

Genealogies of Sequential Monte Carlo Algorithms

Suzie Brown

supervised by Jere Koskela, Adam Johansen, Paul Jenkins, Dario Spanó

Department of Statistics, University of Warwick

Sequential Monte Carlo

Suppose we have a hidden Markov model:

$$\begin{array}{ccccccc} & & Y_{t-1} & & Y_t & & \\ & & \uparrow g & & \uparrow g & & \\ \cdots & \xrightarrow{f} & X_{t-1} & \xrightarrow{f} & X_t & \xrightarrow{f} & \cdots \end{array}$$

where $Y_{1:T}$ are noisy observations of unobservable states $X_{1:T}$. We may want to infer filtering distributions $p(x_t|y_{1:t})$, or smoothing distributions $p(x_{1:t}|y_{1:t})$. These are not available analytically, except in linear Gaussian models.

SMC approximates the posterior distributions by drawing a sample of N particles from the prior distribution and iterating the following steps:

1. **Propagate:** move particles according to the Markov kernel f
2. **Calculate weights:** weight particles according to how likely they are to produce the observations through g
3. **Resample:** duplicate high-weight particles and kill off low-weight particles to obtain a new sample of size N .

The figure shows how the population of particles looks at each step, before resampling. The purple ribbon shows the exact posterior mode and 95% credible interval.

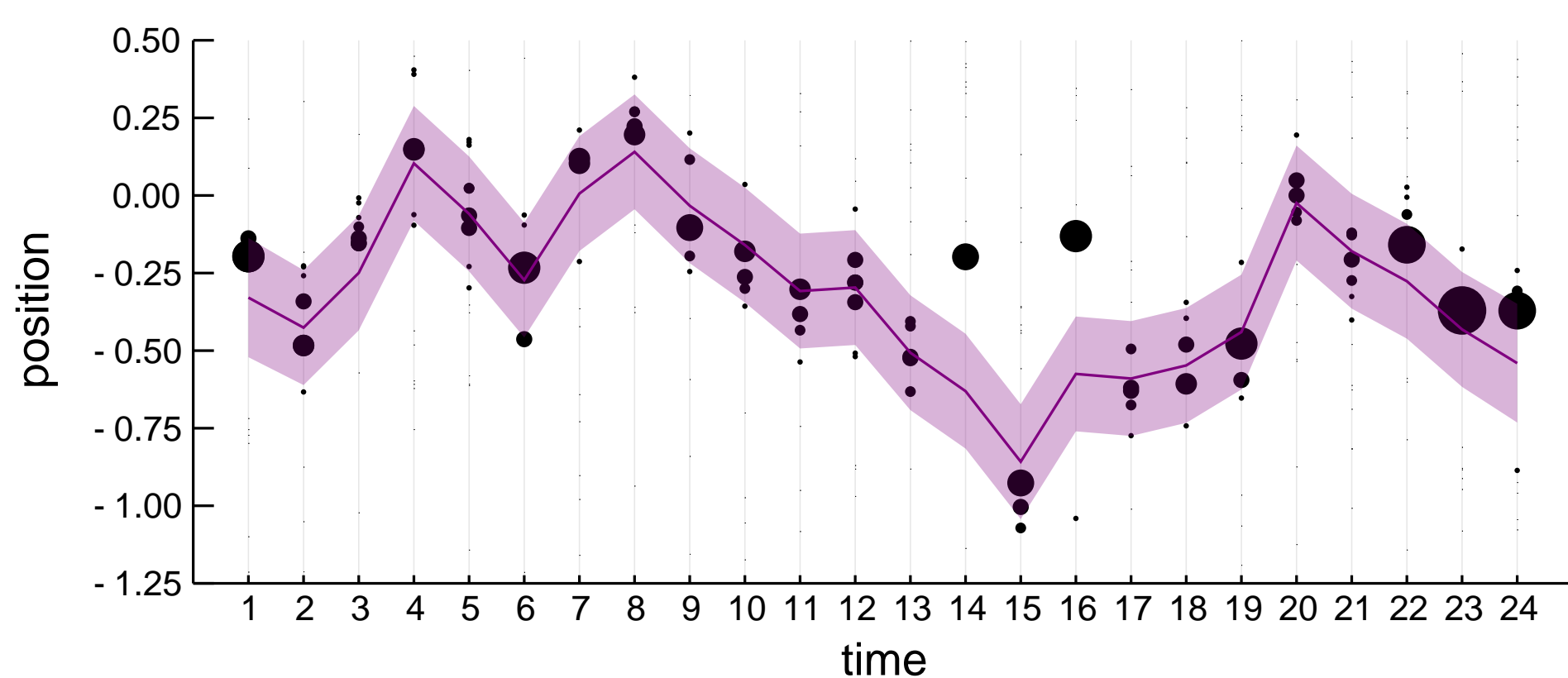


Figure: Exact posterior (purple) and weighted SMC particles before resampling (black) for a linear Gaussian model.

Ancestral Degeneracy

However, these particles cannot approximate the smoothing distributions. For that we need a sample of *trajectories* covering all t time steps, not just a sample of particles at each time step.

We have a sample of N such trajectories: the ancestral trajectories of each of the N particles alive at time t . But due to the resampling mechanism that generates these ancestries, they coalesce backwards in time, leaving much fewer than N distinct samples at time 0.

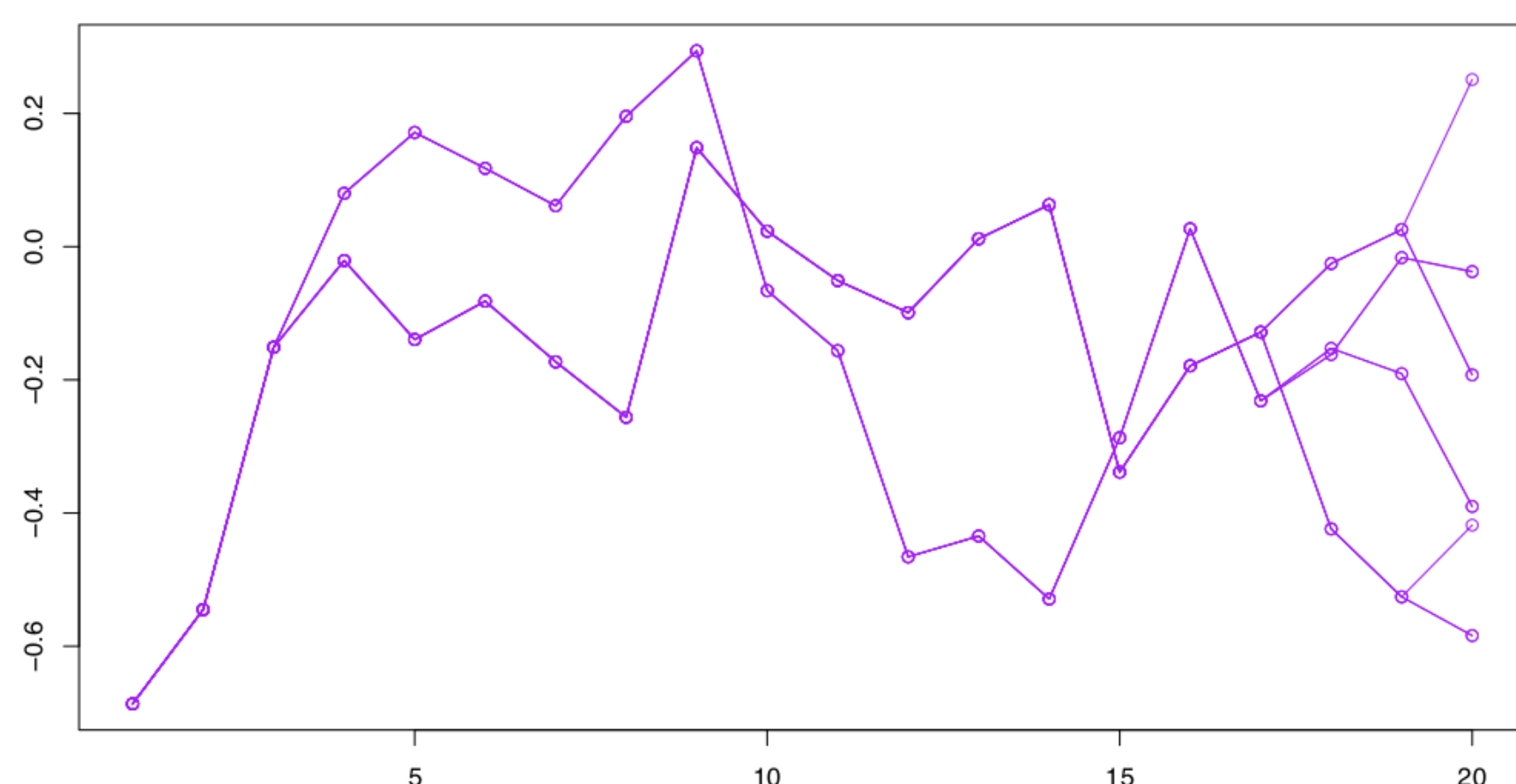


Figure: Trajectories arising from a sample of $N=6$ particles. Except at the last few time steps there are no more than two distinct lineages.

This phenomenon is known as *ancestral degeneracy*. It can be mitigated by changing the resampling procedure, but it cannot be eradicated.

Kingman Coalescent

The Kingman coalescent is a continuous-time process which describes the asymptotic coalescents of many population models.

Each pair of lineages merges with rate 1, meaning that more mergers happen when there are more distinct lineages. The process starts with a population of size N and ends when all of the lineages have merged into one.

The resulting coalescent has many lineages for only a short time, and spends most of its time with just two or three distinct lineages.