

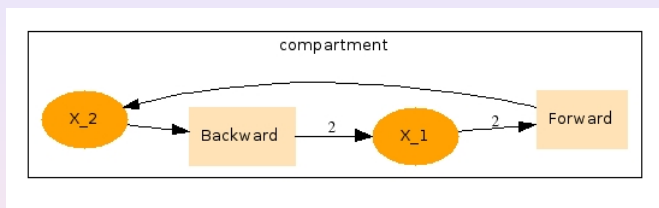
Parameter estimation using moment closure methods

Colin Gillespie

Newcastle University
School of Mathematics & Statistics

March 26, 2008

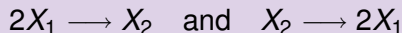
- Start with some kind of picture or diagram for a mechanism



- Turn it into a set of (pseudo-) biochemical reactions:
- Specify the rate laws and rate parameters of the reactions,
- Run some stochastic or deterministic computer simulator of the system dynamics to gain insight into the system.
- But how do you get the parameters in the first place?

- Start with some kind of picture or diagram for a mechanism
- Turn it into a set of (pseudo-) biochemical reactions:

Dimerisation



This will occur when two molecules of X_1 collide to produce a molecule of X_2 .

- Specify the rate laws and rate parameters of the reactions,
- Run some stochastic or deterministic computer simulator of the system dynamics to gain insight into the system.
- But how do you get the parameters in the first place?

- Start with some kind of picture or diagram for a mechanism
- Turn it into a set of (pseudo-) biochemical reactions:
- Specify the rate laws and rate parameters of the reactions,
 - e.g. $\{k_1, k_2\} = \{1, 5\}$ and $X_1(0) = 100$ and $X_2(0) = 0$
- Run some stochastic or deterministic computer simulator of the system dynamics to gain insight into the system.
- But how do you get the parameters in the first place?

- Start with some kind of picture or diagram for a mechanism
- Turn it into a set of (pseudo-) biochemical reactions:
- Specify the rate laws and rate parameters of the reactions,
- Run some **stochastic** or **deterministic** computer simulator of the system dynamics to gain insight into the system.
- But how do you get the parameters in the first place?

- Start with some kind of picture or diagram for a mechanism
- Turn it into a set of (pseudo-) biochemical reactions:
- Specify the rate laws and rate parameters of the reactions,
- Run some stochastic or deterministic computer simulator of the system dynamics to gain insight into the system.
- But how do you get the parameters in the first place?

Simulation techniques

- The Gillespie algorithm is perhaps the most well known.
 - Choose a time when the next reaction occurs.
 - Choose which reaction occurs.
- The Gibson & Bruck method is a more efficient version of the Gillespie algorithm.
- Approximate simulators update more than one event at a time.
- Hybrid methods treat parts of the model as a traditional deterministic ODE and other parts using a stochastic method.
- Use the moment equations to obtain estimates for the mean and variances of the process (very quick).

Moment Equations

- We can formulate the dimer model in terms of moment equations, namely,

$$\frac{dE[X_1]}{dt} = 0.5k_1(E[X_1^2] - E[X_1]) - k_2E[X_1]$$

$$\begin{aligned} \frac{dE[X_1^2]}{dt} &= k_1(E[X_1^2, X_2] - E[X_1, X_2]) + 0.5k_1(E[X_1^2] - E[X_1]) \\ &\quad + k_2(E[X_1] - 2E[X_1^2]) \end{aligned}$$

where $E[X_1]$ is the mean of X_1 and $E[X_1^2] - E[X_1]$ is the variance of X_1 .

Moment Equations

- We can formulate the dimer model in terms of moment equations, namely,

$$\frac{dE[X_1]}{dt} = 0.5k_1(E[X_1^2] - E[X_1]) - k_2E[X_1]$$

$$\begin{aligned}\frac{dE[X_1^2]}{dt} &= k_1(E[X_1^2, X_2] - E[X_1, X_2]) + 0.5k_1(E[X_1^2] - E[X_1]) \\ &\quad + k_2(E[X_1] - 2E[X_1^2])\end{aligned}$$

where $E[X_1]$ is the mean of X_1 and $E[X_1^2] - E[X_1]$ is the variance of X_1 .

- The i^{th} moment equation depends on the $(i+1)^{\text{th}}$ equation.

Moment Equations

- We can formulate the dimer model in terms of moment equations, namely,

$$\frac{dE[X_1]}{dt} = 0.5k_1 E[X_1](E[X_1] - 1) + 0.5k_1 \text{Var}[X_1] - k_2 E[X_1]$$

where $E[X_1]$ is the mean of X_1 and $E[X_1^2] - E[X_1]^2$ is the variance of X_1 .

- The deterministic equation is an approximation to the stochastic mean.

Moment Closure

- For almost all systems, we need to ‘close’ the moment equations.
- The easiest option is to assume an underlying Normal distribution, i.e.

$$E[X_1^3] = 3E[X_1^2]E[X_1] - 2E[X_1]^3$$

- But we could also use, the Poisson

$$E[X_1^3] = E[X_1] + 3E[X_1]^2 + E[X_1]^3$$

or the Lognormal

$$E[X_1^3] = \left(\frac{E[X_1^2]}{E[X_1]} \right)^3$$

- For a model with six chemical species, we would obtain 27 ODEs to solve (means, variances, covariances). These equations can be solved in a few seconds.

Moment Closure

- For almost all systems, we need to 'close' the moment equations.
- The easiest option is to assume an underlying Normal distribution, i.e.

$$E[X_1^3] = 3E[X_1^2]E[X_1] - 2E[X_1]^3$$

- But we could also use, the Poisson

$$E[X_1^3] = E[X_1] + 3E[X_1]^2 + E[X_1]^3$$

or the Lognormal

$$E[X_1^3] = \left(\frac{E[X_1^2]}{E[X_1]} \right)^3$$

- For a model with six chemical species, we would obtain 27 ODEs to solve (means, variances, covariances). These equations can be solved in a few seconds.

Moment Closure

- For almost all systems, we need to 'close' the moment equations.
- The easiest option is to assume an underlying Normal distribution, i.e.

$$E[X_1^3] = 3E[X_1^2]E[X_1] - 2E[X_1]^3$$

- But we could also use, the Poisson

$$E[X_1^3] = E[X_1] + 3E[X_1]^2 + E[X_1]^3$$

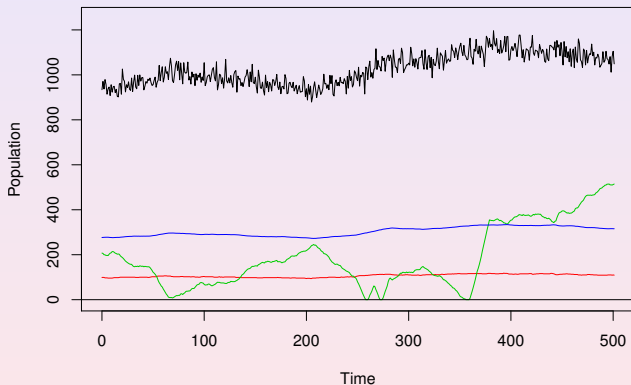
or the Lognormal

$$E[X_1^3] = \left(\frac{E[X_1^2]}{E[X_1]} \right)^3$$

- For a model with six chemical species, we would obtain 27 ODEs to solve (means, variances, covariances). These equations can be solved in a few seconds.

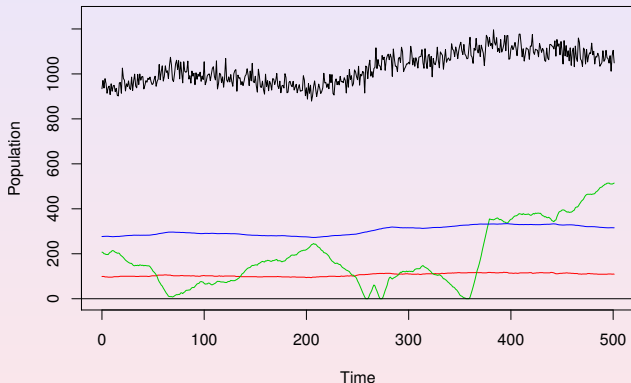
Parameter Inference

If we observe all species at all time points, then estimating the kinetic rate parameters is straightforward



Parameter Inference

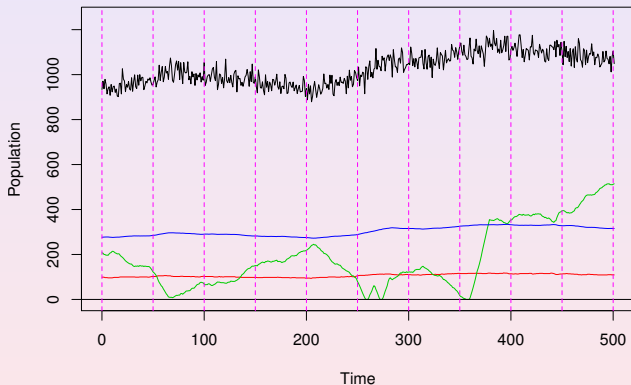
If we observe all species at all time points, then estimating the kinetic rate parameters is straightforward



But this is unlikely to occur.

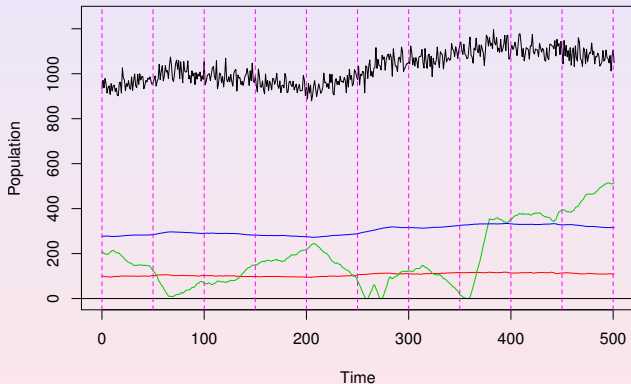
Parameter Inference

Instead, we may we able to observe population levels at discrete time points



Parameter Inference

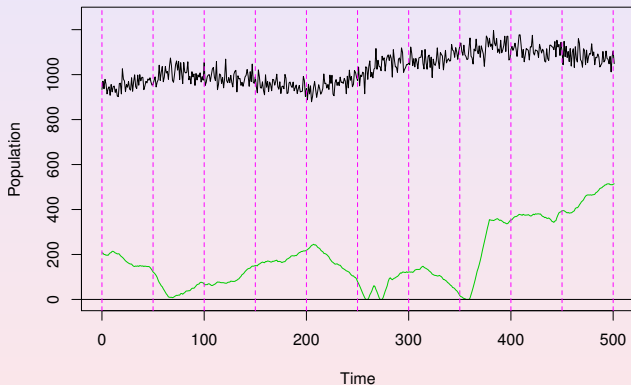
Instead, we may we able to observe population levels at discrete time points



However, this is also a bit hopeful.

Parameter Inference

So we consider the case where we observe a subset of chemical species, at discrete time points



Parameter Inference

- Let $\mathbf{x}(t_i)$ be the i^{th} discrete time observation of the process.
- If we observe all species at each time point and assume a Normal distribution, then

$$\mathbf{x}(t_i) \mid \mathbf{x}(t_{i-1}) \sim N(\mu, \Sigma)$$

where μ and Σ are calculated via the moment closure approximation.

- Hence, we can easily calculate the likelihood to estimate the parameters for a given data set

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^N \Pr[\mathbf{x}(t_i) \mid \mathbf{x}(t_{i-1})]$$

where θ is a parameter vector and $\mathbf{x} = \{\mathbf{x}(t_i) : i = 1, \dots, N\}$.

Technicalities

- We use a simple Metropolis-Hastings sampler to explore the sample space.
- We assume flat priors (on $U[\exp(-10), \exp(10)]$) in the following examples, but for real data some effort should be made to elicit prior information (from previously published models say).
- When there are missing observations, we again use a Metropolis-Hastings sampler to fill in any missing values.
- In the following examples, the observations are made at regular intervals, but this isn't necessary.

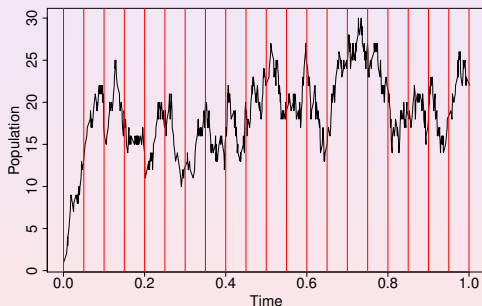
Example 1: Immigration-death model

- The immigration-death model has the following reactions



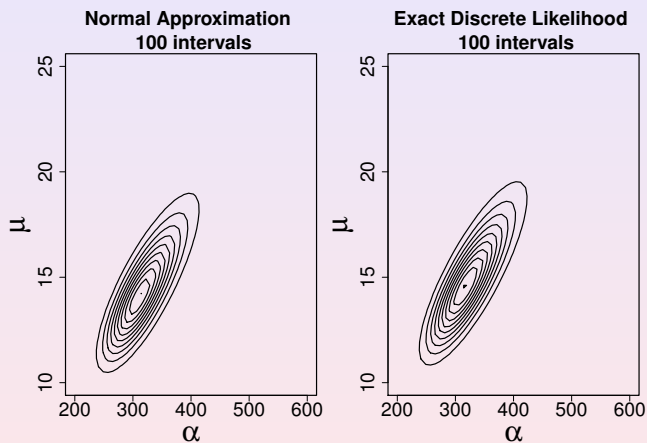
with reaction rates μX and α respectively.

- We observe the following realisation:



where $\alpha = 400$ and $\mu = 20$. This gives a steady state value of 20.

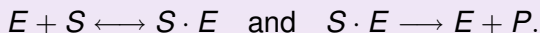
Example 1: Immigration-death model - Results



Time taken to estimate parameters: $\ll 1$ second.

Example 2: Michaelis-Menten

- The Michaelis-Menten system has the following reactions



with reaction rates $c_1 E S$, $c_2 S \cdot E$ and $c_3 S \cdot E$ respectively.

- Due to conservation laws, we have two constant species $\{E, P\}$.
- We observe S at 50 time-points and have no information on $S \cdot E$.

Example 2: Michaelis-Menten inference

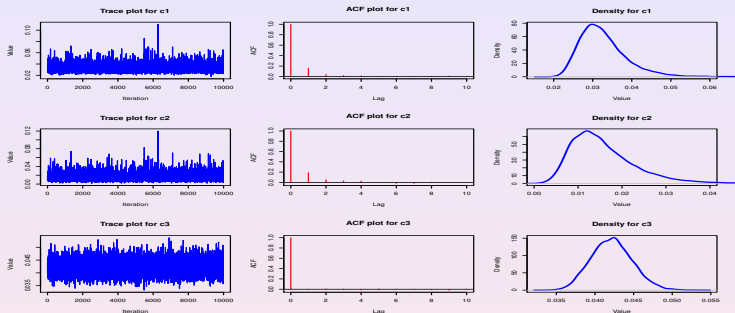


Figure: MCMC output from 5 million iterations, thinned to every 500

	c_1	c_2	c_3
True value	0.03	0.01	0.04
Mode (sd)	0.0295 (0.021)	0.0121 (0.014)	0.0425 (0.0046)

Example 2: Michaelis-Menten inference

- A standard approximation to this system is to use the quasi-steady state assumption



- With reaction rate

$$-\frac{V_{\max}S}{c_m + S}$$

where $V_{\max} = c_3 e_0$ and $c_m = (c_2 + c_3)/c_1$.

- We have reduced the number of species from two to one.

Example 2: Michaelis-Menten QSSA inference

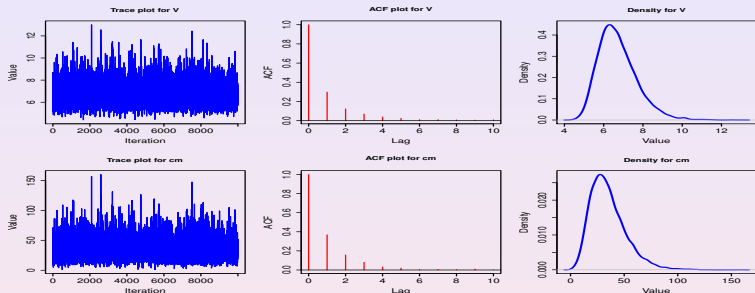


Figure: MCMC output from 500,000 iterations, thinned to every 50

	V_{max}	C_m
True value	6.0	20.0
Mean (sd)	6.25 (0.985)	28.96 (16.76)

Example 3: Prokaryotic Auto-Regulatory gene network

- This model has six species and twelve reactions.
- To estimate the means and variances, we now have to solve 27 ODEs:
 - Six ODEs for the means.
 - Six ODEs for the variances.
 - Fifteen ODEs for the covariances.
- Results similar to Golightly and Wilkinson (2008) who used a diffusion approximation approach.
- The Moment closure approach is about 4-5 times faster.

Conclusions and future work

- The moment closure approach seems to estimate the parameters quickly and correctly.
- By using freely available ODE libraries such as *gsl* and *sundials*, computation is fast and coding is (fairly) straightforward.
- Problems do occur when two (or more) species are highly correlated - but we are working on methods to deal with this problem.
- Software is currently being developed that will take an SBML model.
 - Automatically generate the moment equations (see pysbml.googlecode.com).
 - Infer parameters for a given data set (almost ready).

Acknowledgements

- Peter Milner (Newcastle University)
- Darren Wilkinson (Newcastle University)
- Eric Renshaw (University of Strathclyde)