Bayesian inference for Markov processes with application to biochemical network dynamics

Darren Wilkinson
http://tinyurl.com/darrenjw
School of Mathematics & Statistics,
Newcastle University, UK

PEDS-II
EURANDOM
Eindhoven, Netherlands
5th June, 2012

Systems biology modelling

- Uses accurate high-resolution time-course data on a relatively small number of bio-molecules to parametrise carefully constructed mechanistic dynamic models of a process of interest based on current biological understanding
- Traditionally, models were typically deterministic, based on a system of ODEs known as the Reaction Rate Equations (RREs)
- It is now increasingly accepted that biochemical network dynamics at the single-cell level are intrinsically stochastic
- The theory of stochastic chemical kinetics provides a solid foundation for describing network dynamics using a Markov jump process

Stochastic Chemical Kinetics

Stochastic molecular approach:

- Statistical mechanical arguments lead to a Markov jump process in continuous time whose instantaneous reaction rates are directly proportional to the number of molecules of each reacting species
- Such dynamics can be simulated (exactly) on a computer using standard discrete-event simulation techniques
- Standard implementation of this strategy is known as the "Gillespie algorithm" (just discrete event simulation), but there are several exact and approximate variants of this basic approach

Lotka-Volterra system

Trivial (familiar) example from population dynamics (in reality, the "reactions" will be elementary biochemical reactions taking place inside a cell)

Reactions

$$X \longrightarrow 2X \qquad \text{(prey reproduction)}$$
$$X + Y \longrightarrow 2Y \qquad \text{(prey-predator interaction)}$$
$$Y \longrightarrow \emptyset \qquad \text{(predator death)}$$

- X – Prey, Y – Predator
- We can re-write this using matrix notation

Forming the matrix representation

The L-V system in tabular form

	Rate Law	LHS		RHS		Net-effect	
	$h(\cdot, c)$	X	Y	X	Y	X	Y
R_1	$c_1 x$	1	0	2	0	1	0
R_2	$c_2 xy$	1	1	0	2	-1	1
R_3	$c_3 y$	0	1	0	0	0	-1

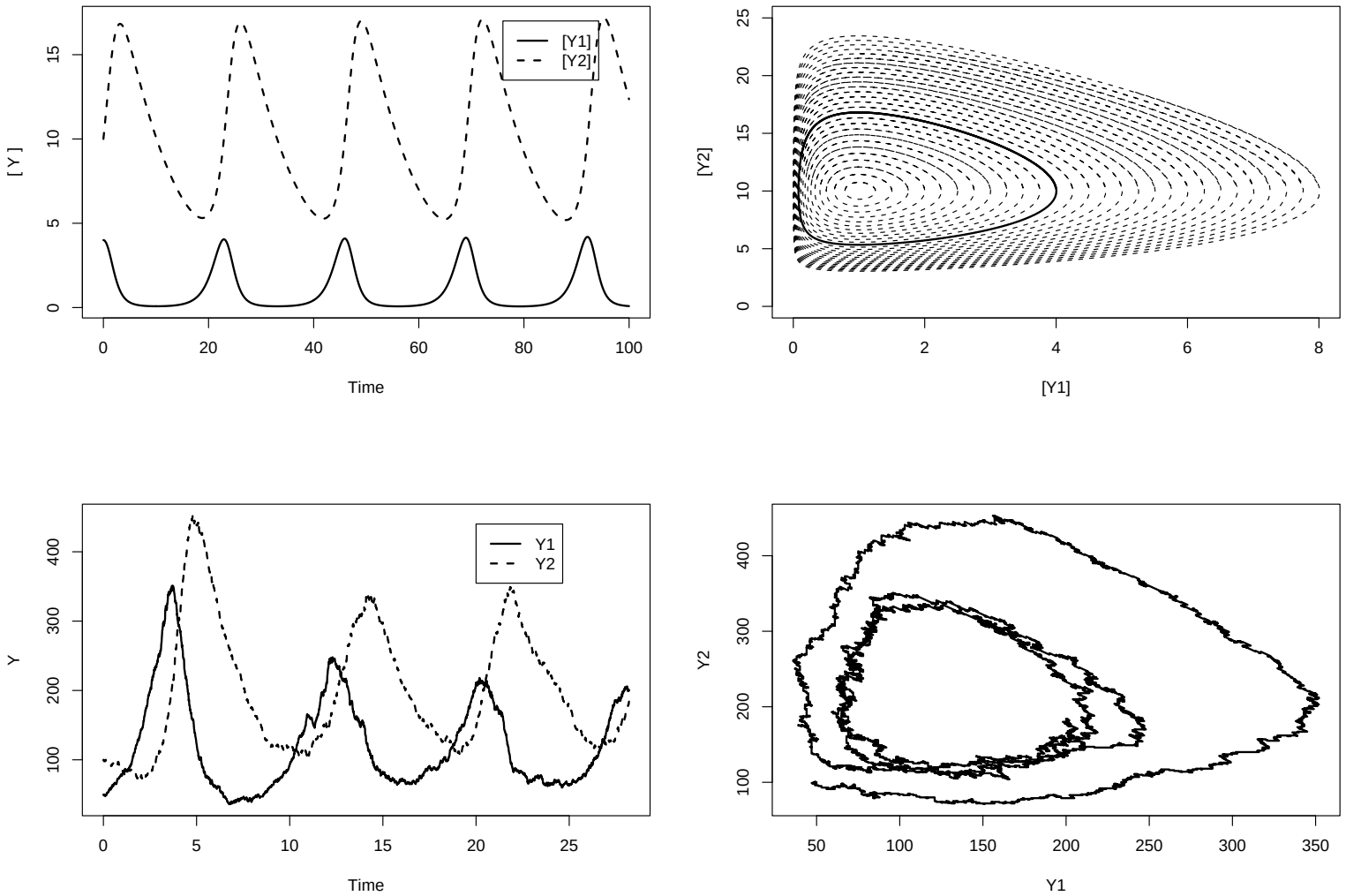
Call the 3×2 net-effect (or reaction) matrix N. The matrix $S = N'$ is the stoichiometry matrix of the system.

Stochastic chemical kinetics

- u species: $\mathcal{X}_1, \ldots, \mathcal{X}_u$, and v reactions: $\mathcal{R}_1, \ldots, \mathcal{R}_v$
- $\mathcal{R}_i: \; p_{i1}\mathcal{X}_1 + \cdots + p_{iu}\mathcal{X}_u \longrightarrow q_{i1}\mathcal{X}_1 + \cdots + q_{iu}\mathcal{X}_u, \; i = 1, \ldots, v$
- In matrix form: $P\mathcal{X} \longrightarrow Q\mathcal{X}$ (P and Q are sparse)
- $S = (Q - P)'$ is the stoichiometry matrix of the system
- X_{jt}: # molecules of $\mathcal{X}_j$ at time t. $X_t = (X_{1t}, \ldots, X_{ut})'$
- Reaction $\mathcal{R}_i$ has hazard (or rate law, or propensity) $h_i(X_t, c_i)$, where c_i is a rate parameter, $c = (c_1, \ldots, c_v)'$, $h(X_t, c) = (h_1(X_t, c_1), \ldots, h_v(X_t, c_v))'$ and the system evolves as a Markov jump process
- For mass-action stochastic kinetics,

$$h_i(X_t, c_i) = c_i \prod_{j=1}^{u} \binom{X_{jt}}{p_{ij}}, \quad i = 1, \ldots, v$$

The Lotka-Volterra model



Time change representation

- R_{it}: # reactions of type $\mathcal{R}_i$ in $(0, t]$, $R_t = (R_{1t}, \ldots, R_{vt})'$
- $X_t - X_0 = SR_t$ (state updating equation)
- For $i = 1, \ldots, v$, $N_i(t)$ are the count functions for independent unit Poisson processes, so

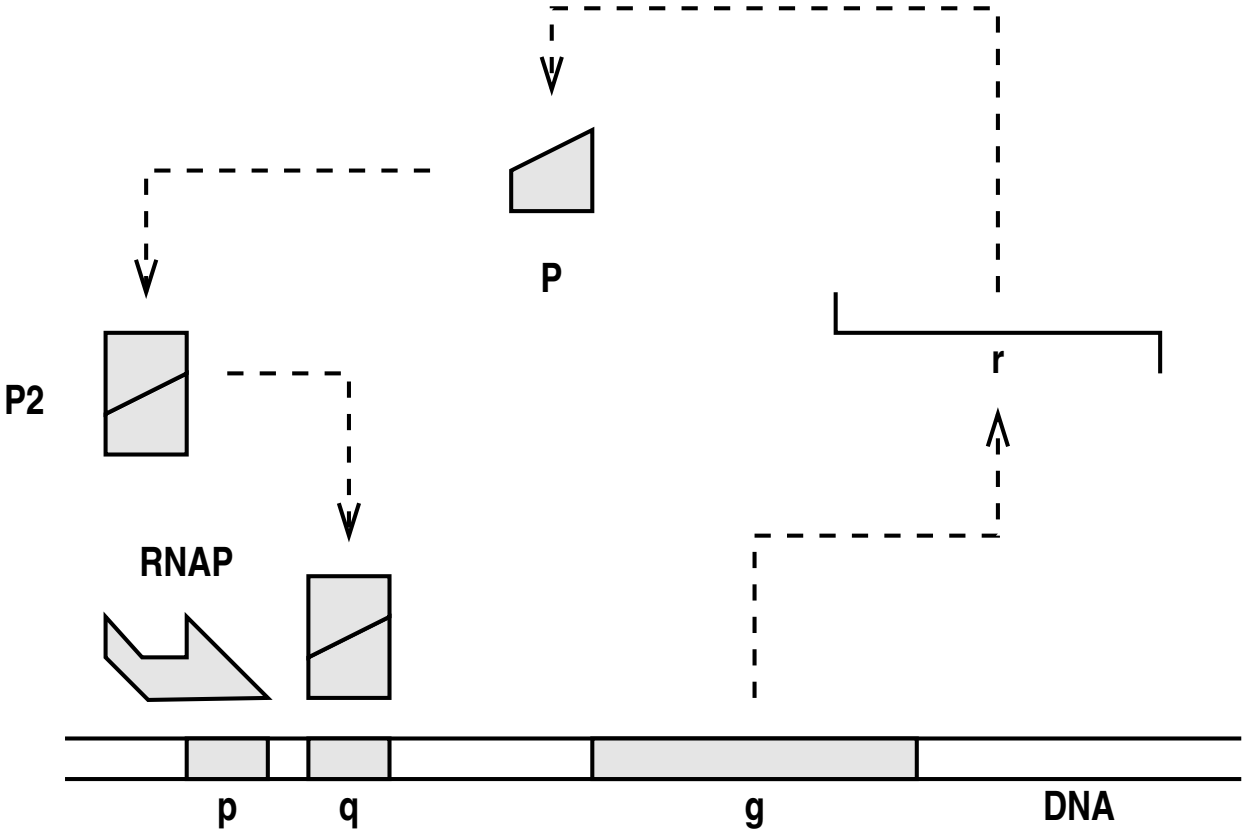
$$R_{it} = N_i \left(\int_0^t h_i(X_\tau, c_i) d\tau \right)$$

- Putting $N(t_1, \ldots, t_v) = (N_1(t_1), \ldots, N_v(t_v))'$, we can write $R_t = N \left(\int_0^t h(X_\tau, c) d\tau \right)$ to get:

Time-change representation of the Markov jump process

$$X_t - X_0 = S\, N \left(\int_0^t h(X_\tau, c) d\tau \right)$$

Example — genetic auto-regulation



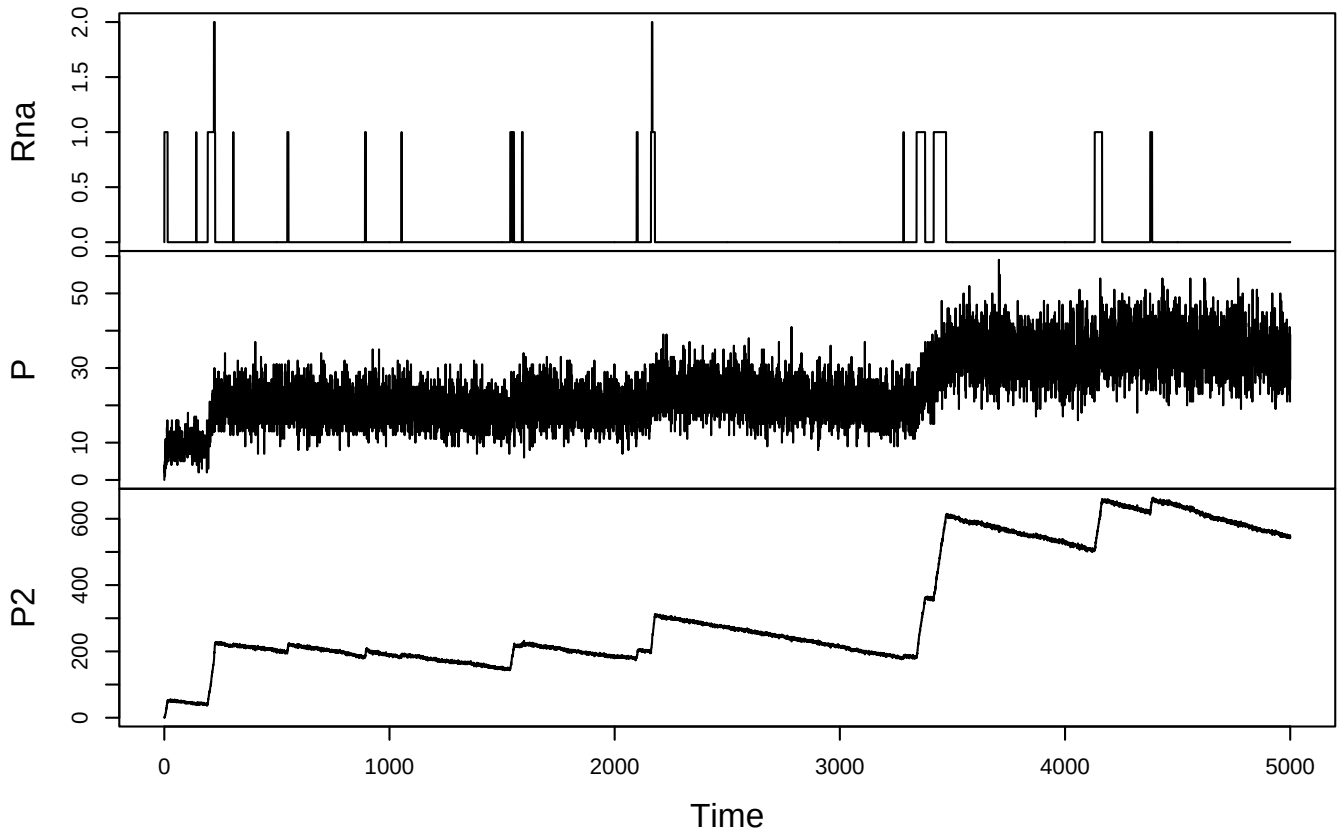
Biochemical reactions

Simplified view:

Reactions

$g + P_2 \longleftrightarrow g \cdot P_2$	Repression
$g \longrightarrow g + r$	Transcription
$r \longrightarrow r + P$	Translation
$2P \longleftrightarrow P_2$	Dimerisation
$r \longrightarrow \emptyset$	mRNA degradation
$P \longrightarrow \emptyset$	Protein degradation

Simulated realisation of the auto-regulatory network



Partially observed Markov process (POMP) models

- Continuous-time Markov process: $\mathbf{X} = \{X_s | s \geq 0\}$ (for now, we suppress dependence on parameters, θ)
- Think about integer time observations (extension to arbitrary times is trivial): for $t \in \mathbb{N}$, $\quad \mathbf{X}_t = \{X_s | t-1 < s \leq t\}$
- Sample-path likelihoods such as $\pi(\mathbf{x}_t | x_{t-1})$ can often (but not always) be computed (but are often computationally difficult), but discrete time transitions such as $\pi(x_t | x_{t-1})$ are typically intractable
- Partial observations: $\mathcal{Y} = \{y_t | t = 1, 2, \ldots, T\}$ where

$$y_t | X_t = x_t \sim \pi(y_t | x_t), \qquad t = 1, \ldots, T,$$

where we assume that $\pi(y_t | x_t)$ can be evaluated directly (simple measurement error model)

Bayesian inference for POMP models

- Most "obvious" MCMC algorithms will attempt to impute (at least) the skeleton of the Markov process: $X_0, X_1, \ldots, X_T$
- This will typically require evaluation of the intractable discrete time transition likelihoods, and this is the problem...
- Two related strategies:
  - Data augmentation: "fill in" the entire process in some way, typically exploiting the fact that the sample path likelihoods are tractable — works in principle, but difficult to "automate", and exceptionally computationally intensive due to the need to store and evaluate likelihoods of cts sample paths
  - Likelihood-free (AKA plug-and-play): exploits the fact that it is possible to forward simulate from $\pi(x_t|x_{t-1})$ (typically by simulating from $\pi(\mathbf{x}_t|x_{t-1})$), even if it can't be evaluated
- Likelihood-free is really just a special kind of augmentation strategy

Bayesian inference

- Let $\pi(\mathbf{x}|c)$ denote the (complex) likelihood of the simulation model
- Let $\pi(\mathcal{Y}|\mathbf{x},\tau)$ denote the (simple) measurement error model
- Put $\theta = (c,\tau)$, and let $\pi(\theta)$ be the prior for the model parameters
- The joint density can be written

$$\pi(\theta,\mathbf{x},\mathcal{Y}) = \pi(\theta)\pi(\mathbf{x}|\theta)\pi(\mathcal{Y}|\mathbf{x},\theta).$$

- Interest is in the posterior distribution $\pi(\theta,\mathbf{x}|\mathcal{Y})$

Particle MCMC (pMCMC)

- Of the various alternatives, pMCMC is the only obvious practical option for constructing global likelihood-free MCMC algorithms which are exact (Andrieu et al, 2010)
- Start by considering a basic marginal MH MCMC scheme with target $\pi(\theta|\mathcal{Y})$ and proposal $f(\theta^\star|\theta)$ — the acceptance probability is $\min\{1, A\}$ where

$$A = \frac{\pi(\theta^\star)}{\pi(\theta)} \times \frac{f(\theta|\theta^\star)}{f(\theta^\star|\theta)} \times \frac{\pi(\mathcal{Y}|\theta^\star)}{\pi(\mathcal{Y}|\theta)}$$

- We can't evaluate the final terms, but if we had a way to construct a Monte Carlo estimate of the likelihood, $\hat{\pi}(\mathcal{Y}|\theta)$, we could just plug this in and hope for the best:

$$A = \frac{\pi(\theta^\star)}{\pi(\theta)} \times \frac{f(\theta|\theta^\star)}{f(\theta^\star|\theta)} \times \frac{\hat{\pi}(\mathcal{Y}|\theta^\star)}{\hat{\pi}(\mathcal{Y}|\theta)}$$

"Exact approximate" MCMC (the pseudo-marginal approach)

- Remarkably, provided only that $\mathrm{E}[\hat{\pi}(\mathcal{Y}|\theta)] = \pi(\mathcal{Y}|\theta)$, the stationary distribution of the Markov chain will be exactly correct (Beaumont, 2003, Andrieu & Roberts, 2009)
- Putting $W = \hat{\pi}(\mathcal{Y}|\theta)/\pi(\mathcal{Y}|\theta)$ and augmenting the state space of the chain to include W, we find that the target of the chain must be

$$\propto \pi(\theta)\hat{\pi}(\mathcal{Y}|\theta)\pi(w|\theta) \propto \pi(\theta|\mathcal{Y})w\pi(w|\theta)$$

and so then the above "unbiasedness" property implies that $\mathrm{E}(W|\theta) = 1$, which guarantees that the marginal for θ is exactly $\pi(\theta|\mathcal{Y})$

Particle marginal Metropolis-Hastings (PMMH)

- Likelihood estimates constructed via importance sampling typically have this "unbiasedness" property, as do estimates constructed using a particle filter
- If a particle filter is used to construct the Monte Carlo estimate of likelihood to plug in to the acceptance probability, we get (a simple version of) the particle Marginal Metropolis Hastings (PMMH) pMCMC algorithm
- The full PMMH algorithm also uses the particle filter to construct a proposal for $\mathbf{x}$, and has target $\pi(\theta, \mathbf{x}|\mathcal{Y})$ — not just $\pi(\theta|\mathcal{Y})$
- The (bootstrap) particle filter relies only on the ability to forward simulate from the process, and hence the entire procedure is "likelihood-free"

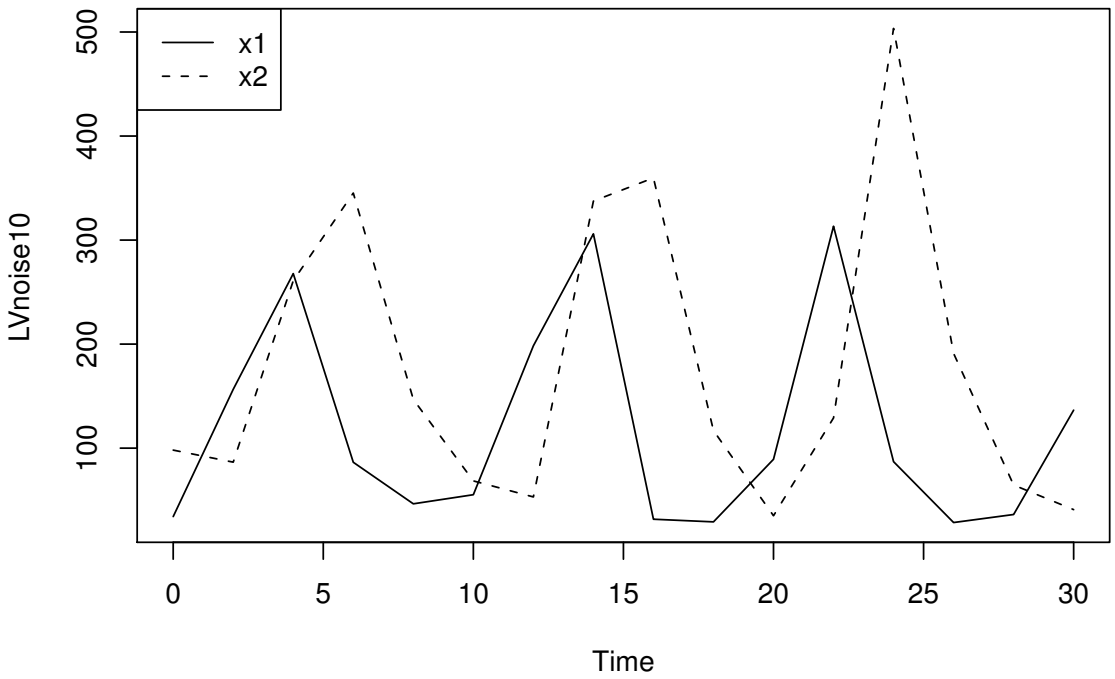
Blog post: http://bit.ly/kvznmq

The bootstrap particle filter

- "Particle cloud": $\mathbf{x}_t = \{x_t^k | k = 1, \ldots, M\}$, $\boldsymbol{\pi}_t = \{\pi_t^k | k = 1, \ldots, M\}$, $\tilde{\mathbf{x}}_t = \{(x_t^k, \pi_t^k) | k = 1, \ldots, M\}$
- Initialise with $\tilde{\mathbf{x}}_0$, where $x_0^k \sim p(x_0)$ and $\pi_0^k = 1/M$ (note that w_0^k is undefined)
- Suppose at time t we have a sample from $p(x_t|y_{1:t})$: $\tilde{\mathbf{x}}_t$
  - Sample $a_t^k \sim \mathcal{F}(a_t^k|\boldsymbol{\pi}_t)$, $k = 1, \ldots, M$
  - Sample $x_{t+1}^k \sim p(x_{t+1}^k | x_t^{a_t^k})$
  - Set $w_{t+1}^k = p(y_{t+1}|x_{t+1}^k)$ and $\pi_{t+1}^k = w_{t+1}^k / \sum_{i=1}^M w_{t+1}^i$
  - Propagate $\tilde{\mathbf{x}}_{t+1}$ to the next step...

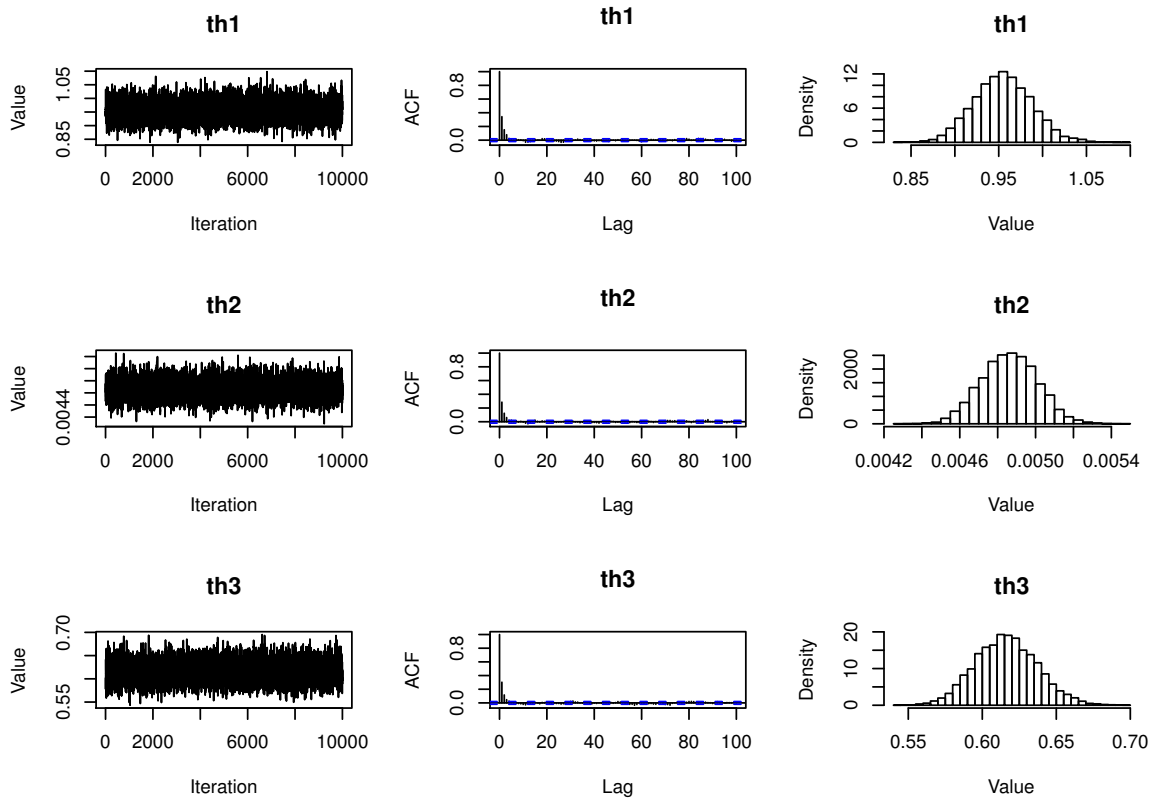
Define $\hat{p}(y_t|y_{1:t-1}) = \frac{1}{M} \sum_{i=1}^M w_t^i$ and $\hat{p}(y_{1:T}) = \prod_{i=1}^T \hat{p}(y_t|y_{1:t-1})$. Clear that $\hat{p}(y_{1:T})$ is a consistent estimator of $p(y_{1:T})$, but not obvious that it is in fact also unbiased (Pitt et al, 2011).

Test problem: Lotka-Volterra model



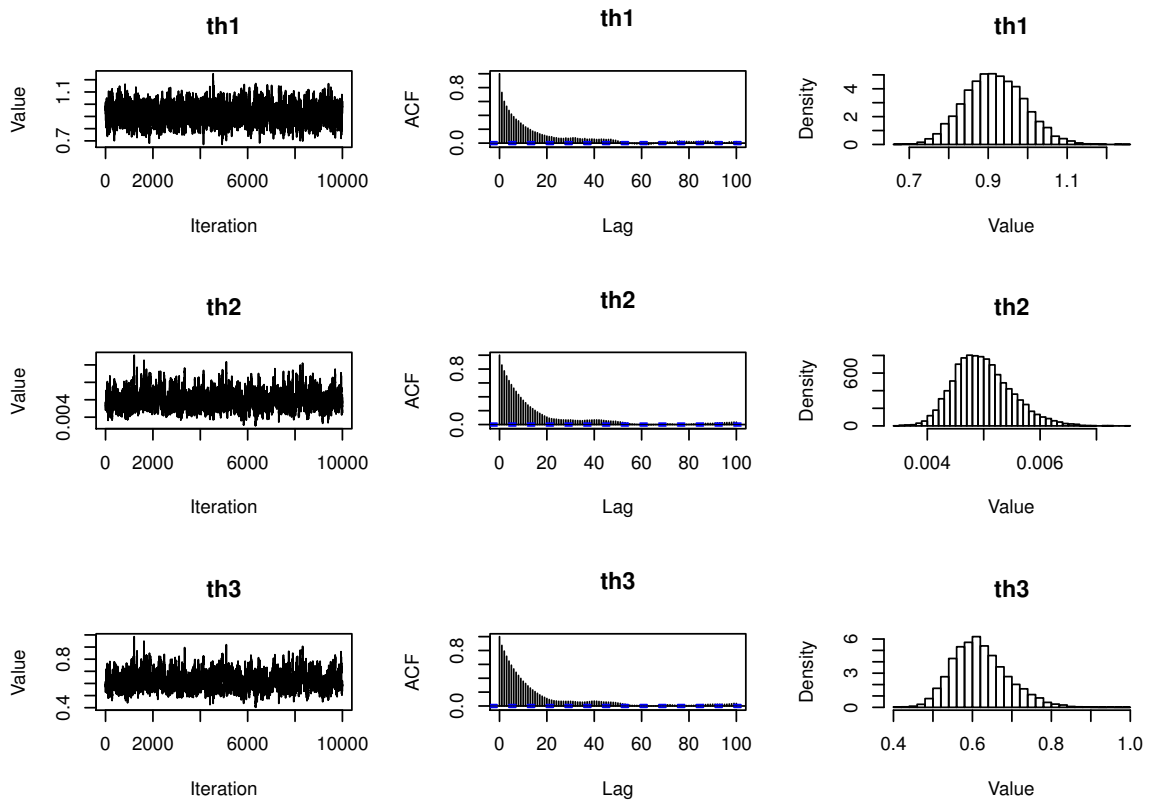
Simulated time series data set consisting of 16 equally spaced observations subject to Gaussian measurement error with a standard deviation of 10.

Marginal posteriors for the Lotka-Volterra model



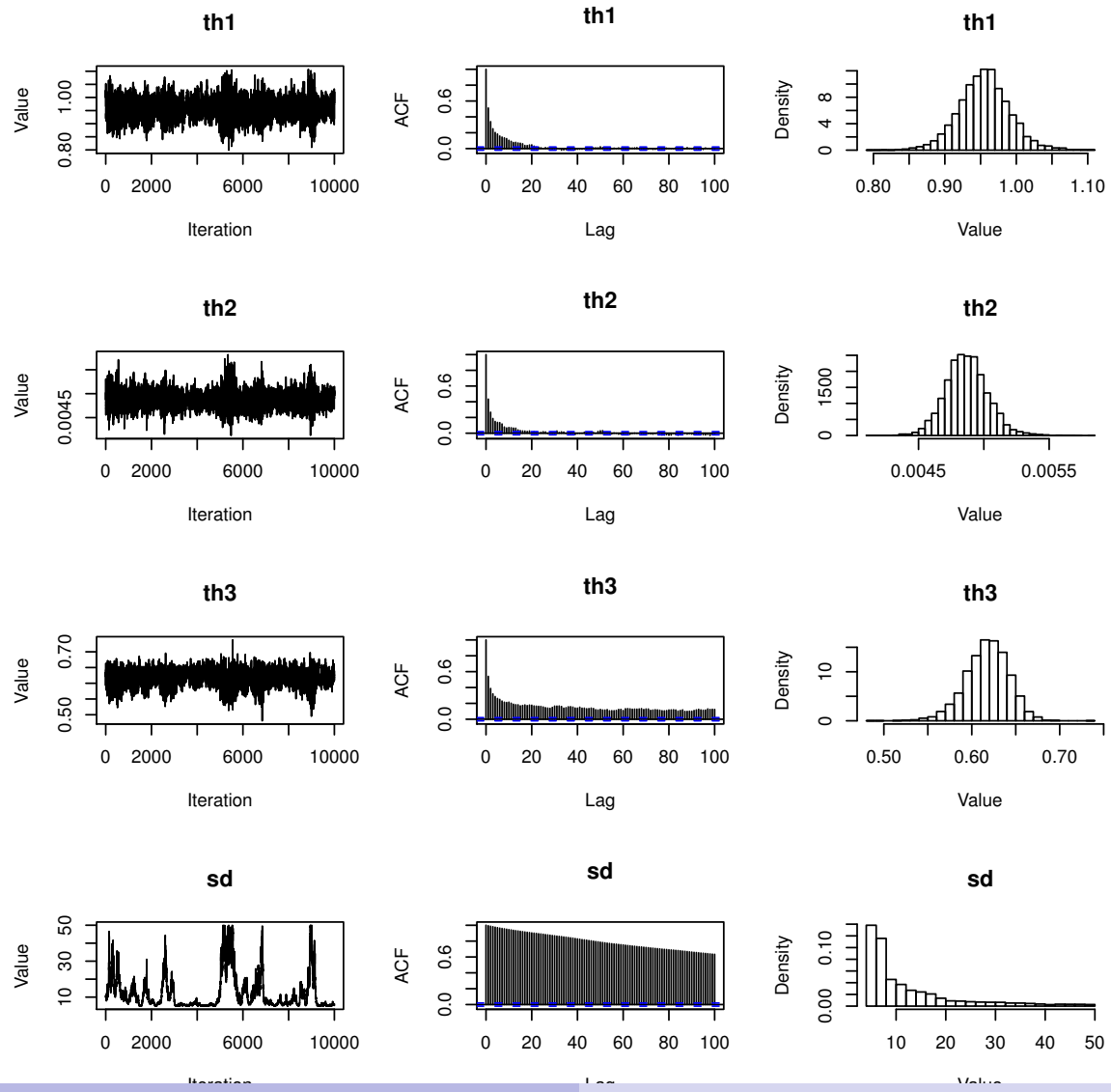
Note that the true parameters, $\theta = (1, 0.005, 0.6)$ are well identified by the data

Marginal posteriors observing only prey



Note that the mixing of the MCMC sampler is reasonable, and that the true parameters, $\theta = (1, 0.005, 0.6)$ are quite well identified by the data

Marginal posteriors for unknown measurement error



R package: smfsb

- Free, open source, well-documented software package for R on CRAN, `smfsb`, associated with Stochastic modelling for systems biology, second edition
- Code for stochastic simulation and of (biochemical) reaction networks (Markov jump processes and chemical Langevin), and pMCMC-based Bayesian inference for POMP models
- Full installation and "getting started" instructions at `http://tinyurl.com/smfsb2e`
- Once the package is installed and loaded, running `demo("PMCMC")` at the R prompt will run a PMMH algorithm for the Lotka-Volterra model discussed here

Hitting the data...

- The "bootstrap" PMMH algorithm works well in many cases, and is extremely general (works for any Markov process)
- In the case of no measurement error, the probability of "hitting" the data (and accepting) is very small (possibly zero), and so the mixing of the MCMC scheme is very poor
- ABC (approximate Bayesian computation) strategy is to accept if

$$\|x^\star_{t+1} - y_{t+1}\| < \varepsilon$$

  but this forces a trade-off between accuracy and efficiency which can be unpleasant (cf. noisy ABC)
- Same problem in the case of low measurement error
- Particularly problematic in the context of high-dimensional data
- Would like a strategy which copes better in this case

The chemical Langevin equation (CLE)

- The CLE is a diffusion approximation to the true Markov jump process
- Start with the time change representation

$$X_t - X_0 = S\,N\left(\int_0^t h(X_\tau, c)d\tau\right)$$

  and approximate $N_i(t) \simeq t + W_i(t)$, where $W_i(t)$ is an independent Wiener process for each i
- Substituting in and using a little stochastic calculus gives:

The CLE as an Itô SDE:

$$dX_t = Sh(X_t, c)\,dt + \sqrt{S\,\mathrm{diag}\{h(X_t, c)\}S'}\,dW_t$$

Improved particle filters for SDEs

- The "bootstrap" particle filter uses blind forward simulation from the model
- If we are able to evaluate the "likelihood" of sample paths, we can use other proposals
- The particle filter weights then depend on the Radon-Nikodym derivative of law of the proposed path wrt the true conditioned process
- For SDEs, the weight will degenerate unless the proposed process is absolutely continuous wrt the true conditioned process
- Ideally we would like to sample from $\pi(\mathbf{x}^\star_{t+1}|c^\star, x^\star_t, y_{t+1})$, but this is not tractable for nonlinear SDEs such as the CLE

Modified diffusion bridge (MDB)

- Need a tractable process $q(\mathbf{x}^\star_{t+1}|c^\star, x^\star_t, y_{t+1})$ that is locally equivalent to $\pi(\mathbf{x}^\star_{t+1}|c^\star, x^\star_t, y_{t+1})$
- Diffusion $dX_t = \mu(X_t)dt + \beta(X_t)^{\frac{1}{2}}dW_t$
- The nonlinear diffusion bridge

$$dX_t = \frac{x_1 - X_t}{1-t}dt + \beta(X_t)^{\frac{1}{2}}dW_t$$

  hits x_1 at $t = 1$, yet is locally equivalent to the true diffusion as it has the same diffusion coefficient
- This forms the basis of an efficient proposal; see Durham & Gallant (2002), Chib, Pitt & Shephard (2004), Delyon & Hu (2006), and Stramer & Yan (2007) for technical details

General SIR particle filter

- At time t, we have (after resampling) an equally weighted sample from $\pi(x_t|y_{1:t})$
- At time $t+1$, we want a weighted sample from $\pi(x_{t+1}|y_{1:t+1})$, though in fact useful to construct a sample from $\pi(x_{t+1}, x_t|y_{1:t+1})$, and then marginalise down as required
- Target $\propto \pi(y_{t+1}|x_{t+1})\pi(x_{t+1}|x_t)\pi(x_t|y_{1:t})$ and proposal is $f(x_{t+1}|x_t, y_{t+1})\pi(x_t|y_{1:t})$, for some $f(\cdot)$, leading to unnormalised weight

$$w_{t+1} = \frac{\pi(y_{t+1}|x_{t+1})\pi(x_{t+1}|x_t)}{f(x_{t+1}|x_t, y_{t+1})}$$

- LF choice is $f(x_{t+1}|x_t, y_{t+1}) = \pi(x_{t+1}|x_t)$, otherwise need to evaluate the discrete time transition density

Weights and RN derivatives

- For diffusions with intractable kernels, make the target $\pi(\mathbf{x}_{t+1}, x_t | y_{1:t+1})$ and then marginalise down to $\pi(x_{t+1}|y_{1:t+1})$ if required
- The proposal path will be of the form $f(\mathbf{x}_{t+1}|x_t, y_{t+1})\pi(x_t|y_{1:t})$, leading to weight

$$w_{t+1} = \pi(y_{t+1}|x_{t+1})\frac{\pi(\mathbf{x}_{t+1}|x_t)}{f(\mathbf{x}_{t+1}|x_t, y_{t+1})}$$

- The expected weight is $\pi(y_{t+1}|y_{1:t})$, as needed for pseudo-marginal MCMC
- Formally,

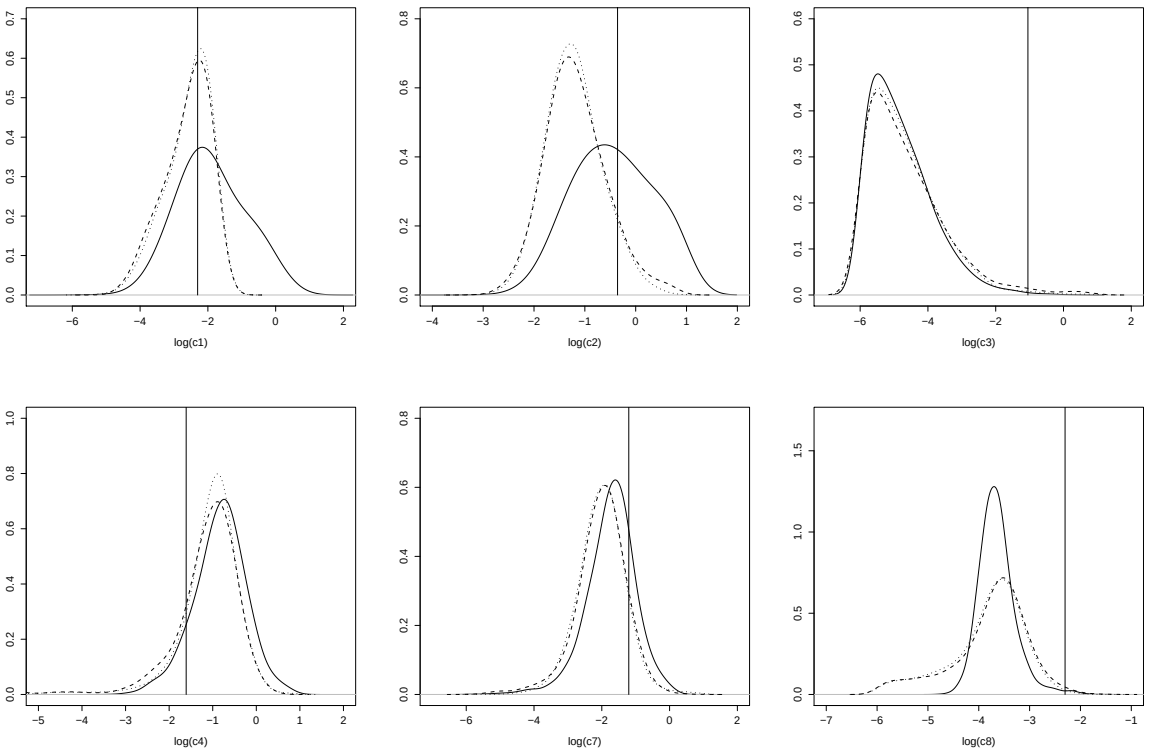
$$w_{t+1} = \pi(y_{t+1}|x_{t+1})\frac{d\mathbb{P}}{d\mathbb{Q}}(\mathbf{x}_{t+1}|x_t),$$

the RN derivative of the true (unconditioned) diffusion wrt the proposal process

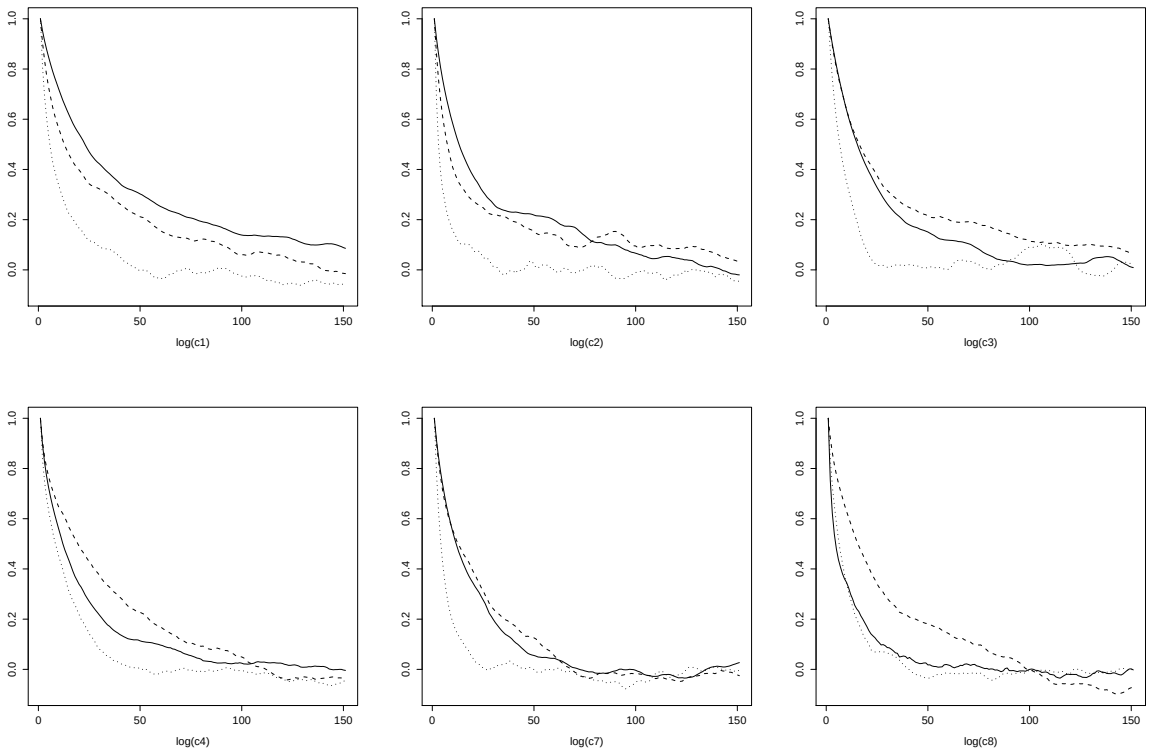
Discrete bridge construction

- In practice, we sample proposed paths using a fine discretisation of the proposal process, and evaluate the (log) likelihoods on the simulated skeleton of the process
- We could simulate from an Euler-Maruyama discretisation of the SDE governing the bridging process, but it turns out to be more efficient to use variants of the modified diffusion bridge (MDB) of Durham and Gallant (2002)
- Details of constructions, and application to PMMH pMCMC for intractable diffusions given in Golightly and W (2011)

Marginal posteriors for the auto-regulatory network



ACFs for the auto-regulatory network



Summary

- Exact LF MCMC inference is possible for intractable nonlinear Markov processes, provided that there is sufficient noise in the observation process
- In low/no noise scenarios, efficient MCMC algorithms for diffusions are still possible — one approach is to use tractable diffusion bridges as proposals in an SMC-within-MCMC algorithm
- This is one approach, but there are many others — exact and approximate
- The relative merits of the various possible approaches are currently poorly understood

References

Golightly, A. and D. J. Wilkinson (2011) Bayesian parameter inference for stochastic biochemical network models using particle MCMC. *Interface Focus*, **1**(6):807–820.

Wilkinson, D. J. (2009) Stochastic modelling for quantitative description of heterogeneous biological systems, *Nature Reviews Genetics*. **10**(2):122-133.

Wilkinson, D. J. (2010) Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology (with discussion), in J.-M. Bernardo et al (eds) *Bayesian Statistics 9*, OUP, pp.679–706.

Wilkinson, D. J. (2011) *Stochastic Modelling for Systems Biology, second edition*. Chapman & Hall/CRC Press.

Contact details...

email: `darren.wilkinson@ncl.ac.uk`

www: `tinyurl.com/darrenjw`