

# Genealogies of sequential Monte Carlo algorithms

Suzie Brown

27 October 2020

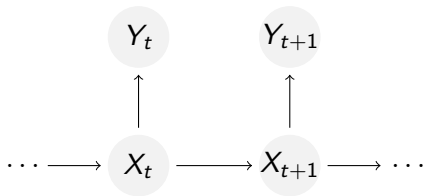
# Outline

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

# Sequential Monte Carlo

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

# State space models

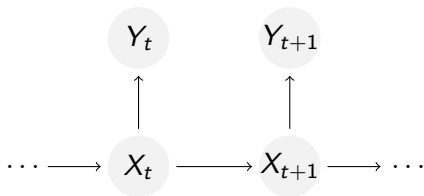


$$X_0 \sim \mu(\cdot)$$

$$X_{t+1} \mid (X_t = x_t) \sim f_t(\cdot \mid x_t)$$

$$Y_t \mid (X_t = x_t) \sim g_t(\cdot \mid x_t)$$

# State space models



$$X_0 \sim \mu(\cdot)$$

$$X_{t+1} \mid (X_t = x_t) \sim f_t(\cdot \mid x_t)$$

$$Y_t \mid (X_t = x_t) \sim g_t(\cdot \mid x_t)$$

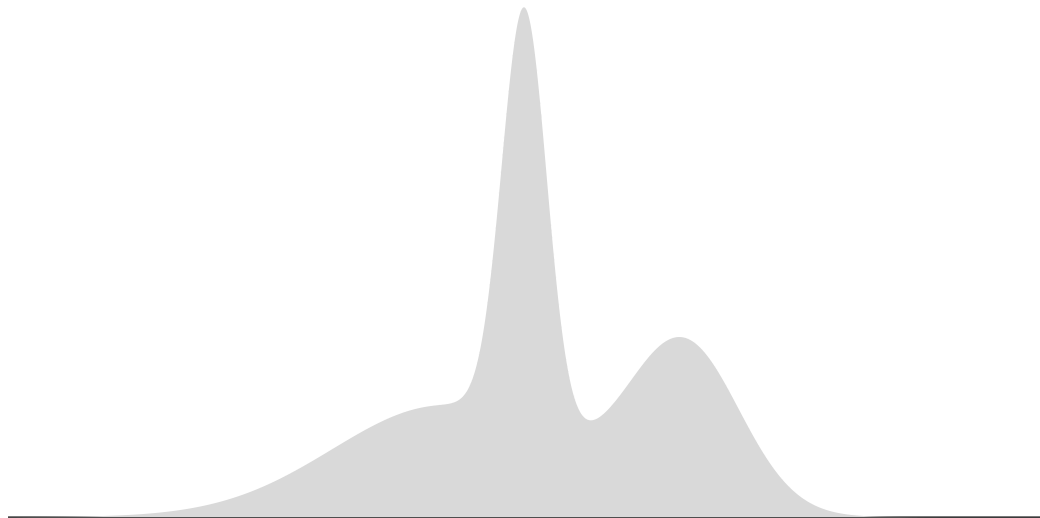
May want to infer ( $s < t$ ):

$p(x_t \mid y_{1:s})$  “prediction”

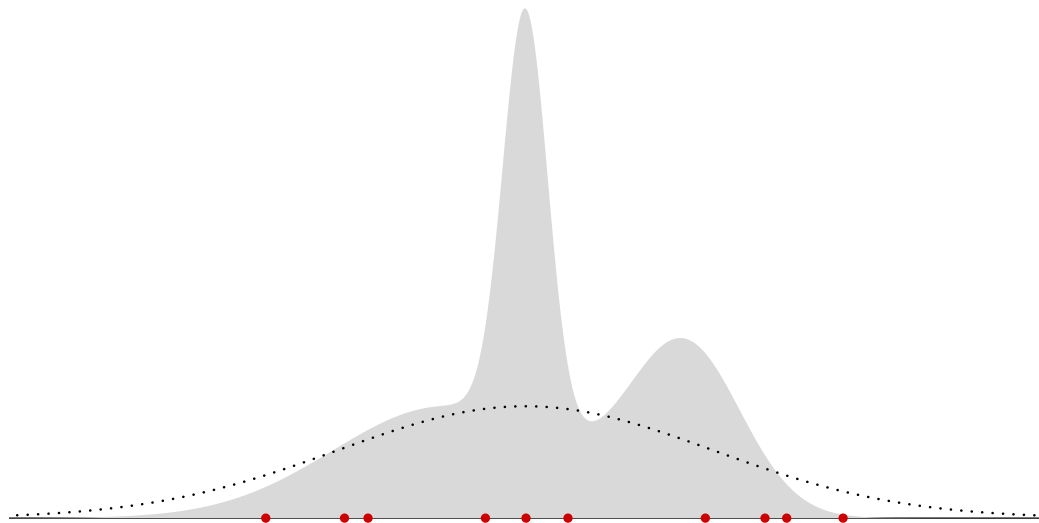
$p(x_t \mid y_{1:t})$  “filtering”

$p(x_s \mid y_{1:t})$  “smoothing”

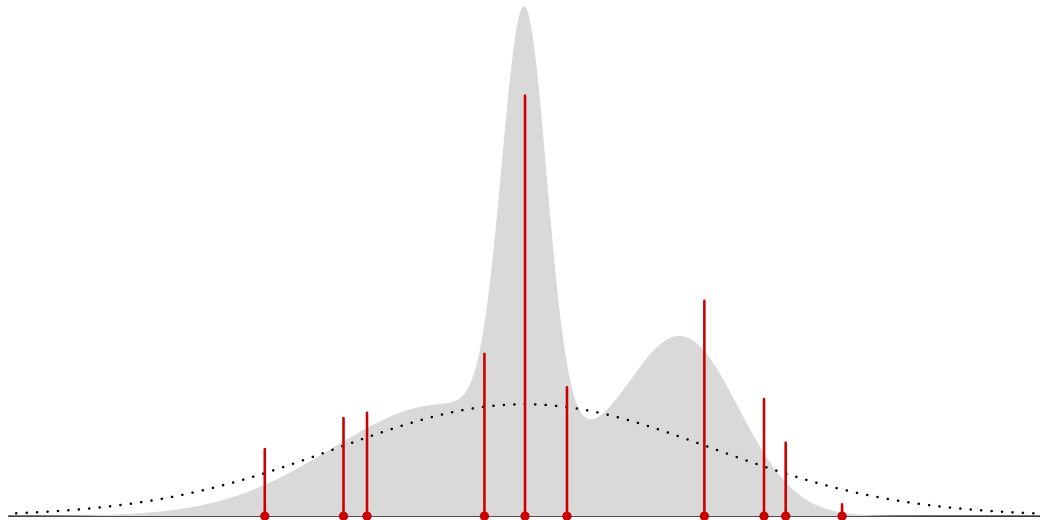
# Importance sampling



# Importance sampling



# Importance sampling





# Sequential importance sampling

- ▶ Idea: use weighted samples from one time step to construct a proposal for the next step

# Sequential importance sampling

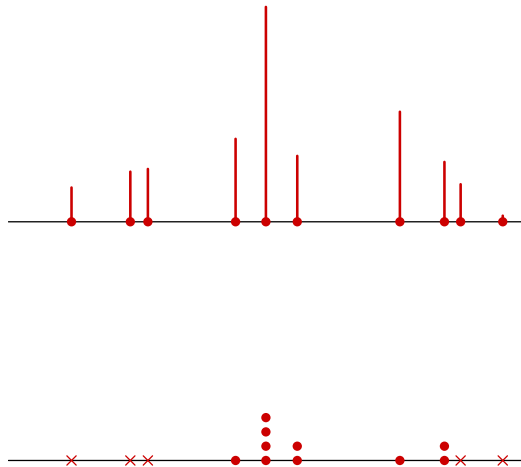
- ▶ Idea: use weighted samples from one time step to construct a proposal for the next step
- ▶ Multiplying weights over time causes *weight degeneracy*

# Sequential importance sampling

- ▶ Idea: use weighted samples from one time step to construct a proposal for the next step
- ▶ Multiplying weights over time causes *weight degeneracy*
- ▶ Can avoid this problem by resampling

# Resampling

- ▶ Transform continuous weights to discrete offspring counts
- ▶ For example, sample counts from Multinomial distribution

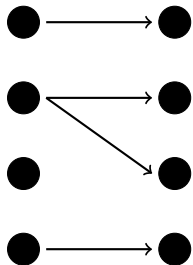


# Sequential Monte Carlo algorithm

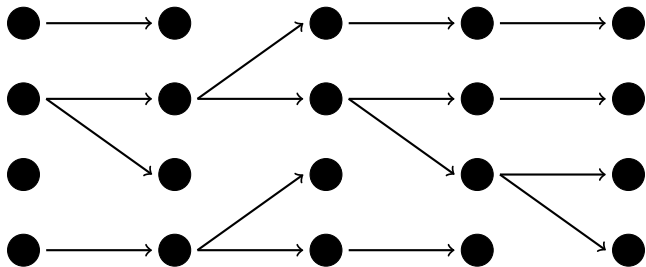
Iterate these steps:

1. Mutate: move particle positions (according to  $f_t$ )
2. Weight: Calculate importance weights for each particle
3. Resample: Duplicate/kill particles according to weights, reset weights to  $1/N$

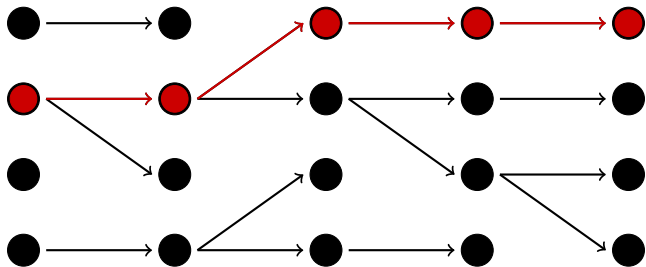
## Ancestral degeneracy



## Ancestral degeneracy

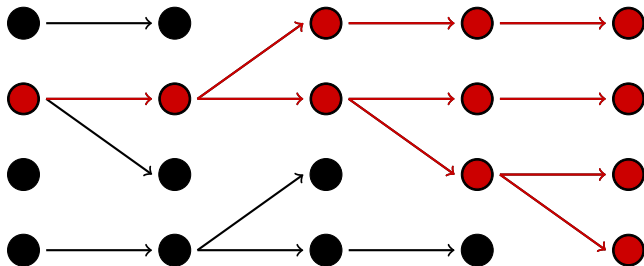


Ancestral degeneracy

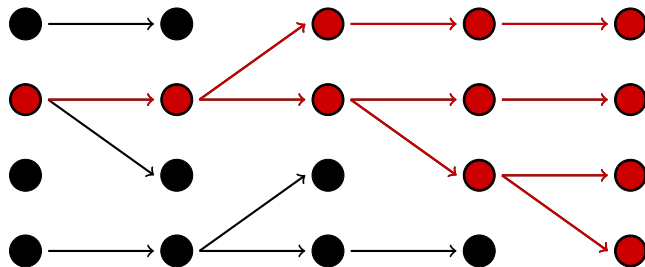




# Ancestral degeneracy



## Ancestral degeneracy



For  $s \ll t$ ,  $p(x_s | y_{1:t})$  is approximated by very few distinct points!

# Ancestral degeneracy

- ▶ Ancestral degeneracy can seriously limit performance for smoothing

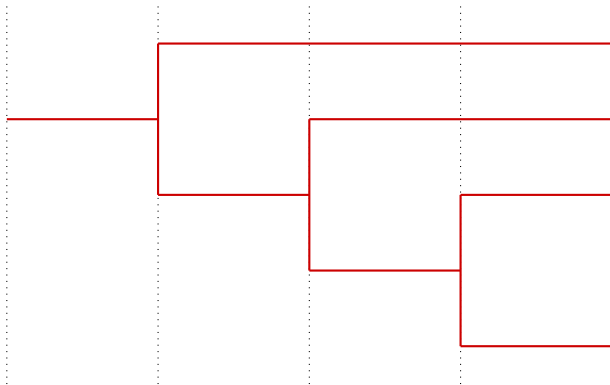
# Ancestral degeneracy

- ▶ Ancestral degeneracy can seriously limit performance for smoothing
- ▶ Resample less often (“adaptive resampling”)
- ▶ Low-variance resampling

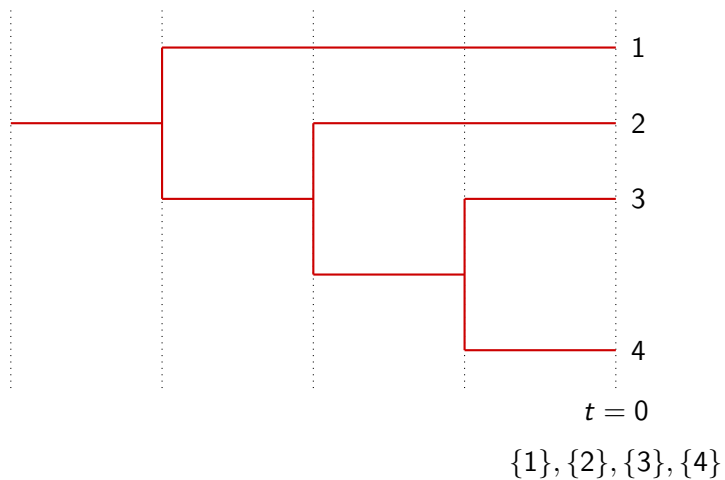
# Ancestral degeneracy

- ▶ Ancestral degeneracy can seriously limit performance for smoothing
- ▶ Resample less often (“adaptive resampling”)
- ▶ Low-variance resampling
- ▶ If we knew how bad ancestral degeneracy would be, we could tune parameters to limit it (e.g. number of particles, threshold for adaptive resampling)
- ▶ Our aim: quantify ancestral degeneracy by analysing the induced genealogy

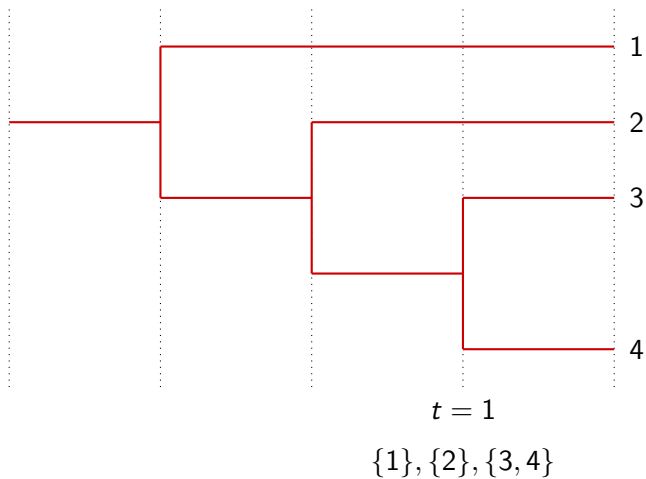
## Encoding genealogies



# Encoding genealogies

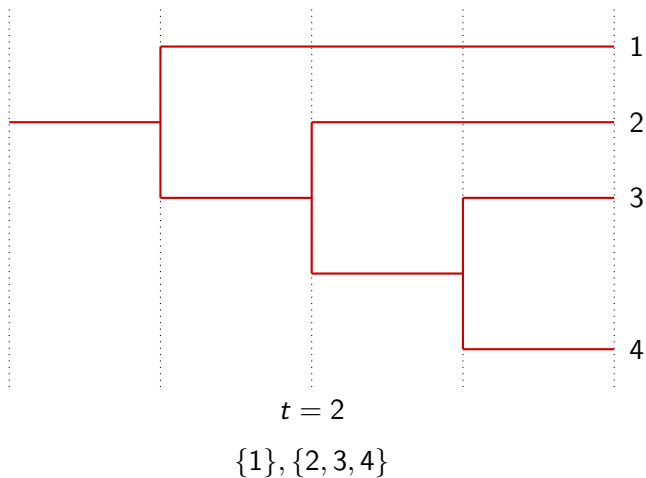


# Encoding genealogies

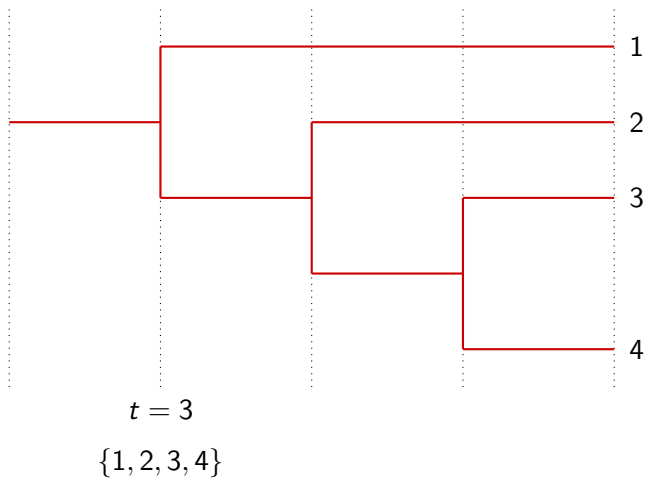




# Encoding genealogies



# Encoding genealogies

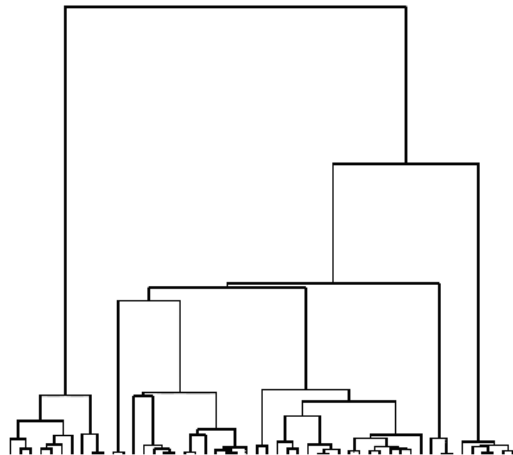


## Asymptotic setting

- ▶ Number of particles  $N \rightarrow \infty$
- ▶ Consider a random sample of  $n \ll N$  particles at time zero
- ▶ Trace these  $n$  lineages back to see the genealogy
- ▶ Rescale time to get a well-defined limiting process

# Kingman's $n$ -coalescent

- ▶ Continuous-time Markov chain on the space of partitions of  $\{1, \dots, n\}$
- ▶ Single pair mergers only
- ▶ Each pair merges independently at rate 1 (total rate  $\binom{k}{2}$  while there are  $k$  distinct lineages)



## Time scale

The probability that a randomly chosen pair of particles at generation  $t$  share a common ancestor at generation  $(t - 1)$ , conditional on offspring counts, is

$$c_N(t) = \frac{1}{(N)_2} \sum_{i=1}^N (\nu_t^{(i)})_2$$

To get an  $n$ -coalescent, this should converge to 1 (the required pair merger rate), so we rescale time by the inverse

$$\tau_N(t) := \min \left\{ s \geq 1 : \sum_{r=1}^s c_N(r) \geq t \right\}$$

# Main theorem

- ▶ Offspring counts  $(\nu_t^{(1)}, \dots, \nu_t^{(N)})$

# Main theorem

- ▶ Offspring counts  $(\nu_t^{(1)}, \dots, \nu_t^{(N)})$
- ▶ Parent-offspring assignments are uniform given offspring counts
- ▶ Time scale does not explode (i.e.  $\mathbb{P}[\tau_N(t) = \infty] = 0$  for all finite  $t$ )
- ▶ There exists a sequence  $(b_N)$  such that  $\lim_{N \rightarrow \infty} b_N = 0$  and

$$\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_3] \leq b_N \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_2]$$

# Main theorem

- ▶ Offspring counts  $(\nu_t^{(1)}, \dots, \nu_t^{(N)})$
- ▶ Parent-offspring assignments are uniform given offspring counts
- ▶ Time scale does not explode (i.e.  $\mathbb{P}[\tau_N(t) = \infty] = 0$  for all finite  $t$ )
- ▶ There exists a sequence  $(b_N)$  such that  $\lim_{N \rightarrow \infty} b_N = 0$  and

$$\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_3] \leq b_N \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_2]$$

Then the finite-dimensional distributions of the time-rescaled genealogies converge to Kingman's  $n$ -coalescent as  $N \rightarrow \infty$ .



We established the theorem in these cases:

- ▶ Multinomial resampling
- ▶ Stochastic rounding
- ▶ Conditional SMC with multinomial resampling

# Multinomial resampling

- ▶ Offspring counts are sampled from Multinomial distribution parametrised by weights
- ▶ Easy to analyse, but doesn't perform well
- ▶ (For the population geneticists): different from Wright-Fisher model because not neutral

# Stochastic rounding

- ▶ Low-variance resampling schemes
- ▶ Offspring count  $\nu_t^{(i)}$  sampled by rounding  $Nw_t^{(i)}$  up or down to an integer
- ▶ Ensure total number of offspring is  $N$
- ▶ (For the SMC specialists): includes systematic, residual-stratified, residual-systematic, SSP resampling, ...

# Conditional SMC

- ▶ Component of particle MCMC
- ▶ One lineage is conditioned to survive all resampling steps
- ▶ Resampling algorithm must deterministically propagate this “immortal lineage”
- ▶ For example: fix one offspring to immortal lineage, multinomial sampling for the remaining counts

# Conditional SMC

- ▶ Component of particle MCMC
- ▶ One lineage is conditioned to survive all resampling steps
- ▶ Resampling algorithm must deterministically propagate this “immortal lineage”
- ▶ For example: fix one offspring to immortal lineage, multinomial sampling for the remaining counts
- ▶ Conjecture: theorem also applies to conditional SMC with stochastic rounding

## In conclusion...

- ▶ Genealogies can help us to analyse performance of smoothing algorithms which suffer ancestral degeneracy
- ▶ We have simple conditions under which these genealogies converge to  $n$ -coalescent
- ▶ These conditions are verified for some important classes of SMC algorithms

## In conclusion...

- ▶ Genealogies can help us to analyse performance of smoothing algorithms which suffer ancestral degeneracy
- ▶ We have simple conditions under which these genealogies converge to  $n$ -coalescent
- ▶ These conditions are verified for some important classes of SMC algorithms

### Open questions

- ▶ Verify theorem for other important resampling schemes (stratified, residual-multinomial)
- ▶ How to estimate the time scale  $c_N$  a priori (since it depends on offspring counts)
- ▶ Weak convergence so we can say more about convergence of expectations
- ▶ Rates of convergence
- ▶ Finite- $N$  behaviour

THE END

For more details see the pre-print  
[arXiv:2007.00096](https://arxiv.org/abs/2007.00096)