Asymptotic genealogies of sequential Monte Carlo algorithms

Suzie Brown
University of Warwick
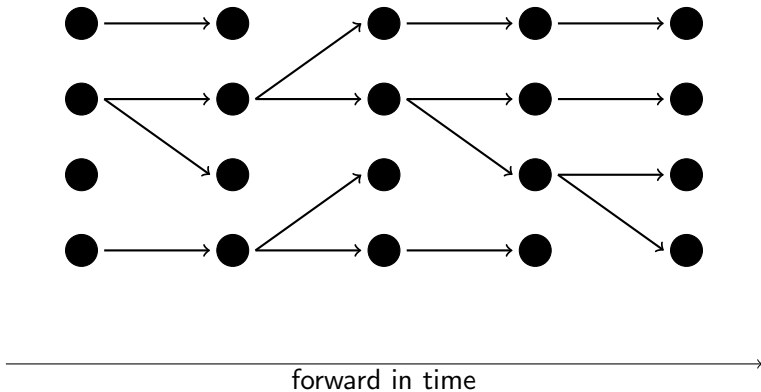with Paul Jenkins, Adam Johansen & Jere Koskela

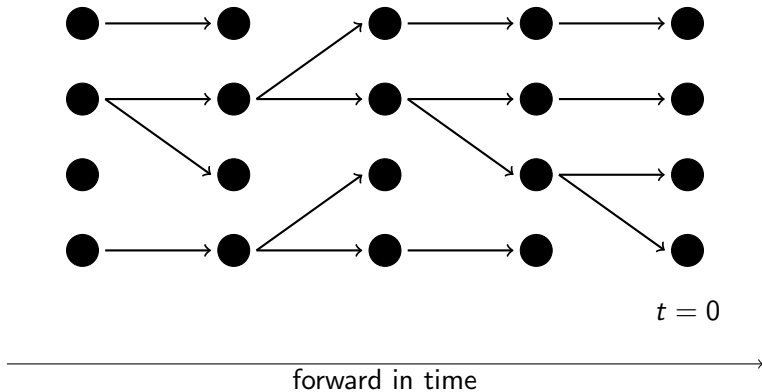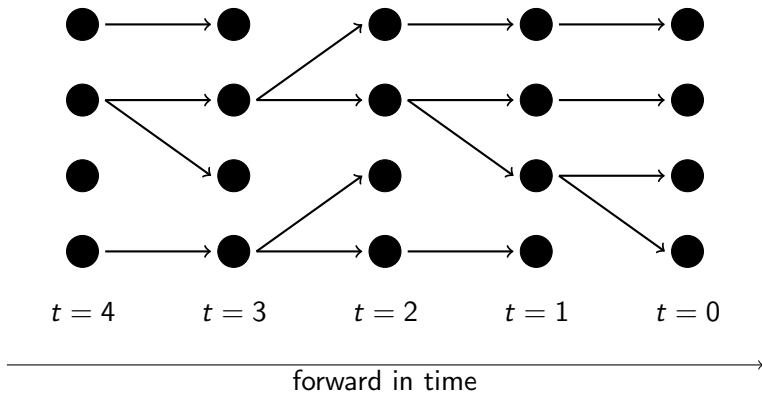18 November 2020

Outline

1. ...

forward in time

Sequential Monte Carlo

$t = 0$

forward in time

Sequential Monte Carlo

$t = 4 \qquad t = 3 \qquad t = 2 \qquad t = 1 \qquad t = 0$

forward in time

Sequential Monte Carlo

$t = 4$ $\quad$ $t = 3$ $\quad$ $t = 2$ $\quad$ $t = 1$ $\quad$ $t = 0$

forward in time

$t = 4 \qquad t = 3 \qquad t = 2 \qquad t = 1 \qquad t = 0$

forward in time

$t = 4$ $t = 3$ $t = 2$ $t = 1$ $t = 0$

forward in time
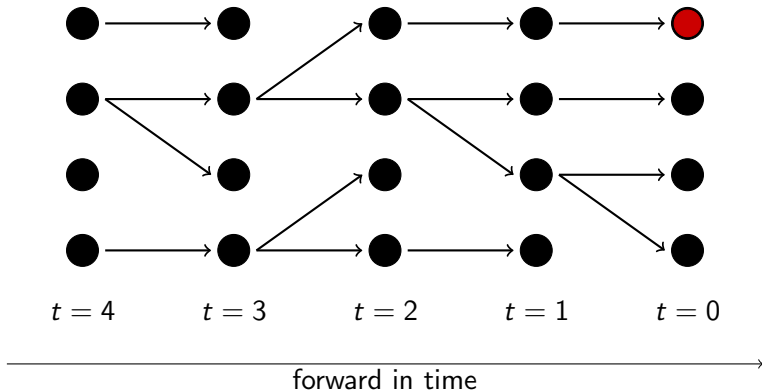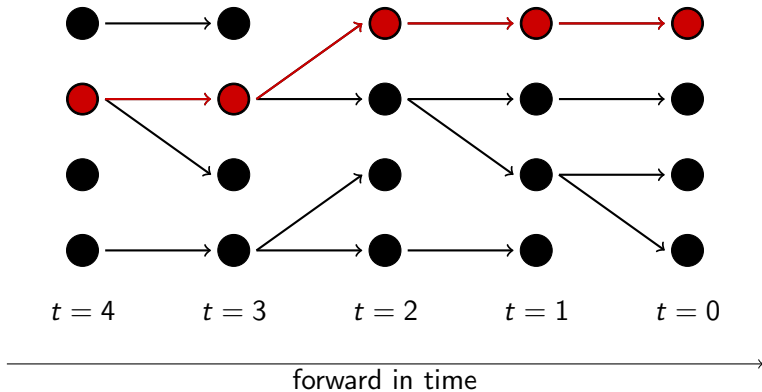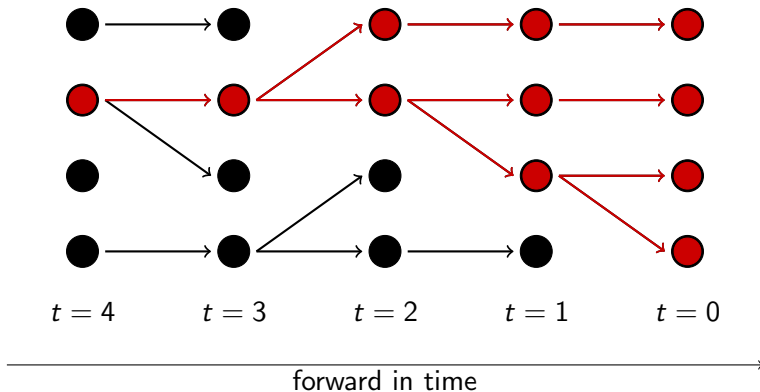
Induced genealogy
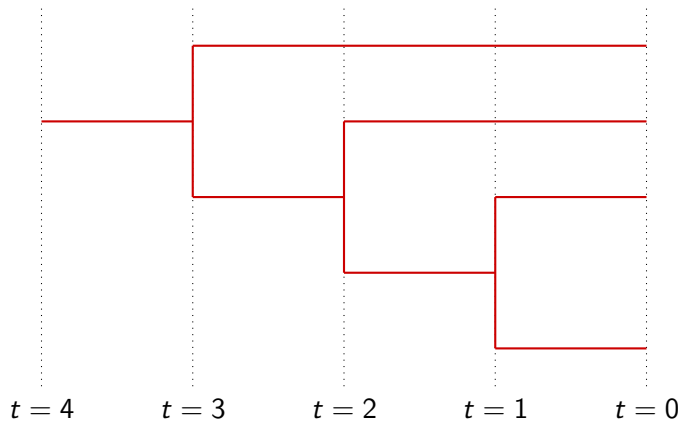
Mathematical formulation

- Population of N particles
- Sample $n \leq N$ terminal particles

Mathematical formulation

- Population of N particles
- Sample $n \leq N$ terminal particles
- Describe genealogy by a partition-valued stochastic process $(G_t^{(n,N)})_{t \in \mathbb{N}_0}$
- Elements i, j are in the same block of the partition $G_t^{(n,N)}$ iff particles i and j share a common ancestor at time t
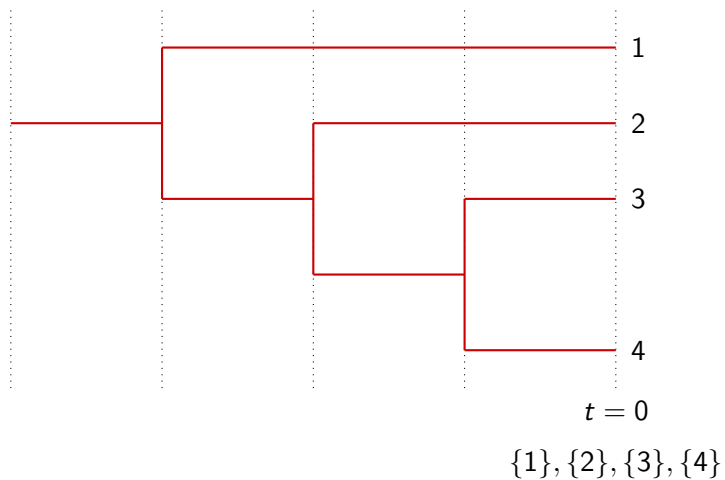
Mathematical formulation

▶ Population of N particles

▶ Sample $n \leq N$ terminal particles

▶ Describe genealogy by a partition-valued stochastic process $(G_t^{(n,N)})_{t \in \mathbb{N}_0}$

▶ Elements i, j are in the same block of the partition $G_t^{(n,N)}$ iff particles i and j share a common ancestor at time t

▶ At time 0, the partition of $\{1, \ldots, n\}$ consisting of singletons: $G_0^{(n,N)} = \{\{1\}, \ldots, \{n\}\}$

▶ The only possible non-identity transitions are those that merge blocks

Example

$t = 0$

$\{1\}, \{2\}, \{3\}, \{4\}$

Example

$t = 1$

$\{1\}, \{2\}, \{3, 4\}$

Example

$t = 2$

$\{1\}, \{2, 3, 4\}$

$t = 3$

$\{1, 2, 3, 4\}$
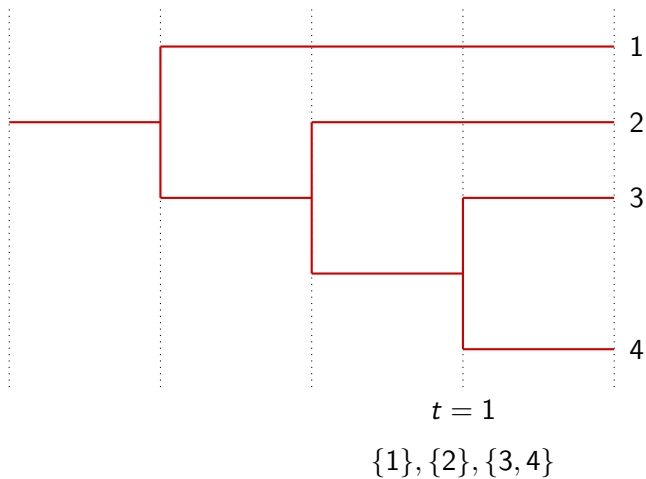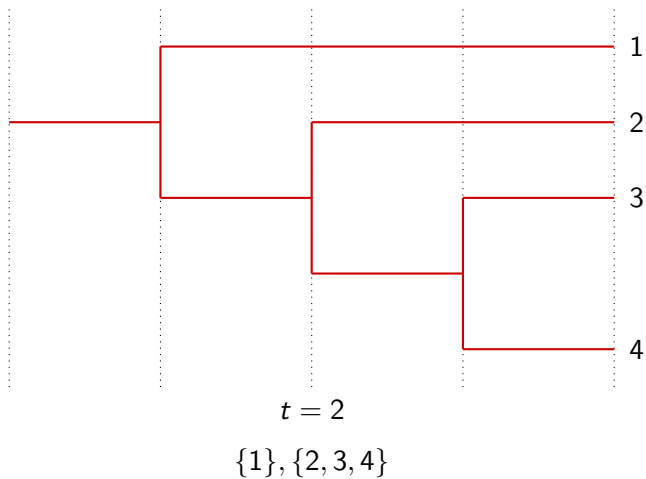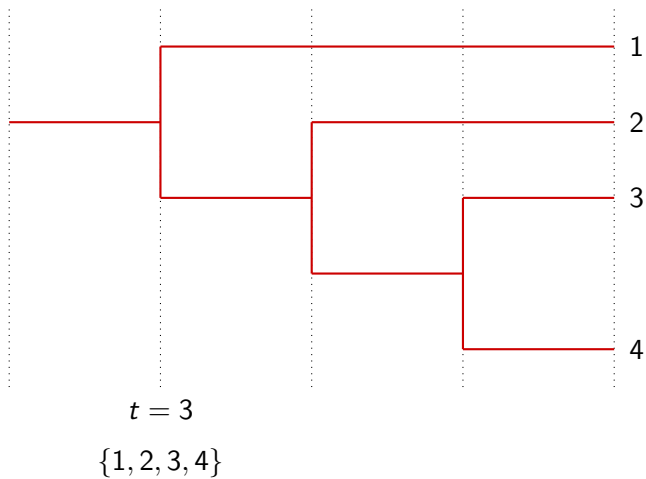
Kingman's *n*-coalescent

Kingman's *n*-coalescent

- *n* terminal particles
- Continuous time
- Each pair of lineages coalesces at rate 1
- No other ($>$ pair) mergers allowed

Notation

- $w_t^{(i)}$ = weight of particle i in generation t
- family size $v_t^{(i)}$ = number of offspring of particle i in generation t

Notation

- $w_t^{(i)}$ = weight of particle i in generation t
- family size $v_t^{(i)}$ = number of offspring of particle i in generation t
- pair merger probability $c_N(t) = \frac{1}{N(N-1)} \sum_{i=1}^{N} v_t^{(i)}(v_t^{(i)} - 1)$

Notation

- $w_t^{(i)}$ = weight of particle i in generation t
- family size $v_t^{(i)}$ = number of offspring of particle i in generation t
- pair merger probability $c_N(t) = \frac{1}{N(N-1)} \sum_{i=1}^{N} v_t^{(i)}(v_t^{(i)} - 1)$
- time scale $\tau_N(t)$ = inverse of $c_N(t)$

A convergence theorem[1]

Suppose:

$$\mathbb{E}[c_N(t)] \to 0$$

$$\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r) \right] \to 0$$

$$\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r)^2 \right] \to 0$$

$$\mathbb{E}[\tau_N(t) - \tau_N(s)] \le C_{t,s} N$$

Then the rescaled genealogical process $(G^{(n,N)}_{\tau_N(t)})_{t \ge 0}$ converges to Kingman's coalescent in the sense of finite-dimensional distributions as $N \to \infty$.

---
[1]Koskela, Jenkins, Johansen, Spanò. *Ann Stat,* to appear.

A simplified statement

A simplified statement

Suppose there exists a deterministic sequence $(b_N)_{N \in \mathbb{N}}$ uniform in $t \geq 1$ such that $b_N \to 0$.

$$\frac{1}{N(N-1)(N-2)} \sum_{i=1}^{N} \mathbb{E}_t[v_t^{(i)}(v_t^{(i)} - 1)(v_t^{(i)} - 2)] \leq b_N \frac{1}{N(N-1)} \sum_{i=1}^{N} \mathbb{E}_t[v_t^{(i)}(v_t^{(i)} - 1)]$$

Then the rescaled genealogical process $(G_{\tau_N(t)}^{(n,N)})_{t \geq 0}$ converges to Kingman's coalescent in the sense of finite-dimensional distributions as $N \to \infty$.

Remarks

▶ We replace four conditions with one

Remarks

- We replace four conditions with one
- We remove condition 4, which we know is not a necessary condition

Remarks

- ▶ We replace four conditions with one
- ▶ We remove condition 4, which we know is not a necessary condition
- ▶ We remove conditions 2 and 3, which contain awkward mixed moments

Remarks

- ▶ We replace four conditions with one
- ▶ We remove condition 4, which we know is not a necessary condition
- ▶ We remove conditions 2 and 3, which contain awkward mixed moments
- ▶ We can verify the new condition in more complicated cases...

Resampling schemes

Problem : (stochastically) map weights $(w_t^{(1)}, \ldots, w_t^{(N)})$ to family sizes $(v_t^{(1)}, \ldots, v_t^{(N)})$.

Resampling schemes

Problem : (stochastically) map weights $(w_t^{(1)}, \ldots, w_t^{(N)})$ to family sizes $(v_t^{(1)}, \ldots, v_t^{(N)})$.
Desirable properties:

▶ Number of particles constant $\sum_{i=1}^{N} v_t^{(i)} = N$
▶ Equal weights after resampling $w_{t+}^{(i)} = 1/N$
▶ Unbiased $\mathbb{E}[v_t^{(i)} | w_t^{(i)}] = N w_t^{(i)}$

- Transform continuous weights to discrete offspring counts
- For example, sample counts from Multinomial distribution
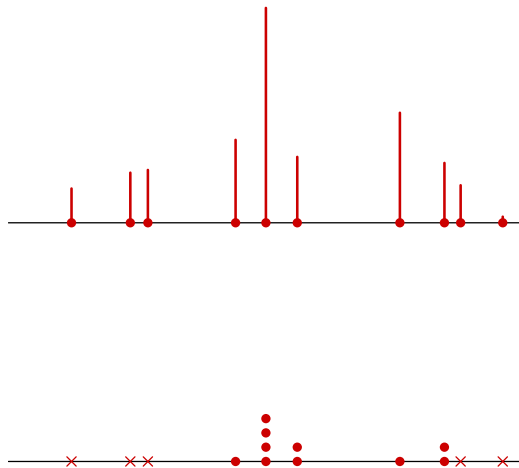
Corollary: multinomial resampling

Consider an SMC algorithm with potential g and transition density q, satisfying

$$\frac{1}{a} \leq g_t(x,x') \leq a$$

$$\varepsilon h(x') \leq q_t(x,x') \leq \frac{1}{\varepsilon} h(x')$$

for constants $0 < \varepsilon \leq 1 \leq a < \infty$, and probablity distribution $h(\cdot)$.

Under multinomial resampling, the rescaled genealogies converge to Kingman's coalescent.

Stochastic roundings

Let $X \geq 0$.
$Y : \mathbb{R}_+ \to \mathbb{N}$ is a *stochastic rounding* of X if Y takes the values

$$Y \mid X = \begin{cases} \lfloor X \rfloor & \text{with probability } 1 - X + \lfloor X \rfloor \\ \lfloor X \rfloor + 1 & \text{with probability } X - \lfloor X \rfloor \end{cases}$$

Stochastic roundings

Let $X \geq 0$.
$Y : \mathbb{R}_+ \to \mathbb{N}$ is a *stochastic rounding* of X if Y takes the values

$$Y \mid X = \begin{cases} \lfloor X \rfloor & \text{with probability } 1 - X + \lfloor X \rfloor \\ \lfloor X \rfloor + 1 & \text{with probability } X - \lfloor X \rfloor \end{cases}$$

Can generalise to multivariate case, with marginals $Y_i \mid X_i$ as above.

Stochastic roundings

- $X_i = N w_t^{(i)}$
- $v_t^{(i)} = Y_i$
- Constraint $Y_1 + \cdots + Y_N = N$

---

[2]Gerber, Whiteley, Chopin. *Ann Stat,* 2019.

- $X_i = N w_t^{(i)}$
- $v_t^{(i)} = Y_i$
- Constraint $Y_1 + \cdots + Y_N = N$
- Examples: systematic resampling, residual-stratified resampling, SSP resampling[2],...

---

[2]Gerber, Whiteley, Chopin. *Ann Stat,* 2019.

Corollary: stochastic rounding -based resampling

Consider an SMC algorithm with potential g and transition density q, satisfying

$$\frac{1}{a} \leq g_t(x, x') \leq a$$
$$\varepsilon h(x') \leq q_t(x, x') \leq \frac{1}{\varepsilon} h(x')$$

for constants $0 < \varepsilon \leq 1 \leq a < \infty$, and probablity distribution $h(\cdot)$.

Under stochastic rounding -based resampling, the rescaled genealogies converge to Kingman's coalescent.

Conditional SMC

- Used for SMC updates in particle MCMC[3]
- Given a fixed "immortal trajectory", conditioned to survive all transition and resampling steps

[3]Andrieu, Doucet, Holenstein. *JRSSB*, 2010.

Corollary: conditional SMC

Consider a conditional SMC algorithm with potential g and transition density q, satisfying

$$\frac{1}{a} \leq g_t(x, x') \leq a$$

$$\varepsilon h(x') \leq q_t(x, x') \leq \frac{1}{\varepsilon} h(x')$$

for constants $0 < \varepsilon \leq 1 \leq a < \infty$, and probablity distribution $h(\cdot)$.

Under multinomial resampling, the rescaled genealogies converge to Kingman's coalescent.

What about finite N?

- As $N \to \infty$ but sample size n fixed, probability of sampling immortal line vanishes, so its effect isn't seen in the limit

What about finite N?

▶ As $N \to \infty$ but sample size n fixed, probability of sampling immortal line vanishes, so its effect isn't seen in the limit
▶ For finite N we have more interesting behaviour
▶ Depends how "typical" the immortal trajectory is

Summary

- We consider genealogies of SMC algorithms
- We provide simple conditions for asymptotically Kingman genealogies
- Satisfied by stochastic rounding -based resampling
- Satisfied by conditional SMC with multinomial resampling...
- ... but this hides some interesting pre-limiting behaviour

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