Bayesian Inference for Stochastic Kinetic Models

Scott Phillips

Durham University

1. Reaction Network

A reaction network [1] is comprised of a set of reactants and products X_i, connected via a set of reactions θ_i. The corresponding network for Lotka-Volterra (LV) with a predator-prey interaction term incorporated is shown below along with a plot for the system:

$$X_1 \xrightarrow{\theta_1} 2X_1, \qquad X_2 \xrightarrow{\theta_2} \emptyset, \qquad X_1 + X_2 \xrightarrow{\theta_3} 2X_2.$$

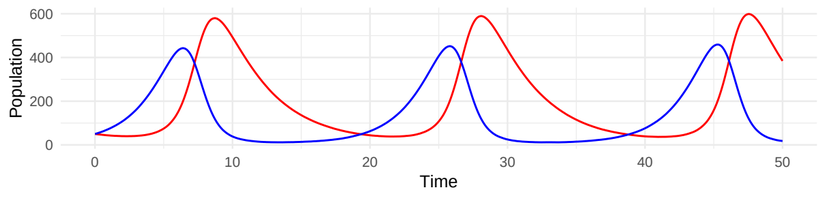

Figure 1. LV Dynamics with parameters $\theta = (0.5, 0.0025, 0.3)$, initialise at $(50, 50)$, blue = X_1, red = X_2.

Where X_1 is the prey and X_2 is the predator. Reaction networks can then be used to construct ODEs for each of the reactants. For the example above the corresponding ODEs are shown below:

$$\frac{dX_1}{dt} = \theta_1 X_1 - \theta_3 X_1 X_2, \qquad \frac{dX_2}{dt} = \theta_3 X_1 X_2 - \theta_2 X_2.$$

Now with some rearrangement, we can generalise these ODEs to SDEs taking the following from:

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} \theta_1 X_1 - \theta_3 X_1 X_2 \\ \theta_3 X_1 X_2 - \theta_2 X_2 \end{bmatrix} dt + \begin{bmatrix} \theta_1 X_1 + \theta_3 X_1 X_2 & \theta_3 X_1 X_2 \\ \theta_3 X_1 X_2 & \theta_3 X_1 X_2 + \theta_2 X_2 \end{bmatrix}^{\frac{1}{2}} dW_t$$

Where W_t represents the Wiener process, a process whose increment over a time interval of length dt are distributed by $N(0, I \cdot dt)$.

2. Euler-Maruyama Approximation

The previously defined SDE is analytically intractable, resulting in any further analysis of the system being impossible analytically. Fortunately, we can use numerical methods to approximate this SDE. The Euler-Maruyama approximation [3] for the system described above is shown to be:

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} X_{1,s} + \theta_1 X_{1,s} - \theta_3 X_{1,s} X_{2,s} \\ X_{2,s} + \theta_3 X_{1,s} X_{2,s} - \theta_2 X_{2,s} \end{bmatrix} (t-s) + \begin{bmatrix} \theta_1 X_{1,s} + \theta_3 X_{1,s} X_{2,s} & \theta_3 X_{1,s} X_{2,s} \\ \theta_3 X_{1,s} X_{2,s} & \theta_3 X_{1,s} X_{2,s} + \theta_2 X_{2,s} \end{bmatrix}^{\frac{1}{2}} dW_{t-s}$$

This is valid for any $t > s$, $\forall t, s \in T$ where T is the total time of the process. Unfortunately, the EM process is inadequate for large time steps.

References

[1] Andrew Golightly and Colin S Gillespie. *Simulation of Stochastic Kinetic Models*, pages 169–187. Humana Press, 2013.

[2] Andrew Golightly and Chris Sherlock. Augmented pseudo-marginal metropolis–hastings for partially observed diffusion processes. *Statistics and Computing*, 32(1):21, 2022.

[3] Lamba H, Mattingly JC, and Stuart AM. An adaptive euler–maruyama scheme for sdes: Convergence and stability. *IMA Journal of Numerical Analysis*, 27(3):479–506, 2006.

[4] Gavin A Whitaker, Andrew Golightly, Richard J Boys, and Chris Sherlock. Improved bridge constructs for stochastic differential equations. *Statistics and Computing*, 27(4):1061–1077, 2017.

[5] Jun Yang, Gareth O Roberts, and Jeffrey S Rosenthal. Optimal scaling of random-walk metropolis algorithms on general target distributions, 2020.

3. Inference Problem

The aim of using stochastic simulation for Stochastic Kinetic Models is estimation of the rate parameters given we have some partially observed data. This is done by simulating draws from the desired posterior $\pi(\theta)$. The nuance being that this posterior is not available to be simulated from directly. A solution to this is utilise a proposal distribution that can be simulated from directly to solve this. Parameter estimations can then be produced on the realisations from the posterior. This is done via MCMC algorithms.

One such algorithm we will be using is the Pseudo-Marginal Metropolis-Hastings algorithm [2]. This is distinct from other Metropolis-Hastings algorithms as this case is applied when the observed data's likelihood is intractable and hence must be approximated. This can be solved by producing an unbiased estimate for the likelihood via an unbiased auxiliary function u approximated with density $g(\cdot)$.

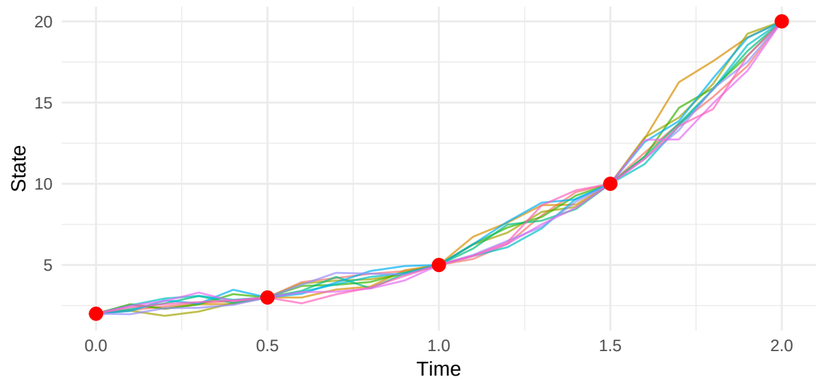

Figure 2. Durham and Gallant bridges between synthetic observed data points with M = 10.

To allow for the use of the EM approximation to the LV model over the time steps chosen by the practitioner, we partition each interval $[t, t+1]$ as $t = t_0 < t_1 < ... < t_M = t + 1$ to improve the EM approximation. Then integrate over the partitions via importance sampling through the use of a bridge construct, used as a proposal mechanism. This approximates the transition densities of the form $p(x_{t+1}|x_t, \theta)$.

4. Psuedo-Marginal Metropolis-Hastings Algorithm

PMMH Algorithm with N Iterations

1. Initialise the chain to $\theta^{(0)} = (\theta_1^{(0)}, ..., \theta_d^{(0)})$ somewhere in the support of $\pi(\theta)$. Draw an inital auxiliary variable $u^{(0)} \sim g(\cdot)$ and compute the noisy target $\hat{\pi}_{u^{(0)}}(\theta^{(0)}) = \pi(\theta^{(0)})u^{(0)}$. Set iteration counter to $j = 1$.

2. Generate a proposed value $\theta^* \sim q(\cdot|\theta)$. Independently draw a new auxiliary variable $u^* \sim g(\cdot)$ and compute a new noisy target $\hat{\pi}_{u^*}(\theta^*) = \pi(\theta^*)u^*$.

3. Evaluate the acceptance probability:

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

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

5. If $j = N$ stop, otherwise put j to $j+1$ and go to step 2.

5. Application

Now applying the PMMH algorithm to the LV model we must specify:

- First, for ease of notation, we take $\alpha(X_t, \theta)$ and $\beta(X_t, \theta)$ to be the drift and diffusion processes respectively for the LV model .
- The target density is taken to be the LV model:

$$p(x_{t_1}, ..., x_{t_M}|x_{t_0}) = \prod_{i=0}^{M-1} Pe(x_{t_{i+1}}|x_{t_i}, \theta)$$
$$\text{Where } Pe(x_{t_{i+1}}|x_{t_i}, \theta) \sim N(x_{t_{i+1}}; x_{t_i} + \alpha(x_{t_i}, \theta)\Delta t, \beta(x_{t_i}, \theta)\Delta t).$$

- The proposal density is taken to be a Durham and Gallant bridge construct [4]

$$q(x_{t_1}, ..., x_{t_{M-1}}|x_{t_0}, x_{t_M}) = \prod_{i=0}^{M-2} q(x_{t_{i+1}}|x_{t_i}, x_{t_M}, \theta)$$
$$\text{Where } q(x_{t_{i+1}}|x_{t_i}, x_{t_M}, \theta) \sim N(x_{t_{i+1}}; x_{t_i} + \frac{x_{t_M} - x_{t_i}}{t_M - t_i}\Delta t, \frac{t_M - t_{i+1}}{t_M - t_i}\beta(x_{t_i}, \theta)\Delta t).$$

- The initial conditions and parameters for this are chosen to be:
  - Observed data is synthetically generated from a LV model with parameters $\theta = (0.5, 0.0025, 0.3)$, results in 10 observed values.
  - Gaussian prior with mean $\theta = (0.7, 0.1, 0.5)$ and variance $0.5 \cdot I_3$.
  - Initial runs variance for theta proposal mechanism = $0.2 \cdot I_3$, the resulting tuning run (Figure 3) is done with the variance of the previous run multiplied by 0.03 [5].
  - Number of bridges = 30.
  - Inter observation time = 0.01.
  - Time simulated over per observed value = 0.1.
  - Number of simulations = 10000.

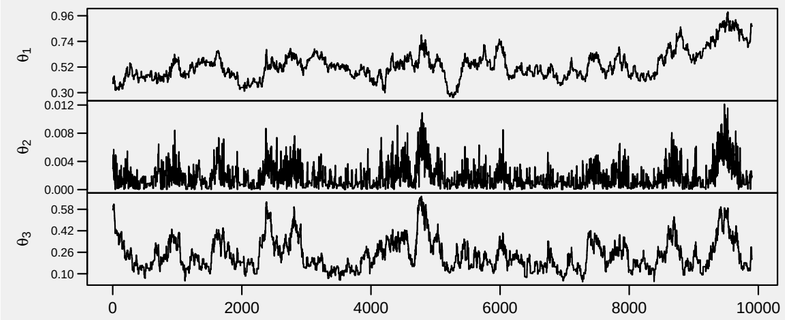

Figure 3. Trace plots for parameters after tuning and discarding 300 burn-in iterations.

6. Further Work

All work that precedes this has been done under the assumption of partially observed data values for **both** species. However, what if we were only able to record one of the species evolution in time. With this in mind we wish to carry out the same PMMH algorithm with the addition of a new proposal mechanism employed for the unobserved species. The end result should be a selection of proposed data points for the unobserved species and the corresponding parameter estimations.

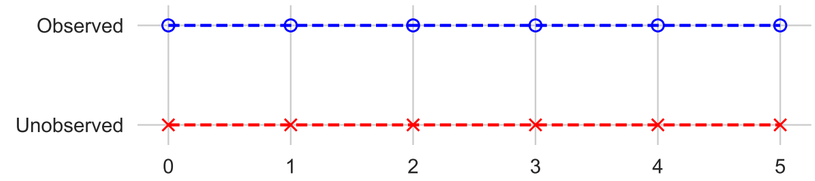

Figure 4. Visualisation of observed data (circles), unobserved data (crosses) and the further partitions of the data represented by dashes.

Supervisor: Professor Andrew Golightly