Genealogies of Sequential Monte Carlo Algorithms

Suzie Brown
with Jere Koskela, Adam Johansen and Paul Jenkins
Department of Statistics, University of Warwick

Sequential Monte Carlo

SMC[1] is an effective Monte Carlo method for models with strong correlation between parameters. For example, inferring a sequence of underlying parameters from an autocorrelated time series.

We approximate posterior distributions by samples of "particles". The algorithm starts with N equally-weighted particles and repeats the steps:
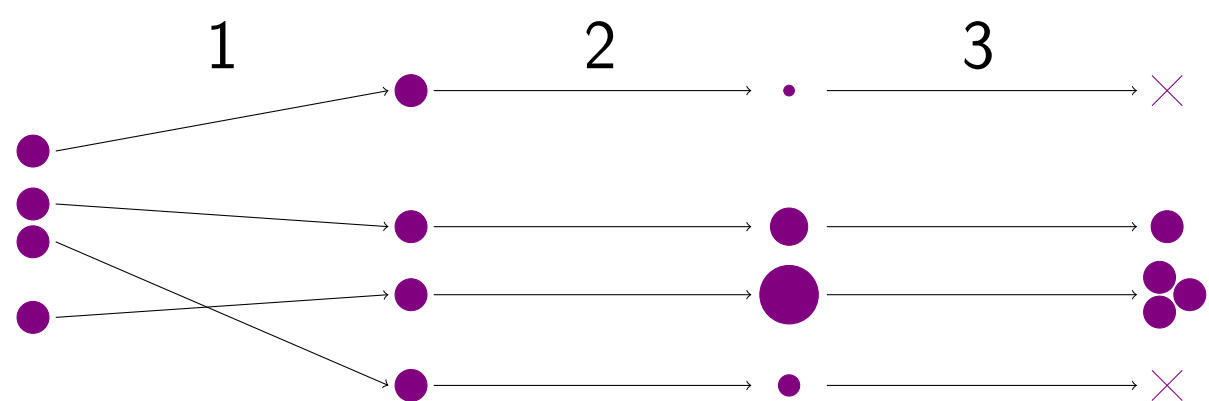
1. **Propagate:** update particle positions for the next parameter
2. **Weight:** weight particles according to agreement with the observation
3. **Resample:** duplicate high-weight particles and kill off low-weight particles to obtain a new sample of N particles.

The figure shows a particle approximation for a tractable model. The exact posterior mode and 95% HPD set are obtained with the Kalman filter.
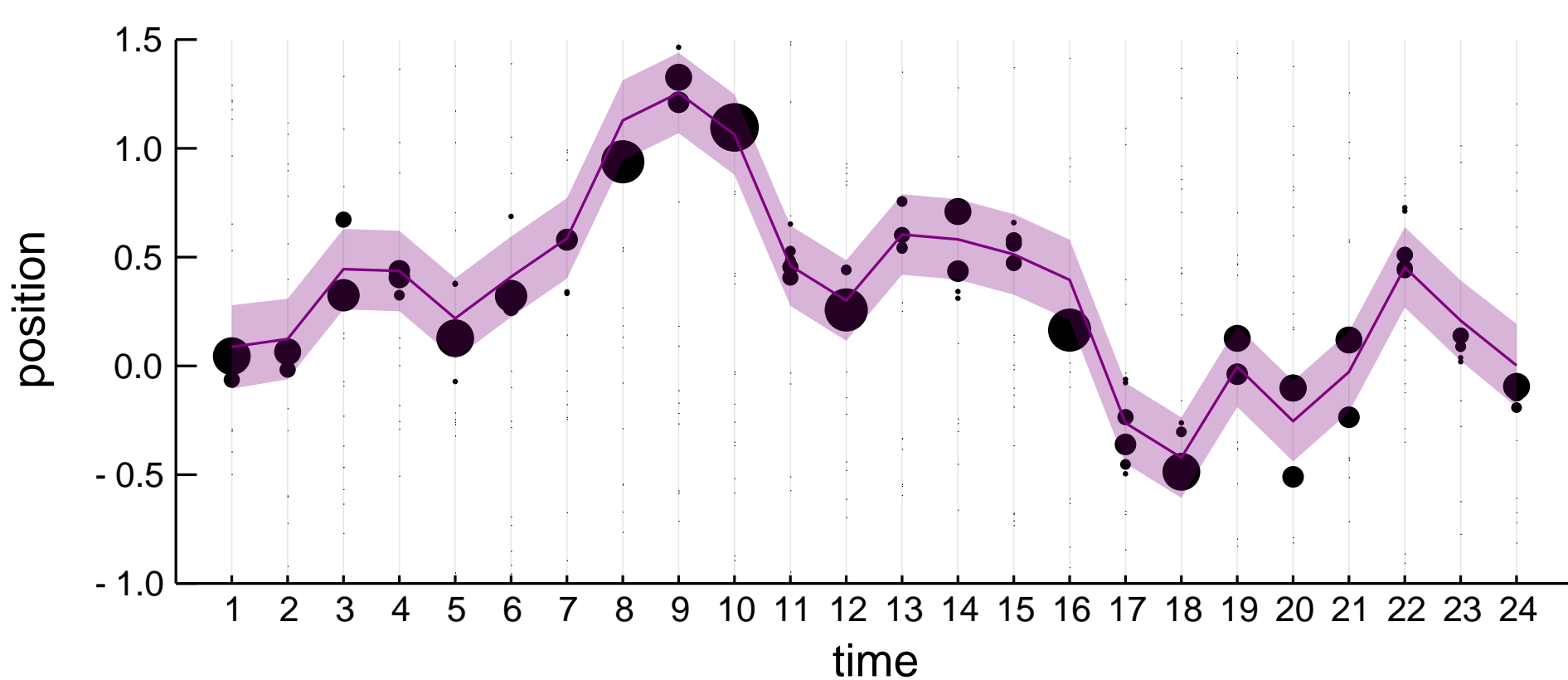

Figure: Exact posterior (purple) and weighted SMC particles before resampling (black) for a linear Gaussian HMM.

Ancestral Degeneracy

For some applications, we want to approximate the *joint* distribution over all parameters. We have N samples from this distribution: the "trajectories" of each of the N particles.

Due to resampling these trajectories overlap, coalescing backwards in time. At some point all the trajectories coalesce, and we approximate that parameter's distribution with just one sample!
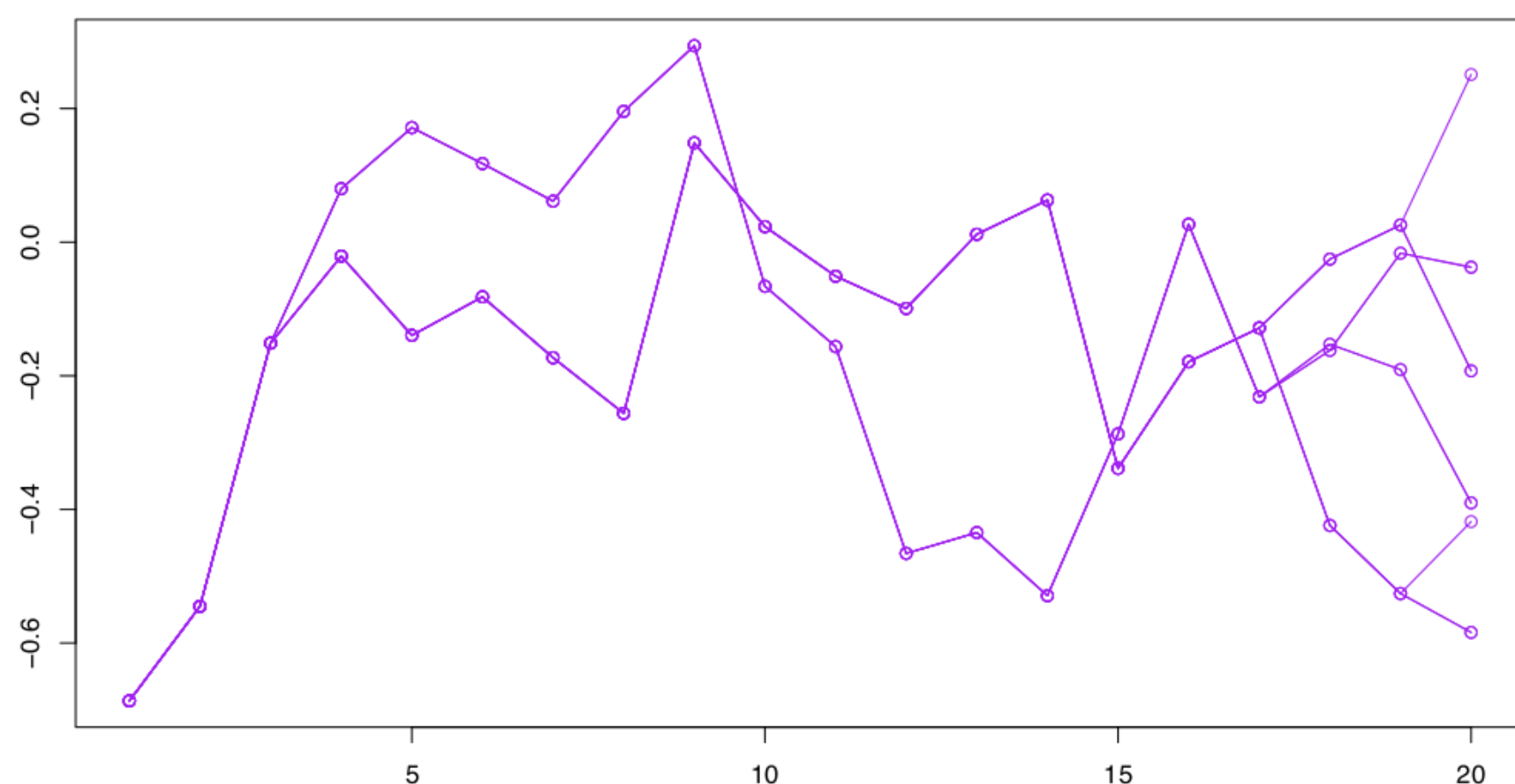

Figure: Trajectories from a sample of N=6 particles. At most points there are just one or two distinct samples.

This phenomenon is known as *ancestral degeneracy*[2], suggesting that the joint structure of the trajectories can be viewed as a *genealogy* of particles. It is one of the main difficulties of SMC, so we would like to characterise it.

The Kingman Coalescent[3]

Kingman's N-coalescent is the coalescent process on a population of size N in which each pair of lineages merges with rate 1.

The Kingman coalescent arises as an extension of the N-coalescent in the $N \to \infty$ limit. It is the limiting genealogy for a large class of population models.
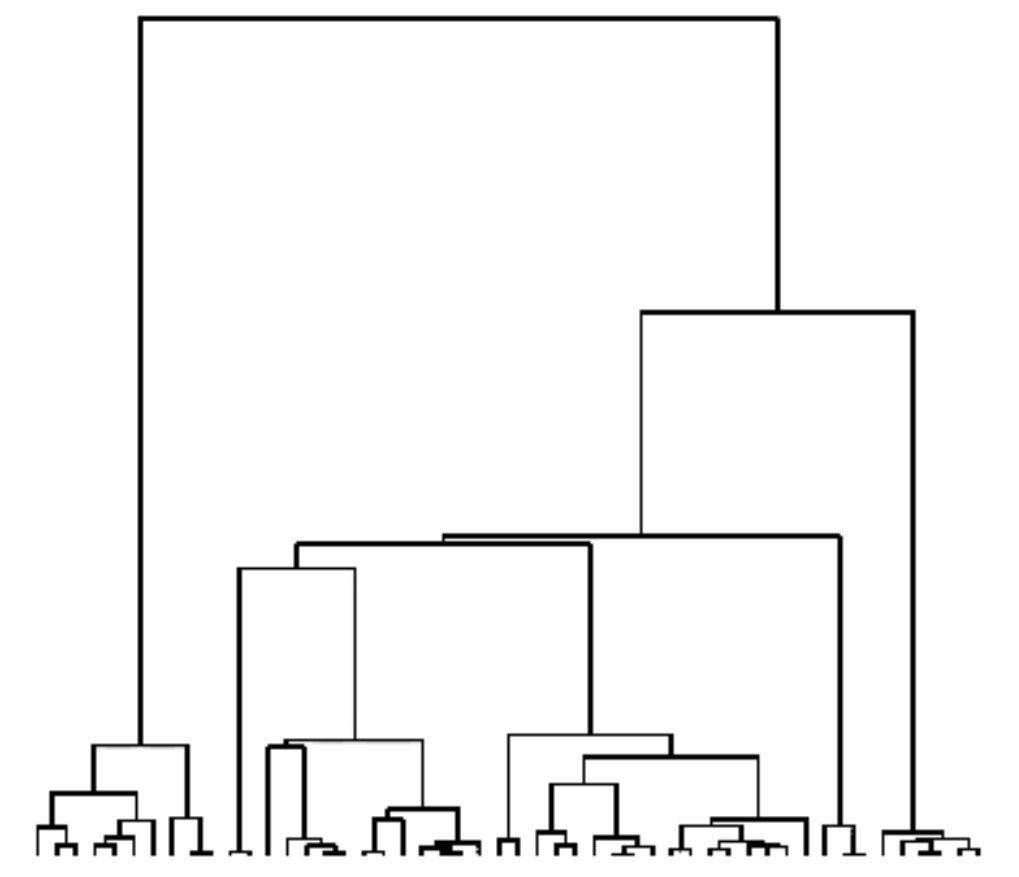
A realisation of the 50-coalescent.
Source: Wikimedia Commons

SMC Genealogies

The genealogical structure revolves around the number of offspring of each particle upon resampling, denoted $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$ for each time t. A key quantity in the asymtotic analysis is

$$c_N(t) := \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}\left[(\nu_t^{(i)})_2\right]$$

which becomes the coalescence rate as $N \to \infty$.
To obtain convergence to the Kingman coalescent, we rescale time by the inverse coalescence rate to achieve the required unit rate.

It has been shown that, under such a rescaling and some standard assumptions, the genealogies induced by multinomial resampling converge (in the sense of finite-dimensional distributions) to the Kingman coalescent[4]. Furthermore, simulations suggest that the same structure appears even for finite N.

We have since extended the result to cover conditional SMC[5] with multinomial resampling. In this algorithm, a certain "immortal trajectory" is conditioned to survive all the resampling steps. Somewhat surprisingly, this still admits the Kingman coalescent as its limiting genealogy.

1. Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE:F* 140(3), 107-113. IET Digital Library.
2. Doucet, A., & Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 656-704. OUP.
3. Kingman, J.F.C. (1982). The coalescent. *Stochastic processes and their applications*, 13(3), 235-248.
4. Koskela, J. et al. (2018). Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo. arXiv:1804.01811.
5. Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *RSS:B*, 72(3), 269-342.