Bayesian inference for stochastic differential random effects models

Gavin Whitaker

School of Mathematics and Statistics
Newcastle University

30th April 2014

Supervisors: Richard Boys and Andrew Golightly

SDE models

- Consider an Itô process $\{\boldsymbol{X}_t, t \geq 0\}$ satisfying

$$d\boldsymbol{X}_t = \boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})dt + \sqrt{\boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})}d\boldsymbol{W}_t$$

  - $\boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})$ is the drift
  - $\boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})$ is the diffusion coefficient
  - $\boldsymbol{W}_t$ is standard Brownian motion

- Seek a numerical solution via the Euler-Maruyama approximation

$$\Delta\boldsymbol{X}_t \equiv \boldsymbol{X}_{t+\Delta t} - \boldsymbol{X}_t = \boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})\Delta t + \sqrt{\boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})}\Delta\boldsymbol{W}_t$$

where $\Delta\boldsymbol{W}_t \sim N(\boldsymbol{0}, \boldsymbol{I}\Delta t)$

Random effects SDE models

- Consider the case where we have ℓ subjects and that each individual can be represented by the same SDE
- Common parameters $\boldsymbol{\theta}$
- Different parameters $\boldsymbol{\psi}^{(i)}$, $i = 1, \ldots, \ell$
- This gives us a stochastic differential random effects model:

$$d\boldsymbol{X}_t^{(i)} = \boldsymbol{\alpha}\left(\boldsymbol{X}_t^{(i)}, \boldsymbol{\theta}, \boldsymbol{\psi}^{(i)}\right) dt + \sqrt{\boldsymbol{\beta}\left(\boldsymbol{X}_t^{(i)}, \boldsymbol{\theta}, \boldsymbol{\psi}^{(i)}\right)} d\boldsymbol{W}_t^{(i)}$$

  for $i = 1, \ldots, \ell$
- Suppose we have data available at times $t_0^{(i)}, t_1^{(i)}, \ldots, t_{n_i}^{(i)}$ for each individual i
- We implement a data augmentation approach

- Consider $\left[t_j^{(i)}, t_{j+1}^{(i)}\right]$ and introduce a partition

$$t_j^{(i)} = \tau_{j,0}^{(i)} < \underbrace{\tau_{j,1}^{(i)} < \ldots < \tau_{j,m_j^{(i)}-1}^{(i)}}_{\text{latent times}} < \tau_{j,m_j^{(i)}}^{(i)} = t_{j+1}^{(i)}$$

- Time step between observations

$$\Delta_{\tau_j}^{(i)} = \frac{t_{j+1}^{(i)} - t_j^{(i)}}{m_j^{(i)}}$$

- Allows for irregularly spaced data for each individual

- For individual i

$$\boldsymbol{d}^{(i)} = \left(\boldsymbol{x}^{(i)}_{t_0}, \boldsymbol{x}^{(i)}_{t_1}, \dots, \boldsymbol{x}^{(i)}_{t_{n_i}}\right)$$

$$\boldsymbol{x}^{(i)} = \left(\boldsymbol{x}^{(i)}_{\tau_{0,1}}, \boldsymbol{x}^{(i)}_{\tau_{0,2}}, \dots, \boldsymbol{x}^{(i)}_{\tau_{0,m_0^{(i)}-1}}, \boldsymbol{x}^{(i)}_{\tau_{1,1}}, \dots, \boldsymbol{x}^{(i)}_{\tau_{n_i-1,m_{n-1}^{(i)}-1}}\right)$$

- $\boldsymbol{x}^{(i)}$ is the values of the skeleton path at times $\tau^{(i)}_{0,1}, \tau^{(i)}_{0,2}$ etc
- Putting these together for ℓ individuals

$$\boldsymbol{d} = \left(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \dots, \boldsymbol{d}^{(\ell)}\right)$$

$$\boldsymbol{x} = \left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(\ell)}\right)$$

Formulate joint posterior for parameters and latent data as

$$\pi\left(\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{x}|\boldsymbol{d}\right) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\psi}|\boldsymbol{\theta})\pi(\boldsymbol{x},\boldsymbol{d}|\boldsymbol{\theta},\boldsymbol{\psi})$$

$$\propto \underbrace{\pi(\boldsymbol{\theta})\pi(\boldsymbol{\psi}|\boldsymbol{\theta})}_{\text{prior}} \times \prod_{i=1}^{\ell}\prod_{j=0}^{n_i-1}\prod_{k=0}^{m_j^{(i)}-1} \underbrace{\pi\left(\boldsymbol{x}^{(i)}_{\tau_{j,(k+1)}}\middle|\boldsymbol{x}^{(i)}_{\tau_{j,k}},\boldsymbol{\theta},\boldsymbol{\psi}^{(i)}\right)}_{\text{Euler density}}$$

where

$$\boldsymbol{x}^{(i)}_{\tau_{j,(k+1)}}\middle|\boldsymbol{x}^{(i)}_{\tau_{j,k}},\boldsymbol{\theta},\boldsymbol{\psi}^{(i)} \sim N\left(\boldsymbol{x}^{(i)}_{\tau_{j,k}}+\boldsymbol{\alpha}^{(i)}_{j,k}\Delta^{(i)}_{t_j},\boldsymbol{\beta}^{(i)}_{j,k}\Delta^{(i)}_{t_j}\right)$$

Note $\boldsymbol{\alpha}^{(i)}_{j,k}=\boldsymbol{\alpha}\left(\boldsymbol{x}^{(i)}_{\tau_{j,k}},\boldsymbol{\theta},\boldsymbol{\psi}^{(i)}\right)$ and $\boldsymbol{\beta}^{(i)}_{j,k}=\boldsymbol{\beta}\left(\boldsymbol{x}^{(i)}_{\tau_{j,k}},\boldsymbol{\theta},\boldsymbol{\psi}^{(i)}\right)$

- The posterior distribution is typically analytically intractable
- Use a Gibbs sampler, alternating between draws of
  - $\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{\psi}$
  - $\boldsymbol{\psi}|\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{\theta}$
  - $\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{\psi}$
- We can update $\boldsymbol{\psi}$ componentwise, for each individual

$$\pi(\boldsymbol{\psi}|\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{\theta}) = \prod_{i=1}^{\ell} \pi\left(\boldsymbol{\psi}^{(i)}|\boldsymbol{x}^{(i)}, \boldsymbol{d}^{(i)}, \boldsymbol{\theta}\right)$$

- Similarly

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{\psi}) = \prod_{i=1}^{\ell} \pi\left(\boldsymbol{x}^{(i)}|\boldsymbol{\theta}, \boldsymbol{d}^{(i)}, \boldsymbol{\psi}^{(i)}\right)$$

- Typically Metropolis within Gibbs updates are needed

Example: orange tree growth

- Picchini and Ditlevsen (2011) discuss a model for the growth of orange trees incorporating random effects. We use an equivalent reparameterisation, written as

$$dX_t^{(i)} = \frac{1}{\phi_1^{(i)}\phi_2^{(i)}} X_t^{(i)} \left(\phi_1^{(i)} - X_t^{(i)}\right) dt + \sigma\sqrt{X_t^{(i)}}dW_t^{(i)}, \quad i = 1, \dots, 100$$

- $\phi_1^{(i)} \sim N(\phi_1, \sigma^2_{\phi_1})$ and $\phi_2^{(i)} \sim N(\phi_2, \sigma^2_{\phi_2})$
- X_t is the circumference (mm)
- t is the number of days since December 31^{st} 1968
- $\phi_1^{(i)}$ is the asymptotic circumference
- $\phi_2^{(i)}$ is the rate of change parameter

$$dX_t^{(i)} = \frac{1}{\phi_1^{(i)}\phi_2^{(i)}} X_t^{(i)} \left(\phi_1^{(i)} - X_t^{(i)}\right) dt + \sigma\sqrt{X_t^{(i)}} dW_t^{(i)}, \quad i = 1, \ldots, 100$$

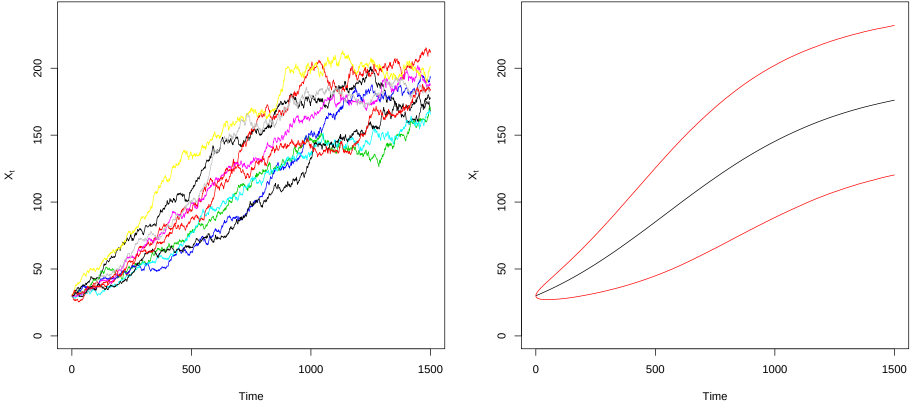
- Common parameters: ϕ_1, ϕ_2, σ_{ϕ_1}, σ_{ϕ_2} and σ
- Tree specific parameters: $\phi_1^{(i)}$ and $\phi_2^{(i)}$
- We repeat Picchini and Ditlevsen's simulation study for 100 trees with $x_0^{(i)} = 30$,

$$(\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2}, \sigma) = (195, 350, 25, 52.5, 0.08),$$

$\phi_1^{(i)} \sim N(195, 25^2)$ and $\phi_2^{(i)} \sim N(350, 52.5^2)$

- This gives us 16 observations at intervals of 100 days on 100 trees

Figure: **left:** 10 simulated skeleton paths
right: mean, upper and lower 2.5 percentiles for 1k realisations



Possible MCMC scheme

- Recall we wish to sample
  - $\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{\psi}$
  - $\boldsymbol{\psi}|\boldsymbol{x}, \boldsymbol{d}, \boldsymbol{\theta}$
  - $\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{\psi}$
- We can formulate the joint posterior as

$$
\begin{aligned}
\pi\left(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{x}|\boldsymbol{d}\right) \propto \pi(\phi_1)\pi(\phi_2)\pi(\sigma_{\phi_1})\pi(\sigma_{\phi_2})\pi(\sigma) \\
\times \pi(\boldsymbol{\phi}_1|\phi_1, \sigma_{\phi_1})\pi(\boldsymbol{\phi}_2|\phi_2, \sigma_{\phi_2}) \\
\times \prod_{i=1}^{100} \underbrace{\pi\left(\boldsymbol{x}^{(i)}|\phi_1^{(i)}, \phi_2^{(i)}, \sigma, \boldsymbol{d}^{(i)}\right)}_{\text{Euler density}}
\end{aligned}
$$

- Sample common parameters $\boldsymbol{\theta}$ via Gibbs updates
- Sample each $\boldsymbol{\psi}^{(i)}$ using random normal proposals
- Use fairly uninformative priors
  - ϕ_1 and ϕ_2 are Normally distributed
  - $\sigma_{\phi_1}^{-2}$, $\sigma_{\phi_2}^{-2}$ and σ^{-2} follow Gamma distributions
- Recall

$$dX_t^{(i)} = \frac{1}{\phi_1^{(i)}\phi_2^{(i)}} X_t^{(i)} \left(\phi_1^{(i)} - X_t^{(i)}\right) dt + \sigma\sqrt{X_t^{(i)}} dW_t^{(i)}, \quad i = 1, \ldots, 100$$

  with $\phi_1^{(i)}$ and $\phi_2^{(i)}$ Normally distributed
- Sample $\boldsymbol{x}$ by updating each path $\boldsymbol{x}^i$ separately, conditional on $\phi_1^{(i)}$, $\phi_2^{(i)}$ and σ

(Modified) Innovation scheme for random effects SDEs

- This type of scheme suffers from intolerably poor mixing as $m \to \infty$
- Caused by relation between the parameters and the path in the quadratic variation $\Rightarrow$ acceptance probability $\to 0$ as $m \to \infty$
- The (Modified) Innovation scheme (Golightly and Wilkinson (2008)) conditions on the Brownian increments

$$\boldsymbol{w} = \left(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}, \ldots, \boldsymbol{w}^{(\ell)}\right)$$

  (which drive a tractable conditioned diffusion) to overcome the dependence between the parameters and the path
- Insight: the quadratic variation of $\boldsymbol{w}$ does not itself determine any of the model parameters and should therefore be effective in decoupling the problematic dependence

  $\Rightarrow$ the scheme should not become degenerate as $m \to \infty$

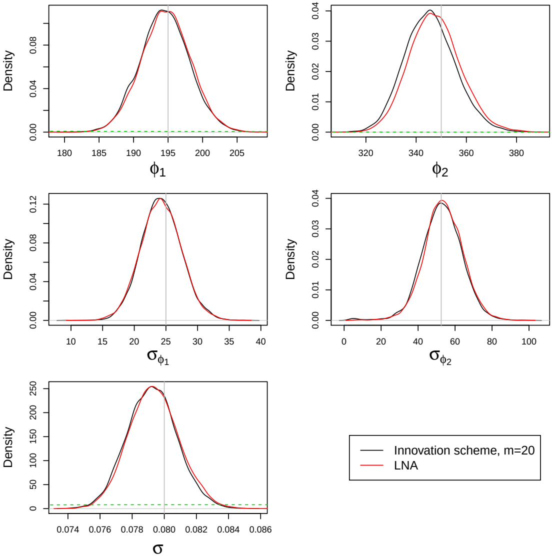
Application: orange tree growth

- Recall we have 16 observations on 100 trees at intervals of 100 days
- Parameter choice:

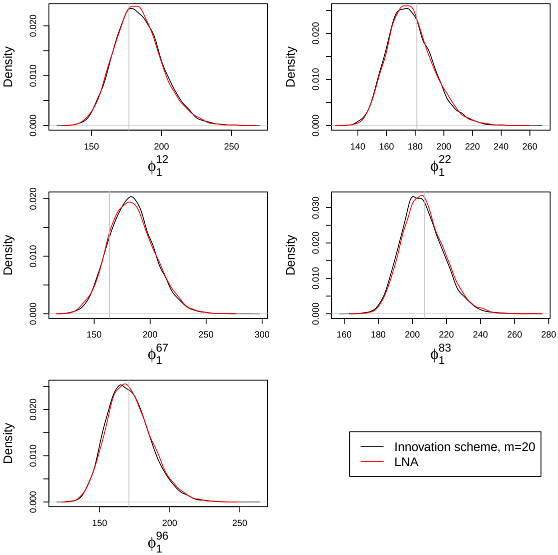
$$(\phi_1, \phi_2, \sigma_{\phi_1}, \sigma_{\phi_2}, \sigma) = (195, 350, 25, 52.5, 0.08)$$
$$\phi_1^{(i)} \sim N(195, 25^2)$$
$$\phi_2^{(i)} \sim N(350, 52.5^2)$$

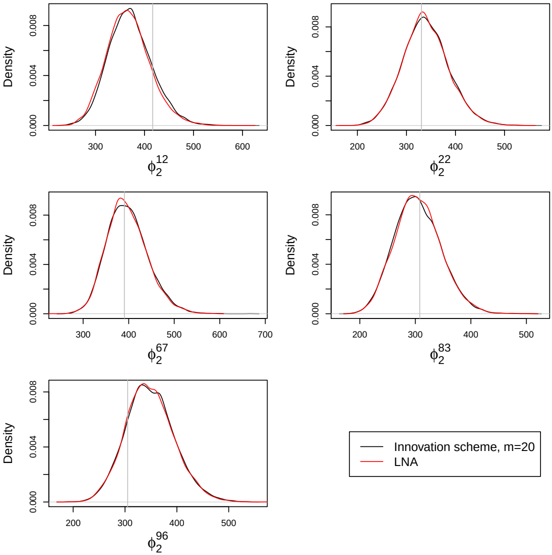
- We run the (Modified) Innovation scheme with $m = 20$, for 1 million iterations with a thin of 100
- We compare the output with that obtained under the linear noise approximation (LNA)

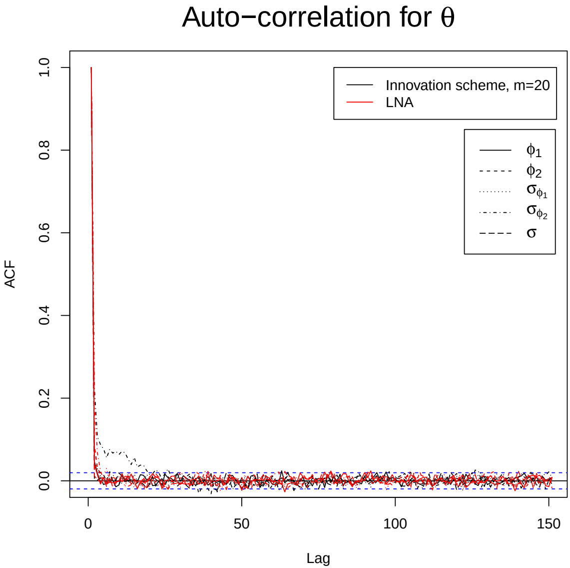
Results



Density
0.00
0.04
0.08
180
185
190
195
200
205
ϕ_1
Density
0.00
0.01
0.02
0.03
0.04
320
340
360
380
ϕ_2
Density
0.00
0.04
0.08
0.12
10
15
20
25
30
35
40
σ_{ϕ_1}
Density
0.00
0.01
0.02
0.03
0.04
0
20
40
60
80
100
σ_{ϕ_2}
Density
0
50
100
150
200
250
0.074
0.076
0.078
0.080
0.082
0.084
0.086
σ
Innovation scheme, m=20
LNA

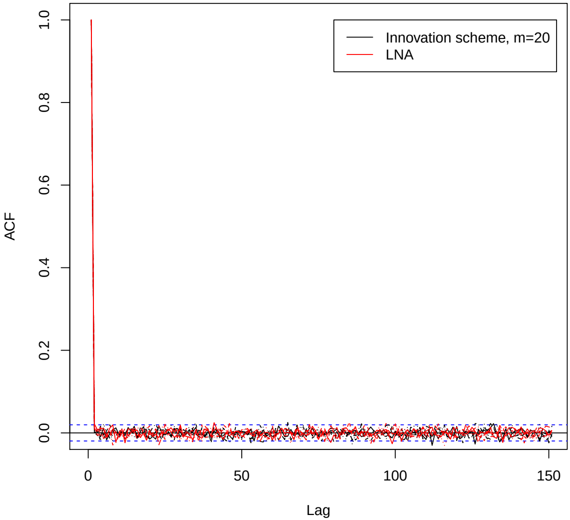

Density
0.000
0.010
0.020
150
200
250
ϕ_1^{12}
140
160
180
200
220
240
260
ϕ_1^{22}
300
ϕ_1^{67}
0.030
280
ϕ_1^{83}
ϕ_1^{96}
Innovation scheme, m=20
LNA


Density
0.000
0.004
0.008
300
400
500
600
ϕ_2^{12}
200
300
400
500
ϕ_2^{22}
300
400
500
600
700
ϕ_2^{67}
200
300
400
500
ϕ_2^{83}
200
300
400
500
ϕ_2^{96}
Innovation scheme, m=20
LNA

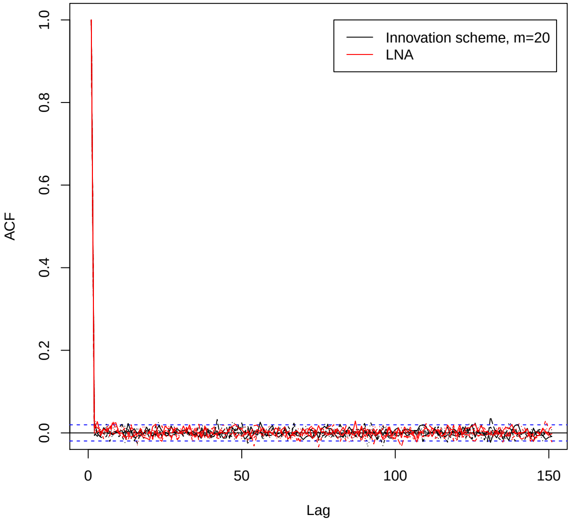


Auto−correlation for θ
Innovation scheme, m=20
LNA
ϕ_1
ϕ_2
σ_{ϕ_1}
σ_{ϕ_2}
σ
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
Lag

Auto−correlation for ϕ_1^i, i=12,22,67,83,96



Auto−correlation for ϕ_2^i, i=12,22,67,83,96



Comparison of the schemes

- Comparing the LNA against the (Modified) Innovation scheme
  - LNA: ESS/sec $= 1.178$
  - (Modified) Innovation scheme: ESS/sec $= 0.057$

  which gives a comparable ESS/sec $= 20.667$
- The gain by using the LNA is slightly exaggerated by the fact that the system of ODEs can be solved analytically
- It should still be advantageous to use the LNA over the (Modified) Innovation scheme on a system that can't be solved analytically
- However it should be noted that the LNA only gives an approximation to the posterior density

Future work

- Extend these methods to data where we only have partial observations
- Examine the case where we have observations observed with error, typically Gaussian error
- Apply these schemes to a more challenging example e.g. larger model
- Compare these methods with pMCMC
- Apply these schemes to lemon trees

References

- Dargatz, C. *Bayesian Inference for Diffusion Processes with Applications in Life Sciences.* PhD thesis, Ludwig–Maximilians–Universität München, 2010
- Fearnhead, P. Giagos, V. and Sherlock, C. *Inference for reaction networks using the linear noise approximation.* http://arxiv.org/abs/1205.6920, 2012
- Golightly, A. and Gillespie, C.S. *Simulation of stochastic kinetic models.* In Silico Systems Biology, 169-187, 2013
- Golightly, A. and Wilkinson, D. J. *Bayesian inference for nonlinear multivariate diffusion models observed with error.* Computational Statistics and Data Analysis, 52 (3) 1674-1693, 2008
- Picchini, U. and Ditlevsen, S. *Practical estimation of high dimensional stochastic differential mixed-effects models.* Computational Statistics and Data Analysis, 55 (E3) 1426-1444, 2011
- Wilkinson, D. J. and Golightly, A. *Markov chain Monte Carlo algorithms for SDE parameter estimation.* Learning and Inference in Computational Systems Biology, 2010