Genealogies of Sequential Monte Carlo Algorithms

Suzie Brown, Jere Koskela, Adam Johansen

Department of Statistics, University of Warwick

Sequential Monte Carlo

Suppose we have a hidden Markov model:

$$\begin{array}{ccc} Y_{t-1} & & Y_t \\ | & & | \\ \cdots \longrightarrow X_{t-1} \longrightarrow X_t \longrightarrow \cdots \end{array}$$

where $Y_{0:T}$ are noisy observations of the unobservable states $X_{0:T}$. We may want to infer the filtering distributions $p(x_t|y_{1:t})$, or the smoothing distributions $p(x_{0:t}|y_{0:t})$. In general, none of these are available analytically, except for linear Gaussian models.

SMC generates samples from the posterior distributions by starting with a sample of N particles from the prior distribution and then iterating the following steps:

1. **Propagate:**
2. **Calculate weights:**
3. **Resample:**

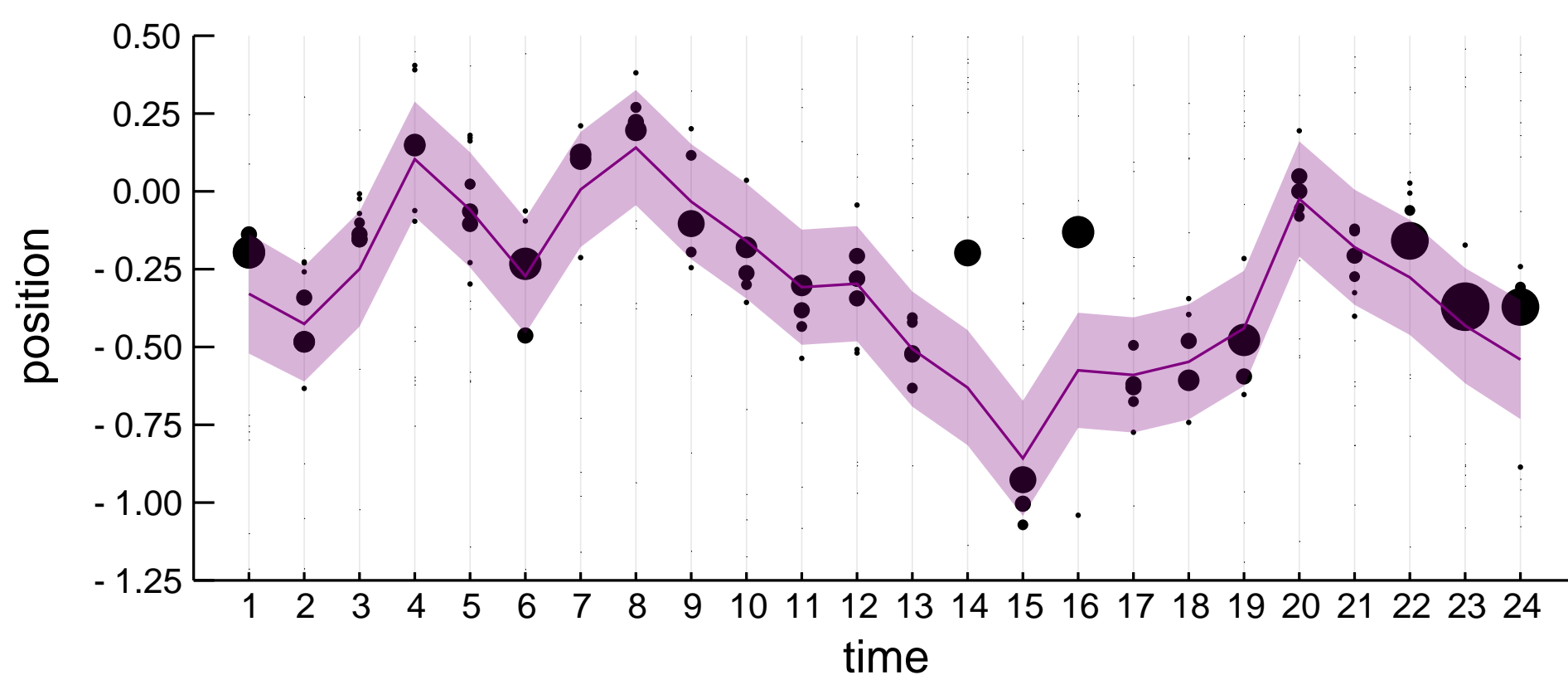The figure shows how the population of particles looks at each step, before resampling.

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.
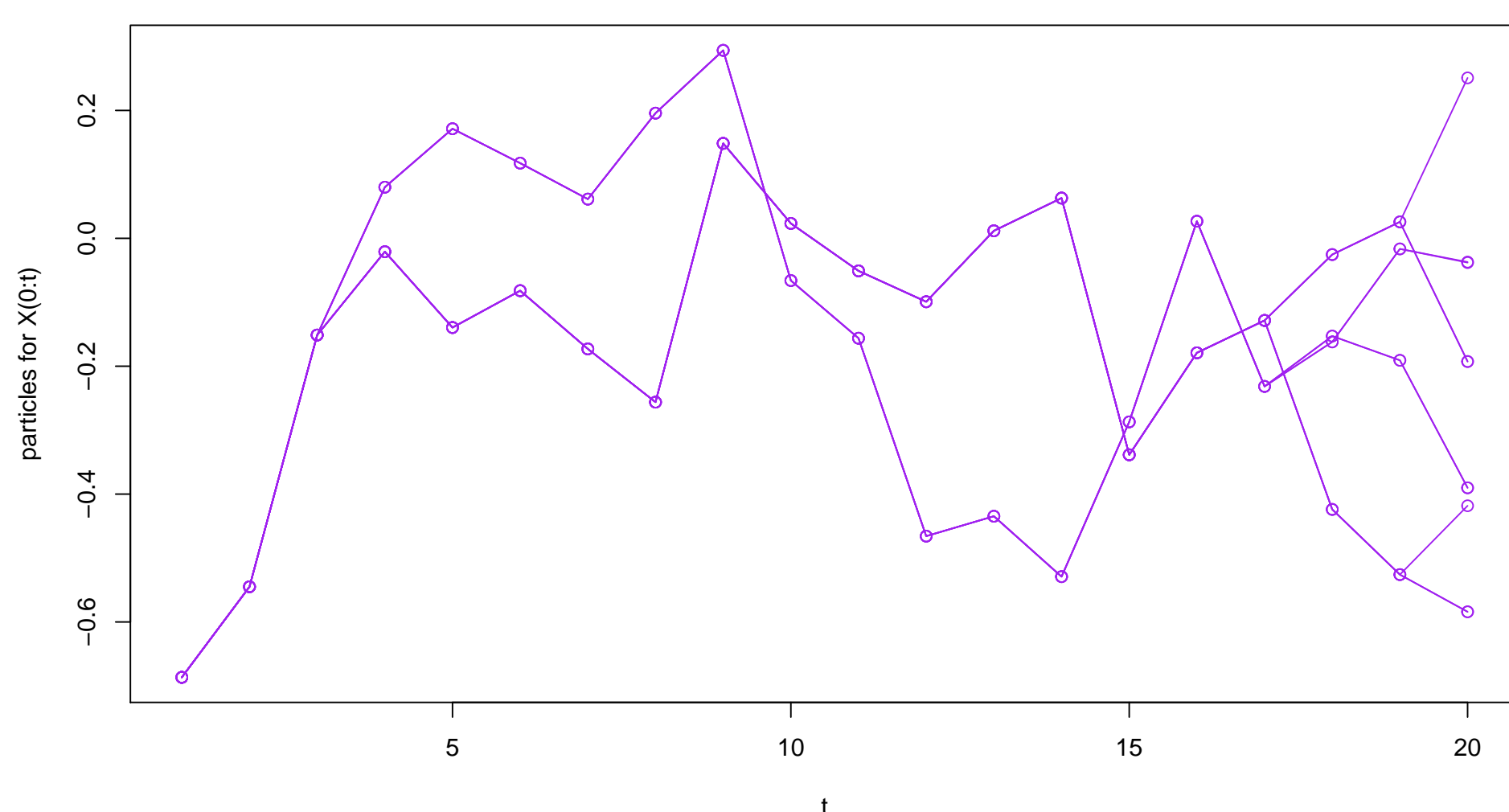

Figure: blah blah

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