frac{12}{\pi}\right)^2\nu_2^2\right)^{-1}\frac{12}{\pi}\times$$

$$\left(\underbrace{\int_{t_{i-1}}^{t} e^{2\theta_1 s}\sin\left(\frac{\pi s}{12}+B\right)\cos\left(\frac{\pi s}{12}+B\right)ds}_{\triangle} - \frac{12}{\pi}\nu_2\underbrace{\int_{t_{i-1}}^{t} e^{2\theta_1 s}\cos\left(\frac{\pi s}{12}+B\right)^2 ds}_{\triangle\triangle}\right)$$

To solve $\triangle$ we first apply the double angle formula (shown in A.2.4) to get

$$\triangle = \frac{1}{2}\int_{t_{i-1}}^{t} e^{2\theta_1 s}\sin\left(2\left[\frac{\pi s}{12}+B\right]\right)ds$$

and the solution to the above integral is defined in A.2.2 to be $g_1(t)$
Hence:

$$\triangle = \frac{1}{2}\left(g_1(t) - g_1(t_{i-1})\right)$$

To solve $\triangle\triangle$ we again apply a double angle formula, as described in section A.2.4. This yields

$$\triangle\triangle = \int_{t_{i-1}}^{t} e^{2\theta_1 s}\left(1+\cos\left(2\left(\frac{\pi s}{12}+B\right)\right)\right)ds$$

Resulting in:

$$\triangle\triangle = \frac{1}{4\theta_1}\left(e^{2\theta_1 t} - e^{2\theta_1 t_{i-1}}\right) + \underbrace{\frac{1}{2}\int_{t_{i-1}}^{t} e^{2\theta_1 s}\cos\left(2\left(\frac{\pi s}{12} + B\right)\right)ds}_{\diamond}$$

To complete the solution for $\triangle\triangle$, we again need to employ repeated integration by parts to solve $\diamond$. In section A.2.3 the solution to $\diamond$ is defined to be $g_2(t)$, hence:

$$\diamond = \frac{1}{2}\left(g_2(t) - g_2(t_{i-1})\right)$$

Pulling this all together for $\star\star\star$ we have

$$\star\star\star = \int_{t_{i-1}}^{t} e^{\nu_1 t} f_2(t,\nu_2)\cos\left(\frac{\pi t}{12} + B\right)dt$$

$$= \left(1 + \left(\frac{12}{\pi}\right)^2\nu_2^2\right)^{-1}\frac{12}{\pi}\times$$

$$\left[\frac{1}{2}\left(g_1(t) - g_1(t_{i-1})\right) - \frac{12}{\pi}\nu_2\left(\frac{1}{4\theta_1}\left(e^{2\theta_1 t} - e^{2\theta_1 t_{i-1}} + \frac{1}{2}\left(g_2(t) - g_2(t_{i-1})\right)\right)\right)\right]$$

Pulling these solutions together for $V_{11}(t)$ we get that

$$V_{11}(t) = V_{11}(t_{i-1})e^{-2\theta_1(t-t_{i-1})}$$

$$+ e^{-2\theta_1 t}\frac{2\pi}{12}V_{12}(t_{i-1})e^{t_{i-1}(\nu_1)}\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{2\theta_3 t_{i-1}}\left[\frac{1}{2}\left(f_1(t,\nu_1)^2 - f_1(t_{i-1},\nu_1)^2\right)\right]$$

$$- 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{2\theta_3 t_{i-1}}f_1(t_{i-1},\nu_1)\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)\left(1 + \left(\frac{12}{\pi}\right)^2\nu_2^2\right)^{-1}\frac{12}{\pi}\times$$

$$\left[\frac{1}{2}\left(g_1(t) - g_1(t_{i-1})\right) - \frac{12}{\pi}\nu_2\left(\frac{1}{4\theta_1}\left(e^{2\theta_1 t} - e^{2\theta_1 t_{i-1}} + \frac{1}{2}\left(g_2(t) - g_2(t_{i-1})\right)\right)\right)\right]$$

$$- \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)f_1(t_{i-1},\nu_2)\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$- \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)e^{2\theta_3 t_{i-1}}\left[\frac{1}{2}\left(f_1(t,\nu_1)^2 - f_1(t_{i-1},\nu_1)^2\right)\right]$$

$$+ \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)e^{2\theta_3 t_{i-1}}f_1(t_{i-1},\nu_1)\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ \frac{\sigma^2}{2\theta_1}\left[e^{2\theta_1 t} - e^{2\theta_1 t_{i-1}}\right].$$

A.1.5 Mean and variance summary

The mean of the process with varying amplitude is governed by:

$$m_1(t) = e^{-\theta_1(t-t_{i-1})}m_1(t_{i-1})$$

$$+ e^{-\theta_1 t}\theta_2\left[e^{\theta_1^t} - e^{\theta_1 t_{i-1}}\right]$$

$$+ e^{-\theta_1 t}\frac{\pi}{12}m_2(t_{i-1})e^{\theta_3(t_{i-1})}\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ e^{-\theta_1 t}\frac{\pi}{12}\theta_4\left[f_1(t,\theta_1) - f_1(t_{i-1},\theta_1)\right]$$

$$- e^{-\theta_1 t}\frac{\pi}{12}\theta_4 e^{\theta_4 t_{i-1}}\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right].$$

$$m_2(t) = m_2(t_{i-1})e^{-\theta_3(t-t_{i-1})} + \theta_4\left(1 - e^{-\theta_3(t-t_{i-1})}\right)$$

The variance of the process with varying amplitude is governed by:

$$V_{11}(t) = V_{11}(t_{i-1})e^{-2\theta_1(t-t_{i-1})}$$

$$+ e^{-2\theta_1 t}\frac{2\pi}{12}V_{12}(t_{i-1})e^{t_{i-1}(\nu_1)}\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{2\theta_3 t_{i-1}}\left[\frac{1}{2}\left(f_1(t,\nu_1)^2 - f_1(t_{i-1},\nu_1)^2\right)\right]$$

$$- 2\left(\frac{\pi}{12}\right)^2 V_{22}(t_{i-1})e^{2\theta_3 t_{i-1}}f_1(t_{i-1},\nu_1)\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)\left(1 + \left(\frac{12}{\pi}\right)^2\nu_2^2\right)^{-1}\frac{12}{\pi}\times$$

$$\left[\frac{1}{2}\left(g_1(t) - g_1(t_{i-1})\right) - \frac{12}{\pi}\nu_2\left(\frac{1}{4\theta_1}\left(e^{2\theta_1 t} - e^{2\theta_1 t_{i-1}} + \frac{1}{2}\left(g_2(t) - g_2(t_{i-1})\right)\right)\right)\right]$$

$$- \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)f_1(t_{i-1},\nu_2)\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$- \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)e^{2\theta_3 t_{i-1}}\left[\frac{1}{2}\left(f_1(t,\nu_1)^2 - f_1(t_{i-1},\nu_1)^2\right)\right]$$

$$+ \frac{\sigma_A^2}{\theta_3}\left(\frac{\pi^2}{288}\right)e^{2\theta_3 t_{i-1}}f_1(t_{i-1},\nu_1)\left[f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right]$$

$$+ \frac{\sigma^2}{2\theta_1}\left[e^{2\theta_1 t} - e^{2\theta_1 t_{i-1}}\right].$$

$$V_{12}(t) = V_{12}(t_{i-1})e^{-(t-t_{i-1})(\theta_1+\theta_3)}$$

$$+ \frac{\pi}{12}V_{22}(t_{i-1})e^{-t(\theta_1+\theta_3)}e^{2\theta_3 t_{i-1}}\left(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right)$$

$$+ \frac{\pi\sigma_A^2}{24\theta_3}e^{-t(\theta_1+\theta_3)}\left[f_1(t,\nu_2) - f_1(t_{i-1},\nu_2) - e^{2\theta_3 t_{i-1}}\left(f_1(t,\nu_1) - f_1(t_{i-1},\nu_1)\right)\right]$$

$$V_{22}(t) = V_{22}(0)e^{-2\theta_3(t-t_{i-1})} + \frac{\sigma_A^2}{2\theta_3}\left(1 - e^{-2\theta_3(t-t_{i-1})}\right)$$

The functions in the above solutions are all defined below.

$$f_1(t,\nu) = \left(1 + \left(\frac{12\nu}{\pi}\right)^2\right)^{-1}\frac{12}{\pi}e^{\pm\nu t}\left(\sin\left(\frac{\pi t}{12} + B\right) \pm \frac{12}{\pi}\nu\cos\left(\frac{\pi t}{12} + B\right)\right)$$

$$g_1(t,\theta = \left(1 + \left(\frac{12\theta_1}{\pi}\right)^2\right)^{-1}\left(\frac{6}{\pi}\right)e^{2\theta_1 t}\left(\frac{12}{\pi}\theta_1\sin\left(2\left[\frac{\pi t}{12} + B\right]\right) - \cos\left(2\left[\frac{\pi t}{12} + B\right]\right)\right)$$

$$g_2(t,\theta) = \left(1 + \left(\frac{12\theta_1}{\pi}\right)^2\right)^{-1}\left(\frac{6}{\pi}\right)e^{2\theta_1 t}\left(\sin\left(2\left[\frac{\pi t}{12} + B\right]\right) + \frac{12}{\pi}\theta_1\cos\left(2\left[\frac{\pi t}{12} + B\right]\right)\right)$$

A.2 Useful integrals

A.2.1 Deriving $f_1(t)$

To solve $\int e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right) dt$ we use integration by parts which states that

$$\int u dv = uv - \int v du$$

with $u = e^{\pm\nu t}$ and $dv = \cos\left(\frac{\pi t}{12} + B\right)$. This results in

$$\int e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right) dt = e^{\pm\nu t} \sin\left(\frac{\pi t}{12} + B\right)\frac{12}{\pi} \mp \frac{12}{\pi}\nu \int e^{\pm\nu t} \sin\left(\frac{\pi t}{12} + B\right) dt$$

Using the technique again for $\int e^{\pm\nu t} \sin\left(\frac{\pi t}{12} + B\right) dt$ gives:

$$\int e^{\pm\nu t} \sin\left(\frac{\pi t}{12} + B\right) dt = \mp e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right)\frac{12}{\pi} + \frac{12}{\pi}\nu \int \pm e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right) dt$$

Combining the two equations above yields:

$$\int e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right) dt = \frac{12}{\pi}e^{\pm\nu t} \sin\left(\frac{\pi t}{12} + B\right) \pm \left(\frac{12}{\pi}\right)^2 \nu e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right)$$
$$- \left(\frac{12\nu}{\pi}\right)^2 \int e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right) dt$$

We then add the last term to both sides of the above equation and then divide through by $\left(1 + \left(\frac{12\nu}{\pi}\right)^2\right)$ we are left with:

$$\int e^{\pm\nu t} \cos\left(\frac{\pi t}{12} + B\right) dt = \left(1 + \left(\frac{12\nu}{\pi}\right)^2\right)^{-1}\frac{12}{\pi}e^{\pm\nu t}\left(\sin\left(\frac{\pi t}{12} + B\right) \pm \frac{12}{\pi}\nu \cos\left(\frac{\pi t}{12} + B\right)\right)$$

Which we shall define as $f_1(t)$:

$$f_1(t) = \int e^{\theta_1 t} \cos\left(\frac{\pi t}{12} + B\right) dt$$

A.2.2 Deriving $g_1(t)$

To solve $\int e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$ we use integration by parts with $u = e^{2\theta_1 t}$ and $dv = \sin\left(2\left[\frac{\pi t}{12} + B\right]\right)$. This results in:

$$\int e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) dt = -\frac{6}{\pi}e^{2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right) + \frac{12}{\pi}\theta_1 \int e^{2\theta_1 s} \cos\left(2\left[\frac{\pi s}{12} + B\right]\right) ds$$
$$\text{(A.2)}$$

Again, we have an integral in the above solution that requires the use of integration by parts to solve it. We will use $u = e^{2\theta_1 t}$ and $dv = \cos\left(2\left[\frac{\pi t}{12} + B\right]\right)$.

$$\int e^{2\theta_1 s} \cos\left(2\left[\frac{\pi s}{12} + B\right]\right) ds = \frac{6}{\pi} e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) - \frac{12\theta_1}{\pi} \int e^{2\theta_1 s} \sin\left(2\left[\frac{\pi s}{12} + B\right]\right) ds$$
(A.3)

Substituting in what we have for $\int e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$ and rearranging gives us

$$\int e^{2\theta_1 s} \sin\left(2\left[\frac{\pi s}{12} + B\right]\right) ds + \left(\frac{12}{\pi}\right)^2 \theta_1^2 \int e^{2\theta_1 s} \sin\left(2\left[\frac{\pi s}{12} + B\right]\right) ds$$

$$= \left(\frac{72}{\pi^2}\right) e^{2\theta_1 t} \theta_1 \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) - \frac{6}{\pi} e^{2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right)$$

Dividing through by $\left(1 + \left(\frac{12}{\pi}\right)^2 \theta_1^2\right)$ leaves us with

$$\int e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$$

$$= \left(1 + \left(\frac{12}{\pi}\right)^2 \theta_1^2\right)^{-1} \left(\frac{6}{\pi}\right) e^{2\theta_1 t} \left(\frac{12}{\pi} \theta_1 \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) - \cos\left(2\left[\frac{\pi t}{12} + B\right]\right)\right)$$

Which we define as $g_1(t)$

$$g_1(t) = \int e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$$

A.2.3 Deriving $g_1 2(t)$

To solve $\int e^{2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$ we use the result from A.3. This results in:

$$\int e^{2\theta_1 s} \cos\left(2\left[\frac{\pi s}{12} + B\right]\right) ds = \frac{6}{\pi} e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) - \frac{12\theta_1}{\pi} \int e^{2\theta_1 s} \sin\left(2\left[\frac{\pi s}{12} + B\right]\right) ds$$
(A.4)

In the above result we have the same integral as A.2, so substituting that result in an rearranging give us

$$\int e^{2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right) dt + \left(\frac{12}{\pi}\right)^2 \theta_1^2 \int e^{2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$$

$$= \frac{6}{\pi} e^{2\theta_1 t} \sin\left(2\left[\frac{\pi t}{12} + B\right]\right) + \frac{72\theta_1}{\pi^2} e^{2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right)$$

Dividing through by $(1 + \left(\frac{12}{\pi}\right)^2 \theta_1^2)$ leaves us with

$$\int e^{-2\theta_1 t} \frac{1}{2} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right) dt$$
$$= \left(1 + \left(\frac{12}{\pi}\right)^2 \theta_1^2\right)^{-1} \left(\frac{6}{\pi}\right) e^{2\theta_1 t} \left(\sin\left(2\left[\frac{\pi t}{12} + B\right]\right) + \frac{12}{\pi} \theta_1 \cos\left(2\left[\frac{\pi t}{12} + B\right]\right)\right)$$

Which we define as $g_2(t)$

$$g_2(t) = \int e^{-2\theta_1 t} \cos\left(2\left[\frac{\pi t}{12} + B\right]\right) dt.$$

A.2.4   Double angle formulae

The double angle formula can be stated in in multiple ways

Double angle formula for functions of the form $\sin(2A)$

$$2\sin(A)\cos(A) = \sin(2A)$$

Double angle formula for functions of the form $\cos(2A)$

$$\cos(2A) = \cos^2(A) - \sin^2(A)$$

Bibliography

AIT-SAHALIA, Y. 2008 Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics* **36** (2), 906–937.

ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. 2009 Particle Markov chain Monte Carlo for efficient numerical simulation. In *Monte Carlo and Quasi-Monte Carlo Methods 2008* (ed. P. L'Ecuyer & A. B. Owen), pp. 45–60. Spinger-Verlag Berlin Heidelberg.

ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. 2010 Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc. B* **72** (3), 1–269.

BÉRARD, J., DEL-MORAL, P. & DOUCET, A. 2013 A lognormal central limit theorem for particle approximations of normalizing constants.

BLACK, F. & SCHOLES, M. 1973 The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637–654.

BOTHA, I., KOHN, R. & DROVANDI, C. 2019 Particle methods for stochastic differential equation mixed effects models.

BUCY, R. & JOSEPH, P. 2005 *Filtering for Stochastic Processes with Applications to Guidance*, 2nd edn. John Wiley & Sons.

CHOPPALA, P., GUNAWAN, D., CHEN, J., TRAN, M.-N. & KOHN, R. 2016 Bayesian inference for state space models using block and correlated pseudo marginal methods. Available from `http://arxiv.org/abs/1311.3606`.

COTTER, S. L., ROBERTS, G. O., STUART, A. M. & WHITE, D. 2013 MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science* **28**, 424–446.

DAHLIN, J., LINDSTEN, F., KRONANDER, J. & SCHON, T. B. 2015 Accelerating pseudo-marginal Metropolis-Hastings by correlating auxiliary variables. Available from `https://arxiv.1511.05483v1`.

DEL MORAL, P. 2004 *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer.

DEL MORAL, P., DOUCET, A. & JASRA, A. 2006 Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **68**, 411–436.

DELATTRE, M. & LAVIELLE, M. 2013 Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and its interface* **6** (4), 519–532.

DELIGIANNIDIS, G., DOUCET, A. & PITT, M. K. 2018 The correlated pseudo-marginal method. *J. R. Statist. Soc.B* **80**, 839–870.

DONNET, S., FOULLEY, J.-L. & SAMSON, A. 2010 Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics* **66** (3), 733–741.

DONNET, S. & SAMSON, A. 2013*a* A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Advanced drug delivery reviews* **65** (7), 929–939.

DONNET, S. & SAMSON, A. 2013*b* Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. *Journal de la Société Française de Statistique* **155** (1), 49–72.

DOUCET, A., PITT, M. K. & KOHN, R. 2015 Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102**, 295–313.

DUFFY, P. H., FEUERS, R. J., LEAKEY, J. E. A., NAKAMURA, K. D., TURTURRO, A. & HART, R. W. 1989 Effects of chronic caloric restriction on physiological variables related to energy metabolism in the male fischer 344 rat. *Mechanisms of Ageing and Development* **48**, 117–133.

ELF, J. & EHRENBERG, M. 2003 Fast evolution of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.* **13** (11), 2475–2484.

FUCHS, C. 2013 *Inference for diffusion processes with applications in Life Sciences*. Springer.

GAMERMAN, H. & LOPES, H, F. 2006 *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall / CRC Texts in Statistical Science.

GELFAND, A, E. & SMITH, A F, M. 1990 Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* pp. 390–409.

GEMEN, S. & GEMEN, D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 721–741.

GERBER, M. & CHOPIN, N. 2015 Sequential quasi Monte Carlo (with discussion). *Journal of the Royal Statistical Society, Series B* **77**, 509–579.

GOLIGHTLY, A., BOYS, R. J., CAMERON, K. M. & ZGLINICKI, T. V. 2012 The effect of late onset, short-term caloric restriction on the core temperature and physical activity in mice. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61** (5), 733–751.

GOLIGHTLY, A., BRADLEY, E., LOWE, T. & GILLESPIE, C. S. 2019 Correlated pseudo-marginal schemes for time-discretised stochastic kinetic models. *CSDA to appear* .

GOLIGHTLY, A., HENDERSON, D. A. & SHERLOCK, C. 2015 Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing* **25**, 1039–1055.

GOLIGHTLY, A. & WILKINSON, D. J. 2011 Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* **1** (6), 807–820.

GORDON, N. J., SALMOND, D. J. & SMITH, A. F. M. 1993 Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140**, 107–113.

HASTINGS, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1), 97–109.

KLOEDEN, P. E. & PLATEN, E. 1992 *Numerical solution of stochastic differential equations*. Springer.

KOMOROWSKI, M., FINKENSTADT, B., HARPER, C. & RAND, D. 2009 Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics* **10** (1), 343.

KÜNSCH, H. R. 2013 Particle filters. *Bernoulli* **19**, 1391–1403.

LAVIELLE, M. 2014 *Mixed effects models for the population approach: models, tasks, methods and tools*. Chapman and Hall/CRC.

LIAO, C. Y., RIKKE, B. A., JOHNSON, T. E., DIAZ, V. & NELSON, J. F. 2010 Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. *Aging Cell* **9** (1), 92–95.

MERTON, R. C. 1973 Theory of rational option pricing. *The Bell Journal of Economics and Management Science* **4**, 144–183.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. 1953 Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21** (6), 1087–1092.

MURRAY, L., LEE, A. & JACOB, P. E. 2016 Parallel resampling in the particle filter. *Journal of Computational and Graphical Statistics* **25** (3), 789–805.

ØKSENDAL, B. 1995 *Stochastic differential equations: An introduction with applications*, 6th edn. Springer-Verlag, Berlin Heidelberg New York.

OVERGAARD, R. V., JONSSON, N., TORNØE, C. W. & MADSEN, H. 2005 Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *Journal of pharmacokinetics and pharmacodynamics* **32** (1), 85–107.

OWEN, J., WILKINSON, D. J. & GILLESPIE, C. S. 2015 Scalable inference for Markov processes with intractable likelihoods. *Statistics and Computing* **25** (1), 145–156.

PICCHINI, U., DE GAETANO, A. & DITLEVSEN, S. 2010 Stochastic differential mixed-effects models. *Scandinavian Journal of statistics* **37** (1), 67–90.

PICCHINI, U. & DITLEVSEN, S. 2011 Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55** (3), 1426–1444.

PICCHINI, U. & FORMAN, J. L. 2019 Bayesian inference for stochastic differential equation mixed effects models of a tumor xenography study. *Journal of the Royal Statistical Society: Series C* Doi:10.1111/rssc.12347.

PITT, M. K., DOS SANTOS SILVA, R., GIORDANI, P. & KOHN, R. 2012 On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Econometrics* **171** (2), 134–151.

PLUMMER, M., BEST, N., COWLES, K. & VINES, K. 2006 CODA: convergence diagnosis and output analysis for MCMC. *R News* **6** (1), 7–11.

PUGH, T. D., OBERLEY, T. D. & WEINDRUCH, R. 1999 Dietary intervention at middle age. *Cancer Research* **59** (7), 1642–1648.

ROBERTS, G. O. & ROSENTHAL, J, S. 2001 Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.* **16** (4), 351–367.

124

ROTH, G. S., LANE, M. A., INGRAM, D. K., MATTISON, J. A., ELAHI, D. J., TOBIN, D. MULLER, D. & METTER, E. J. 2002 Biomarkers of caloric restriction may predict longevity in humans. *Science* **297**, 297–811.

SAARINEN, A., LINNE, M.-L. & YLI-HARJA, O. 2006 Modeling single neuron behavior using stochastic differential equations. *Neurocomputing* **69** (10–12), 1091–1096.

SHERLOCK, C., THIERY, A., ROBERTS, G. O. & ROSENTHAL, J. S. 2013 On the efficiency of pseudo-marginal random walk Metropolis algorithms. Available from http://arxiv.org/abs/1309.7209.

SHERLOCK, C., THIERY, A., ROBERTS, G. O. & ROSENTHAL, J. S. 2015 On the efficiency of pseudo-marginal random walk Metropolis algorithms. *The Annals of Statistics* **43** (1), 238–275.

SØRENSEN, H. 2004 Parametric inference for diffusion processes observed at discrete points in time. *International Statistical Review* **72** (3), 337–354.

SPINDLER, S. R. 2005 Rapid and reversible induction of the longevity, anticancer and genomic effects of caloric restriction. *Mechanisms of Ageing and Development* **126**, 960–966.

STEELE, J. 2012 *Stochastic calculus and financial applications*. Springer Science & Business Media.

TRAN, M. N., KOHN, R., QUIROZ, M. & VILLANI, M. 2017 The block pseudo-marginal sampler. Available from https://arxiv.org/abs/1603.02485.

VOLK, M. J., PUGH, T. D., KIM, M., FRITH, C. H., DAYNES, R. A., ERSHLER, W. B. & WEINDRUCH, R. 1994 Dietary restriction from middle age attenuates age-associated lymphoma development and interleukin 6 dysregulation in c57bl/6 mice. *Cancer Reseach* **54**, 3054–61.

WEINDRUCH, R. H. & WALFORD, R. L. 1988 *The retardation of aging and disease by dietary restriction*. Springfield, IL, USA: Charles C. Thomas.

WHITAKER, G. A. 2016 Bayesian inference for stochastic differential mixed-effects models. PhD thesis, Newcastle University.

WHITAKER, G. A., GOLIGHTLY, A., BOYS, R. J. & SHERLOCK, C. 2017 Bayesian inference for diffusion driven mixed-effects models. *Bayesian Analysis* **12**, 435–463.

WILKINSON, D. J. 2018 *Stochastic Modelling for Systems Biology*, 3rd edn. Boca Raton, Florida: Chapman & Hall/CRC Press.

YU, B. P., MASORO, E. J. & MCMAHAN, C. A. 1985 Nutritional influences on aging of fischer 344 rats: I. physical, metabolic, and longevity characteristics. *Journal of Gerontology* **40**, 657–670.