MCMC schemes for partially observed diffusions: challenges and solutions

Gavin Whitaker

5th March 2014

Supervisors: RJB and AG
Outline

- Reminder of Stochastic Differential Equations (SDE)
- The Aphid model
- Bayesian inference for random effects SDEs
- The Aphid model - results
- Challenges
  - Working with the Euler approximation of non-negative processes
- Possible solutions
Consider an Itô process $\{X_t, t \geq 0\}$ satisfying

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

- $\alpha(X_t, \theta)$ is the drift
- $\beta(X_t, \theta)$ is the diffusion coefficient
- $W_t$ is standard Brownian motion

Euler-Maruyama approximation:

$$\Delta X_t \equiv X_{t+\Delta t} - X_t = \alpha(X_t, \theta) \Delta t + \sqrt{\beta(X_t, \theta)} \Delta W_t$$

where $\Delta W_t \sim N(0, I \Delta t)$
Aphid growth model

Also known as plant lice, or greenfly
They are small sap sucking insects
Female aphids can reproduce without mating
Some species of ants “farm” aphids for the honeydew they release. These “dairying ants” “milk” the aphids by stroking them
Reaction 1: \( \mathcal{N} \xrightarrow{\lambda} 2\mathcal{N} + C \)

Reaction 2: \( \mathcal{N} + C \xrightarrow{\mu} C \)

The mass action SDE representation of the system dynamics is

\[
\begin{pmatrix}
\frac{dN_t}{dt} \\
\frac{dC_t}{dt}
\end{pmatrix} = 
\begin{pmatrix}
\lambda N_t - \mu N_tC_t \\
\lambda N_t
\end{pmatrix} dt + 
\begin{pmatrix}
\lambda N_t + \mu N_tC_t & \lambda N_t \\
\lambda N_t & \lambda N_t
\end{pmatrix}^{1/2} dW_t
\]

- \( N_t \) is the aphid population size at time \( t \)
- \( C_t \) is the cumulative population at time \( t \)
Figure: Skeleton path for Aphid model, $\lambda = 1.75$, $\mu = 0.001$
Consider the case where we have $\ell$ subjects and that each individual can be represented by the same SDE.

Different parameters $\phi^{(i)}$, $i = 1, \ldots, \ell$

This gives us a stochastic differential random effects model:

$$dX^{(i)}_t = \alpha \left( X^{(i)}_t, \phi^{(i)} \right) dt + \sqrt{\beta \left( X^{(i)}_t, \phi^{(i)} \right) } dW^{(i)}_t$$

for $i = 1, \ldots, \ell$

Random effects distributions depend on $\theta$

Suppose we have data available at times $t^{(i)}_0, t^{(i)}_1, \ldots, t^{(i)}_{n_i}$ for each individual $i$
Aphid growth model - random effects

SDE for population $i$

$$
\begin{align*}
\frac{dX_t^{(i)}}{dt} &= \left( \lambda^{(i)} N_t^{(i)} - \mu^{(i)} N_t^{(i)} C_t^{(i)} \lambda^{(i)} N_t^{(i)} \right) dt \\
&+ \left( \begin{array}{cc}
\lambda^{(i)} N_t^{(i)} + \mu^{(i)} N_t^{(i)} C_t^{(i)} & \lambda^{(i)} N_t^{(i)} \\
\mu^{(i)} N_t^{(i)} & \lambda^{(i)} N_t^{(i)} \\
\end{array} \right)^{\frac{1}{2}} dW_t^{(i)}
\end{align*}
$$

- $\lambda^{(i)} \sim LN(\lambda, \sigma_{\lambda}^2)$ and $\mu^{(i)} \sim LN(\mu, \sigma_{\mu}^2)$
- Common parameters: $\theta = (\lambda, \mu, \sigma_\lambda, \sigma_\mu)^T$
- Population specific parameters: $\phi^{(i)} = (\lambda^{(i)}, \mu^{(i)})^T$
- We look at a simulation study involving 20 populations
Figure: Skeleton paths for 20 Aphid populations
Consider $\left[ t_j^{(i)}, t_{j+1}^{(i)} \right]$ and introduce a partition

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

\[ \text{latent times} \]

- **Time step between observations**

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

- **Allows for irregularly spaced data for each individual**
For individual $i$

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

$$x^{(i)} = \left( x^{(i)}_{\tau_{0,1}}, x^{(i)}_{\tau_{0,2}}, \ldots, x^{(i)}_{\tau_{0,m(i)-1}}, x^{(i)}_{\tau_{1,1}}, \ldots, x^{(i)}_{\tau_{n_i-1,m(i)-1-1}} \right)$$

Putting these together for $\ell$ individuals

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

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

Formulate joint posterior for parameters and latent data as

$$\pi(\theta, \phi, x | d) \propto \pi(\theta) \pi(\phi | \theta) \pi(x, d | \theta, \phi)$$

$$\propto \pi(\theta) \pi(\phi | \theta) \times \prod_{i=1}^{\ell} \prod_{j=0}^{n_i-1} \prod_{k=0}^{m(i)-1} \pi \left( x^{(i)}_{\tau_{j,(k+1)}}, x^{(i)}_{\tau_{j,k}}, \theta, \phi^{(i)} \right)$$
The (Modified) Innovation scheme conditions on the Brownian increments

\[ w = \left( w^{(1)}, w^{(2)}, \ldots, w^{(\ell)} \right) \]

(which drive a tractable conditioned diffusion) to overcome the dependence between the parameters and the path.

The posterior distribution is typically analytically intractable.

Use a Gibbs sampler, alternating between draws of

- \( \theta|w, d, \phi \)
- \( \phi|w, d, \theta \)
- \( w|\theta, d, \phi \)

See previous SBSSB talks for details on these updates.
We take observations on $N$ for 20 populations at intervals of 0.25, giving us partial observations.

Parameter choice:

$$(\lambda, \mu, \sigma_\lambda, \sigma_\mu)^T = (\log 1.75, \log 0.00095, 0.2, 0.1)^T$$

$$\lambda^{(i)} \sim LN\left(\log 1.75, 0.2^2\right)$$

$$\mu^{(i)} \sim LN\left(\log 0.00095, 0.1^2\right)$$

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

Figure: **Black:** (Modified) Innovation scheme, $m = 20$

**Red:** LNA

![Graphs showing density plots for $\lambda$, $\mu$, $\sigma_\lambda$, and $\sigma_\mu$.]

- **$\lambda$**: Density plots for $\lambda$ with black and red curves, illustrating the distribution of $\lambda$.
- **$\mu$**: Density plots for $\mu$ with black and red curves, illustrating the distribution of $\mu$.
- **$\sigma_\lambda$**: Density plots for $\sigma_\lambda$ with black and red curves, illustrating the distribution of $\sigma_\lambda$.
- **$\sigma_\mu$**: Density plots for $\sigma_\mu$ with black and red curves, illustrating the distribution of $\sigma_\mu$. 

Gavin Whitaker  
SBSSB: 5th March 2014  14/33
Figure: **Black:** (Modified) Innovation scheme, $m = 20$

**Red:** LNA

\[ \begin{align*}
\lambda^{13} & \quad \text{Density} \\
0.00095 & \quad 0.00100 & \quad 0.00105 & \quad 0.00110 \\
0 & \quad 5000 & \quad 15000 \\
\mu^{13} & \quad \text{Density} \\
0.00095 & \quad 0.0010 & \quad 0.00105 & \quad 0.0011 \\
0 & \quad 5000 & \quad 15000
\end{align*} \]
Figure: **Black:** (Modified) Innovation scheme, $m = 20$

**Red:** LNA

- **Top Diagram:** Graph showing density against $\lambda^1$ with two curves.
- **Bottom Diagram:** Graph showing density against $\mu^1$ with two curves.
What’s going wrong!!!
How was the mixing?

Auto-correlation for $\lambda^i$

- Innovation scheme, $m=20$
- LNA

- Population 1
- Population 6
- Population 11
- Population 13
Initial exploratory analysis

How was the mixing?

Auto-correlation for $\mu_i$

- Innovation scheme, $m=20$
- LNA
- Population 1
- Population 6
- Population 11
- Population 13
Initial exploratory analysis

How was the mixing?
- Over some paths the scheme mixes well
- Over others, it does **not**!

Any problematic data traces?
- Looked at the data traces corresponding to the worst mixing
- These data traces appear to be those which go closest to 0

Thoughts
- What do the latent paths look like as observations approach 0?
- What is the acceptance probability over these intervals?
- Let’s consider a problem data trace...
Figure: Skeleton path for data trace 1, $\lambda^{(1)} = 2.497$, $\mu^{(1)} = 0.001$
Figure: Skeleton path for data trace 1
Consider an interval of length $2m + 1$

See my SBSSB talk on 5th December 2012 for details on the bridge

In the following we consider fixed parameters
Figure: 10 000 iterations, the acceptance probability is 0.1568
Figure: 10 000 iterations, the acceptance probability is 0.1935
Figure: 10 000 iterations, the acceptance probability is 0.1053
Figure: 10 000 iterations, the acceptance probability is 0.0147
As we approach observations around 0, the path acceptance probability declines and we observe a radical drop if the interval contains an observation very close to 0.

This is with fixed parameters, when we also update the parameters (along with the path) the acceptance probability over an interval will worsen!

As the process can’t go negative (but the numerical solution can) we reject any path that goes negative by ascribing a likelihood of 0.

⇒ the scheme becomes degenerate as the path is not updated.
Possible solutions

▷ Why don’t we log transform the process?
▷ Using the multivariate Itô formula where $Y_t^{(i)} = \log \left( X_t^{(i)} \right)$:

\[
dY_t^{(i)} = \left( \lambda_t^{(i)} - \mu_t^{(i)} \exp \left( Y_{2,t}^{(i)} \right) - \frac{1}{2} \exp \left( -Y_{1,t}^{(i)} \right) \left[ \lambda_t^{(i)} + \mu_t^{(i)} \exp \left( Y_{2,t}^{(i)} \right) \right] \right) dt \\
+ \left( \lambda_t^{(i)} \exp \left( -Y_{1,t}^{(i)} \right) + \mu_t^{(i)} \exp \left( Y_{2,t}^{(i)} - Y_{1,t}^{(i)} \right) \right) \left( \lambda_t^{(i)} \exp \left( -Y_{2,t}^{(i)} \right) \right)^{1/2} dW_t^{(i)}
\]

▷ What do the skeleton paths look like under the log SDE?
▷ Using the same $\lambda_t^{(i)}$ and $\mu_t^{(i)}$ as above, we get ...
Figure: Skeleton paths for 20 Aphid populations, log SDE
We observe similar patterns in the path acceptance probability under the log SDE as we observed under the SDE.

Using the log SDE also caused the ODE solver used in the LNA to break.

What about multiple path updates each iteration?

We can update problem intervals (those with a low acceptance probability) multiple times, such that there is a reasonable chance the path will be updated.

We run the (Modified) Innovation scheme with $m = 20$ (implementing 50 path updates per parameter update) and the LNA for the conditions described above.
Figure: **Black:** (Modified) Innovation scheme, $m = 20$, 50 path updates  
**Red:** LNA
Future work

- Apply these methods to actual Aphid data
  - Added challenge - efficient sampling of conditioned diffusions
  - To bridge the distance between the real observations we introduce latent values

\[
\begin{array}{cccccc}
N_0 & N_1 \\
\bullet & \times & \times & \times & \bullet & \ldots
\end{array}
\]

- We then bridge between these latent values to give us a path between the real observations
- So that these latent values do not remain fixed we need to slightly alter our bridging scheme to include 2 passes

- Compare this new bridging scheme against the LNA for the actual Aphid data
Tease - a new bridging scheme

Pass 1:

Pass 2:

D&G bridge

partial D&G bridge


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