Asymptotic analysis of genealogies induced by sequential Monte Carlo algorithms

Suzie Brown

March 19, 2019

1 Introduction

- organisation of the report

Sequential Monte Carlo has become a popular tool, particularly in applications such as target tracking, where there is a natural sequential component and we wish to infer underlying states from noisy observations. While particle methods can be very effective for filtering, it is more difficult to apply them to smoothing because they typically suffer very badly from ancestral degeneracy in the particle genealogies.

When attempting to mitigate this problem, one often encounters a trade-off between ancestral degeneracy (arising from resampling) and weight degeneracy (arising from sequential importance sampling). However, while weight degeneracy is a reasonably well-quantified problem, there exists little in the way of tools for quantifying ancestral degeneracy a priori. There have been some simulation studies attempting to cast light on the magnitude of this problem, but analytical findings remain elusive, since the complexity of the most commonly used particle methods makes it difficult to obtain any rigorous results. Consequently, there is a wealth of pertinent open questions in this area. This work attempts to extend a first result for a standard class of SMC algorithms to the more sophisticated algorithms which are typically used in practice.

Throughout this document we will use the compact notation $X_{m:n}$ as shorthand for $X_m, X_{m+1}, \dots, X_n$, as well as $X_{-n} := X_0, \dots, X_{n-1}, X_{n+1}, \dots X_N$. We denote falling factorial powers $(x)_a := x(x-1)\dots(x-a+1)$, with the convention $(x)_0 = 1$.

2 Sequential Monte Carlo

References for this section are Doucet et al. (2001), Del Moral et al. (2006), and Doucet and Johansen (2011).

2.1 Class of models

Although sequential Monte Carlo (SMC) methods can be applied in a much more general setting, they are particularly easy to motivate in the setting of state space models, where the "sequential" nature follows naturally from the discrete time steps present in the model. For the purposes of presenting the algorithm, let us consider a time-homogeneous state space model consisting of an unobservable discrete-time Markov process $X_{0:T}$ and observables $Y_{0:T}$, satisfying the conditional independence structure

$$(X_{t+1:T} \perp\!\!\!\perp X_{0:t-1}) \mid X_t$$
$$(Y_t \perp\!\!\!\perp Y_{-t}, X_{-t}) \mid X_t$$

for all $t \in \{0, 1, \dots, T\}$, as represented by the graphical model below.

$$\begin{tikzcd}
 & Y_{t-1} & Y_t & \\
\cdots \arrow[r, "f"] & X_{t-1} \arrow[u, "g"] \arrow[r, "f"] & X_t \arrow[u, "g"] \arrow[r, "f"] & \cdots
\end{tikzcd}$$

We assume for notational convenience that $x_0, \dots, x_T$ take values in a common state space $\mathcal{X}$, and $y_0, \dots, y_T$ in a common state space $\mathcal{Y}$, but these assumptions can be dropped.

Suppose we have the following model:

$$\begin{aligned}
&X_0 \sim \mu(\cdot) \\
&X_{t+1} \mid (X_t = x_t) \sim f(\cdot|x_t) \qquad t = 0, \dots, T-1 \\
&Y_t \mid (X_t = x_t) \sim g(\cdot|x_t) \qquad t = 0, \dots, T
\end{aligned}$$

where $(X_t)_{t=0}^T$ is an unobservable discrete-time Markov process and the observables $(Y_t)_{t=0}^T$ satisfy $Y_t \perp\!\!\!\perp \{Y_{-t}, X_{-t}\} \mid X_t$.

We assume that the *transition* and *emission* kernels have densities which are denoted by f and g respectively, but this is not necessary in general. We only require that we can sample from $\mu(\cdot)$ and $f(\cdot|x)$, and calculate *unnormalised* potentials $g(y|x)$, for all x, y.

As a concrete example, let us consider the application of target tracking. Suppose we are using radar to track the position of an aeroplane. The true trajectory of the aeroplane is unknown and is represented by $X_{0:T}$ (perhaps a sequence of positions in $\mathbb{R}^3$), with f encoding our model for how an aeroplane moves (perhaps some differential equations). What we observe is the output $Y_{0:T}$ of our radar equipment, which has some measurement uncertainty that is encoded in g.

2.2 Inference in state space models

Suppose we are in a Bayesian setting, where μ is our prior distribution at time 0, observations y_t arrive sequentially, and we want to infer information about the hidden states (either on- or off-line). The three main inference problems are:

Filtering (where is it now?) $p(x_t|y_{0:t})$

Prediction (where will it go next?) $p(x_{t+1}|y_{0:t})$

Smoothing (where has it been?) $p(x_{0:t}|y_{0:t})$

In the on-line setting, we take as our prior the posterior distribution from the previous time step $t-1$, and update it using the new observation y_t. The inference must be fast enough to keep up with the rate of arrival of observations, so in particular the complexity of the update must not increase with T. In the off-line setting, we take μ as the prior distribution, and infer the set of posteriors once all $T+1$ observations have arrived.

Prediction and filtering are essentially equivalent, because given a filtering distribution, the corresponding predictive distribution can be obtained by applying the transition kernel f. Smoothing is considered a harder task because it requires us to infer many more parameters from the same amount of information; indeed the dimension of the problem increases linearly with T.

In the case of linear Gaussian state space models, the posterior distributions of interest are available analytically, by way of the Kalman filter (Kalman, 1960) and Rauch-Tung-Striebel (RTS) smoother recursions (Rauch et al., 1965). Recursions are also available for some other conjugate models: see for example Vidoni (1999). The other analytic case occurs if the state space of $(X_t)_{t=0}^\infty$ is finite, in which case the forward-backward algorithm (Baum et al., 1970) yields the exact posteriors.

2.3 Particle approximation

In more complex models such techniques are not feasible, and we are forced to resort to Monte Carlo methods. For state space models, Markov chain Monte Carlo methods are not very effective due to the high dimension of the parameter space. But we can exploit the sequential nature of the underlying dynamics to decompose the problem into a sequence of inferences of more manageable dimension. This is the motivation behind sequential Monte Carlo (SMC) methods.

The conditional independence structure in the model implies that the (joint) marginal distribution of the hidden states $X_{0:t}$ is given by

$$p(x_{0:t}) = \mu(x_0) \prod_{i=1}^{t} f(x_i \mid x_{i-1})$$

and that the likelihood of the observations $y_{0:t}$ given the underlying states $x_{0:t}$ takes the form

$$p(y_{0:t} \mid x_{0:t}) = \prod_{i=0}^{t} g(y_i \mid x_i).$$

The smoothing distribution $p(x_t|y_{0:T})$ is obtained from $p(x_{0:T}|y_{0:T})$ by marginalising. Using the conditional independence structure, we can write

$$\begin{aligned} p(x_{0:t}|y_{0:t}) &\propto g(y_t|x_t) f(x_t|x_{t-1}) p(x_{0:t-1}|y_{0:t-1}) && (1) \\ &\propto \mu(x_0) g(y_0|x_0) \prod_{i=1}^{t} f(x_i|x_{i-1}) g(y_i|x_i) && (2) \end{aligned}$$

for $t = 0, \ldots, M$, where the one-step recursion (1) is obtained using Bayes rule, and (2) is obtained by applying (1) t times. The filtering distribution $p(x_t|y_{0:t})$ can be obtained from (1) by marginalising out $x_{0:t-1}$, which is straightforward if Monte Carlo samples are available. The predictive distributions can also be derived from the smoothing distributions using

$$p(x_{t+1}|y_{0:t}) \propto f(x_{t+1}|x_t) p(x_{0:t}|y_{0:t}).$$

SMC provides a particle method to approximate to (1), given a model specification and a sequence of observations. Like the underlying process, the algorithm proceeds sequentially, returning its approximation to the smoothing distribution at each time step. This approximation is the empirical distribution of the particles:

$$\hat{p}(x_{0:t}|y_{0:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \tag{3}$$

The particle approximation is justified by various convergence results - see for example Del Moral (2013) for details.

A generic SMC algorithm is presented in Algorithm 1. For the state space model described above, we can take $K_{t+1}(x_t, \cdot) \equiv f(\cdot|x_t)$ and $g_{t+1}(x_t, x_{t+1}) \equiv g(y_{t+1}|x_{t+1})$.

Figure 1 illustrates the particle approximation arising from such an algorithm on a linear Gaussian model, with the exact posterior for reference.

Algorithm 1 Standard SMC

Require: $N, T, \mu, \{K_t\}, \{g_t\}, y_{0:T}$
for $i \in \{1, \ldots, N\}$ **do**
  Sample $X_0^{(i)} \sim \mu(\cdot)$ ▷ initialise
  $w_0^{(i)} \leftarrow \frac{g_0(X_0^{(i)})}{\sum_{j=1}^{N} g_0(X_0^{(j)})}$
end for
for $t \in \{0, \ldots, T-1\}$ **do**
  Sample $a_t^{(1:N)} \sim \text{RESAMPLE}(\{1, \ldots, N\}, w_t^{(1:N)})$ ▷ resample particles
  for $i \in \{1, \ldots, N\}$ **do**
    Sample $X_{t+1}^{(i)} \sim K_{t+1}(X_t^{(a_t^{(i)})}, \cdot)$ ▷ propagate particles
    $w_{t+1}^{(i)} \leftarrow g_{t+1}(X_t^{(a_t^{(i)})}, X_{t+1}^{(i)})$ ▷ calculate weights
  end for
  $W \leftarrow \sum_{j=1}^{N} w_{t+1}^{(j)}$
  for $i \in \{1, \ldots, N\}$ **do**
    $w_{t+1}^{(i)} \leftarrow \frac{1}{W} w_{t+1}^{(i)}$ ▷ normalise weights
  end for
end for

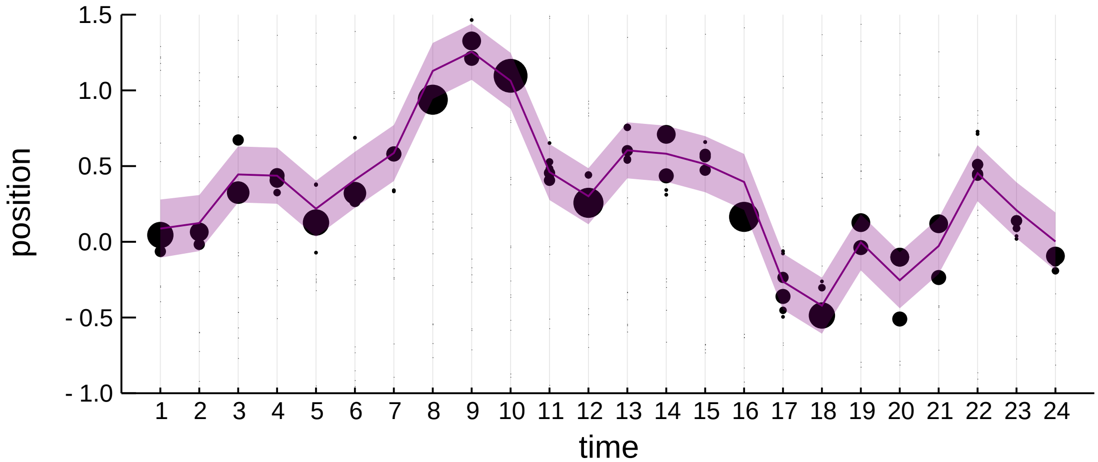

Figure 1: SMC particles before resampling for a linear Gaussian model. The purple ribbon shows the exact posterior mode and 95% credible interval, computed using the Kalman filter and RTS smoother. The black dots show the positions of the SMC particles, with size proportional to weight. After resampling all particles have equal weights but some are duplicated.

If only the latest filtering distribution is required, we can marginalise out $\mathbf{x}_{0:t-1}$ at each step by simply throwing away the particle histories and keeping only the particle approximation $\mathbf{x}_t$ to the filtering distribution at the current time t. The algorithm progresses in a Markovian fashion, only ever referring to the particles at the immediately previous step, so filtering distributions can be approximated with minimal memory usage. If, say, the mean and variance of $X_t \mid y_{0:t}$ at each time t are required, we can store just these summary statistics, plus the two most recent generations of particles, and throw away all other information about the particles at previous time steps. This is vital if one wishes to carry out filtering in an on-line fashion, as it prevents the memory requirements accumulating more than necessary.

The form of the RESAMPLE function in Algorithm 1 is discussed in Section 5.

2.4 Ancestral degeneracy

- the problem with smoothing: ancestral vs. weight degeneracy

3 SMC genealogies as coalescents

- pop gen literature about large population cts time limits of various models
- resampling viewed backwards in time: branching process $\to$ coalescent process
- asymptotic properties of SMC lit review: CLT, path storage, coalescence etc.
- the gap in knowledge that we aim to fill

3.1 Kingman's coalescent

Imagine we have a population with fixed size N over discrete generations, where each individual is descended from one randomly chosen individual of the previous generation. Then for each individual in the present generation, we can trace their *lineage* back through the generations. If we trace two lineages back in time, at some generation they may descend from the same individual, at which point we say they have *coalesced*. Once two lineages have coalesced they will stay together going backwards in time. The combined lineages of $n \leq N$ of the present individuals therefore forms a tree, or several non-overlapping trees, the entirety of which we refer to as the *ancestry* or *genealogy* of those n individuals.

Kingman's n-*coalescent* provides a model for such genealogies. Kingman showed in (Kingman, 1982*a*,*b*,*c*) that the n-coalescent is the limiting process for samples from a wide class of population models as $N \to \infty$.

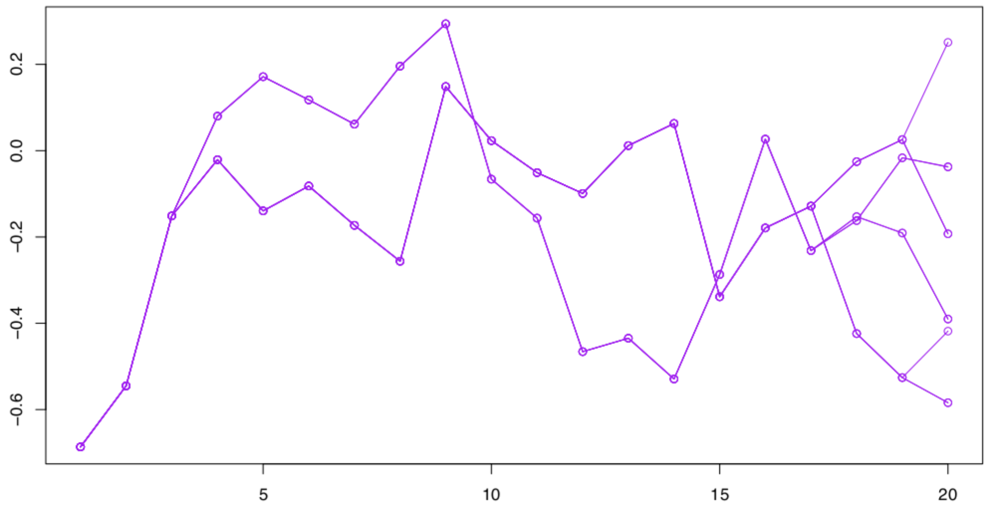

Figure 2: A sample of $N = 6$ trajectories, illustrating ancestral degeneracy. At the "present" time there are six distinct lineages, but just three steps back they have coalesced onto only two lineages, and it takes less than 20 steps back before only one lineage remains.

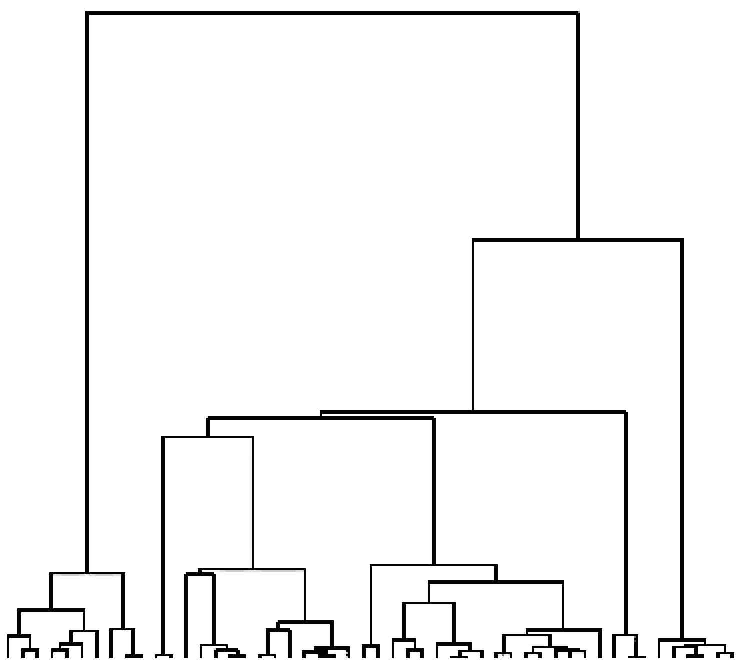

Figure 3: A realisation of Kingman's n-coalescent for a sample of size $n = 50$. At first there are many distinct lineages, and mergers happen rapidly. Once there are fewer distinct lineages left, they take longer to merge. The process spends about half of its time with just two or three distinct lineages. *(Source: Wikimedia Commons)*

The defining feature of the model is that each pair of lineages merges with unit rate. This means that many coalescences occur while there are many distinct lineages present. In particular, the n-coalescent can be formulated as a Poisson process where pairs of lineages coalesce independently at rate 1, with the pair to coalesce being chosen uniformly at random (Wakeley, 2009, Section 3.2).

In the notation of Wakeley (2009), let T_i; $i = 2, \dots, n$ be the i^{th} coalescence time, that is, the length of time for which there are exactly i branches in the sample genealogy. The n-coalescent is the process in which these times are distributed as independent Exponentials with rate $\binom{i}{2}$.

Möhle (1998) writes the same process in terms of the infinitesimal generator Q of a Markov process on the set of equivalence relations on n elements, having entries

$$q_{\xi\eta} = \begin{cases} -\binom{b}{2} & \text{if } \xi = \eta \\ 1 & \text{if } \xi \prec \eta \\ 0 & \text{otherwise} \end{cases}$$

where b is the number of equivalence classes of ξ, and $\xi \prec \eta$ means that η is a state with exactly one more pair of lineages coalesced compared to ξ.

The *Kingman coalescent* is the process on the whole population of size $N \to \infty$, such that the genealogy of any sample of size $n < N$ individuals from the present generation is an n-coalescent.

3.2 Existing results

The first results showing convergence of population models to the Kingman coalescent appear in Kingman's original paper (Kingman, 1982*c*) introducing the Kingman coalescent. This includes, but is not limited to, the neutral Wright-Fisher model (Fisher, 1923, 1930; Wright, 1931) and the Moran model (Moran, 1958). It is known that a general class of exchangeable models known as neutral Cannings models converge to the Kingman coalescent (Etheridge, 2011, Section 2.2). Möhle (1998) proved convergence for a larger class, including some non-exchangeable models.

Koskela et al. (2018) presents the first application of this type of analysis to SMC genealogies. Their result relies heavily on the methods introduced by Möhle (1998). They were able to prove convergence to the Kingman coalescent for genealogies induced by standard SMC algorithms with multinomial resampling. In the following sections we attempt to extend their result to cover some other SMC algorithms.

4 Conditional SMC

Conditional SMC differs from the standard algorithm in that one predetermined trajectory (that is, a sequence of particle positions and the corresponding ancestral line) is conditioned to survive all of the propagation and resampling steps. We will refer to this sequence as the *immortal trajectory*, following the terminology used for conditioned Galton-Watson processes, and the *immortal particle* will refer to the particle in a particular generation that is part of the immortal trajectory.

The conditional SMC algorithm was proposed by Andrieu et al. (2010) for use in the *particle Gibbs* sampler, which they introduce as part of a more general class of particle MCMC methods. In the particle Gibbs sampler, the standard SMC algorithm does not admit the desired target distribution, so this conditional version must be used instead.

When used as a component of the particle Gibbs algorithm, the immortal trajectory $x^*_{0:T}$ for each SMC run is sampled from the trajectories output from the previous run (Andrieu et al., 2010, Section 2.4.3). However, for our purposes we just consider a single SMC run for which the immortal trajectory is fixed.

A conditional SMC algorithm employing multinomial resampling is described in Algorithm 2.

In the particle Gibbs sampler, it is crucial that the conditional SMC output maintains at least two distinct trajectories. The immortal trajectory will of course be among the surviving trajectories, but additionally, the new immortal trajectory (for the next SMC run) is chosen from among the surviving trajectories. Thus if all the trajectories coalesce onto the immortal trajectory, we are forced to choose the same immortal trajectory for the next run, at least for some early time steps. One can imagine that if there was a high probability of full coalescence on each run, we could easily end up with samples from $p(x_{0:T}|y_{0:T})$ that are identical in some coordinates $0 : t$, which would not lead to good results overall.

Algorithm 2 Conditional SMC with multinomial resampling

Require: $N, T, \mu, \{K_t\}, \{g_t\}, y_{0:T}, x^*_{0:T}$

for $i \in \{1, \ldots, N\}$ **do**
    Sample $X_0^{(i)} \sim \mu(\cdot)$ ▷ initialise
end for
Sample $a_0^* \sim \text{Uniform}(\{1, \ldots, N\})$
$X_0^{(a_0^*)} \leftarrow x_0^*$
for $i \in \{1, \ldots, N\}$ **do**
    $w_0^{(i)} \leftarrow \frac{g_0(X_0^{(i)})}{\sum_{j=1}^N g_0(X_0^{(j)})}$
end for
for $t \in \{0, \ldots, T-1\}$ **do**
    Sample $a_t^{(1:N)} \sim \text{Categorical}(\{1, \ldots, N\}, w_t^{(1:N)})$ ▷ resample particles
    Sample $a_{t+1}^* \sim \text{Uniform}(\{1, \ldots, N\})$
    $a_t^{(a_{t+1}^*)} \leftarrow a_t^*$
    for $i \in \{1, \ldots, N\}$ **do**
        Sample $X_{t+1}^{(i)} \sim K_{t+1}(X_t^{(a_t^{(i)})}, \cdot)$ ▷ propagate particles
    end for
    $X_{t+1}^{(a_{t+1}^*)} \leftarrow X_{t+1}^*$
    for $i \in \{1, \ldots, N\}$ **do**
        $w_{t+1}^{(i)} \leftarrow g_{t+1}(X_t^{(a_t^{(i)})}, X_{t+1}^{(i)})$ ▷ calculate weights
    end for
    $W \leftarrow \sum_{j=1}^N w_{t+1}^{(j)}$
    for $i \in \{1, \ldots, N\}$ **do**
        $w_{t+1}^{(i)} \leftarrow \frac{1}{W} w_{t+1}^{(i)}$ ▷ normalise weights
    end for
end for

The problem can be avoided by using a sufficiently large number of particles for the fixed time window T of the conditional SMC runs. This would require a priori knowledge of the coalescence mechanism, which is not available. However, Corollary 1 could possibly provide such knowledge. If, say, we want to ensure that the probability of all N lineages coalescing is below a certain threshold, all of the relevant information is encoded in the distribution of the time to MRCA of the genealogical process. For the Kingman coalescent this distribution is known, and Corollary 1 states that as $N \to \infty$ the genealogy is a Kingman coalescent. The remaining question is whether the Kingman coalescent provides a reasonable approximation outside of the asymptotic regime - since in reality we simulate finitely many particles. We intend to investigate this question by way of a simulation study.

4.1 Genealogies of conditional SMC algorithms

In this section we calculate various quantities related to the genealogical process induced by conditional SMC with multinomial resampling. By writing these in terms of the corresponding quantities for standard SMC with multinomial resampling, we are able to apply results from Koskela et al. (2018). In this way we will show that the genealogical process converges to the Kingman coalescent, in the sense of finite-dimensional distributions, as the number of particles $N \to \infty$.

The derivations of the expressions (4), (5), (6), along with details of the application of results from Koskela et al. (2018), are relegated to the appendix. Below is an overview of the proof. To prove convergence to the Kingman coalescent, we must control the rates of different types of mergers. In particular, we ensure that in the large population limit (under an appropriate time-scaling), pairwise mergers happen at the correct rate, and larger mergers never occur.

Throughout the following we use tilde to indicate the conditional SMC versions of the untilded quantities relating to standard SMC, always with multinomial resampling.

Firstly, we have the expected coalescence rate:

$$\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] = \frac{N-2}{N}\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + \frac{2}{N}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \tag{4}$$

Then the expected rate of super-binary mergers (that is, more than two lineages merging simultaneously into one or more lineages) is bounded above by:

$$\begin{aligned}\mathbb{E}[\tilde{D}_N(t)|\mathcal{F}_{t-1}] \leq \mathbb{E}[D_N(t)|\mathcal{F}_{t-1}] &+ \frac{3}{N}\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + \frac{4}{N^2}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\ &+ \frac{4}{N}\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}] + \frac{2}{N^2}\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}w_t^{(i)}|\mathcal{F}_{t-1}] + \frac{1}{N^2}\sum_{i=2}^{N}\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\end{aligned} \tag{5}$$

And lastly the expectation of the squared coalescence rate is bounded above by:

$$\begin{aligned}\mathbb{E}[\tilde{c}_N(t)^2|\mathcal{F}_{t-1}] \leq \mathbb{E}[c_N(t)^2|\mathcal{F}_{t-1}] &+ \frac{4}{N}\mathbb{E}[(w_t^{(1)})^3|\mathcal{F}_{t-1}] + \frac{12}{N^2}\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + \frac{4}{N(N)_2}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\ &+ \frac{4}{N}\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}]\end{aligned} \tag{6}$$

We then apply Lemma 3 of Koskela et al. (2018) to obtain the more tractable expressions

$$\begin{aligned}&\frac{\varepsilon^4}{Na^4} + O(N^{-2}) \leq \mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] \leq \frac{a^4}{N\varepsilon^4} + O(N^{-2}) \\ &\mathbb{E}[\tilde{D}_N(t)|\mathcal{F}_{t-1}] \leq \frac{C}{N}\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \\ &\mathbb{E}[\tilde{c}_N(t)^2|\mathcal{F}_{t-1}] \leq \frac{C}{N}\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] + O(N^{-3})\end{aligned}$$

and define the time-scaling

$$\tilde{\tau}_N(t) := \min\left\{ s \geq 1 : \sum_{r=1}^{s} \tilde{c}_N(r) \geq t \right\}. \tag{7}$$

Then, using Koskela et al. (2018, Lemma 2), which readily generalises to our modified quantities, we are able to verify the four conditions of Koskela et al. (2018, Theorem 1). Finally we are able to conclude the following.

Corollary 1. *Under the conditions of Koskela et al. (2018, Lemma 3), the genealogy of any n particles from a conditional SMC algorithm with multinomial resampling converges to Kingman's n-coalescent in the sense of finite-dimensional distributions, under the time-scaling defined in* (7).

5 Alternative resampling schemes

- overview of the main variance-reducing schemes
- results: theorem for residual resampling (hopefully)
- maybe results for other schemes

There is a great deal of flexibility in the function referred to as RESAMPLE in Algorithm 1. The most straightforward choice is multinomial resampling (Efron and Tibshirani, 1994), which is also the easiest to analyse. However, multinomial resampling is well known to be sub-optimal in terms of the resulting Monte Carlo variance, and is rarely used in practice. For instance, Douc et al. (2005) proves that both residual resampling and stratified resampling yield lower variance. In this section we will present some resampling schemes that claim to perform better than multinomial resampling.

6 Discussion

- results so far
- impact of this work: to practitioners, to enriching the SMC literature, interpretation within pop gen.
- future directions

A Proof of Corollary 1

In the derivation of (4) – (6) we will make extensive use of the formula for factorial moments of the multinomial distribution given in Mosimann (1962, p.67):

$$\mathbb{E}[(X_i)_a(X_j)_b] = (n)_{a+b}\, p_i^a p_j^b \tag{8}$$

where $(X_1, \dots X_k) \sim \text{Multinomial}(n, \mathbf{p})$. To apply this formula we need to write everything in terms of falling factorial powers. The required conversions are summarised in Table 1.

In standard SMC with multinomial resampling, the marginal offspring distributions, conditioned on the filtration $\mathcal{F}_{t-1}$ generated by the previous offspring counts, are

$$v_t^{(i)} \overset{d}{=} \text{Binomial}(N, w_t^{(i)}), \qquad i = 1, \dots, N$$

where $v_t^{(i)}$ is the number of offspring in generation $t+1$ of the ith particle in generation t, N is the number of particles and $w_t^{(i)}$ is the weight associated with the ith particle in generation t.

In conditional SMC we condition on the immortal trajectory surviving each resampling step. By exchangeability we can set without loss of generality that the immortal trajectory consists of particle 1 in each generation. At each resampling step, particle 1 must therefore choose particle 1 as its parent, while the remaining $N-1$ offspring are assigned multinomially to the N possible parents. The marginal offspring distributions are then

$$\tilde{v}_t^{(1)} \overset{d}{=} 1 + \text{Binomial}(N-1, w_t^{(1)})$$
$$\tilde{v}_t^{(i)} \overset{d}{=} \text{Binomial}(N-1, w_t^{(i)}), \qquad i = 2, \dots, N.$$

First let us consider the pair-merger rate

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

x	$=$	$(x)_1$
x^2	$=$	$(x)_2 + (x)_1$
x^3	$=$	$(x)_3 + 3(x)_2 + (x)_1$
x^4	$=$	$(x)_4 + 6(x)_3 + 7(x)_2 + (x)_1$
xy	$=$	$(x)_1(y)_1$
x^2y	$=$	$(x)_2(y)_1 + (x)_1(y)_1$
xy^2	$=$	$(x)_1(y)_2 + (x)_1(y)_1$
x^2y^2	$=$	$(x)_2(y)_2 + (x)_2(y)_1 + (x)_1(y)_2 + (x)_1(y)_1$
$(x+1)_2$	$=$	$(x)_2 + 2(x)_1$
$(x+1)^2$	$=$	$(x)_2 + 3(x)_1 + 1$
$(x+1)_2(x+1)$	$=$	$(x)_3 + 5(x)_2 + 4(x)_1$
$(x+1)_2^2$	$=$	$(x)_4 + 8(x)_3 + 14(x)_2 + 4(x)_1$

Table 1: Conversion of ordinary powers into falling factorial powers

For standard SMC the expected value is, using the tower rule,

$$\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] = \frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}\left[\mathbb{E}[(v_t^{(i)})_2]|\mathcal{F}_{t-1}\right] = \frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}\left[(N)_2(w_t^{(i)})^2|\mathcal{F}_{t-1}\right] = \sum_{i=1}^{N}\mathbb{E}\left[(w_t^{(i)})^2|\mathcal{F}_{t-1}\right]$$

as stated in Koskela et al. (2018, Remark 3). In the case of conditional SMC we separate the first term (corresponding to the immortal particle) from the sum to get

$$\begin{aligned}
\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] &= \frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}\left[(\tilde{v}_t^{(i)})_2|\mathcal{F}_{t-1}\right] = \frac{1}{(N)_2}\mathbb{E}\left[(\tilde{v}_t^{(1)})_2|\mathcal{F}_{t-1}\right] + \frac{1}{(N)_2}\sum_{i=2}^{N}\mathbb{E}\left[(\tilde{v}_t^{(i)})_2|\mathcal{F}_{t-1}\right] \\
&= \frac{1}{(N)_2}\left\{(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + 2(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] + \sum_{i=2}^{N}(N-1)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\right\} \\
&= \frac{(N-1)_2}{(N)_2}\sum_{i=1}^{N}\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}] + \frac{2(N-1)}{(N)_2}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&= \frac{N-2}{N}\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + \frac{2}{N}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}]
\end{aligned}$$

which gives us (4).

An upper bound on the rate of super-binary mergers is given by

$$D_N(t) := \frac{1}{N(N)_2}\sum_{i=1}^{N}(v_t^{(i)})_2\left(v_t^{(i)} + \frac{1}{N}\sum_{j\neq i}(v_t^{(j)})^2\right).$$

In the standard case this quantity has expectation

$$\begin{aligned}
\mathbb{E}[D_N(t)|\mathcal{F}_{t-1}] = \frac{1}{N(N)_2}\sum_{i=1}^{N}&\left\{(N)_3\mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2(N)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\right\} \\
&+ \frac{1}{N^2(N)_2}\sum_{i=1}^{N}\sum_{j\neq i}\left\{(N)_4\mathbb{E}[(w_t^{(i)})^2(w_t^{(j)})^2|\mathcal{F}_{t-1}] + (N)_3\mathbb{E}[(w_t^{(i)})^2 w_t^{(j)}|\mathcal{F}_{t-1}]\right\}
\end{aligned}$$

while in the conditional case, again separating the terms involving particle 1,

$$
\begin{aligned}
\tilde{D}_N(t) &= \frac{1}{N(N)_2}(\tilde{v}_t^{(1)})_2 \left(\tilde{v}_t^{(1)} + \frac{1}{N}\sum_{j\neq 1}(\tilde{v}_t^{(j)})^2 \right) + \frac{1}{N(N)_2}\sum_{i\neq 1}(\tilde{v}_t^{(i)})_2 \left(\tilde{v}_t^{(i)} + \frac{1}{N}(\tilde{v}_t^{(1)})^2 + \frac{1}{N}\sum_{1\neq j\neq i}(\tilde{v}_t^{(j)})^2 \right) \\
&= \frac{1}{N(N)_2}\left\{ (\tilde{v}_t^{(1)})_2 v_t^{(1)} + \frac{1}{N}\sum_{j\neq 1}(\tilde{v}_t^{(1)})_2(\tilde{v}_t^{(j)})^2 + \frac{1}{N}\sum_{i\neq 1}(\tilde{v}_t^{(i)})_2(\tilde{v}_t^{(1)})^2 \right\} \\
&\quad + \frac{1}{N(N)_2}\sum_{i\neq 1}\left\{ (\tilde{v}_t^{(i)})_2\tilde{v}_t^{(i)} + \frac{1}{N}\sum_{1\neq j\neq i}(\tilde{v}_t^{(i)})_2(\tilde{v}_t^{(j)})^2 \right\} \\
&= \frac{1}{N(N)_2}\left\{ (\tilde{v}_t^{(1)})_2\tilde{v}_t^{(1)} + \frac{1}{N}\sum_{i\neq 1}\Big((\tilde{v}_t^{(1)})_2(\tilde{v}_t^{(i)})^2 + (\tilde{v}_t^{(i)})_2(\tilde{v}_t^{(1)})^2 \Big) \right\} \\
&\quad + \frac{1}{N(N)_2}\sum_{i\neq 1}\left\{ (\tilde{v}_t^{(i)})_3 + 2(\tilde{v}_t^{(i)})_2 + \frac{1}{N}\sum_{1\neq j\neq i}\Big((\tilde{v}_t^{(i)})_2(\tilde{v}_t^{(j)})_2 + (\tilde{v}_t^{(i)})_2\tilde{v}_t^{(j)} \Big) \right\}
\end{aligned}
$$

and so by applying the moments from (8) and Table 1 we find the expectation

$$
\begin{aligned}
&\mathbb{E}[\tilde{D}_N(t)|\mathcal{F}_{t-1}] = \\
&= \frac{1}{N(N)_2}\Big\{(N-1)_3\mathbb{E}[(w_t^{(1)})^3|\mathcal{F}_{t-1}] + 5(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + 4(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N^2(N)_2}\sum_{i=2}^{N}\Big\{2(N-1)_4\mathbb{E}[(w_t^{(1)})^2(w_t^{(i)})^2|\mathcal{F}_{t-1}] + (N-1)_3\mathbb{E}[(w_t^{(1)})^2 w_t^{(i)}|\mathcal{F}_{t-1}] \\
&\qquad + 5(N-1)_3\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}] + 2(N-1)_2\mathbb{E}[w_t^{(1)}w_t^{(i)}|\mathcal{F}_{t-1}] + (N-1)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N(N)_2}\sum_{i=2}^{N}\Big\{(N-1)_3\mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2(N-1)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N^2(N)_2}\sum_{i=2}^{N}\sum_{1\neq j\neq i}\Big\{(N-1)_4\mathbb{E}[(w_t^{(i)})^2(w_t^{(j)})^2|\mathcal{F}_{t-1}] + (N-1)_3\mathbb{E}[(w_t^{(i)})^2 w_t^{(j)}|\mathcal{F}_{t-1}]\Big\} \\
&= \frac{1}{N(N)_2}\sum_{i=1}^{N}\Big\{(N-1)_3\mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2(N-1)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N^2(N)_2}\sum_{i=1}^{N}\sum_{j\neq i}\Big\{(N-1)_4\mathbb{E}[(w_t^{(i)})^2(w_t^{(j)})^2|\mathcal{F}_{t-1}] + (N-1)_3\mathbb{E}[(w_t^{(i)})^2 w_t^{(j)}|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N(N)_2}\Big\{3(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + 4(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N^2(N)_2}\sum_{i=2}^{N}\Big\{4(N-1)_3\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}] + 2(N-1)_2\mathbb{E}[w_t^{(1)}w_t^{(i)}|\mathcal{F}_{t-1}] + (N-1)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\Big\} \\
&\leq \mathbb{E}[D_N(t)|\mathcal{F}_{t-1}] + \frac{1}{N(N)_2}\Big\{3(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + 4(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}]\Big\} \\
&\quad + \frac{1}{N^2(N)_2}\sum_{i=2}^{N}\Big\{4(N-1)_3\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}] + 2(N-1)_2\mathbb{E}[w_t^{(1)}w_t^{(i)}|\mathcal{F}_{t-1}] + (N-1)_2\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\Big\} \\
&\leq \mathbb{E}[D_N(t)|\mathcal{F}_{t-1}] + \frac{3}{N}\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + \frac{4}{N^2}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&\quad + \frac{4}{N}\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}] + \frac{2}{N^2}\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}w_t^{(i)}|\mathcal{F}_{t-1}] + \frac{1}{N^2}\sum_{i=2}^{N}\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]
\end{aligned}
$$

The second line of the first equality relies on multiplying the relevant terms in Table 1. For the second equality we recombine the terms in particle 1 into the sum. The inequalities follow by bounding e.g. $N-1$ by N, and identifying the first two lines with $\mathbb{E}[D_N(t)|\mathcal{F}_{t-1}]$. This gives us the inequality (5).

We also need control of the squared coalescence rate:

$$
c_N(t)^2 = \frac{1}{(N)_2^2}\left(\sum_{i=1}^{N}(v_t^{(i)})_2\right)^2 = \frac{1}{(N)_2^2}\left\{\sum_{i=1}^{N}\mathbb{E}[(v_t^{(i)})_2^2] + \sum_{i=1}^{N}\sum_{j\neq i}\mathbb{E}[(v_t^{(i)})_2(v_t^{(j)})_2]\right\}
$$

A bound on its expected value is proved in Koskela et al. (2018), but here we will use a different, more explicit bound to allow direct comparison between the standard and conditional cases. For standard SMC we have:

$$\begin{aligned}
\mathbb{E}[c_N(t)^2|\mathcal{F}_{t-1}] = \frac{1}{(N)_2^2} &\left\{ (N)_4 \sum_{i=1}^N \mathbb{E}[(w_t^{(i)})^4|\mathcal{F}_{t-1}] + 4(N)_3 \sum_{i=1}^N \mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2(N)_2 \sum_{i=1}^N \mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}] \right\} \\
&+ \frac{1}{(N)_2^2}(N)_4 \sum_{i=1}^N \sum_{j\neq i} \mathbb{E}[(w_t^{(i)})^2 (w_t^{(j)})^2|\mathcal{F}_{t-1}] \\
= \frac{1}{(N)_2} &\left\{ (N-2)_2 \sum_{i=1}^N \mathbb{E}[(w_t^{(i)})^4|\mathcal{F}_{t-1}] + 4(N-2) \sum_{i=1}^N \mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2 \sum_{i=1}^N \mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}] \right. \\
&\left. \qquad + (N-2)_2 \sum_{i=1}^N \sum_{j\neq i} \mathbb{E}[(w_t^{(i)})^2 (w_t^{(j)})^2|\mathcal{F}_{t-1}] \right\}
\end{aligned}$$

For conditional SMC, we again separate the terms involving particle 1:

$$\begin{aligned}
\tilde{c}_N(t)^2 &= \frac{1}{(N)_2^2} \left\{ \sum_{i=1}^N \mathbb{E}[(\tilde{v}_t^{(i)})_2^2] + \sum_{i=1}^N \sum_{j\neq i} \mathbb{E}[(\tilde{v}_t^{(i)})_2 (\tilde{v}_t^{(j)})_2] \right\} \\
&= \frac{1}{(N)_2^2} \left\{ \sum_{i=2}^N \mathbb{E}[(\tilde{v}_t^{(i)})_2^2] + \mathbb{E}[(\tilde{v}_t^{(1)})_2^2] + \sum_{i=2}^N \sum_{1\neq j\neq i} \mathbb{E}[(\tilde{v}_t^{(i)})_2 (\tilde{v}_t^{(j)})_2] + 2\sum_{i=2}^N \mathbb{E}[(\tilde{v}_t^{(1)})_2 (\tilde{v}_t^{(i)})_2] \right\}
\end{aligned}$$

and then use the same techniques as for $\tilde{D}_N(t)$ to calculate the expectation:

$$
\begin{aligned}
&\mathbb{E}[\tilde{c}_N(t)^2|\mathcal{F}_{t-1}] = \\
&= \frac{1}{(N)_2^2}\left\{(N-1)_4\sum_{i=2}^{N}\mathbb{E}[(w_t^{(i)})^4|\mathcal{F}_{t-1}] + 4(N-1)_3\sum_{i=2}^{N}\mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2(N-1)_2\sum_{i=2}^{N}\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\right\} \\
&\quad + \frac{1}{(N)_2^2}\left\{(N-1)_4\mathbb{E}[(w_t^{(1)})^4|\mathcal{F}_{t-1}] + 8(N-1)_3\mathbb{E}[(w_t^{(1)})^3|\mathcal{F}_{t-1}] + 14(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}]\right\} \\
&\quad + \frac{1}{(N)_2^2}4(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] + \frac{1}{(N)_2^2}(N-1)_4\sum_{i=2}^{N}\sum_{1\neq j\neq i}\mathbb{E}[(w_t^{(i)})^2(w_t^{(j)})^2|\mathcal{F}_{t-1}] \\
&\quad + \frac{2}{(N)_2^2}\sum_{i=2}^{N}\Big((N-1)_4\mathbb{E}[(w_t^{(1)})^2(w_t^{(i)})^2|\mathcal{F}_{t-1}] + 2(N-1)_3\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}]\Big) \\
&= \frac{1}{(N)_2^2}\left\{(N-1)_4\sum_{i=1}^{N}\mathbb{E}[(w_t^{(i)})^4|\mathcal{F}_{t-1}] + 4(N-1)_3\sum_{i=1}^{N}\mathbb{E}[(w_t^{(i)})^3|\mathcal{F}_{t-1}] + 2(N-1)_2\sum_{i=1}^{N}\mathbb{E}[(w_t^{(i)})^2|\mathcal{F}_{t-1}]\right\} \\
&\quad + \frac{(N-1)_4}{(N)_2^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbb{E}[(w_t^{(i)})^2(w_t^{(j)})^2|\mathcal{F}_{t-1}] \\
&\quad + \frac{1}{(N)_2^2}\left\{4(N-1)_3\mathbb{E}[(w_t^{(1)})^3|\mathcal{F}_{t-1}] + 12(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + 4(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}]\right\} \\
&\quad + \frac{1}{(N)_2^2}\left\{4(N-1)_3\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}]\right\} \\
&\leq \mathbb{E}[c_N(t)^2|\mathcal{F}_{t-1}] + \frac{1}{(N)_2^2}\left\{4(N-1)_3\mathbb{E}[(w_t^{(1)})^3|\mathcal{F}_{t-1}] + 12(N-1)_2\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}]\right\} \\
&\quad + \frac{1}{(N)_2^2}\left\{4(N-1)\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] + 4(N-1)_3\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}]\right\} \\
&\leq \mathbb{E}[c_N(t)^2|\mathcal{F}_{t-1}] + \frac{4}{N}\mathbb{E}[(w_t^{(1)})^3|\mathcal{F}_{t-1}] + \frac{12}{N^2}\mathbb{E}[(w_t^{(1)})^2|\mathcal{F}_{t-1}] + \frac{4}{N(N)_2}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&\quad + \frac{4}{N}\sum_{i=2}^{N}\mathbb{E}[w_t^{(1)}(w_t^{(i)})^2|\mathcal{F}_{t-1}]
\end{aligned}
$$

The conditions (18) and (19) of Koskela et al. (2018, Lemma 3) give us control over the weights so that we have $w_t^{(i)} = O(1)$ for all i. Under these conditions, in the limit as $N \to \infty$, the three modified expectations derived above simplify to:

$$
\begin{aligned}
&\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] \leq \mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + O(N^{-2}) \\
&\mathbb{E}[\tilde{D}_N(t)|\mathcal{F}_{t-1}] \leq \mathbb{E}[D_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \\
&\mathbb{E}[\tilde{c}_N(t)^2|\mathcal{F}_{t-1}] \leq \mathbb{E}[c_N(t)^2|\mathcal{F}_{t-1}] + O(N^{-3})
\end{aligned}
$$

This shows that each of these quantities for conditional SMC is bounded above by the corresponding standard SMC quantity, plus some vanishing error term. This will allow us to apply Koskela et al. (2018, Theorem 1), as we will show in the following.

Next we apply the result of Koskela et al. (2018, Lemma 3), so that for our modified quantity $\tilde{c}_N(t)$ we have the upper bound:

$$\begin{aligned}
\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] &= \frac{N-2}{N}\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + \frac{2}{N}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&\leq \mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + \frac{2}{N}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&\leq \frac{a^4}{N\varepsilon^4} + \frac{2}{N}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&= \frac{a^4}{N\varepsilon^4} + O(N^{-2}) \qquad (9)
\end{aligned}$$

and lower bound:

$$\begin{aligned}
\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] &= \frac{N-2}{N}\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + \frac{2}{N}\mathbb{E}[w_t^{(1)}|\mathcal{F}_{t-1}] \\
&\geq \frac{N-2}{N}\frac{\varepsilon^4}{Na^4} + O(N^{-2}) \\
&= \frac{\varepsilon^4}{Na^4} - \frac{2\varepsilon^4}{N^2a^4} + O(N^{-2}) \\
&= \frac{\varepsilon^4}{Na^4} + O(N^{-2}) \qquad (10)
\end{aligned}$$

corresponding to (22) in Koskela et al. (2018), except for the addition of a vanishing error term. Furthermore, we obtain for the other quantities (where the constant C may change from one line to the next):

$$\begin{aligned}
\mathbb{E}[\tilde{D}_N(t)|\mathcal{F}_{t-1}] &\leq \mathbb{E}[D_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \\
&\leq \frac{C}{N}\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \\
&= \frac{C}{N}\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \qquad (11)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\tilde{c}_N(t)^2|\mathcal{F}_{t-1}] &\leq \mathbb{E}[c_N(t)^2|\mathcal{F}_{t-1}] + O(N^{-3}) \\
&\leq \frac{C}{N}\mathbb{E}[c_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \\
&= \frac{C}{N}\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] + O(N^{-3}) \qquad (12)
\end{aligned}$$

according to equations (20) and (21) in Koskela et al. (2018), again with additional error terms.

Now let us define the time-scaling:

$$\tilde{\tau}_N(t) := \min\left\{ s \geq 1 : \sum_{r=1}^{s} \tilde{c}_N(r) \geq t \right\}$$

which is a generalised inverse of $\tilde{c}_N(t)$ and thus satisfies the property:

$$t - s - 1 \leq \sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{c}_N(r) \leq t - s + 1. \qquad (13)$$

We are finally ready to verify the conditions of Koskela et al. (2018, Theorem 1). The conditions are the following.

(Standing Assumption) The conditional distribution of parental indices $a_t^{(1:N)}$ given offspring counts $v_t^{(1:N)}$ is uniform over all valid assignments.

(A) $\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{D}_N(r)\right] = 0$

(B) $\lim_{N\to\infty} \mathbb{E}[\tilde{c}_N(t)] = 0$

(C) $\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{c}_N(r)^2\right] = 0$

(D) $\mathbb{E}[\tilde{\tau}_N(t) - \tilde{\tau}_N(s)] \leq C_{t,s} N$

These five conditions are verified below.

(Standing Assumption) This holds by the conditional exchangeability of offspring assignments arising from Algorithm 2, where "valid assignments" require that the resampling preserves the immortal line (i.e. the immortal child is assigned to the immortal parent).

(B) Using (9) and applying the tower rule, we find

$$\mathbb{E}[\tilde{c}_N(t)] = \mathbb{E}[\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}]] \leq \frac{a^4}{N\varepsilon^4} + O(N^{-2}) \overset{N\to\infty}{\longrightarrow} 0$$

(C) Using Koskela et al. (2018, Lemma 2) along with (12) and the upper bound in (13),

$$\begin{aligned}
\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{c}_N(r)^2\right] &= \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \mathbb{E}[\tilde{c}_N(r)^2|\mathcal{F}_{t-1}]\right] \leq \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \left(\frac{C}{N}\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] + O(N^{-3})\right)\right] \\
&= \frac{C}{N}\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}]\right] + O(N^{-2}) = \frac{C}{N}\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{c}_N(r)\right] + O(N^{-2}) \\
&\leq \frac{C}{N}(t-s+1) + O(N^{-2}) \overset{N\to\infty}{\longrightarrow} 0
\end{aligned}$$

(A) The above calculation replacing (12) with (11) yields

$$\begin{aligned}
\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{D}_N(r)\right] &= \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \mathbb{E}[\tilde{D}_N(r)|\mathcal{F}_{t-1}]\right] \\
&\leq \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \left(\frac{C}{N}\mathbb{E}[\tilde{c}_N(t)|\mathcal{F}_{t-1}] + O(N^{-3})\right)\right] \overset{N\to\infty}{\longrightarrow} 0
\end{aligned}$$

(D) Using (10), the upper bound in (13) and Koskela et al. (2018, Lemma 2),

$$\begin{aligned}
\mathbb{E}[\tilde{\tau}_N(t) - \tilde{\tau}_N(s)] &= \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} 1\right] = \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \frac{\mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]}{\mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]}\right] \leq \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \frac{\mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]}{\frac{\varepsilon^4}{Na^4} + O(N^{-2})}\right] \\
&= \frac{1}{\frac{\varepsilon^4}{Na^4} + O(N^{-2})}\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]\right] = \frac{1}{\frac{\varepsilon^4}{Na^4} + O(N^{-2})}\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{c}_N(r)\right] \\
&\leq \frac{t-s+1}{\frac{\varepsilon^4}{Na^4} + O(N^{-2})} = \frac{(t-s+1)a^4 N}{\varepsilon^4 + O(N^{-1})} = (t-s+1)\frac{a^4}{\varepsilon^4}N + O(1)
\end{aligned}$$

where the last equality follows by a Taylor expansion of $(\frac{\varepsilon^4}{Na^4} + O(N^{-2}))^{-1}$.
Similarly we derive a lower bound using (9), the lower bound in (13) and Koskela et al. (2018, Lemma

2):

$$
\begin{aligned}
\mathbb{E}[\tilde{\tau}_N(t)-\tilde{\tau}_N(s)] &= \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} 1\right] = \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \frac{\mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]}{\mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]}\right] \geq \mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \frac{\mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]}{\frac{a^4}{N\varepsilon^4}+O(N^{-2})}\right] \\
&= \frac{1}{\frac{a^4}{N\varepsilon^4}+O(N^{-2})}\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \mathbb{E}[\tilde{c}_N(r)|\mathcal{F}_{t-1}]\right] = \frac{1}{\frac{a^4}{N\varepsilon^4}+O(N^{-2})}\mathbb{E}\left[\sum_{r=\tilde{\tau}_N(s)+1}^{\tilde{\tau}_N(t)} \tilde{c}_N(r)\right] \\
&\geq \frac{t-s-1}{\frac{a^4}{N\varepsilon^4}+O(N^{-2})} = \frac{(t-s-1)\varepsilon^4 N}{a^4+O(N^{-1})} = (t-s-1)\frac{\varepsilon^4}{a^4}N + O(1)
\end{aligned}
$$

Therefore we have as required

$$\mathbb{E}[\tilde{\tau}_N(t)-\tilde{\tau}_N(s)] \sim C_{t,s}N$$

as $N \to \infty$.

This conclude the proof of Corollary 1. □

References

Andrieu, C., Doucet, A. and Holenstein, R. (2010), 'Particle Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**(3), 269–342.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970), 'A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains', *The Annals of Mathematical Statistics* **41**, 164–171.

Del Moral, P. (2013), *Mean Field Simulation for Monte Carlo Integration*, Chapman and Hall/CRC.

Del Moral, P., Doucet, A. and Jasra, A. (2006), 'Sequential Monte Carlo samplers', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(3), 411–436.

Douc, R., Cappé, O. and Moulines, E. (2005), Comparison of resampling schemes for particle filtering, *in* 'Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on', IEEE, pp. 64–69.

Doucet, A., De Freitas, N. and Gordon, N. (2001), An introduction to sequential Monte Carlo methods, *in* 'Sequential Monte Carlo methods in Practice', Springer, pp. 3–14.

Doucet, A. and Johansen, A. M. (2011), A tutorial on particle filtering and smoothing: Fifteen years later, *in* 'Handbook of nonlinear filtering', OUP, pp. 656–704.

Efron, B. and Tibshirani, R. J. (1994), *An Introduction to the Bootstrap*, CRC press.

Etheridge, A. (2011), *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*, Springer.

Fisher, R. A. (1923), 'On the dominance ratio', *Proceedings of the Royal Society of Edinburgh* **42**, 321–341.

Fisher, R. A. (1930), 'The distribution of gene ratios for rare mutations', *Proceedings of the Royal Society of Edinburgh* **50**, 205–220.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering* **82**(1), 35–45.

Kingman, J. (1982*a*), 'The coalescent', *Stochastic Processes and Their Applications* **13**(3), 235–248.

Kingman, J. (1982*b*), Exchangeability and the evolution of large populations, *in* 'Proceedings of the International Conference on Exchangeability in Probability and Statistics, Rome, 6th-9th April, 1981, in Honour of Professor Bruno de Finetti', North-Holland, Amsterdam.

Kingman, J. (1982*c*), 'On the genealogy of large populations', *Journal of Applied Probability* **19**(A), 27–43.

Koskela, J., Jenkins, P. A., Johansen, A. M. and Spanò, D. (2018), 'Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo', *arXiv preprint arXiv:1804.01811* .

Möhle, M. (1998), 'Robustness results for the coalescent', *Journal of Applied Probability* **35**(2), 438–447.

Moran, P. A. P. (1958), Random processes in genetics, *in* 'Mathematical Proceedings of the Cambridge Philosophical Society', Vol. 54, Cambridge University Press, pp. 60–71.

Mosimann, J. E. (1962), 'On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions', *Biometrika* **49**(1/2), 65–82.

Rauch, H. E., Striebel, C. and Tung, F. (1965), 'Maximum likelihood estimates of linear dynamic systems', *AIAA Journal* **3**(8), 1445–1450.

Vidoni, P. (1999), 'Exponential family state space models based on a conjugate latent process', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**(1), 213–221.

Wakeley, J. (2009), *Coalescent Theory: An Introduction*, Roberts & Co. Publishers.

Wright, S. (1931), 'Evolution in mendelian populations', *Genetics* **16**(2), 97–159.