(Particle) Markov chain Monte Carlo

T.J. McKinley (t.mckinley@exeter.ac.uk)

Inference and prediction

Once the model is fitted and the model fit assessed, we can use the model / parameter estimates in various ways:

- **Inference**: interpreting the parameter estimates.
- **Prediction**: to predict what might happen if the outbreak were to occur under the same conditions again.
- **Forecasting**: to predict what might happen in the future, based on data available now.

So what’s the problem?

This is straightforward right?

So what's the problem?

This is straightforward right?

The likelihood for compartmental models relies on having **exact** observations of event **times** and **types**. In practice events are rarely observed in detail:

- **Surveillance**: e.g. under-reporting, imperfect coverage, imperfect diagnosis, mis-diagnosis;
- **Rounding error**: e.g. data often collated daily / weekly;
- **Hidden states**: some epidemiological processes never observed (e.g. you might know *roughly* when you started feeling sick with flu, but not when you were infected or when you became infectious).

So what's the problem?

This is straightforward right?

The likelihood for compartmental models relies on having **exact** observations of event **times** and **types**. In practice events are rarely observed in detail:

- **Surveillance**: e.g. under-reporting, imperfect coverage, imperfect diagnosis, mis-diagnosis;
- **Rounding error**: e.g. data often collated daily / weekly;
- **Hidden states**: some epidemiological processes never observed (e.g. you might know *roughly* when you started feeling sick with flu, but not when you were infected or when you became infectious).

Dealing with these challenges is **hard**! (But we will have a go!)

Intractable likelihoods

To deal with the **partially observed** data, we can introduce a set of **latent** variables, $\mathbf{x} = (\mathbf{t}, \delta)$, where $\mathbf{t}$ is a vector of **hidden** event *times*, and δ is a vector of **hidden** event *types*.

Then the **likelihood** can be expressed as:

$$f(\mathbf{y} \mid \theta) = \int_{\mathcal{X}} f(\mathbf{y} \mid \mathbf{x}, \theta) f(\mathbf{x} \mid \theta) d\mathbf{x},$$

where

- $f(\mathbf{y} \mid \mathbf{x}, \theta)$ is an **observation** process (or **measurement error** / **model discrepancy**);
- $f(\mathbf{x} \mid \theta)$ is the **likelihood function** based on the **latent** variables $\mathbf{x}$.

Intractable likelihoods

$$f(\mathbf{y} \mid \theta) = \int_{\mathcal{X}} f(\mathbf{y} \mid \mathbf{x}, \theta) f(\mathbf{x} \mid \theta) d\mathbf{x},$$

This **marginalises** (*averages*) across the hidden variables $\mathbf{x}$.

This is a complex integral, over all possible combinations of events, and all possible event times consistent with the data.

It may also be the case that the **number** of hidden events is **unknown**, in which can we have to repeat the integration for every possible number of hidden events.

Data augmentation

One approach is therefore to include the **hidden** variables $\mathbf{x}$ as **additional parameters** in the model.

We can then estimate the ***joint* posterior** distribution for $(\theta, \mathbf{x})$, and then derive the ***marginals*** for the parameters of interest (θ) *numerically*.

This is usually done using MCMC methods; an approach known as **data-augmented MCMC** (e.g. Gibson and Renshaw 1998; Philip D. O'Neill and Roberts 1999; Jewell et al. 2009).

It is very powerful, but difficult to code, scale and optimise.

Simulation-based approaches

Alternatively, we can build inference algorithms around **simulating** directly from the model-of-interest, and then searching for parameter sets that are more consistent with the **observed data**.

These **simulation-based methods** are also powerful and flexible:

- Don't have to store all of the latent variables (so memory requirements are lower).
- Are often straightforward to parallelise.
- Simulation can often be easier than calculating the likelihood.
- Implementation often easier than DA (e.g. "plug-and-play")

However, there are also practical difficulties:

- The probability of matching the data exactly (i.e. getting a non-zero likelihood) is often very low.
- Often require some form of approximation to obtain a match.

Alternative fitting methods

Examples of latent variable methods:

- **Data-augmented MCMC** (e.g. Gibson and Renshaw 1998; Philip D. O'Neill and Roberts 1999; S. Cauchemez and Ferguson 2008; Jewell et al. 2009)
- **Sequential Monte Carlo** (Simon Cauchemez et al. 2008)

Examples of simulation-based methods:

- **Maximum likelihood via iterated filtering** (Ionides, Bretó, and King 2006)
- **Approximate Bayesian Computation** (e.g. Toni et al. 2009; McKinley, Cook, and Deardon 2009; Conlan et al. 2012; Brooks Pollock, Roberts, and Keeling 2014)
- **Pseudo-marginal methods** (e.g. P. D. O'Neill et al. 2000; Beaumont 2003; Andrieu and Roberts 2009; McKinley et al. 2014)
- **Particle MCMC** (Andrieu, Doucet, and Holenstein 2010; Drovandi, Pettitt, and McCutchan 2016)
- **Synthetic likelihood** (Wood 2010)
- **History matching** (with **emulation**) (e.g. Andrianakis et al. 2015; McKinley et al. 2018)

Pseudo-marginal MCMC

Require: $\theta^{(0)}$.

for $i = 1, \ldots, n$ **do**

Propose **candidate** $\theta' \sim q\left(\cdot \mid \theta^{(i-1)}\right)$.

Calculate the **acceptance probability**:

$$\alpha = \min\left(1, \frac{\hat{f}\left(\mathbf{y} \mid \theta'\right) f\left(\theta'\right)}{\hat{f}\left(\mathbf{y} \mid \theta^{(i-1)}\right) f\left(\theta^{(i-1)}\right)} \times \frac{q\left(\theta^{(i-1)} \mid \theta'\right)}{q\left(\theta' \mid \theta^{(i-1)}\right)}\right)$$

Sample $u \sim U(0, 1)$

if $u < \alpha$ **then**

$\theta^{(i)} = \theta'$

else

$\theta^{(i)} = \theta^{(i-1)}$

end if

end for

One option is to simply plug this **estimate** into a standard Metropolis-Hastings algorithm in place of the true likelihood.

Remarkably, as long as this estimate is **unbiased**, this will still converge to the **true** posterior.

This approach is known as **pseudo-marginal MCMC**.

Beaumont (2003); Andrieu and Roberts (2009).

Simulation-based approximations

One option is to replace the likelihood, $f(\mathbf{y} \mid \theta)$, by a **Monte Carlo** estimate:

$$
\begin{aligned}
f(\mathbf{y} \mid \theta) &= \int_{\mathcal{X}} f(\mathbf{y} \mid \mathbf{x}, \theta) f(\mathbf{x} \mid \theta) d\mathbf{x} \\
&\approx \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{y} \mid \mathbf{x}_i, \theta),
\end{aligned}
$$

where $\mathbf{x}_i \sim f(\mathbf{x} \mid \theta)$ are simulations from the underlying model.

This provides an **unbiased** estimate for $f(\mathbf{y} \mid \theta)$.

Efficiency of pseudo-marginal MCMC

The efficiency (i.e. **mixing**) of pseudo-marginal MCMC relies on the **variance** of the **estimator** $\hat{f}(\mathbf{y} \mid \theta)$.

- If the variance is **small**, then mixing will be **improved**.
- If the variance is **large**, then mixing will be **poor**.

We can reduce the variance by:

- increasing the number of simulations $M \rightarrow$ higher computational burden;
- improving the estimator.

Particle MCMC

This leads on to the idea of **particle MCMC** (Andrieu, Doucet, and Holenstein 2010).

In essence this aims to use **Sequential Monte Carlo**[†] to produce an **unbiased** estimate of the likelihood that has **lower variance** than a vanilla Monte Carlo estimate.

One of the earliest and most widely used particle filters is known as the **bootstrap particle filter** (Gordon, Salmond, and Smith 1993).

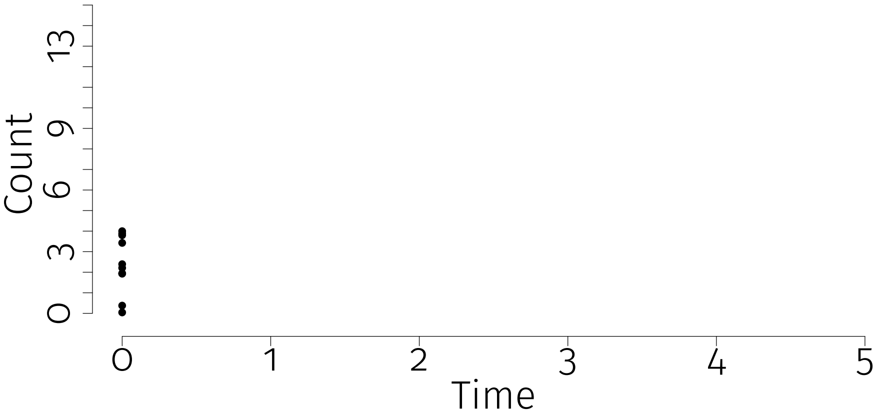
[†] i.e. **particle filtering**

Bootstrap particle filter

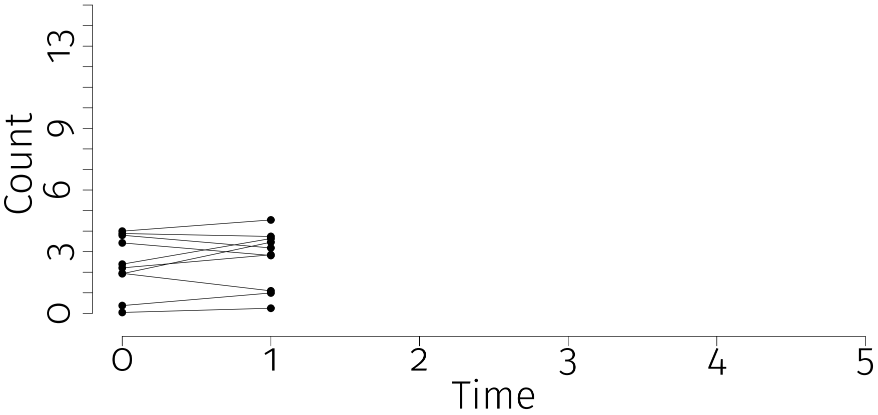
Each **particle** now corresponds to the **unobserved states** of the system at time 0, $\mathbf{x}_0 = (\mathbf{x}_0^1, \ldots, \mathbf{x}_0^M)$. The parameters are **fixed**.

1. Each particle m is propagated forwards in time by **simulating** from the model $\mathbf{x}_1^m \sim f(\mathbf{x} \mid \mathbf{x}_0^m, \theta)$.
2. Each new particle is **weighted** according to the **observation process**, $f(\mathbf{y} \mid \mathbf{x}_1^m, \theta)$.
3. These weights are **normalised**, and a **re-sampling** step undertaken.
4. The new set of particles are propagated forwards to time $t + 1$ and so on...

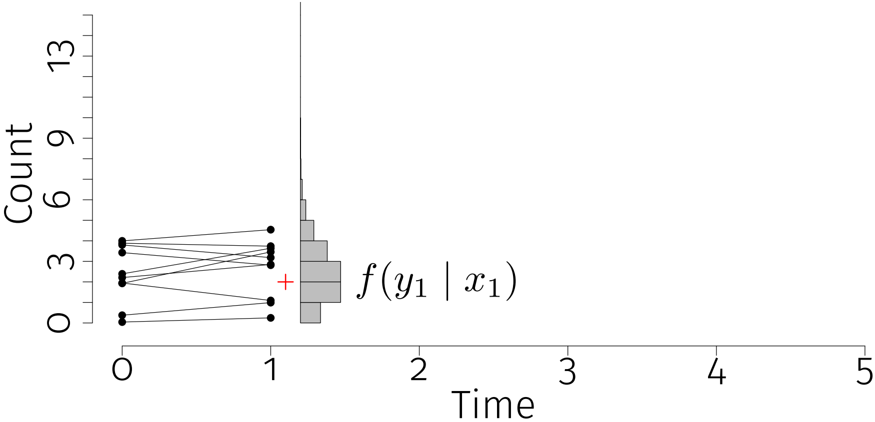
Bootstrap particle filter



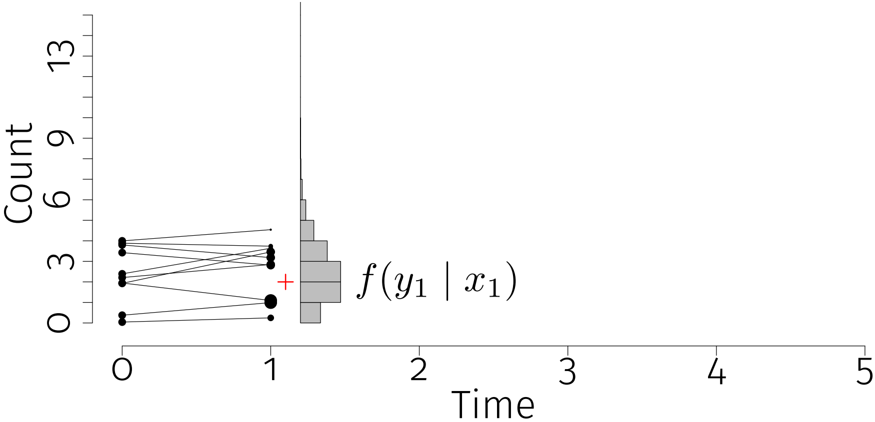
Bootstrap particle filter



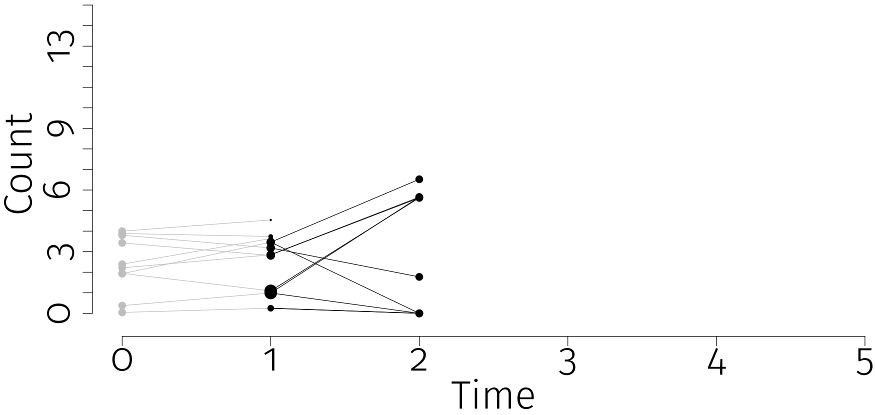
Bootstrap particle filter



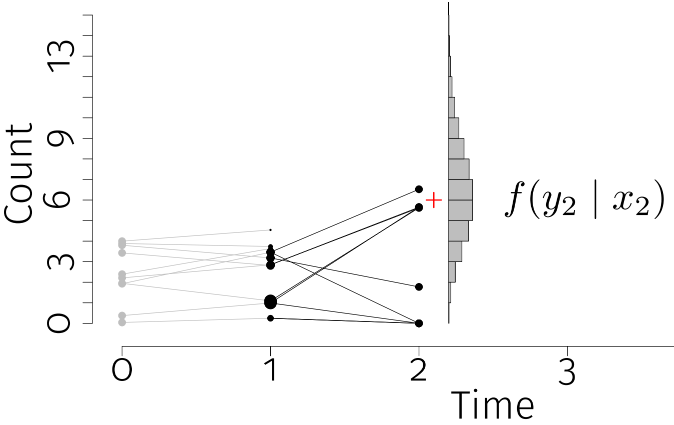
Bootstrap particle filter



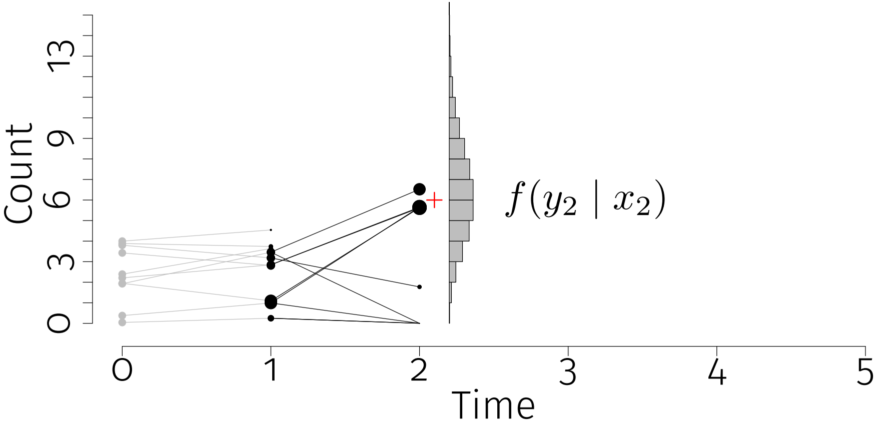
Bootstrap particle filter



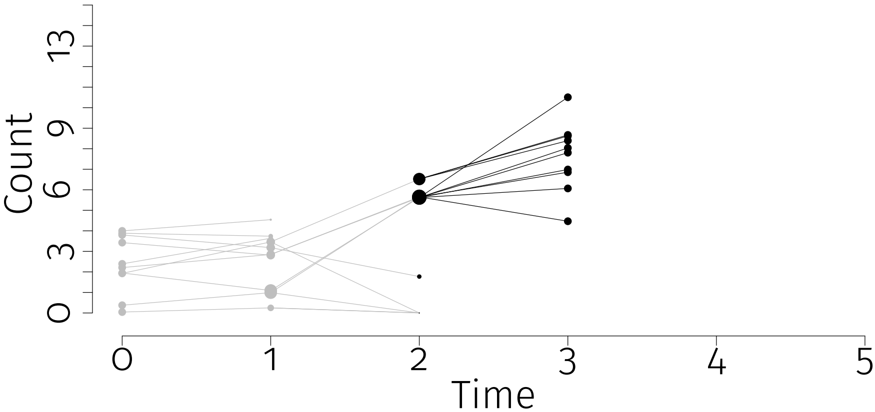
Bootstrap particle filter



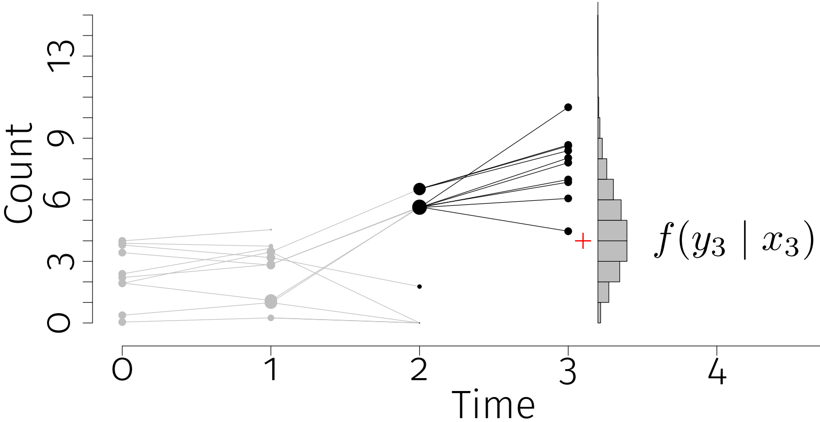
Bootstrap particle filter



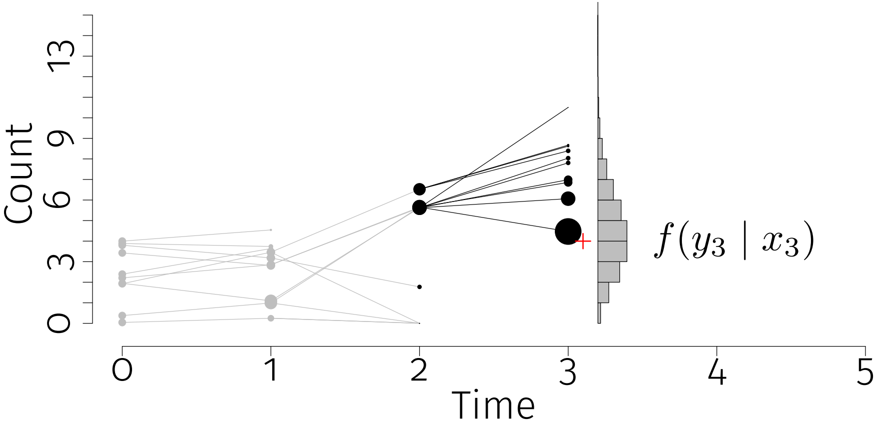
Bootstrap particle filter



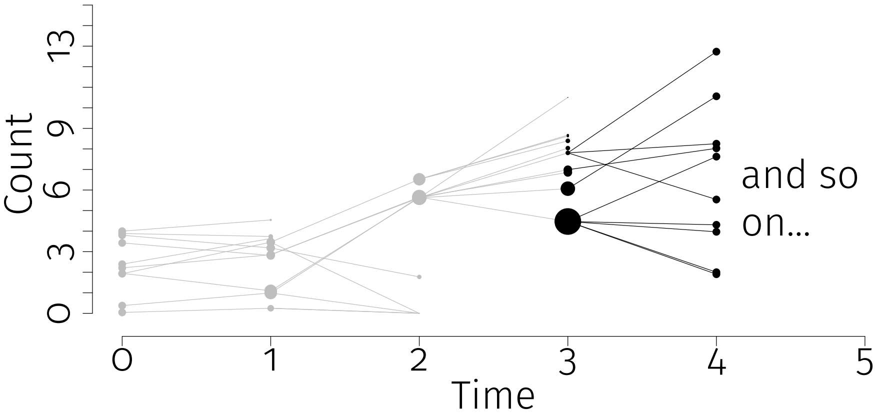
Bootstrap particle filter



Bootstrap particle filter



Bootstrap particle filter



Bootstrap particle filter

We can generate an **unbiased estimate** of the conditional densities:

$$\hat{f}\left(y_t \mid y_{0:(t-1)}\right) = \frac{1}{M}\sum_{m=1}^{M} f\left(y_t \mid \mathbf{x}_t^m, \theta\right),$$

where $y_{0:(t-1)}$ corresponds to the observed time-series counts at time $t_0, t_1, \dots, t_{t-1}$.

It turns out that we can also derive an **unbiased** estimate of the overall **likelihood** as:

$$\hat{f}\left(\mathbf{y} \mid \theta\right) = f\left(y_0\right)\prod_{t=1}^{T}\hat{f}\left(y_t \mid y_{0:(t-1)}\right).$$

Particle MCMC

Hence we can generate an **unbiased** estimate of the likelihood which **numerically** integrates over the **hidden states**.

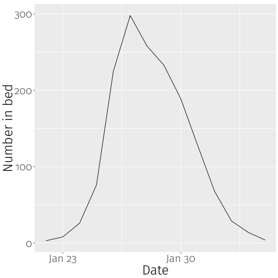
We can then plug this estimate into a standard Metropolis-Hastings algorithm to produce a **pseudo-marginal** MCMC routine that will converge to the *correct posterior distribution in probability*.

This approach only requires a **simulation** model, and an **observation process**.

The bootstrap particle filter we've used is defined for **time-series** counts, and can be extended in various ways.

Example: flu in boarding school

To illustrate some of these ideas we can use a case study of influenza in a boarding school. These data are from a paper in the BMJ in 1978 (Anonymous 1978) and provided in the `outbreaks` package. We use a simple $SIRR_1$ model:



The event probabilities are:

$$P\left[S_{t+\delta t} = S_t - 1, I_{t+\delta t} = I_t + 1\right] \approx \beta SI/N$$
$$P\left[I_{t+\delta t} = I_t - 1, R_{t+\delta t} = R_t + 1\right] \approx \gamma I$$
$$P\left[R_{t+\delta t} = R_t - 1, R_{1,t+\delta t} = R_{1,t} + 1\right] \approx \gamma_1 R$$

Example: flu in boarding school

Here we will place a Poisson error process around the R curve, such that:

$$R_t \sim \text{Po}(R'_t + 10^{-6}),$$

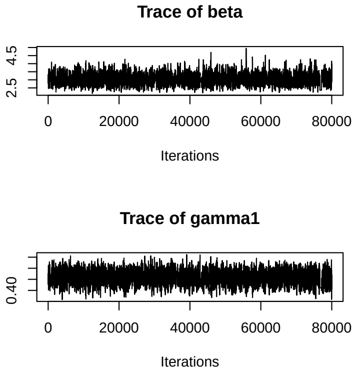
where R_t is the **observed** R count at time t, R'_t is the simulated count[†].

The initial population size is 763 pupils, and we assume an initial introduction of infection of a single child at day 0.

[†] see e.g. Funk et al. (2016) or here for similar ideas in practice

Example: flu in boarding school

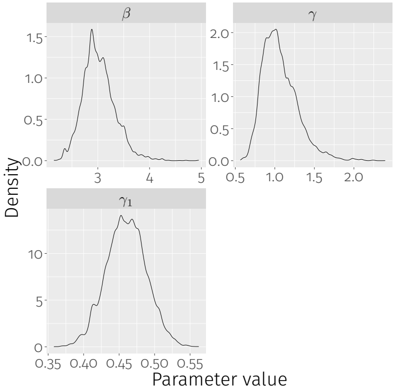
We ran a PMCMC algorithm for 100,000 iterations, discarding the first 20,000 as burn-in. We used 75 particles for the particle filter.



Example: flu in boarding school

Summaries of the marginal posterior distributions are:

Parameter	Mean	2.5%	97.5%
β	3	2.5	3.7
γ	1.1	0.73	1.6
γ_1	0.46	0.41	0.52



Summary

Particle MCMC is a powerful approach for inference in **partially observed** systems (see e.g Wilkinson 2012 or his associated blog for fantastic explanations of these methods).

It is often used when there is some form of **stochastic** discrepancy / observation process mapping the **hidden** states to the **observed** states.

Other particle filters exist, such as the **Alive Particle Filter** (Jasra et al. 2013), and the system can be extended to the **ABC** setting, where approximate matching around data points is used (Drovandi, Pettitt, and Lee 2014; McKinley et al. 2020).

Software

Partially Observed Markov Processes:

- `pomp`
- `SimBIID`†
- `SimInf`‡
- `nimble`§
- `hmer`¶

†designed mostly for teaching purposes, but should work for simple models
‡now implements ABC-SMC (e.g. Toni et al. 2009; McKinley, Cook, and Deardon 2009)
§now supports state-space models (although I've not used it for these)
¶hot-off-the-press! Implements emulation and history matching for epidemic models

References i

Andrianakis, Ioannis, Ian Vernon, Nicky McCreesh, Trevelyan J. McKinley, Jeremy E. Oakley, Rebecca N. Nsubuga, Michael Goldstein, and Richard G. White. 2015. "Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda." *PLoS Computational Biology* 11 (1): e1003968.

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein. 2010. "Particle Markov Chain Monte Carlo Methods." *Journal of the Royal Statistical Society, Series B (Methodological)* 72 (3): 269–342.

Andrieu, Christophe, and Gareth O. Roberts. 2009. "The Pseudo-Marginal Approach for Efficient Monte Carlo Simulation." *The Annals of Statistics* 37 (2): 697–725.

References ii

Anonymous. 1978. “Influenza in a Boarding School.” *British Medical Journal* 1: 578.

Beaumont, Mark A. 2003. “Estimation of Population Growth and Decline in Genetically Monitored Populations.” *Genetics* 164: 1139–60.

Brooks Pollock, Ellen, Gareth O. Roberts, and Matt J. Keeling. 2014. “A Dynamic Model of Bovine Tuberculosis Spread and Control in Great Britain.” *Nature* 511: 228–31. https://doi.org/10.1038/nature13529.

Cauchemez, S., and Neil M. Ferguson. 2008. “Likelihood-Based Estimation of Continuous-Time Epidemic Models from Time-Series Data: Application to Measles Transmission in London.” *Journal of the Royal Society Interface* 5 (25): 885–97.

Cauchemez, Simon, Alain-Jacques Valleron, Pierre-Yves Boëlle, Antoine Flahault, and Neil M. Ferguson. 2008. “Estimating the Impact of School Closure on Influenza Transmission from Sentinel Data.” *Nature* 452: 750–55. https://doi.org/10.1038/nature06732.

Conlan, Andrew J. K., Trevelyan J. McKinley, Katerina Karolemeas, Ellen Brooks Pollock, Anthony V. Goodchild, Andrew P. Mitchell, Colin P. D. Birch, Richard S. Clifton-Hadley, and James L. N. Wood. 2012. “Estimating the Hidden Burden of Bovine Tuberculosis in Great Britain.” *PLoS Computational Biology* 8 (10): e1002730.

Drovandi, Christopher C., Anthony N. Pettitt, and Anthony Lee. 2014. “Bayesian Indirect Inference Using a Parametric Auxiliary Model.” *Statistical Science* 30 (1): 72–95.

References iv

Drovandi, Christopher C., Anthony N. Pettitt, and Roy A. McCutchan. 2016. "Exact and Approximate Bayesian Inference for Low Integer-Valued Time Series Models with Intractable Likelihoods." *Bayesian Analysis* 11 (2): 325–52.

Funk, Sebastian, Adam J. Kucharski, Anton Camacho, Rosalind M. Eggo, Laith Yakob & Lawrence M. Murray, and W. John Edmunds. 2016. "Comparative Analysis of Dengue and Zika Outbreaks Reveals Differences by Setting and Virus." *PLoS Neglected Tropical Diseases* 10 (12): e0005173.

Gibson, Gavin J., and Eric Renshaw. 1998. "Estimating Parameters in Stochastic Compartmental Models Using Markov Chain Methods." *IMA Journal of Mathematics Applied in Medicine and Biology* 15: 19–40.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith. 1993. "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation." *Radar and Signal Processing, IEE Proceedings F.* 140 (2): 107–13. https://doi.org/10.1049/ip-f-2.1993.0015.

Ionides, E. L., C. Bretó, and A. A. King. 2006. "Inference for Nonlinear Dynamical Systems." *Proceedings of the National Academy of Sciences USA* 103: 18438–43.

Jasra, Ajay, Anthony Lee, Christopher Yau, and Xiaole Zhang. 2013. "The Alive Particle Filter." https://arxiv.org/abs/1304.0151.

Jewell, Chris P., Theodore Kypraios, Peter Neal, and Gareth O. Roberts. 2009. "Bayesian Analysis for Emerging Infectious Diseases." *Bayesian Analysis* 4 (4): 465–96.

McKinley, Trevelyan J., Alex R. Cook, and Robert Deardon. 2009. "Inference in Epidemic Models Without Likelihoods." *The International Journal of Biostatistics* 5 (1). https://doi.org/10.2202/1557-4679.1171.

McKinley, Trevelyan J., Peter Neal, Simon E. F. Spencer, Andrew J. K. Conlan, and Laurence Tiley. 2020. "Efficient Bayesian Model Choice for Partially Observed Processes: With Application to an Experimental Transmission Study of an Infectious Disease." *Bayesian Analysis* 15 (3): 839–70. https://doi.org/10.1214/19-BA1174.

McKinley, Trevelyan J., Joshua V. Ross, Rob Deardon, and Alex R. Cook. 2014. "Simulation-Based Bayesian Inference for Epidemic Models." *Computational Statistics and Data Analysis* 71: 434–47.

VACCINE IMPACT MODELLING CONSORTIUM

McKinley, Trevelyan J., Ian Vernon, Ioannis Andrianakis, Nicky McCreesh, Jeremy E. Oakley, Rebecca N. Nsubuga, Michael Goldstein, and Richard G. White. 2018. "Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models." *Statistical Science* 33 (1): 4–18. https://doi.org/10.1214/17-STS618.

O'Neill, P. D., D. J. Balding, N. G. Becker, M. Eerola, and D. Mollison. 2000. "Analyses of Infectious Disease Data from Household Outbreaks by Markov Chain Monte Carlo Methods." *Applied Statistics* 49: 517–42.

O'Neill, Philip D., and Gareth O. Roberts. 1999. "Bayesian Inference for Partially Observed Stochastic Epidemics." *Journal of the Royal Statistical Society. Series A (General)* 162: 121–29.

Toni, Tina, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P. H. Strumpf. 2009. "Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems." *Journal of the Royal Society Interface* 6: 187–202.

Wilkinson, Darren J. 2012. *Stochastic Modelling for Systems Biology*. 2nd ed. Chapman; Hall / CRC.

Wood, Simon N. 2010. "Statistical Inference for Noisy Nonlinear Ecological Dynamic Systems." *Nature* 466: 1102–4.