Asymptotic genealogies of sequential Monte Carlo algorithms

Suzie Brown
with Paul Jenkins, Adam Johansen & Jere Koskela

4 December 2020

Outline

1. Sequential Monte Carlo
2. Resampling and degeneracy
3. SMC genealogies
4. Examples

Sequential Monte Carlo

Target a sequence of distributions by simulating a population of particles evolving in time

Sequential Monte Carlo

Target a sequence of distributions by simulating a population of particles evolving in time

Iterate these steps:

1. **Mutate** particles via Markov transition density q_t
2. **Weight** particles by potential function g_t
3. **Resample** particles in proportion to their weights

Resampling

Stochastically map continuous weights $(w_t^{(1)}, \ldots, w_t^{(N)})$ to discrete offspring counts $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$
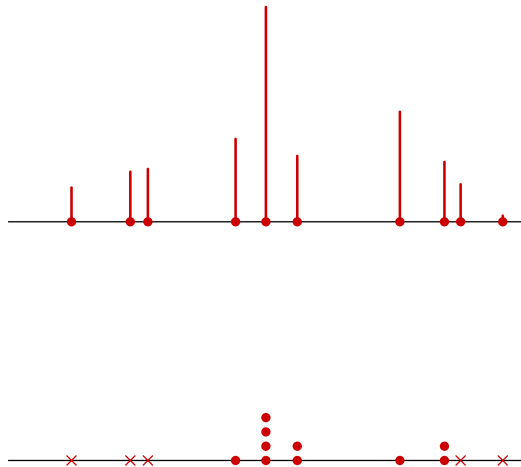
Resampling

Stochastically map continuous weights
$(w_t^{(1)}, \ldots, w_t^{(N)})$ to discrete offspring
counts $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$
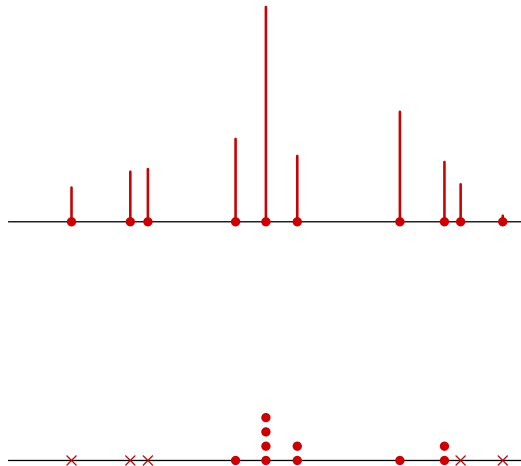
Properties:

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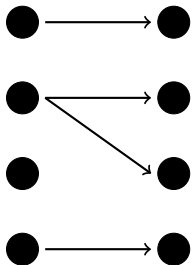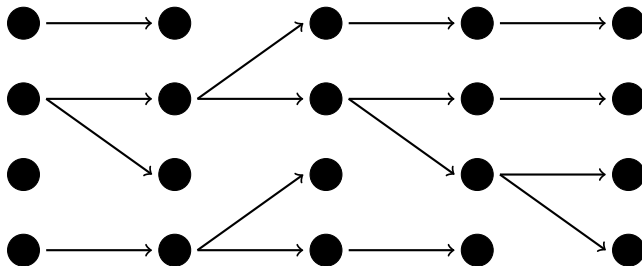 $w_{t+}^{(i)} = 1/N$
- Unbiased
  $\mathbb{E}[\nu_t^{(i)} | w_t^{(i)}] = N w_t^{(i)}$

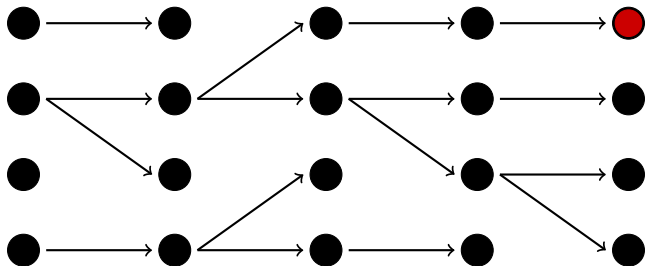# Resampling induces a genealogy
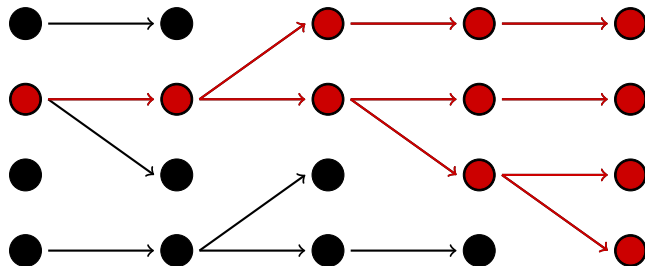
Ancestral degeneracy: for $t << T$, few distinct samples are available

Mitigating ancestral degeneracy

Resample less often?
Adaptive resampling. Only resample when effective sample size falls below some threshold.
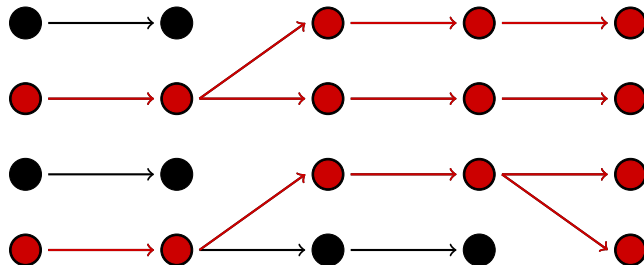
Mitigating ancestral degeneracy

Resample more cleverly?

Low-variance resampling. Resample in a way that limits the amount of extra randomness added by the resampling step.
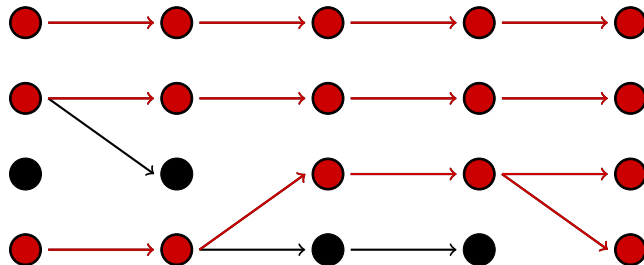
Mitigating ancestral degeneracy

Make use of killed samples?

Backward simulation. Use a backward pass to sample new ancestors for the terminal particles.
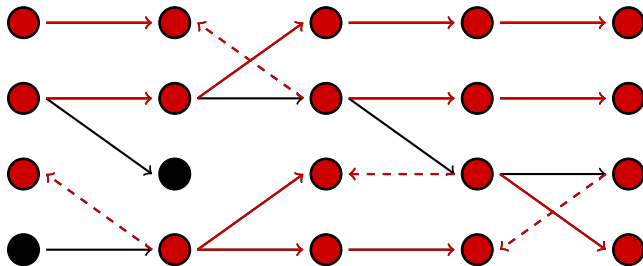
Encoding genealogies

▶ Label time in reverse

Encoding genealogies

▶ Label time in reverse
▶ Population of N particles
▶ Sample $n \leq N$ terminal particles

Encoding genealogies

- ▶ Label time in reverse
- ▶ Population of N particles
- ▶ Sample $n \leq N$ terminal particles
- ▶ Describe genealogy by stochastic process $(G_t^{(n,N)})_{t \in \mathbb{N}_0}$ on space of partitions of $\{1, \ldots, n\}$
- ▶ 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

Encoding genealogies

- ▶ Label time in reverse
- ▶ Population of N particles
- ▶ Sample $n \leq N$ terminal particles
- ▶ Describe genealogy by stochastic process $(G_t^{(n,N)})_{t \in \mathbb{N}_0}$ on space of partitions of $\{1, \ldots, n\}$
- ▶ 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
- ▶ Initially $G_0^{(n,N)} = \{\{1\}, \ldots, \{n\}\}$
- ▶ The only possible non-identity transitions are those that merge blocks
- ▶ The trivial partition $\{\{1, \ldots, n\}\}$ is an absorbing state

Encoding genealogies
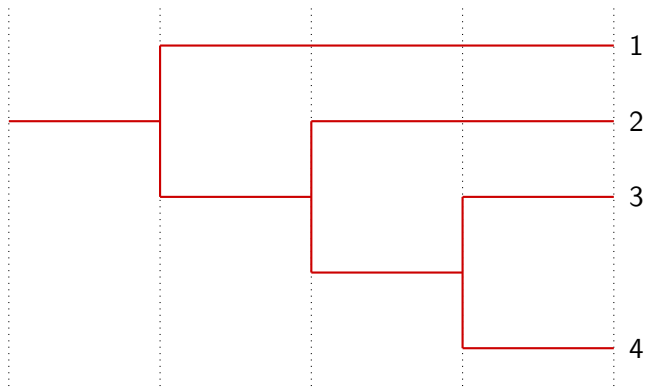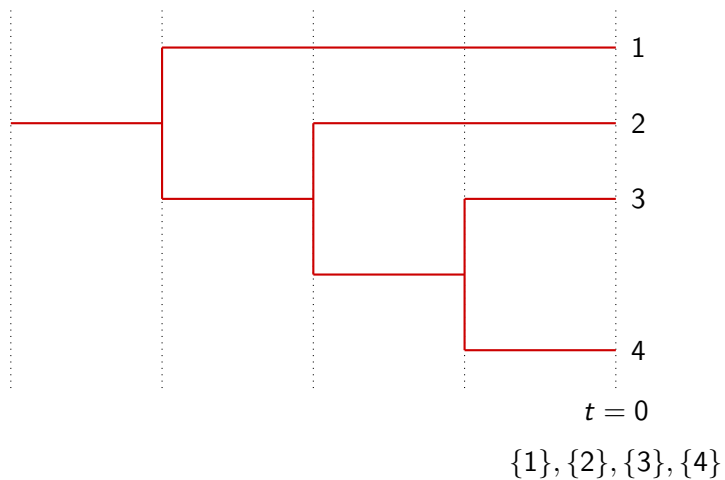
$t = 0$

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

Encoding genealogies

$t = 1$

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

Encoding genealogies

$t = 2$

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

Encoding genealogies

$t = 3$

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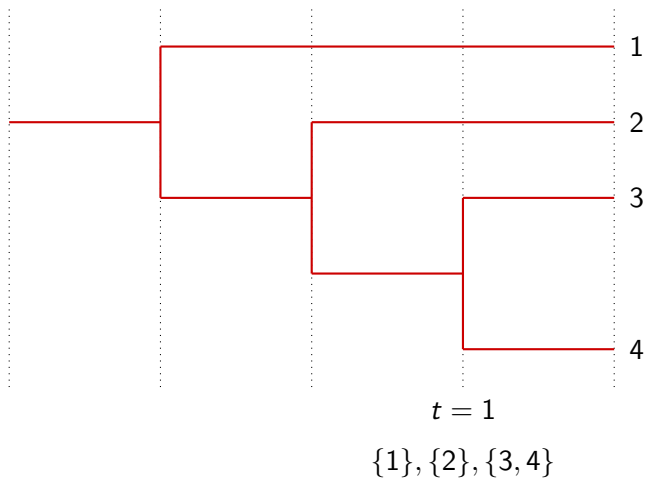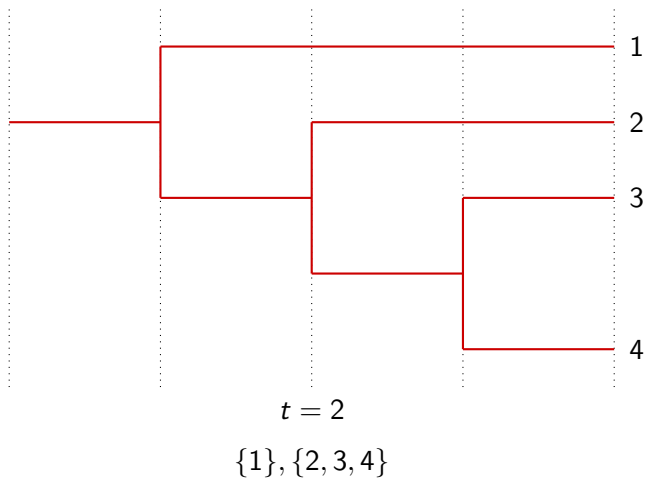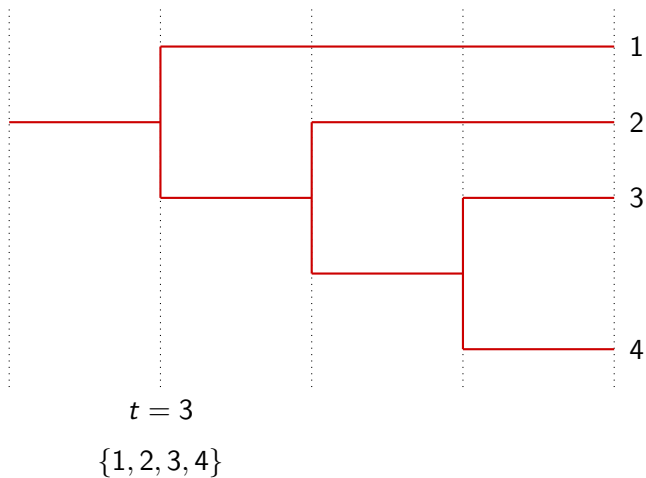
Kingman's *n*-coalescent[1]

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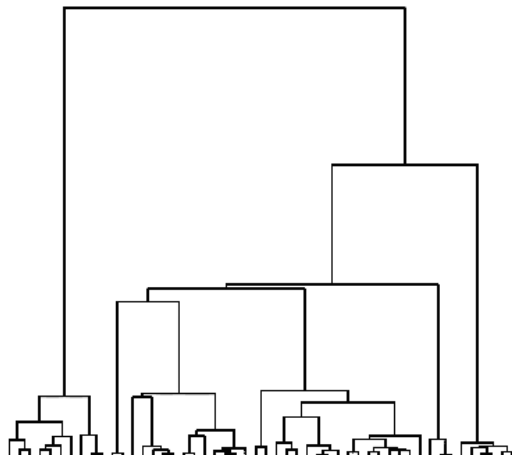
image: Wikimedia Commons

---

[1] J F C Kingman, *Stochastic Processes & their Applications*, 1982.

Time scale

Pair merger probability conditional on $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$:

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

Time scale

Pair merger probability conditional on $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$:

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

Rescale time by inverse:

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

Main theorem

Assumptions:

▶ Parent-offspring assignments are uniform given offspring counts

Main theorem

Assumptions:

▶ 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)

Main theorem

Assumptions:

- 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 $\lim_{N \to \infty} b_N = 0$ and

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

Main theorem

Assumptions:

▶ 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 $\lim_{N \to \infty} b_N = 0$ 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 $\left(G_{\tau_N(t)}^{(n,N)} \right)_{t \geq 0}$ converge to Kingman's *n*-coalescent as $N \to \infty$.

Examples

We have established the theorem in these cases:

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

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 $h(\cdot)$.

---

[2] J Koskela, P A Jenkins, A M Johansen, D Spanò. *Annals of Statistics*, 2020.

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:

$$(\nu_t^{(1)}, \ldots, \nu_t^{(N)}) \sim \text{Multinomial}\left(N, (w_t^{(1)}, \ldots, w_t^{(N)})\right)$$

the rescaled genealogies converge to Kingman's coalescent.[2]

---

[2] J Koskela, P A Jenkins, A M Johansen, D Spanò. *Annals of Statistics*, 2020.

Stochastic rounding

$\mathbf{Y} : \mathbb{R}_+^N \to \mathbb{N}^N$ is a *stochastic rounding* of $\mathbf{X}$ if for $i = 1, \ldots, N$

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

---

[3] D Crisan, T Lyons. *Probability Theory & Related Fields*, 1997.
[4] M Gerber, N Whiteley, N Chopin. *Annals of Statistics*, 2019.

Stochastic rounding

$\mathbf{Y} : \mathbb{R}_+^N \to \mathbb{N}^N$ is a *stochastic rounding* of $\mathbf{X}$ if for $i = 1, \ldots, N$

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

▶ We can construct low-variance resampling schemes using stochastic rounding
▶ Take $X_i = N w_t^{(i)}$ and $Y_i = \nu_t^{(i)}$
▶ Require further constraint $Y_1 + \cdots + Y_N = N$

---

[3] D Crisan, T Lyons. *Probability Theory & Related Fields*, 1997.
[4] M Gerber, N Whiteley, N Chopin. *Annals of Statistics*, 2019.

Stochastic rounding

$\mathbf{Y} : \mathbb{R}_+^N \to \mathbb{N}^N$ is a *stochastic rounding* of $\mathbf{X}$ if for $i = 1, \ldots, N$

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

▶ We can construct low-variance resampling schemes using stochastic rounding
▶ Take $X_i = N w_t^{(i)}$ and $Y_i = \nu_t^{(i)}$
▶ Require further constraint $Y_1 + \cdots + Y_N = N$
▶ **Examples:** systematic, residual-stratified, branching system[3], SSP resampling[4]

---

[3] D Crisan, T Lyons. *Probability Theory & Related Fields*, 1997.
[4] M Gerber, N Whiteley, N Chopin. *Annals of Statistics*, 2019.
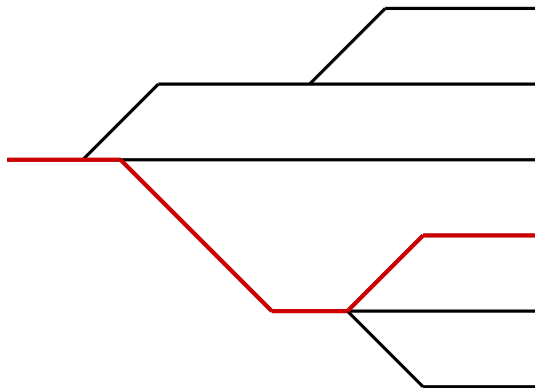
Stochastic rounding

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

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

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

for constants $0 < \varepsilon \leq 1 \leq a < \infty$.

Resampling via stochastic rounding, the rescaled genealogies converge to Kingman's coalescent.

Conditional SMC

- Used for SMC updates in particle MCMC[5]
- One fixed "immortal lineage" is conditioned to survive all mutation and resampling steps
- Resampling must deterministically propagate the immortal lineage

---

[5]C Andrieu, A Doucet, R Holenstein. *Journal of the Royal Statistical Society B*, 2010.

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 probability distribution $h(\cdot)$.

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

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 probability distribution $h(\cdot)$.

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

Conjecture: the theorem also applies to conditional SMC with stochastic rounding

In conclusion...

- Genealogies can help us to analyse performance of SMC algorithms which suffer ancestral degeneracy
- We have simple conditions under which these genealogies converge to Kingman's 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 τ_N a priori (since it depends on offspring counts)
- Weak convergence
- Rates of convergence
- Finite-N behaviour

Thank you!

pre-print at arXiv:2007.00096