Parameter inference for small biochemical systems using likelihood-free MCMC

Eric Kernfeld[1]

[1]University of Washington Department of Statistics

Paper

Darren Wilkinson's "Parameter inference for stochastic kinetic models of bacterial gene regulation," a chapter of the proceedings in [1].

Cast of characters

- Simulation begins at time 0 and proceeds in continuous time.
- $\mathcal{R}_j$ is a chemical reaction.
- $R_j(t) \in \mathbb{Z}_{\geq 0}$ counts occurrences of $\mathcal{R}_j$ up to time t.
- $\theta_j \in \mathbb{R}_{\geq 0}$ is the rate at which $\mathcal{R}_j$ happens. More on this in future slides.
- $X_i(t) \in \mathbb{Z}_{\geq 0}$ is the number of molecules of type i
- $\mathcal{D}_t \in \mathbb{R}$ is an incomplete observation of $X(t)$ with error.

Wilkinson's eXample–reaction requirement

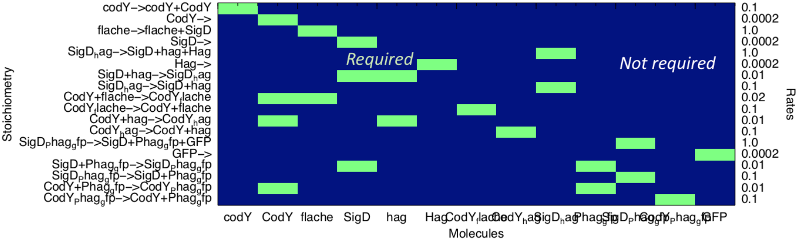


Figure : Roles of chemicals in various reactions. Green means the molecule is involved and blue means no involvement.

$$R_1(t) \sim PP(\theta_1 X_1(t))$$

$$R_5(t) \sim PP(\theta_5 X_4(t) X_5(t))$$

Wilkinson's eXample–net change

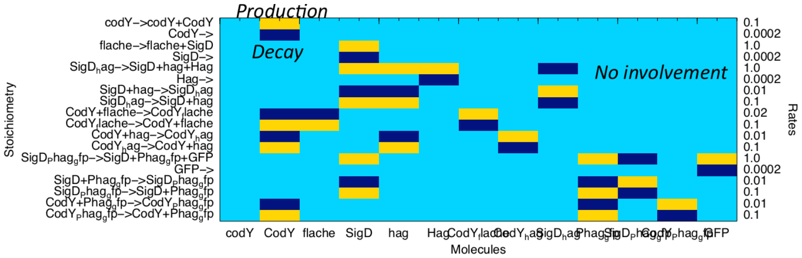


Figure : Roles of chemicals in various reactions. Green means the molecule is involved and blue means no involvement.

$$R_1(t) \sim PP(\theta_1 X_1(t))$$

$$R_5(t) \sim PP(\theta_5 X_4(t) X_5(t))$$

Some notes

- Assume reaction $\mathcal{R}_j$ occurs independently of the others (given the state).
- Priors on θ and $X(0)$, plus the dynamics already mentioned, determine the whole system.
- EXact forward simulations are easy.

Forward simulation

Given:

- Duration T and initial particle counts $X(0)$
- $S_{i,j}$, net change in molecules of type i in a reaction of type j, and
- $P_{i,j}$ number of molecules of type i entering a reaction of type j
- θ, a vector of reaction rates

Do this:

- Initialize X to $X(0)$ and t to 0.
- While true:
  - Calculate $\alpha_j = \theta_j \prod_i \binom{X_i}{P_{ij}}$
  - Increment t by EXponential(rate=$\sum_j \alpha_j$)
  - If $t > T$, quit and return X.
  - Otherwise, choose an integer j with probability $\frac{\alpha_j}{\sum_j \alpha_j}$.
  - Increment X by adding column j of S.

Likelihoods

Likelihood if reaction ν_i happens at t_i:

$$\prod_{i=1}^{\text{events}} \theta_{\nu_i} \prod_{j=1}^{\text{rXn types}} \binom{X_j(t_{i-1})}{p_{\nu_i j}} \exp\left(-\theta_{\nu_i}(t_i - t_{i-1}) \binom{X_j(t_{i-1})}{p_{\nu_i j}}\right)$$

Likelihood from observing $X(t_i)$. Sum is over "eligible" paths for ν and integral is over a simpleX of possible wait-time tuples.

$$\sum_{\nu} \int_t \prod_{i=1}^{\text{events in } \nu} \theta_{\nu_i} \prod_{j=1}^{\text{rXn types}} \binom{X_j(t_{i-1})}{p_{\nu_i j}} \exp\left(\text{same as above}\right).$$

Angles of attack

Likelihood

$$\sum_{\nu} \int_{t} \prod_{i=1}^{\text{events in } \nu} \theta_{\nu_i} \prod_{j=1}^{\text{rXn types}} \binom{X_j(t_{i-1})}{p_{\nu_i j}} \exp\left(-\theta_{\nu_i}(t_i - t_{i-1}) \binom{X_j(t_{i-1})}{p_{\nu_i j}}\right)$$

You might try:

- EM–still requires nasty sum.
- Metropolis-Hastings on θ–still requires likelihood evaluation.
- Gibbs sampling–how do you update ν?
- Rejection sampling–most samples get rejected.

Solutions will need to either approXimate the likelihood or avoid it completely.

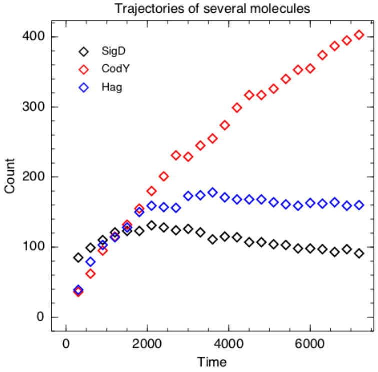
A complication: measurement error

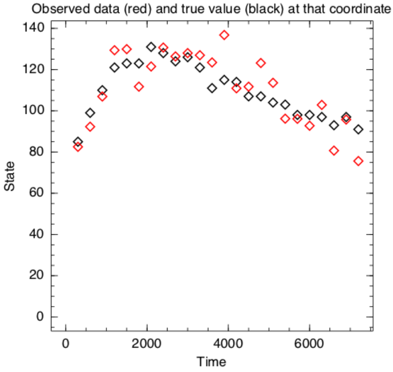
If reactions occur at times t_i and observations occur at times s_k, the likelihood becomes a sum over possible values of X at times s_0, s_1, ... $s_{(\text{num obs})}$ of:

$P(\mathcal{D}|X)\times$

$$\sum_{\nu}\int_{t}\prod_{i=1}^{\text{events in }\nu}\theta_{\nu_i}\prod_{j=1}^{\text{rXn types}}\binom{X_j(t_{i-1})}{p_{\nu_i j}}\exp\left(-\theta_{\nu_i}(t_i-t_{i-1})\binom{X_j(t_{i-1})}{p_{\nu_i j}}\right.$$

Another view of the problem





Likelihood free MCMC intro–the M-H recipe

To produce a chain of samples from $P(\theta|D)$, using a proposal $q(\theta^*|\theta)$, accept with probability $p_{rej}(\theta^*|\theta) \equiv \min\{1, A\}$ if

$$A = \frac{q(\theta, X|\theta^*, X^*)}{q(\theta^*, X^*|\theta, X)} \times \frac{P(\theta^*, X^*|\mathcal{D})}{P(\theta, X|\mathcal{D})}$$

Can just as well use $\frac{P(\theta^*, X^*, \mathcal{D})}{P(\theta, X, \mathcal{D})}$.

Likelihood Free MCMC

$$
\begin{aligned}
&\frac{q(\theta^*, X^*|\theta, X)}{q(\theta, X|\theta^*, X^*)} \times \frac{P(X|\theta)}{P(X^*|\theta^*)} \frac{P(\theta)}{P(\theta^*)} \frac{P(\mathcal{D}|X, \theta)}{P(\mathcal{D}|X^*, \theta^*)} \\
&= \frac{f(\theta^*|\theta)}{f(\theta|\theta^*)} \frac{P(X^*|\theta^*)}{P(X|\theta)} \times \frac{P(X|\theta)}{P(X^*|\theta^*)} \frac{P(\theta)}{P(\theta^*)} \frac{P(\mathcal{D}|X, \theta)}{P(\mathcal{D}|X^*, \theta^*)} \\
&= \frac{f(\theta^*|\theta)}{f(\theta|\theta^*)} \times \frac{P(\theta)}{P(\theta^*)} \frac{P(\mathcal{D}|X, \theta)}{P(\mathcal{D}|X^*, \theta^*)}.
\end{aligned}
$$

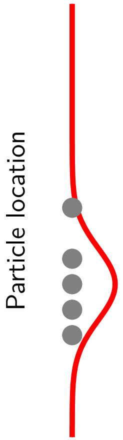
This approach, from 2003, is due to Marjoram et al. (paper title: "MCMC Without Likelihoods") [2].

Wilkinson's adaptation of LF-MCMC

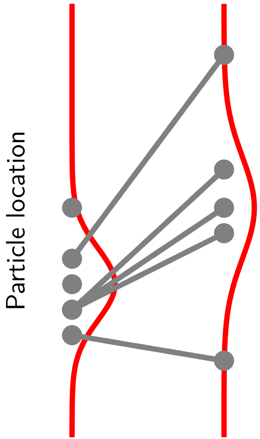
LF-MCMC fails for the same reason as everything else: $P(\mathcal{D}|X,\theta)$ is tiny for almost all X resulting from simulations.

The graphic on the following slides is from [3].

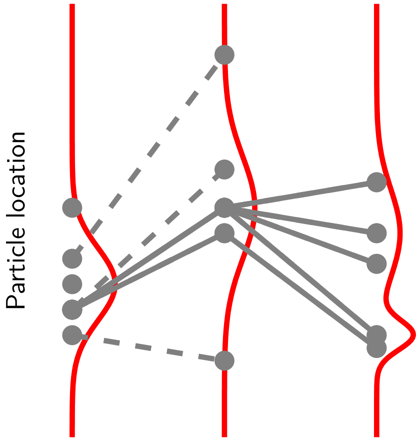
SMC: Sketch



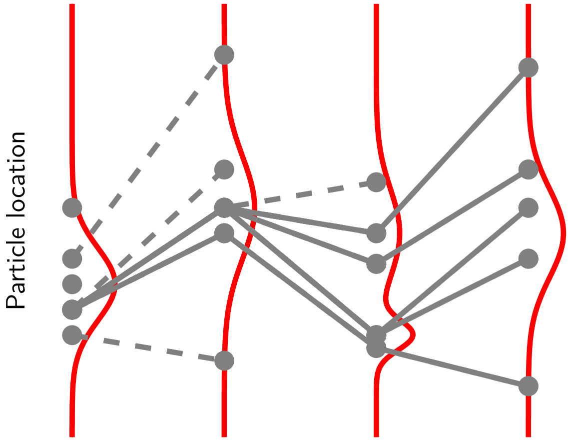
SMC: Sketch



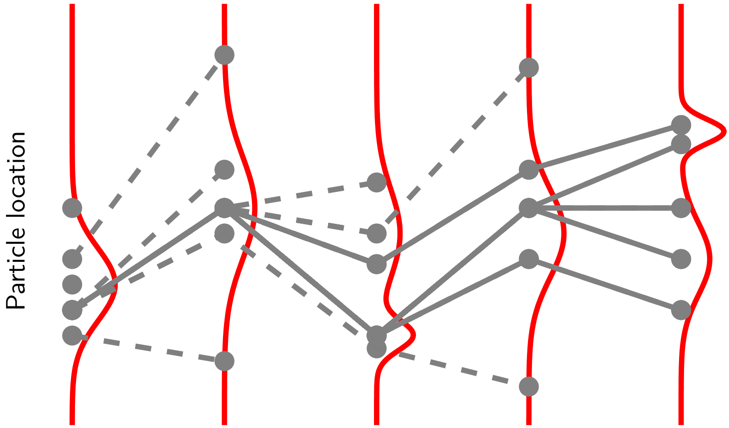
SMC: Sketch



SMC: Sketch



SMC: Sketch



Likelihood Free Particle MCMC

Given a hidden continuous-time Markov process $\{X_t\}_{t=0}^T$ with:

- Unknown parameters θ and known initial state X_0
- Data points $\mathcal{D}_{t_i}$ at times t_i, $i \in \{1, ...I\}$
- A simple, tractable error model $P(\mathcal{D}_{t_i}|X_{t_i}, \theta)$
- A simulator for paths of X given θ
- An array B_0 of 1,000,000 samples from a prior on θ, X_0
- Empty arrays B_i of the same length

For each time point (for $i \in \{1, ...I\}$):

- Until B_i is full:
  - Draw $(\theta^*, X^*_{t_{i-1}})$ from B_{i-1} or a KDE of its contents
  - Using $(\theta^*, X^*_{t_{i-1}})$, simulate up to $X^*_{t_i}$, the state at time t_i
  - Set $A = \min(1, \frac{P(\mathcal{D}_{t_i}|X^*_{t_i}, \theta^*)}{P(\mathcal{D}_{t_i}|X_{t_i}, \theta)})$
  - With probability A, overwrite (θ, X_{t_i}) with $(\theta^*, X^*_{t_i})$
  - After burn-in and thinning, add (θ, X_{t_i}) to B_i

Likelihood Free Particle MCMC

In place of the posterior, use this joint pdf: $P(\mathcal{D}_{t_i}, x(t_{1:i}), \theta | \mathcal{D}_{t_{i:i-1}})$.

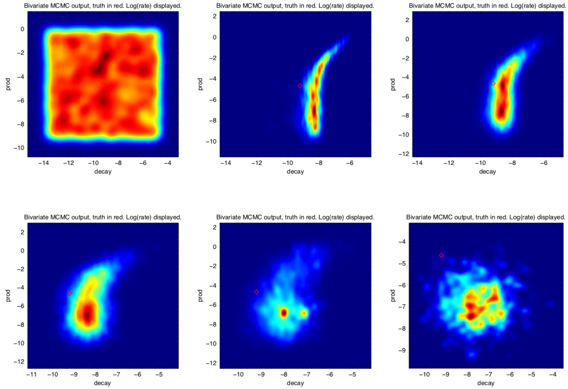
$$\overbrace{\frac{P(X(t_{1:i-1})^*, \theta^* | \mathcal{D}_{t_{1:i-1}})}{P(X(t_{1:i-1}), \theta | \mathcal{D}_{t_{1:i-1}})}}^{\text{induction hypothesis}} \underbrace{\frac{P(X(t_i)^* | X(t_{1:i-1})^*, \theta^*, \mathcal{D}_{t_{1:i-1}})}{P(X(t_i) | X(t_{1:i-1}), \theta, \mathcal{D}_{t_{1:i-1}})}}_{\text{forward simulation}} \times$$

$$\overbrace{\frac{P(X(t_{1:i-1}), \theta | \mathcal{D}_{t_{1:i-1}})}{P(X(t_{1:i-1})^*, \theta^* | \mathcal{D}_{t_{1:i-1}})}}^{\text{For that joint PDF, start here...}} \underbrace{\frac{P(X(t_i) | X(t_{1:i-1}), \theta, \mathcal{D}_{t_{1:i-1}})}{P(X(t_i)^* | X(t_{1:i-1})^*, \theta^*, \mathcal{D}_{t_{1:i-1}})}}_{\text{...then look here...}}$$

$$\underbrace{\frac{P(\mathcal{D}_{t_i} | X(t_i), \theta, \mathcal{D}_{t_{1:i-1}})}{P(\mathcal{D}_{t_i} | X(t_i)^*, \theta^*, \mathcal{D}_{t_{1:i-1}})}}_{\text{...and finally, look here.}}.$$

Does it mix?

Bivariate marginals of the distribution of production (vertical) and decay (horizontal) log rates in a simple system. From left to right, distributions condition on zero data points, one, two, three, four, and 24.



Sample Impoverishment

Questions? I

- Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M.:
  Ninth Valencia international meeting on Bayesian statistics, Benidorm, Spain, 03-08.06.2010.
  Oxford U.P., Oxford (2012)
- Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S.:
  Markov chain monte carlo without likelihoods.
  Proceedings of the National Academy of Sciences **100**(26) (2003) 15324–15328
- Finke, A.:
  Introduction to sequential monte carlo and particle mcmc methods (July 2013)

Questions? II