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].

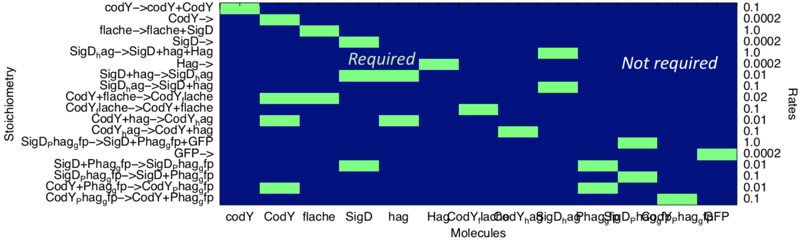
Cast of characters

- 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 reactions.
- z_i categorizes the ith reaction.
- The ith reaction occurs at t_i.
- $\theta_j \in \mathbb{R}_{\geq 0}$ is the rate of $\mathcal{R}_j$.
- $X_i(t) \in \mathbb{Z}_{\geq 0}$ counts type i molecules.
- $\mathcal{D}_{s_k} \in \mathbb{R}$ is an incomplete observation of $X(s_k)$ with error.

Of these, the only known quantity is $\mathcal{D}_{s_k}$ for some set of times $\{s_k\}$, $k \in \{0, ...n\}$. But, we also know all the stoichiometry.

Wilkinson's example–reactants and rates

Roles of chemicals in various reactions. Green means the molecule is in the Poisson intensity and blue means it is not. Reactions requiring multiple copies of reactants use binomial coeficients.



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

$$R_7(t) \sim PP(\theta_7 X_4(t) X_5(t))$$

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 (Gillespie method)

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.

Forward simulation: computation

Simulate in five minute (300-second) intervals. Reaction rates peak at $1/$second. Reactant amounts go up to 200. That makes for 60,000 exponential and categorical draws, plus some FLOPs along with each.

Likelihoods

Likelihood if reaction z_i happens at t_i:

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

Likelihoods

Likelihood if reaction z_i happens at t_i:

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

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

$$\sum_z \int_t \prod_{i=1}^{\text{events in } z} \theta_{z_i} \prod_{j=1}^{\text{rxn types}} \binom{X_j(t_{i-1})}{p_{z_i j}} \exp\left(\text{same as above}\right).$$

It's not just hard analytically: rejection sampling fails, too.

Methods for this problem

Approximate likelihoods:
Diffusion approximations, the LNA, moment-matching

Issue:

Methods for this problem

Approximate likelihoods:
Diffusion approximations, the LNA, moment-matching

Issue: approximations break down for small systems
Endpoint-conditioned simulation:
EM, Gibbs, LF-MCMC, simulated maximum likelihood, PMMH

Recurring issue:

Methods for this problem

Approximate likelihoods:
Diffusion approximations, the LNA, moment-matching

Issue: approximations break down for small systems
Endpoint-conditioned simulation:
EM, Gibbs, LF-MCMC, simulated maximum likelihood, PMMH

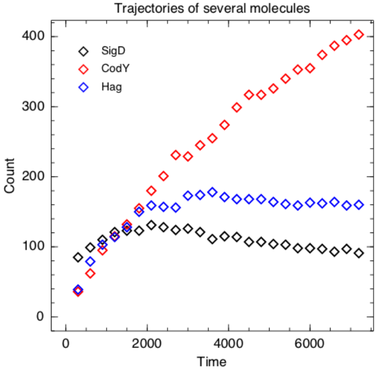
Recurring issue: forward simulations land too far from data.

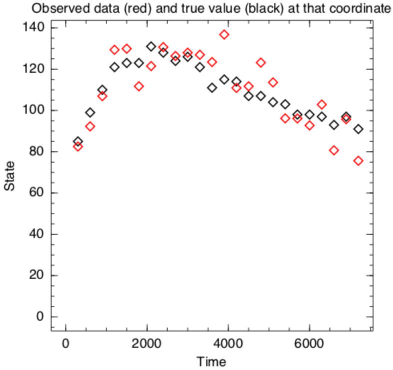
Compromises:
ABC to initialize MCMC, "tweaked" rejection sampling

Oddballs:
Variational inference

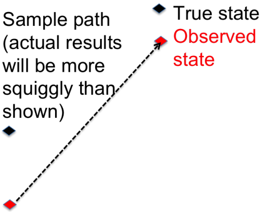
Another view of the problem

Plots of simulated reactions. Scale changes left to right; black diamonds are identical.



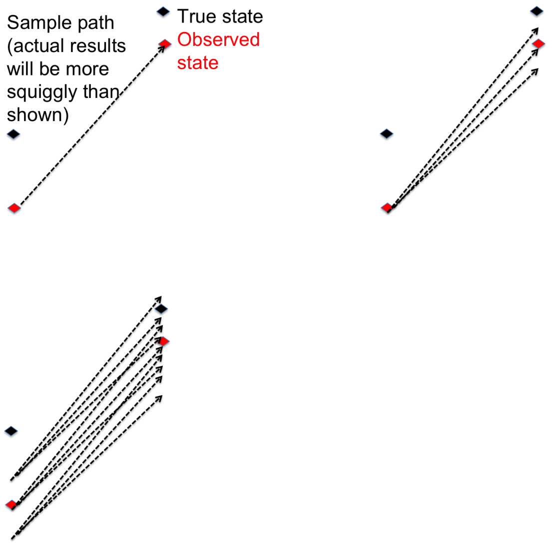


Another view of the problem

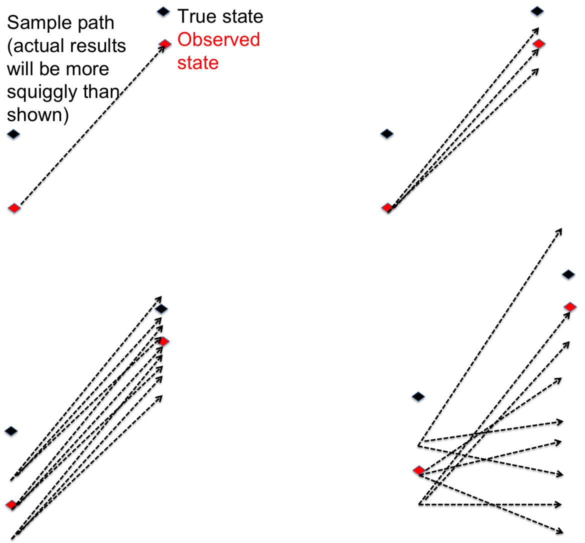


Another view of the problem

Another view of the problem



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

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

Likelihood Free MCMC

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

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

Likelihood Free MCMC

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

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

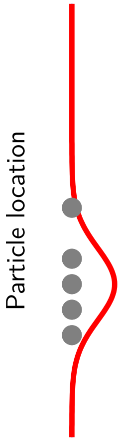
X here means $\{X(t)|t \in [0, T]\}$, so z would be redundant. This approach, from 2003, is due to Marjoram et al. (paper title: "MCMC Without Likelihoods") [2].

Wilkinson's adaptation of LF-MCMC

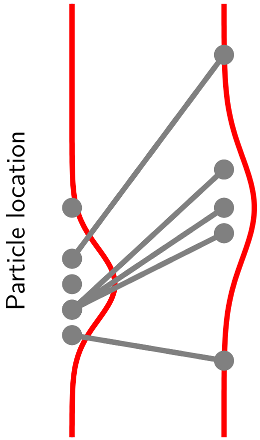
LF-MCMC fails because $P(\mathcal{D}|X, \theta)$ is tiny for almost all X resulting from simulations.

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

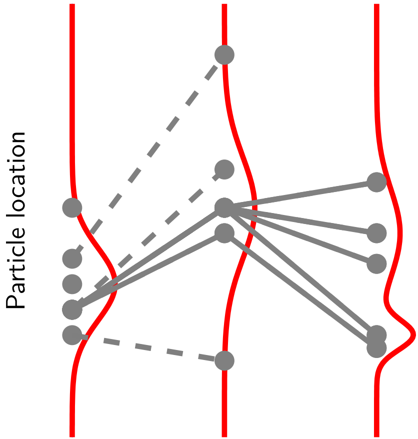
SMC: Sketch



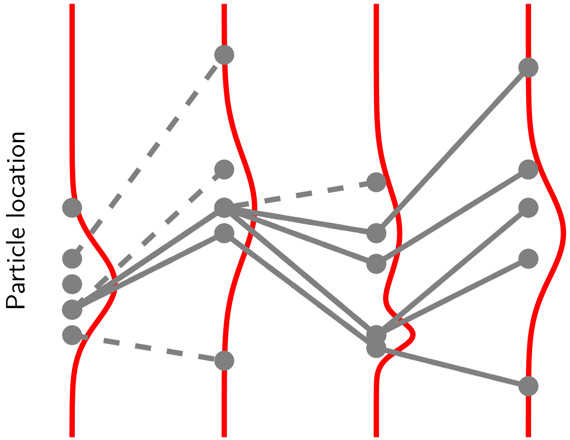
SMC: Sketch



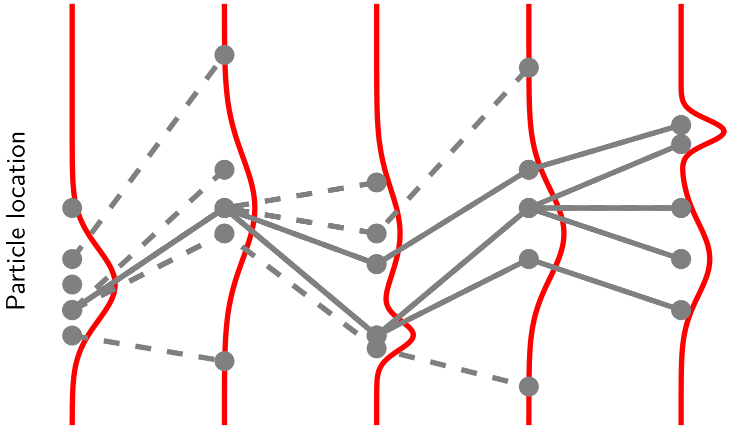
SMC: Sketch



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

LF-pMCMC computation

$$\underbrace{\frac{300\text{ sec}}{\text{simulation}} \times \frac{1\text{ reaction}}{\text{sec} \times \text{molecule}} \times 200\text{ molecules}}_{60{,}000\text{ reactions per simulation}}$$

$$\times \frac{24\text{simulations}}{\text{particle}} \times 5,000,000\text{particles}$$

The product is 7.2e12, which is 7.2 trillion. It runs in two days.

Likelihood Free Particle MCMC

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

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

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

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

Will it mix?

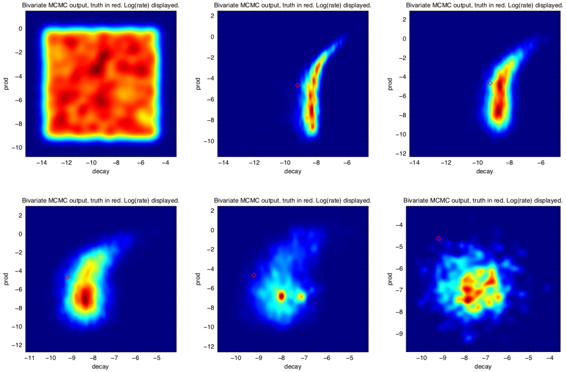
Bivariate marginals of the distribution of production (vertical) and decay (horizontal) log rates in a simple system. From left to right, these condition on zero data points, one, two, three, four, and 24.

Will it mix?

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

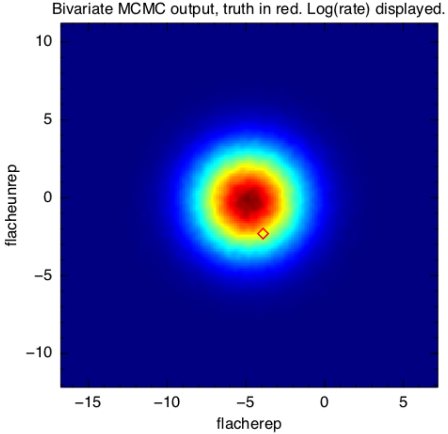
Data used:
0, 1, 2

Data used:
3, 4, 24

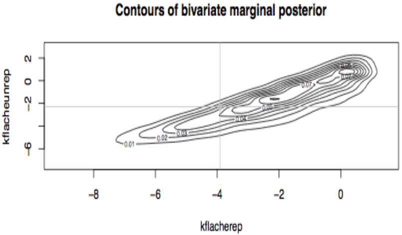


The red diamond is the true value.

Will it mix?



A replication attempt using Wilkinson's settings: a 13-molecule, 18-reaction system with 1,000,000 particles. Log rates displayed.

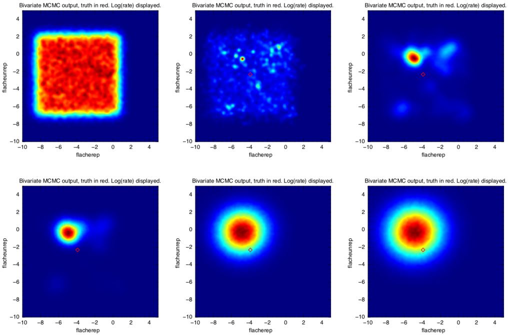


Will it mix?

Log parameters at some intermediate stages. From left to right, these condition on zero data points, one, two, three, 12, and 16.

Data used:
0, 1, 2

Data used:
3, 12, 16



The red diamond is the true value.

Sample Impoverishment

Repeats come from common ancestry or from rejection events.

Sample Impoverishment

Repeats come from common ancestry or from rejection events. How many unique particles persist by the end?

Sample Impoverishment

Repeats come from common ancestry or from rejection events. How many unique particles persist by the end?

Keeping every fifth sample, the event $E_i = \{$ten consecutive rejections$\}$ is always enough to get a repeat.

Sample Impoverishment

Repeats come from common ancestry or from rejection events. How many unique particles persist by the end?

Keeping every fifth sample, the event $E_i = \{$ten consecutive rejections$\}$ is always enough to get a repeat. So, if the stage-i rejection rate is r, and consecutive rejections are independent, $\Pr(E_i^C)$ is $a_i = 1 - r_i^{10}$. In turn, the (asymptotic) proportion of unique particles surviving is below $\Pr(E^C)$.

Sample Impoverishment

Repeats come from common ancestry or from rejection events. How many unique particles persist by the end?

Keeping every fifth sample, the event $E_i = \{$ten consecutive rejections$\}$ is always enough to get a repeat. So, if the stage-i rejection rate is r, and consecutive rejections are independent, $\Pr(E_i^C)$ is $a_i = 1 - r_i^{10}$. In turn, the (asymptotic) proportion of unique particles surviving is below $\Pr(E^C)$.

For the samples I showed, the product of empirical estimates for a_i was $300/1,000,000$.

Sample Impoverishment

Repeats come from common ancestry or from rejection events. How many unique particles persist by the end?

Keeping every fifth sample, the event $E_i = \{$ten consecutive rejections$\}$ is always enough to get a repeat. So, if the stage-i rejection rate is r, and consecutive rejections are independent, $\Pr(E_i^C)$ is $a_i = 1 - r_i^{10}$. In turn, the (asymptotic) proportion of unique particles surviving is below $\Pr(E^C)$.

For the samples I showed, the product of empirical estimates for a_i was $300/1,000,000$. That ignores correlation in the rejection events and common ancestry.

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