

MCMC schemes for partially observed diffusions: challenges and solutions

Gavin Whitaker

5th March 2014

Supervisors: RJB and AG

- ▷ 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 $\{\mathbf{X}_t, t \geq 0\}$ satisfying

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

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

- ▷ Euler-Maruyama approximation:

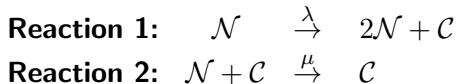
$$\Delta\mathbf{X}_t \equiv \mathbf{X}_{t+\Delta t} - \mathbf{X}_t = \boldsymbol{\alpha}(\mathbf{X}_t, \boldsymbol{\theta})\Delta t + \sqrt{\boldsymbol{\beta}(\mathbf{X}_t, \boldsymbol{\theta})}\Delta\mathbf{W}_t$$

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

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

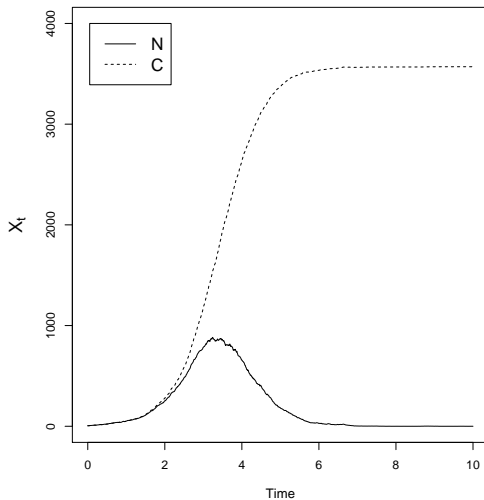


The mass action SDE representation of the system dynamics is

$$\begin{pmatrix} dN_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \lambda N_t - \mu N_t C_t \\ \lambda N_t \end{pmatrix} dt + \begin{pmatrix} \lambda N_t + \mu N_t C_t & \lambda N_t \\ \lambda N_t & \lambda N_t \end{pmatrix}^{\frac{1}{2}} d\mathbf{W}_t$$

- ▷ 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, $\lambda = 1.75$, $\mu = 0.001$



Random effects SDE models

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

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

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

- ▶ Random effects distributions depend on $\boldsymbol{\theta}$
- ▶ Suppose we have data available at times $t_0^{(i)}, t_1^{(i)}, \dots, t_{n_i}^{(i)}$ for each individual i

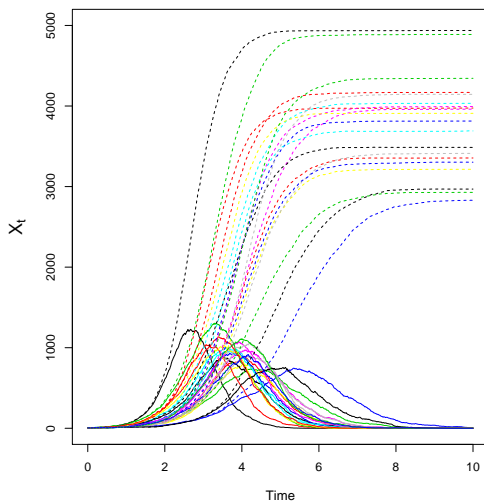
Aphid growth model - random effects

SDE for population i

$$d\mathbf{X}_t^{(i)} = \begin{pmatrix} \lambda^{(i)} N_t^{(i)} - \mu^{(i)} N_t^{(i)} C_t^{(i)} \\ \lambda^{(i)} N_t^{(i)} \end{pmatrix} dt + \begin{pmatrix} \lambda^{(i)} N_t^{(i)} + \mu^{(i)} N_t^{(i)} C_t^{(i)} & \lambda^{(i)} N_t^{(i)} \\ \lambda^{(i)} N_t^{(i)} & \lambda^{(i)} N_t^{(i)} \end{pmatrix}^{\frac{1}{2}} d\mathbf{W}_t^{(i)}$$

- ▶ $\lambda^{(i)} \sim LN(\lambda, \sigma_\lambda^2)$ and $\mu^{(i)} \sim LN(\mu, \sigma_\mu^2)$
- ▶ Common parameters: $\boldsymbol{\theta} = (\lambda, \mu, \sigma_\lambda, \sigma_\mu)^T$
- ▶ Population specific parameters: $\boldsymbol{\phi}^{(i)} = (\lambda^{(i)}, \mu^{(i)})^T$
- ▶ We look at a simulation study involving 20 populations

Figure: Skeleton paths for 20 Aphid populations



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

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

- ▶ Time step between observations

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

- ▶ Allows for irregularly spaced data for each individual

- ▷ For individual i

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

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

- ▷ Putting these together for ℓ individuals

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

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

- ▷ Formulate joint posterior for parameters and latent data as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x} | \mathbf{d}) \propto \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\phi} | \boldsymbol{\theta}) \pi(\mathbf{x}, \mathbf{d} | \boldsymbol{\theta}, \boldsymbol{\phi})$$

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

(Modified) Innovation scheme for random effects SDEs

- ▶ The (Modified) Innovation scheme conditions on the Brownian increments

$$\mathbf{w} = \left(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots, \mathbf{w}^{(\ell)} \right)$$

(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
 - $\theta | \mathbf{w}, \mathbf{d}, \phi$
 - $\phi | \mathbf{w}, \mathbf{d}, \theta$
 - $\mathbf{w} | \theta, \mathbf{d}, \phi$
- ▶ See previous SBSSB talks for details on these updates

Application: Aphid growth

- ▶ 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)

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

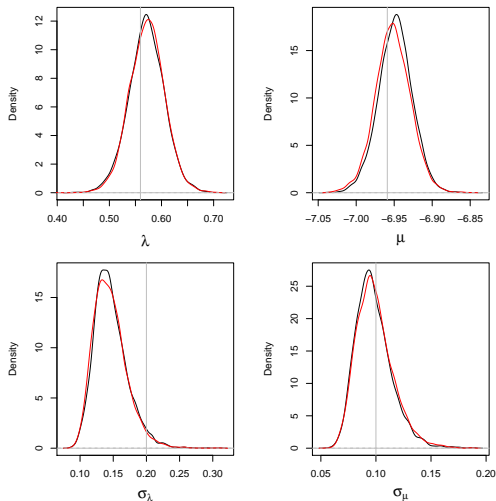


Figure: **Black:** (Modified) Innovation scheme, $m = 20$
Red: LNA

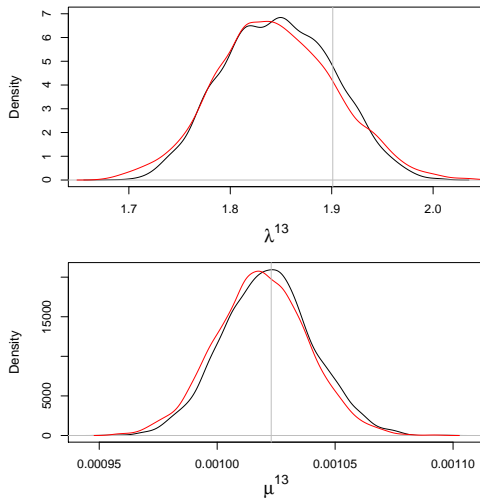
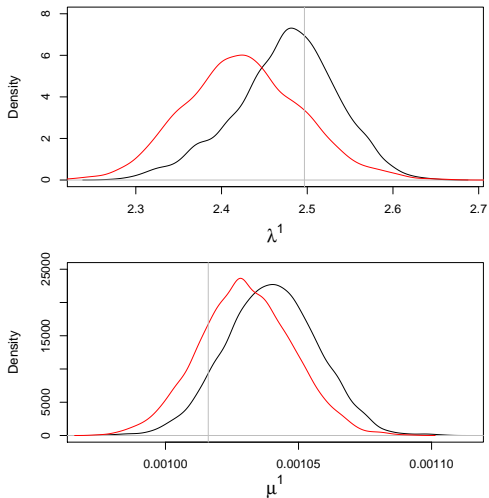


Figure: **Black:** (Modified) Innovation scheme, $m = 20$
Red: LNA

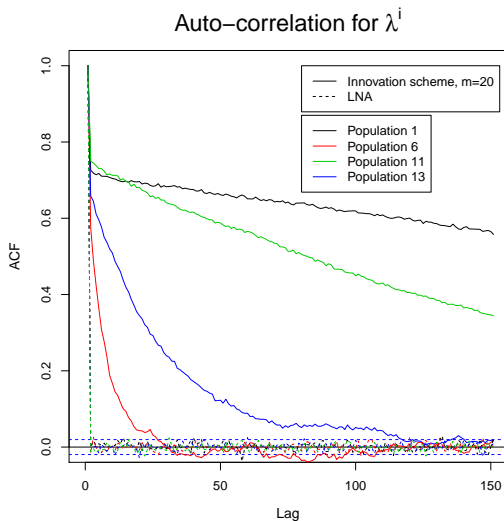


What's going wrong!!!



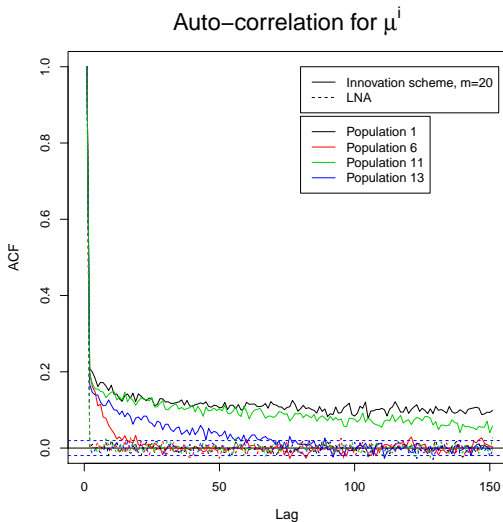
Initial exploratory analysis

How was the mixing?



Initial exploratory analysis

How was the mixing?



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...

Approaching zero - data trace 1

Figure: Skeleton path for data trace 1, $\lambda^{(1)} = 2.497$, $\mu^{(1)} = 0.001$

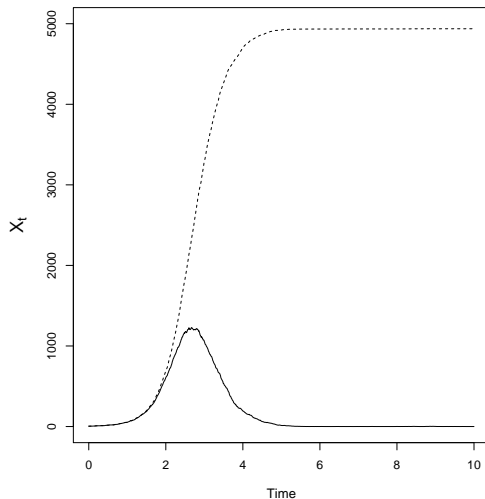
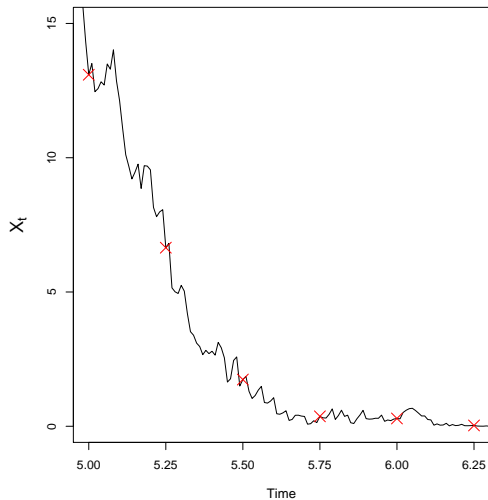
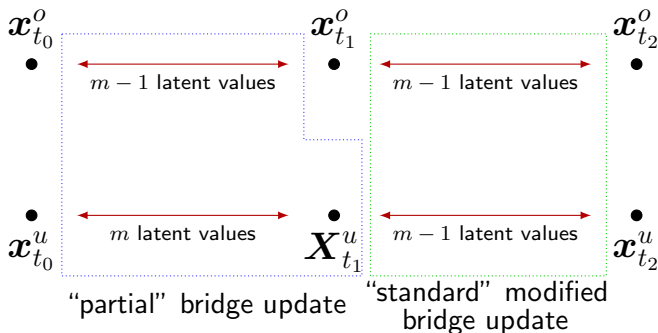


Figure: Skeleton path for data trace 1



The bridging process

- ▶ 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

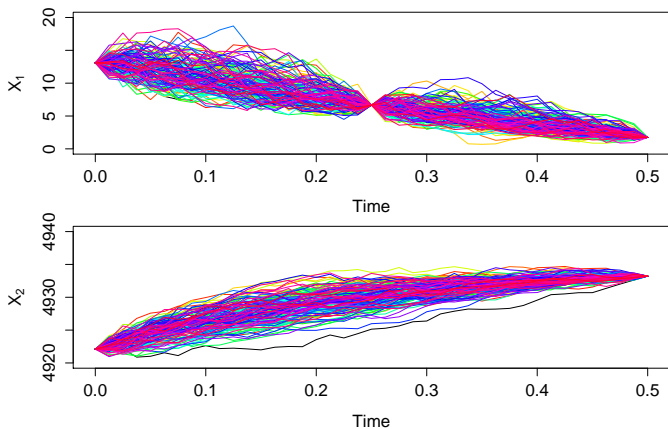


Figure: 10 000 iterations, the acceptance probability is 0.1935

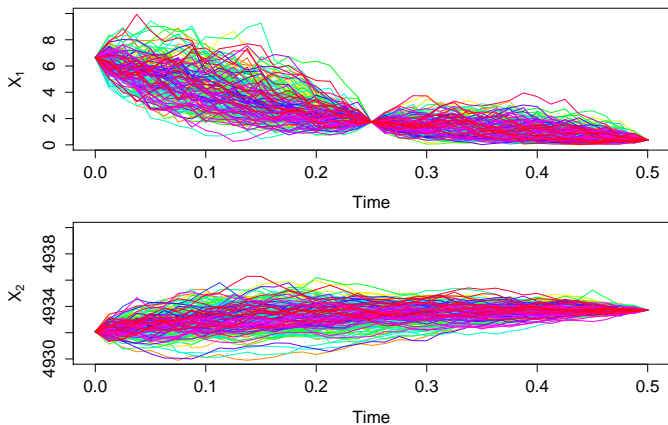


Figure: 10 000 iterations, the acceptance probability is 0.1053

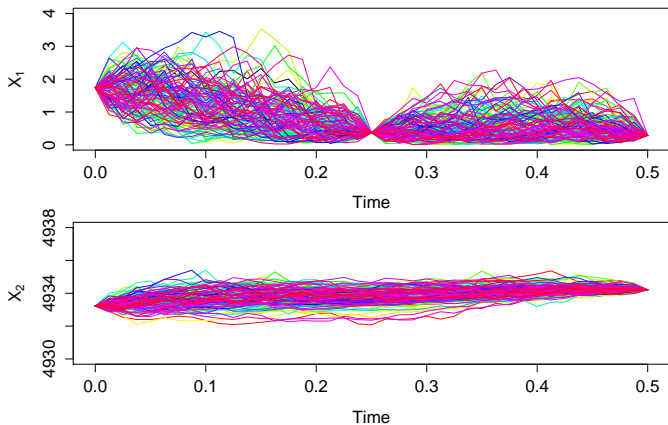
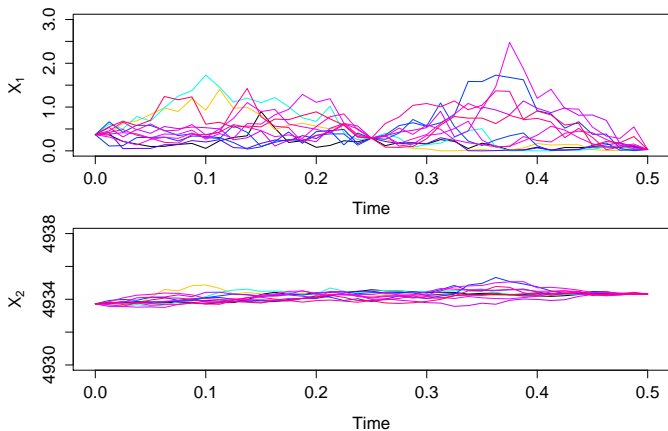


Figure: 10 000 iterations, the acceptance probability is 0.0147



- ▷ 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

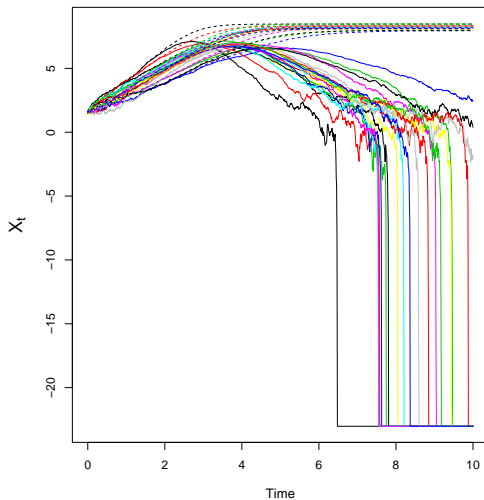
Possible solutions

- ▶ Why don't we log transform the process?
- ▶ Using the multivariate Itô formula where $\mathbf{Y}_t^{(i)} = \log(\mathbf{X}_t^{(i)})$:

$$d\mathbf{Y}_t^{(i)} = \begin{pmatrix} \lambda^{(i)} - \mu^{(i)} \exp(Y_{2,t}^{(i)}) - \frac{1}{2} \exp(-Y_{1,t}^{(i)}) [\lambda^{(i)} + \mu^{(i)} \exp(Y_{2,t}^{(i)})] \\ \lambda^{(i)} \exp(Y_{1,t}^{(i)} - Y_{2,t}^{(i)}) [1 - \frac{1}{2} \exp(-Y_{2,t}^{(i)})] \end{pmatrix} dt \\ + \begin{pmatrix} \lambda^{(i)} \exp(-Y_{1,t}^{(i)}) + \mu^{(i)} \exp(Y_{2,t}^{(i)} - Y_{1,t}^{(i)}) & \lambda^{(i)} \exp(-Y_{2,t}^{(i)}) \\ \lambda^{(i)} \exp(-Y_{2,t}^{(i)}) & \lambda^{(i)} \exp(Y_{1,t}^{(i)} - 2Y_{2,t}^{(i)}) \end{pmatrix}^{\frac{1}{2}} d\mathbf{W}_t^{(i)}$$

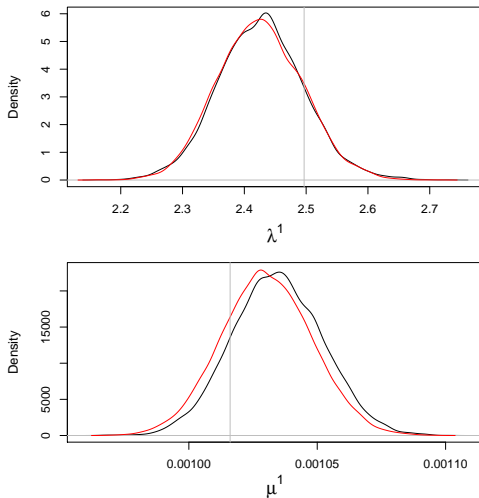
- ▶ What do the skeleton paths look like under the log SDE?
- ▶ Using the same $\lambda^{(i)}$ and $\mu^{(i)}$ as above, we get ...

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



- ▷ 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

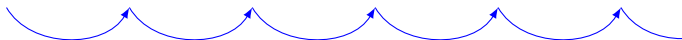


- We then bridge between these latent values to give us a path between the real observations
 - So that these latent values do not remain fixed we need to slightly alter our bridging scheme to include 2 passes
- ▷ Compare this new bridging scheme against the LNA for the actual Aphid data

Tease - a new bridging scheme



Pass 1:



Pass 2:



updates

C_1



D&G bridge



partial D&G bridge

- ▶ Durham, G. B. and Gallant, A. R. *Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes*. Journal of Business and Economic Statistics, 20 297-338, 2001
- ▶ Fearnhead, P. Giagos, V. and Sherlock, C. *Inference for reaction networks using the linear noise approximation*. <http://arxiv.org/abs/1205.6920>, 2012
- ▶ Gillespie, C. S. and Golightly, A. *Bayesian inference for generalised stochastic population growth models with application to aphids*. JRSS Series C, Applied Statistics, 59 (2) 341-357, 2010
- ▶ 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