Genealogies of sequential Monte Carlo algorithms

Suzie Brown

27 October 2020

Outline

1. Sequential Monte Carlo
2. Resampling and degeneracy
3. Genealogies

Sequential Monte Carlo

- Want to sample from a sequence of intractable target distributions
- Typical setting: dimension of target increases in time, or strong dependence between consecutive targets (so MCMC is impractical)
- SMC can obtain exact draws, and thus approximate expectations

State space models

$$
\begin{tikzcd}
 & Y_t & Y_{t+1} & \\
\cdots \arrow[r] & X_t \arrow[u] \arrow[r] & X_{t+1} \arrow[u] \arrow[r] & \cdots
\end{tikzcd}
$$

$$
\begin{aligned}
&X_0 \sim \mu(\cdot) \\
&X_{t+1} \mid (X_t = x_t) \sim f_t(\cdot|x_t) \\
&Y_t \mid (X_t = x_t) \sim g_t(\cdot|x_t)
\end{aligned}
$$

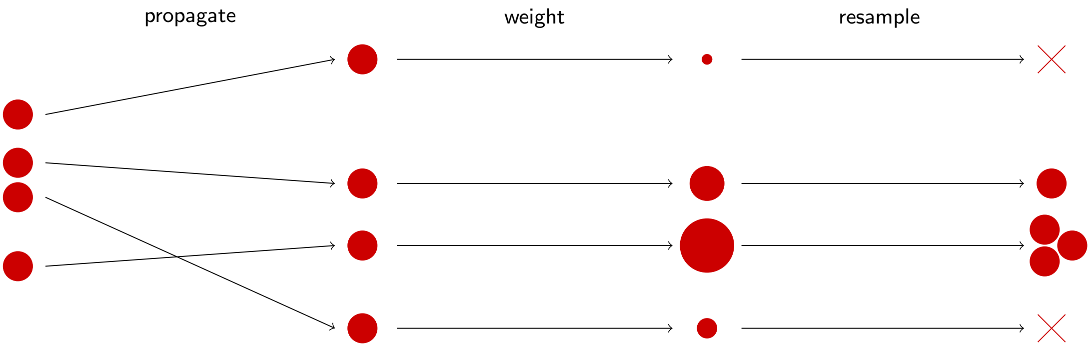
May want to infer $(t < T)$:

$p(x_{1:T} \mid y_{1:t})$	"prediction"
$p(x_{1:t} \mid y_{1:t})$	"filtering"
$p(x_{1:t} \mid y_{1:T})$	"smoothing"

Importance Sampling

Sequential Monte Carlo

Illustration



Resampling and Genealogies

- Resampling creates a genealogy (family tree) of particles
- Properties of the genealogy affect performance of the SMC algorithm
- Different resampling schemes give different forms of genealogies
- Basic quantity for analysing genealogies is the pair coalescence probability

Coalescence Probability

Definition

The probability that a randomly chosen pair of particles at generation t share a common ancestor at generation $(t-1)$

$$c_N = \frac{1}{N(N-1)} \sum_{i=1}^{N} v_i(v_i - 1)$$

Coalescence Probability

Example

Consider the case where we have only two particles ($N = 2$)

$$c_2 = \frac{1}{2}\left[v_1(v_1 - 1) + v_2(v_2 - 1)\right]$$

The expectation of c_2 conditional on knowing the weights (w_1, w_2) is

$$\begin{aligned} c_2 &= \frac{1}{2}\mathbb{E}[v_1(v_1 - 1) \mid w_{1:2}] + \frac{1}{2}\mathbb{E}[v_2(v_2 - 1) \mid w_{1:2}] \\ &= \mathbb{P}[v_1 = 2 \mid w_{1:2}] + \mathbb{P}[v_2 = 2 \mid w_{1:2}] \end{aligned}$$

Coalescence Probability

Example

- We proved that asymptotically (as $N \to \infty$) residual resampling dominates multinomial in terms of expected coalescence probability
- We also proved it in cases $N = 2$ and $N = 3$

- We proved that asymptotically (as $N \to \infty$) residual resampling dominates multinomial in terms of expected coalescence probability
- We also proved it in cases $N = 2$ and $N = 3$
- We conjecture that it holds for all finite N too
- It just remains to prove it for $N = 4, 5, \ldots$

- We proved that asymptotically (as $N \to \infty$) residual resampling dominates multinomial in terms of expected coalescence probability
- We also proved it in cases $N = 2$ and $N = 3$
- We conjecture that it holds for all finite N too
- It just remains to prove it for $N = 4, 5, \dots$
- We proved that systematic resampling (and some others) dominate multinomial in expected coalescence probability, for all N.

THE END