Supplemental material

Moment closure based parameter inference of stochastic kinetic models

Authors:
Peter Milner, Colin S. Gillespie and Darren J. Wilkinson

S.1 Updating the latent predator state

To update the latent predator state Y_2, we use a Metropolis-Hastings step. We first initialise the unobserved data to a valid sample path from the set of all plausible sample paths, for this problem we sampled from a uniform $U(1, 500)$ to initialise the latent data points. To update the data we use the following algorithm for each unobserved data point $Y_2(t_i)$:

1. Propose
$$Y_2(t_i|t_{i-1}, t_{i+1})^* \sim N(\mu_i^\dagger, \sigma_i^\dagger) \tag{1}$$
where
$$\mu_i^\dagger = \frac{1}{2}[Y_2(t_{i-1}) + Y_2(t_{i+1})] \quad \text{and} \quad \sigma_i^\dagger = \frac{1}{2}\text{Var}[Y_2(t_i)] \ .$$
Since our observations are on a regularly spaced grid, this proposal uses the average of the neighbouring data points. If observations were irregularly spaced, then we could use a simple linear interpolation scheme. The value $\text{Var}[Y_2(t_i)]$ is estimated from the the moment equations for time t_{i-1}.

2. Accept $Y_2(t_i)^*$ with probability $\alpha = \min\{1, a\}$ where
$$a = \frac{\pi[Y_2(t_i)^* | \, Y_2(t_{i-1}), Y_2(t_{i+1})]}{\pi[Y_2(t_i) | \, Y_2(t_{i-1}), Y_2(t_{i+1})]} \times \frac{q[Y_2(t_i) | \, Y_2(t_{i-1}), Y_2(t_{i+1})]}{q[Y_2(t_i)^* | \, Y_2(t_{i-1}), Y_2(t_{i+1})]}$$

where

$$\begin{aligned}\pi[Y_2(t_i)^* | \, Y_2(t_{i-1}), Y_2(t_{i+1})] &= p[Y_2(t_i)^* | \, Y_2(t_{i-1})] \, p[Y_2(t_{i+1}) | \, Y_2(t_i)^*] \\ &= \phi(Y_2(t_i)^*; \psi_{i-1}, \Sigma_{i-1}) \phi(Y_2(t_{i+1}); \psi_i^*, \Sigma_i^*)\end{aligned}$$

and

$$q[Y_2(t_i) | \, Y_2(t_{i-1}), Y_2(t_{i+1})] = \phi(Y_2(t_i); \mu_i^\dagger, \sigma_i^\dagger) \ .$$

S.2 Prokaryotic auto regulatory gene network

This is a simple model of a prokaryotic auto regulation network taken from Golightly and Wilkinson, 2006. It contains twelve reactions and six species. In this model a protein $\mathtt{I}$ coded for by a gene $\mathtt{i}$ represses its own transcription and also the transcription of another gene, $\mathtt{g}$ by binding to a regulatory region upstream of the gene. This can be simplified to the following reactions

$$R_1 : \ \mathtt{I} + \mathtt{i} \rightarrow \mathtt{I} \cdot \mathtt{i}, \qquad R_2 : \ \mathtt{I} \cdot \mathtt{i} \rightarrow \mathtt{I} + \mathtt{i},$$
$$R_3 : \ \mathtt{I} + \mathtt{g} \rightarrow \mathtt{I} \cdot \mathtt{g}, \qquad R_4 : \ \mathtt{I} \cdot \mathtt{g} \rightarrow \mathtt{I} + \mathtt{g}.$$

The transcription of $\mathtt{i}$ and $\mathtt{g}$ and the translation of mRNA $\mathtt{r_i}$ and $\mathtt{r_g}$ are represented by

$$R_5 : \ \mathtt{i} \rightarrow \mathtt{i} + \mathtt{r_i}, \qquad R_6 : \ \mathtt{r_i} \rightarrow \mathtt{r_i} + \mathtt{I},$$
$$R_7 : \ g \rightarrow \mathtt{g} + \mathtt{r_g}, \qquad R_8 : \ \mathtt{r_g} \rightarrow \mathtt{r_g} + \mathtt{G}.$$

We also have mRNA degradation

$$R_9 : \ \mathtt{r_i} \rightarrow \emptyset, \qquad R_{10} : \ \mathtt{r_g} \rightarrow \emptyset,$$

and protein degradation

$$R_{11} : \ \mathtt{I} \rightarrow \emptyset, \qquad R_{12} : \ \mathtt{G} \rightarrow \emptyset.$$

Each reaction i has a stochastic rate constant c_i. There are two conservation laws in the model

$$\mathtt{I} \cdot \mathtt{i} + \mathtt{i} = K_1, \qquad \mathtt{I} \cdot \mathtt{g} + \mathtt{g} = K_2,$$

where K_1 and K_2 are conservation constants. If K_1 and K_2 are known, then we can simplify the model using the conservation laws to remove $\mathtt{I} \cdot \mathtt{i}$ and $\mathtt{I} \cdot \mathtt{g}$. This simplification reduces the model to six dimensions

$$\mathbf{y} = (\mathtt{I}, \mathtt{G}, \mathtt{i}, \mathtt{g}, \mathtt{r_i}, \mathtt{r_g})^{\mathsf{T}} .$$

The reaction hazards for R_1 and R_2 are $h_1(\mathbf{y}, c_1) = c_1 \mathtt{I}\mathtt{i}$ and $h_2(\mathbf{y}, c_2) = c_2 \mathtt{I} \cdot \mathtt{i} = c_2(K_1 - \mathtt{i})$ respectively. Hazards for R_3 and R_4 are calculated similarly. The remaining hazards are first order reactions.

The CME for the auto regulatory gene network is

$$\frac{d}{dt} p(\mathbf{y}; t) = \sum_{i=1}^{12} \{ h_i(\mathbf{y} - A_i^{\mathsf{T}}, c_i) p(\mathbf{y} - A_i^{\mathsf{T}}; t) - h_i(\mathbf{y}, c_i) p(\mathbf{y}; t) \}, \qquad (2)$$

where

$$A^{\mathsf{T}} = \begin{pmatrix} -1 & 1 & -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \end{pmatrix}.$$

A is known as the reaction matrix and each column of is $\mathbf{s_L}$. See section 2.1 of the main paper for further details.

Multiplying equation (2) by the generating function (3) and extracting coefficients of θ gives the moment equations. Assuming an underlying Gaussian distribution gives us a closed set of twenty-seven ODEs. Six ODEs for the means, six ODEs for the variances and fifteen ODEs for the covariance terms.