Simple conditions for convergence of sequential Monte Carlo genealogies with applications

Suzie Brown* Paul A. Jenkins*† Adam M. Johansen*† Jere Koskela*

Abstract

We present simple conditions under which the limiting genealogical process associated with a class of interacting particle systems with non-neutral selection mechanisms, as the number of particles grows, is a time-rescaled Kingman coalescent. Sequential Monte Carlo algorithms are popular methods for approximating integrals in problems such as non-linear filtering and smoothing which employ this type of particle system. Their performance depends strongly on the properties of the induced genealogical process. We verify the conditions of our main result for standard sequential Monte Carlo algorithms with a broad class of low-variance resampling schemes, as well as for conditional sequential Monte Carlo with multinomial resampling.

Keywords: coalescent; interacting particle system; particle filter; resampling; selection.

MSC2020 subject classifications: 60J90; 60J95; 65C05; 65C35.

Submitted to EJP on July 13, 2020, final version accepted on November 18, 2020.

Supersedes arXiv:2007.00096v1.

1 Introduction

Interacting particle systems (IPSs) are used as models in many situations in which one wants to study the macroscopic effects of a system defined via a large number of microscopic interactions. Such models have been used to explain phenomena as diverse as tumour growth, behavioural systems, and epidemics. See [23] for an extensive treatment of such systems in continuous time and [6] for a class of discrete time systems of interest in this article. As particles reproduce and die out, there naturally arises an embedded genealogical history of the system, and this genealogy is of interest both in its own right and as a tool for understanding the evolution of the system. This is particularly true in population genetics, where a broad class of neutral IPS models for
Simple conditions for convergence of SMC genealogies

an evolving natural population converge, in the infinite population limit and after a suitable time-rescaling, to a well-studied random genealogical process, the so-called \( n \)-coalescent first described by [19]. One might ask: in what other contexts does the \( n \)-coalescent, or another genealogical process, arise as the scaling limit of a particular IPS? Our goal in this paper is to answer this question within the context of sequential Monte Carlo (SMC), a broad class of stochastic algorithms used for computational statistical inference. These algorithms are best known for their application in non-linear filtering and smoothing [15], but can be applied generally to mean-field approximation of Feynman-Kac flows; see [6] for more background. These methods have found diverse applications throughout signal processing, statistics, econometrics, biology, and many other disciplines, and understanding their properties is of widespread importance; see, for example, [4, 8, 11, 27] for recent surveys.

In SMC, a population of particles evolves in discrete time. The algorithm proceeds by iterating through two steps: a mutation step in which the positions of the particles are updated by applying some Markov kernel; and a selection step in which the particles are weighted by some potential function and resampled to form the next generation. Resampling stochastically duplicates high-weight particles and removes low-weight particles. The duplicates of a particle in the following generation are termed its offspring. A succession of resampling steps induces a genealogy, that is, a process recording the parent-offspring relationships between particles at consecutive generations.

SMC genealogies are important because they capture the phenomenon of ancestral degeneracy, which has a substantial impact on the performance of the algorithm. Due to resampling, the number of distinct ancestors whose descendants comprise the particles at the terminal time decays rapidly as the time horizon increases, a well known consequence of which is that path-based Monte Carlo estimators typically have high variance (see, for example, [2, 11]). The theoretical analysis of SMC genealogies is a useful tool for addressing a number of practical questions in SMC methods. It provides a framework for answering questions such as: for a fixed number of particles, what is the maximum time horizon over which smoothing estimators are reliable? This is important for tuning the parameters in settings such as particle MCMC [1], fixed-lag smoothing, and stable variance estimation [28]. It also provides estimates of the memory cost associated with path storage, building on the results of [18] and [20, Corollary 2].

In this paper we provide simple sufficient conditions under which the genealogy of an SMC algorithm converges to a time-rescaled \( n \)-coalescent [19] in the large population limit. We require control over only the second and third moments of the marginal family size of each parent. This builds upon [21], a slight error in which was corrected in [20]. Ours is a substantial improvement over that work, which requires additional control over fourth moments, including cross-terms, to obtain the same convergence result. That work also imposes a condition on the speed of convergence that is violated for instance by the neutral Moran model, any finite sample of which is known to have a Kingman genealogy in the large population limit (see [9, p. 47]). Our result covers algorithms that the existing results do not and also relies on simple conditions which admit verification for a broad class of algorithms.

The result is known to apply to standard SMC with multinomial resampling [20, Corollary 1]. We additionally prove convergence for any resampling scheme based on stochastic rounding, described in Definition 4.1. This includes low-variance schemes such as systematic resampling, residual resampling with stratified residuals, and the more exotic schemes proposed in [5] and [14]. The results presented in this paper therefore provide the first a priori characterization of the genealogy of an SMC algorithm for resampling schemes that are widely used in practice, offering quantitative insight into the nature of ancestral degeneracy and into the relative performance of different resampling
Simple conditions for convergence of SMC genealogies

schemes. We also show that our result applies to conditional SMC, a variant that forms a building block of the particle Gibbs algorithm [1], in which it is important to maintain at least two distinct ancestors across a fixed time window. That our results apply to conditional SMC is important because for many practical statistical problems comprising a high-dimensional latent state space model, particle Gibbs is one of the few algorithms that is able to explore parameter posteriors in reasonable time, so understanding the factors affecting its mixing time is of great interest.

2 Sequential Monte Carlo

Algorithm 1 describes sequential Monte Carlo with \( N \) particles over a fixed time window \( T \). The initial proposal distribution is \( \mu, (K_t)_{t=1}^T \) is a sequence of Markov transition kernels with respective transition densities \( (q_t)_{t=0}^T \) and \( (g_t)_{t=0}^T \) is a sequence of potential functions. For simplicity these may be assumed to exist on a common state space that is a subspace of \( \mathbb{R}^d \), so \( g_0 : \mathbb{R}^d \to \mathbb{R}_+ \) and, for \( t > 0 \), \( q_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+ \) and \( g_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+ \), although the state spaces can in general be any sequence of Polish spaces. We allow potential functions at time \( t \geq 1 \) to depend upon both the current and previous value of the state to allow our results to directly cover several classes of algorithm which occur in practice, including particle filters in which a proposal other than the transition density is used (see, for example, [8, Section 4.1], [4, Section 10.3.2]) and sequential Monte Carlo samplers [7]. At generation \( t \), \( w_t^{(1:N)} = (w_t^{(1)}, \ldots, w_t^{(N)}) \) is the vector of particle weights, and \( a_t^{(1:N)} = (a_t^{(1)}, \ldots, a_t^{(N)}) \) is the vector of resampled parental indices.

The output of Algorithm 1 will depend on the particular application, but might include the complete collection of sampled state values, \( \{X_t^{(i)} : t \in \{0, \ldots, T\}, i \in \{1, \ldots, N\}\} \) and parental indices \( \{a_t^{(i)} : t \in \{0, \ldots, T - 1\}, N \in \{1, \ldots, N\}\} \). These algorithms are widely used in particular for inference in general state space hidden Markov models where the principal estimation problems are filtering, using \( \sum_{i=1}^N w_i^{(t)} g_t \delta_{X_t^{(i)}} \) at time \( t \); smoothing using \( \sum_{i=1}^N w_i^{(t)} g_t \delta_{\tilde{X}_t^{(i)}} \) where the path-particles \( \tilde{X}_t^{(i)} \) are vectors in \( \mathbb{R}^{d \times (t+1)} \) obtained by setting \( \tilde{X}_t^{(i)} = X_t^{(i)} \) and setting \( \tilde{X}_s^{(i)} \) for each \( s < t \) to the state associated with the time \( s \) ancestor of particle \( i \) at time \( t \); and computation of the marginal likelihood via

\[
\hat{Z}_T := \frac{1}{N} \sum_{i=1}^N g_0(X_0^{(i)}) \prod_{t=1}^T \frac{1}{N} \sum_{i=1}^N g_t(X_t^{(a_t^{(i-1)})}, X_t^{(i)}).
\]

Of these, smoothing most explicitly depends upon samples from the full path \( X_{0:T} \), but information about the paths may also be retained in the other settings in order to estimate variance [22, 28] and the associated genealogies influence the sampling properties of all such estimators.

The variety of possible procedures for the RESAMPLE step is an active area of research. Some important examples are explored in Section 4. We take valid resampling schemes to be those where: the total number of resampled offspring is \( N \); the expected number of offspring of particle \( i \) conditional on the weights is \( N w_t^{(i)} \); and each offspring is assigned equal weight after resampling.

Under the standing assumption stated below, it is sufficient for our purposes to consider the vector \( \nu_t^{(1:N)} = (\nu_t^{(1)}, \ldots, \nu_t^{(N)}) \) of offspring counts, where \( \nu_t^{(i)} = |\{j \in \{1, \ldots, N\} : a_t^{(j)} = i\}| \).

Standing Assumption: The conditional distribution of parental indices \( a_t^{(1:N)} \) given offspring counts \( \nu_t^{(1:N)} \) is uniform over all valid assignments.
Simple conditions for convergence of SMC genealogies

Input: $N, T, \mu, (K_t)_{t=0}^T, (g_t)_{t=0}^T$

for $i \in \{1, \ldots, N\}$ do Sample $X_0^{(i)} \sim \mu(\cdot)$

for $i \in \{1, \ldots, N\}$ do $w_0^{(i)} \leftarrow \left\{ \sum_{j=1}^N g_0(X_0^{(j)}) \right\}^{-1} g_0(X_0^{(i)})$

for $t \in \{0, \ldots, T-1\}$ do

  Sample $a_1^{(1:N)} \sim \text{RESAMPLE}([1, \ldots, N], w_1^{(1:N)})$

  for $i \in \{1, \ldots, N\}$ do Sample $X_t^{(i)} \sim K_{t+1}(X_t^{(a_i)} \cdot)$

  for $i \in \{1, \ldots, N\}$ do $w_{t+1}^{(i)} \leftarrow \left\{ \sum_{j=1}^N g_{t+1}(X_t^{(a_j)}, X_{t+1}^{(j)}) \right\}^{-1} g_{t+1}(X_t^{(a_i)}, X_{t+1}^{(i)})$

end

Algorithm 1: Sequential Monte Carlo.

The standing assumption is a weaker condition than exchangeability [25, p. 446]. Any resampling scheme can be made to satisfy it by applying an additional permutation of the offspring indices after selecting the parents (see [1, p. 290]).

3 A limit theorem for sequential Monte Carlo genealogies

For convenience, we will henceforth measure time backwards, with the terminal particles at time 0 and the initial particles at time $T$. The forward-time process of particles replicating or dying induces a coalescent process when viewed in backwards time. The full forward-time process is Markovian, but this no longer holds after integrating out the positions of particles and their weights. Thus, the reverse-time process of ancestral lineages without position or weight information is not Markovian either.

We will analyse an asymptotic regime in which the total number of particles $N \to \infty$, and consider the finite-dimensional restriction to a sample of $n \leq N$ terminal particles. The genealogy of such a sample is described by a partition-valued stochastic process $(G_t^{(n,N)})_{t=0}^T$. At time 0 its value is the set of singletons $\{(1), \ldots, (n)\}$. At each time $t$, $i \neq j \in \{1, \ldots, n\}$ belong to the same partition block in $G_t^{(n,N)}$ if and only if terminal particles $i$ and $j$ share a common ancestor at time $t$. We will take $N \to \infty$ and show that the $n$-coalescent [19] is the correct limiting object. Our limit theorem will apply after a rescaling of time in which the genealogy is viewed over an infinite time horizon; that is, $T \to \infty$.

Definition 3.1. The $n$-coalescent is the homogeneous continuous-time Markov process on the set of partitions of $\{1, \ldots, n\}$ with infinitesimal generator $Q$ having entries

$$
(Q)_{\xi, \eta} = \begin{cases} 
1 & \xi \prec \eta \\
-|\xi|(|\xi| - 1)/2 & \xi = \eta \\
0 & \text{otherwise}
\end{cases}
$$

where $\xi$ and $\eta$ are partitions of $\{1, \ldots, n\}$, $|\xi|$ denotes the number of blocks in $\xi$, and $\xi \prec \eta$ means that $\eta$ is obtained from $\xi$ by merging exactly one pair of blocks.

Throughout the following, falling factorials are denoted by $(a)_b = a(a-1) \cdots (a-b+1)$. We denote by $F_s = \sigma(\nu_s^{1:N})$ the filtration generated by offspring counts, and use the shorthand $E[s;] \equiv E[s; F_{s-1}]$ for filtered expectations. Since time is labelled in reverse, the filtrations contain information about the future of the original system rather than the past.

A central quantity for analysing convergence of these coalescent processes is the pair merger rate, which can be interpreted as the probability that a randomly chosen
Simple conditions for convergence of SMC genealogies

pair of particles at time $t - 1$ have a common ancestor at time $t$. Conditional on $\mathcal{F}_t$, this probability is

$$c_N(t) := \frac{1}{(N)^2} \sum_{i=1}^{N} (\nu_i(t))_2. \quad (3.1)$$

In the $n$-coalescent the pair merger rate is equal to 1. Thus, as $N \to \infty$, the time scaling required to possibly obtain a Kingman coalescent limit is the generalized inverse

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

Following [25], we exclude the case where $P[\tau_N(t) = \infty] > 0$ for finite $t$. This is a very mild assumption which in reality is violated only by pathological (from this perspective) models, for example where minimum-variance resampling is used and the potentials are constant. Settings in which this assumption are violated are unlikely to be encountered in any practical application of SMC. For the classes of algorithms given as examples in Section 4 we provide tractable sufficient conditions under which the assumption holds, and these are explicitly verified in Appendix B.

In the $n$-coalescent, there are almost surely no mergers involving more than two lineages at a time. As shown in [20, Lemma 1, Case 3], an upper bound on the conditional probability that more than two lineages merge at time $t$ in the pre-limiting model is

$$D_N(t) := \frac{1}{N(N)^2} \sum_{i=1}^{N} (\nu^{(i)}_t + \frac{1}{N} \sum_{j \neq i} (\nu^{(j)}_t)^2).$$

This includes both the possibility of three or more lineages merging into one, and of two or more simultaneous pair mergers. Theorem 3.2, our main result, gives a simple sufficient condition controlling these merger rates to yield a Kingman limit for particle genealogies.

**Theorem 3.2.** Let $\nu^{(1:N)}_i$ denote the offspring numbers in an IPS satisfying the standing assumption and such that, for any $N$ sufficiently large, $P[\tau_N(t) = \infty] = 0$ for all finite $t$. Suppose that there exists a deterministic sequence $(b_N)_{N \geq 1}$ such that $\lim_{N \to \infty} b_N = 0$ and

$$\frac{1}{N(N)^3} \sum_{i=1}^{N} \mathbb{E}_t \left[ (\nu^{(i)}_t)_3 \right] \leq b_N \frac{1}{N(N)^2} \sum_{i=1}^{N} \mathbb{E}_t \left[ (\nu^{(i)}_t)_2 \right] \quad (3.2)$$

for all $N$, uniformly in $t \geq 1$. Then the rescaled genealogical process $(G^{(n,N)}_{\tau^{(n)}_t})_{t \geq 0}$ converges in the sense of finite-dimensional distributions to Kingman’s $n$-coalescent as $N \to \infty$.

**Proof outline.** Here we present the high level argument. In the interests of clarity, details are deferred to a sequence of lemmata which follow this result, and to Section 3.1.

Theorem 3.2 has the same conclusion as [20, Theorem 1] for which the conditions are

$$E[c_N(t)] \to 0, \quad (3.3)$$

$$E \left[ \sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r) \right] \to 0, \quad (3.4)$$

$$E \left[ \sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r)^2 \right] \to 0, \quad (3.5)$$

$$E[\tau_N(t) - \tau_N(s)] \leq C_{t,s} N; \quad (3.6)$$
as $N \to \infty$, for some strictly positive constant $C_{t,s}$ that does not depend on $N$. We show that the more tractable conditions of the present theorem are sufficient for the result.

The series of Lemmata 3.3–3.5 below show that the assumptions (3.3)–(3.5) follow from (3.2). Lemma 3.6 allows us to remove condition (3.6) by improving upon some arguments from the proof of [20, Theorem 1]; this argument is presented in detail in Section 3.1.

Our result improves on [20] by eliminating the restrictive condition (3.6), which is shown in Lemma 3.6 to be unnecessary. This allows our result to apply to some models not previously included; for example the neutral Moran model of population genetics violates (3.6) but is covered by Theorem 3.2. In neutral models the straightforward analogue of (3.2) is both necessary and sufficient [24, Theorem 5.4], suggesting that in general this condition is not significantly stronger than (3.3)–(3.5) combined.

**Lemma 3.3.** (3.4) ⇒ (3.5).

**Proof.** It is sufficient to show that $c_N(t)^2 \leq D_N(t)N/(N - 1)$. We have

$$c_N(t)^2 = \frac{1}{N(N - 1)(N)^2} \sum_{i=1}^{(N)} (\nu_t(i)_2) \left\{ \nu_t(i)/(N - 1) + \frac{1}{N - 1} \sum_{j=1}^{(N)} (\nu_t(j)_2) \right\}$$

$$= \frac{1}{N(N)^2} \sum_{i=1}^{N} (\nu_t(i)_2) \left\{ \nu_t(i)/(N - 1) + \frac{1}{N - 1} \sum_{j=1}^{N} (\nu_t(j)_2) \right\}$$

$$\leq \frac{1}{N(N)^2} \sum_{i=1}^{N} (\nu_t(i)_2) \left\{ \nu_t(i) + \frac{1}{N - 1} \sum_{j=1}^{N} (\nu_t(j)_2) \right\}$$

$$\leq \frac{1}{N(N)^2} \sum_{i=1}^{N} (\nu_t(i)_2) \left\{ \nu_t(i) + \frac{N/(N - 1)}{N} \sum_{j=1}^{N} (\nu_t(j)_2) \right\}$$

$$\leq \frac{N/(N - 1)}{N(N)^2} \sum_{i=1}^{N} (\nu_t(i)_2) \left\{ \nu_t(i) + \frac{1}{N - 1} \sum_{j=1}^{N} (\nu_t(j)_2) \right\} = \frac{N}{N - 1} D_N(t)$$

which concludes the proof. □

**Lemma 3.4.** (3.2) ⇒ (3.3).

**Proof.** Following the proof of [24, Lemma 5.5], we fix $\epsilon > 0$ and define the event $A_i = \{\nu_t(i) \leq N\epsilon\}$. Now

$$E_t[c_N(t)] = \frac{1}{(N)^2} \sum_{i=1}^{N} E_t \left\{ (\nu_t(i)_2) \right\} = \frac{1}{(N)^2} \sum_{i=1}^{N} \left\{ E_t \left\{ (\nu_t(i)_2 I_{A_i}) \right\} + E_t \left\{ (\nu_t(i)_2 I_{A_i^c}) \right\} \right\}$$

$$\leq \frac{\epsilon}{N - 1} \sum_{i=1}^{N} E_t \left\{ \nu_t(i) I_{A_i} \right\} + \sum_{i=1}^{N} E_t \left\{ I_{A_i^c} \right\}$$

$$\leq \{1 + O(N^{-1})\} \epsilon + \sum_{i=1}^{N} P \left\{ \nu_t(i) > N\epsilon \mid F_{t-1} \right\}.$$  

(3.7)
Simple conditions for convergence of SMC genealogies

For $N \geq 3/\epsilon$, Markov’s inequality yields

$$\sum_{i=1}^{N} P \left[ \nu_{t}^{(i)} > N\epsilon \mid \mathcal{F}_{t-1} \right] \leq \frac{1}{(N\epsilon)^3} \sum_{i=1}^{N} E_{t} \left[ (\nu_{t}^{(i)})^{3} \right] = \frac{1 + O(N^{-1})}{\epsilon^{3}(N)^{3}} \sum_{i=1}^{N} E_{t} \left[ (\nu_{t}^{(i)})^{3} \right]$$

$$\leq \{ 1 + O(N^{-1}) \} \frac{b_{N}}{\epsilon^{3}} E_{t}[c_{N}(t)]. \quad (3.8)$$

Substituting (3.8) into (3.7) and using $c_{N}(t) \leq 1$ results in

$$E_{t}[c_{N}(t)] \leq \{ 1 + O(N^{-1}) \} \left( \epsilon + \frac{b_{N}}{\epsilon^{3}} \right) \rightarrow \epsilon$$

since $b_{N} \rightarrow 0$. As $\epsilon > 0$ was arbitrary, we have $E[c_{N}(t)] = E[E_{t}[c_{N}(t)]] \rightarrow 0$ as $N \rightarrow \infty$. □

**Lemma 3.5.** (3.2) $\Rightarrow$ (3.4).

**Proof.** We decompose $D_{N}(t)$ as the sum of two terms and consider their filtered expectations. The first is

$$\frac{1}{N(N)^{2}} \sum_{i=1}^{N} E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} + (\nu_{t}^{(i)})_{2}^{3} \right] = \frac{1}{N(N)^{2}} \sum_{i=1}^{N} E_{t} \left[ 2(\nu_{t}^{(i)})_{2}^{2} + (\nu_{t}^{(i)})_{3}^{2} \right]$$

$$\leq \frac{2}{N} E_{t}[c_{N}(t)] + \frac{1}{(N)^{3}} \sum_{i=1}^{N} E_{t} \left[ (\nu_{t}^{(i)})_{3}^{3} \right]$$

$$\leq \left( \frac{2}{N} + b_{N} \right) E_{t}[c_{N}(t)]. \quad (3.9)$$

The second is

$$\frac{1}{N^{2}(N)^{2}} \sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(j)})_{2}^{2} + (\nu_{t}^{(j)})_{2}^{3} \right] = \frac{1}{N^{2}(N)^{2}} \sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(j)})_{2}^{2} + (\nu_{t}^{(j)})_{2}^{3} \right]$$

$$\leq \frac{1}{N^{2}(N)^{2}} \sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(j)})_{2}^{2} + (\nu_{t}^{(j)})_{2}^{3} \right] + \frac{E_{t}[c_{N}(t)]}{N}. \quad (3.10)$$

Now, with $A_{t}$ defined as in Lemma 3.4,

$$\sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} (\nu_{t}^{(j)})_{2} \right] = \sum_{j=1}^{N} \sum_{i \neq j} \left\{ E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} (\nu_{t}^{(j)})_{2} I_{A_{t}} \right] + E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} (\nu_{t}^{(j)})_{2} I_{\bar{A}_{t}} \right] \right\}$$

$$\leq N\epsilon \sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} (\nu_{t}^{(j)})_{2} I_{A_{t}} \right] + N\epsilon \sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} I_{\bar{A}_{t}} \right]$$

$$\leq N^{2}(N)^{2} \epsilon E_{t}[c_{N}(t)] + N^{4} \sum_{i=1}^{N} P \left[ \nu_{t}^{(i)} > N\epsilon \mid \mathcal{F}_{t-1} \right]. \quad (3.11)$$

Substituting (3.8) into (3.11) yields

$$\sum_{j=1}^{N} \sum_{i \neq j} E_{t} \left[ (\nu_{t}^{(i)})_{2}^{2} (\nu_{t}^{(j)})_{2} \right] \leq N^{4} \{ 1 + O(N^{-1}) \} \left( \epsilon + \frac{b_{N}}{\epsilon^{3}} \right) E_{t}[c_{N}(t)]. \quad (3.12)$$
and substituting (3.12) into (3.10) gives
\[
\frac{1}{N^2(N)_2} \sum_{j=1}^{N} \sum_{i \neq j} E_t \left[ (\nu_t^{(i)})_2 (\nu_t^{(j)})^2 \right] \leq \left\{ 1 + O(N^{-1}) \right\} \left( \epsilon + \frac{b_N}{c^3} \right) + \frac{1}{N} \sum_{i=1}^{\lfloor \epsilon t \rfloor} E_t[c_N(t)].
\] (3.13)

Combining (3.9) and (3.13), we have that
\[
E_t[D_N(t)] \leq \left\{ 1 + O(N^{-1}) \right\} \left( \epsilon + \frac{b_N}{c^3} \right) + \frac{3}{N} b_N \sum_{i=1}^{\lfloor \epsilon t \rfloor} E_t[c_N(t)].
\]

Finally, invoking [20, Lemma 2] twice gives
\[
E \left[ \sum_{r=\tau_N(t)+1}^{\tau_N(t)} D_N(r) \right] = E \left[ \sum_{r=\tau_N(s)+1}^{\tau_N(t)} E[c_N(r)] \right] \leq \left\{ 1 + O(N^{-1}) \right\} \left( \epsilon + \frac{b_N}{c^3} \right) + \frac{3}{N} b_N \sum_{i=1}^{\lfloor \epsilon t \rfloor} E_t[c_N(t)] \left( t - s + 1 \right) \rightarrow_{N \to \infty} \epsilon(t - s + 1),
\]
and recalling that \( \epsilon > 0 \) was arbitrary concludes the proof. \( \square \)

For Lemma 3.6, we introduce the quantity \( p_{\xi t}(t) \). For any fixed \( n \) and \( N \), \( p_{\xi t}(t) \) is the time \( t - 1 \) conditional transition probability of the genealogical process from \( \xi \) to \( \eta \), where \( \xi \) and \( \eta \) are partitions of \( \{1, \ldots, n\} \). The transition probability is non-zero only when \( \eta = \xi \) or \( \eta \) can be obtained from \( \xi \) by merging some blocks. Let \( b_j \) \((j = 1, \ldots, \lfloor \xi \rfloor)\) denote the number of blocks in \( \xi \) that merged to form block \( j \) of \( \eta \).

**Lemma 3.6.**
\[
p_{\xi t}(t) \geq 1 - B_{\lfloor \xi \rfloor} \left[ 1 + O(N^{-1}) \right] D_N(t) - \left( \frac{\lfloor \xi \rfloor}{2} \right) \left[ 1 + O(N^{-1}) \right] c_N(t),
\] (3.14)
for a constant \( B_{\lfloor \xi \rfloor} > 0 \) increasing in \( \lfloor \xi \rfloor \) that does not depend on \( N \).

**Proof.** Let \( \kappa_i = \{|j : b_j = i\}| \) denote the multiplicity of mergers of size \( i \), with the slight abuse of terminology that \( \kappa_1 \) counts non-merger events. In particular, we have that \( \kappa_1 + 2\kappa_2 + \cdots + \lfloor \xi \rfloor \kappa_{\lfloor \xi \rfloor} = \lfloor \xi \rfloor \). Now
\[
p_{\xi t}(t) = 1 - \frac{1}{(N)_{\lfloor \xi \rfloor}} \sum_{k=1}^{\lfloor \xi \rfloor - 1} \sum_{b_1 \geq \cdots \geq b_k = |\xi|} \frac{|\xi|!}{\prod_{j=1}^{\lfloor \xi \rfloor} (j)!^{\kappa_j}} \sum_{b_1 + \cdots + b_k = |\xi|} \sum_{\substack{i_1, \ldots, i_k \geq 1 \\ i_1 \neq \cdots \neq i_k}} \prod_{j=1}^{\lfloor \xi \rfloor} (\nu_{t}^{(i_j)})_{b_j},
\]
because the right hand side subtracts the probabilities of all possible merger events. See [13, Eq. (11)] for the combinatorial factor, which gives the number of partitions of a sequence of length \( \lfloor \xi \rfloor \) having \( \kappa_j \) subsequences of length \( j \) for each \( j \). The omitted \( k = |\xi| \) summand would correspond to the probability of an identity transition. The non-increasing ordering of \((b_1, \ldots, b_k)\) in the sum is arbitrary, but without loss of generality: choosing any ordering of the same set of merger sizes would give the same result.
Simple conditions for convergence of SMC genealogies

Firstly, we separate out the $k = |\xi| - 1$ term, which covers isolated binary mergers, and note that in that case the only possible $b$-vector is $(2, 1, \ldots, 1)$, for which

$$\frac{|\xi|!}{\prod_{j=1}^{|\xi|} (j!)^{s_j \kappa_j}} = \frac{|\xi|!}{2^{|\xi| - 2}!} = \binom{|\xi|}{2}$$

and a multinomial expansion argument yields

$$\sum_{i_1 \neq \ldots \neq i_{|\xi|-1} = 1}^N (\nu_t^{(i_1)})_2 \nu_t^{(i_2)} \ldots \nu_t^{(i_{|\xi|-1})} \leq N^{|\xi|-2} \sum_{i=1}^N (\nu_t^{(i)})_2.$$ 

Thus

$$p_{\xi}(t) \geq 1 - \left( \frac{|\xi|}{2} \right) \frac{1 + O(N^{-1})}{(N)_2} \sum_{i=1}^N (\nu_t^{(i)})_2$$

$$- \frac{1}{(N)!|\xi|} \sum_{k=1}^{|\xi| - 2} \sum_{b_1 \leq \ldots \leq b_k = 1}^{b_1 + \ldots + b_k = |\xi|} \frac{|\xi|!}{\prod_{j=1}^{|\xi|} (j!)^{s_j \kappa_j}} \sum_{i_1 \neq \ldots \neq i_k = 1 \text{ all distinct}}^N (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_k)})_{b_k}.$$

For the other summands, we have

$$\frac{|\xi|!}{\prod_{j=1}^{|\xi|} (j!)^{s_j \kappa_j}} \leq |\xi|!$$

and, similarly to [20, Lemma 1, Case 3],

$$\sum_{i_1 \neq \ldots \neq i_k = 1 \text{ all distinct}}^N (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_k)})_{b_k} \leq \sum_{i=1}^N (\nu_t^{(i)})_2 \left( N^{|\xi|-2} - \sum_{j_1 \neq \ldots \neq j_{|\xi|-2} = 1 \text{ all distinct and } \neq i}^N \nu_t^{(j_1)} \ldots \nu_t^{(j_{|\xi|-2})} \right)$$

$$\leq \sum_{i=1}^N (\nu_t^{(i)})_2 \left( N^{|\xi|-2} - (N - \nu_t^{(i)}) |\xi|-2 + \left( \frac{|\xi|-2}{2} \right) \sum_{j \neq i} (\nu_t^{(j)})_2 \left( \sum_{k \neq i} (\nu_t^{(k)}) \right) \right)$$

$$\leq \sum_{i=1}^N (\nu_t^{(i)})_2 \left( (|\xi|-2)\nu_t^{(i)} N^{|\xi|-3} + \left( \frac{|\xi|-2}{2} \right) \sum_{j \neq i} (\nu_t^{(j)})_2 N^{|\xi|-4} \right),$$

where the last step uses $(N - x)_b \geq N^b - bx N^{b-1}$ for $x \leq N, b \geq 0$. Overall

$$p_{\xi}(t) \geq 1 - \left( \frac{|\xi|}{2} \right) \frac{1 + O(N^{-1})}{(N)_2} \sum_{i=1}^N (\nu_t^{(i)})_2$$

$$- \frac{1}{(N)!|\xi|} \sum_{k=1}^{|\xi| - 2} \sum_{b_1 \leq \ldots \leq b_k = 1}^{b_1 + \ldots + b_k = |\xi|} |\xi|! \sum_{i=1}^N (\nu_t^{(i)})_2 \left( (|\xi|-2)\nu_t^{(i)} N^{|\xi|-3} + \left( \frac{|\xi|-2}{2} \right) \sum_{j \neq i} (\nu_t^{(j)})_2 N^{|\xi|-4} \right).$$

The summand in the third term depends neither on $k$ nor on $b_1, \ldots, b_k$, and the number of terms in those sums is bounded above by $\gamma_{|\xi|-2} (|\xi|-2)$, where $\gamma_n$ is the number of integer
partitions of \( n \). By [16, Section 2], \( \gamma_n < Ke^{2\sqrt{2\pi}}/n \) for a constant \( K > 0 \) independent of \( n \). Thus, for \( |\xi| > 2 \),

\[
p_{\xi\xi}(t) \geq 1 - \frac{|\xi|}{2} 1 + O(N^{-1}) (N - 1) \sum_{i=1}^{N} (\nu_{i}^{(i)})_{2} \\
- K e^{2\sqrt{2(|\xi| - 2)|\xi|}} \left( \frac{|\xi| - 1}{2} \right)^{N|\xi| - 3} \sum_{i=1}^{N} (\nu_{i}^{(i)})_{2} \left\{ \nu_{i}^{(i)} + \frac{1}{N} \sum_{j \neq i} (\nu_{j}^{(i)})_{2} \right\}
\]

\[
= 1 - \left( \frac{|\xi|}{2} \right) \{ 1 + O(N^{-1}) \} c_{N}(t) - B_{|\xi|} \{ 1 + O(N^{-1}) \} D_{N}(t),
\]

where \( B_{|\xi|} > 0 \) depends on \( |\xi| \) but not on \( N \). When \( |\xi| \leq 2 \), there are no higher order interactions and the result is immediate. \( \square \)

Using (3.14) in place of the lower bound used in the proof of [20, Theorem 1] facilitates a modification of that proof such that (3.6) is not needed. Details of this argument are in the following subsection.

### 3.1 Proof of Theorem 3.2 without Assumption (3.6)

The proof of [20, Theorem 1] proceeds in three parts. The first is a vanishing upper bound on finite-dimensional distributions of the genealogical process when the path of the process involves either multiple simultaneous mergers or any merger involving more than two particles. The second is showing that the finite-dimensional distributions of the \( n \)-coalescent upper bound those of the genealogical process in the limit \( N \to \infty \) when the path of the genealogy consists of only isolated binary mergers. The final piece is a similar lower bound, which establishes convergence of the finite-dimensional distributions. Only the third part makes use of assumption (3.6). Hence, it suffices to show that Lemma 3.6 can be used to obtain the same lower bound without making use of (3.6).

**Proof.** Let \( \chi_{\alpha}^{d} \) be the conditional transition probability of a transition from state \( \eta_{d-1} \) to state \( \eta_{d} \) at times \( \tau_{N}(t_{d-1}) \) and \( \tau_{N}(t_{d}) \) respectively, conditional on the offspring counts between those times \( \nu_{1}^{(1:N)}, \ldots, \nu_{\tau_{N}(t_{d})}^{(1:N)} \). This transition can happen via any valid path of merger events, but we restrict to paths involving binary mergers only, and denote by \( \chi_{d} \) the conditional transition probability subject to this restriction. Considering the Proof of Theorem 1 in [20], the derivation of an upper bound on \( \chi_{d} \) holds without modification in our context; while the first step in the derivation of a lower bound which can be found on page 14 of that work involves the application of its Lemma 1 to bound \( \chi_{d} \) from below. Instead, we apply Lemma 3.6 to obtain

\[
\chi_{d} \geq \sum_{s_{1} < \ldots < s_{n} = \tau_{N}(t_{d-1})+1}^{\tau_{N}(t_{d})} (\hat{Q}^{\alpha})_{\eta_{d-1}\eta_{d}} \left( \prod_{r=1}^{\alpha} c_{N}(s_{r}) - \left( \frac{n-2}{2} \right) \{ 1 + O(N^{-1}) \} D_{N}(s_{r}) \right)
\times \prod_{r = \tau_{N}(t_{d-1})+1}^{\tau_{N}(t_{d})} \left[ 1 - B_{n} \{ 1 + O(N^{-1}) \} D_{N}(r) \right]
\left( |\eta_{d-1}| - \left| \left\{ i : s_{i} < r \right\} \right| \right) \{ 1 + O(N^{-1}) \} c_{N}(r).
\]

Here \( \hat{Q} \) is the matrix obtained from the generator \( Q \) of Kingman’s \( n \)-coalescent (see Definition 3.1) by setting the diagonal entries to 0. The number of pair-merger steps
required to transition from $\eta_{d-1}$ to $\eta_d$ is $\alpha = |\eta_{d-1}| - |\eta_d|$. The sequences $s_1, \ldots, s_\alpha$ denote the times at which these pair-mergers happen. At the remaining times $r$ the partition is unchanged, and the bound in Lemma 3.6 has been applied to the one-step transition probabilities corresponding to these identity transitions. The constant $B_n$ is that appearing in Lemma 3.6, where we replace $|\eta_d|$ by its upper bound $n$. A sum over an index vector of length zero should be interpreted as the identity operator here and in the following.

The rest of the proof proceeds as in [20], albeit from this modified initial lower bound. A multinomial expansion of the product on the second line yields

$$
\chi_d \geq \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{(\lambda, \mu) \in B_2([\alpha+\beta])} \left\{ 1 + O(N^{-1}) \right\}^\beta 
\times \sum_{s_1 < \ldots < s_{n+\beta} = \tau_N(t_{d-1})+1} \prod_{r \in \lambda} c_N(s_r) - \left( \frac{n-2}{2} \right) \{ 1 + O(N^{-1}) \} D_N(s_r) 
\times \prod_{r \in \mu} \left\{ - \left( \frac{|\eta_d-1| - |\{ i \in \lambda : i < r \}|}{2} \right) c_N(s_r) - B_n D_N(s_r) \right\}
$$

where $\Pi_i([\nu])$ denotes the set of partitions of $\{1, \ldots, n\}$ into exactly $i$ blocks. Expanding the product over $\lambda$ gives

$$
\chi_d \geq \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{(\lambda, \mu, \pi) \in B_3([\alpha+\beta])} \left( \frac{n-2}{2} \right) \{ 1 + O(N^{-1}) \}^\beta \| \pi \|
\times \sum_{s_1 < \ldots < s_{n+\beta} = \tau_N(t_{d-1})+1} \prod_{r \in \lambda} c_N(s_r) \prod_{r \in \pi} D_N(s_r) 
\times \prod_{r \in \mu} \left\{ - \left( \frac{|\eta_d-1| - |\{ i \in \lambda \cup \pi : i < r \}|}{2} \right) c_N(s_r) - B_n D_N(s_r) \right\}
$$

and expanding the product over $\mu$ results in

$$
\chi_d \geq \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{(\lambda, \mu, \pi, \sigma) \in B_4([\alpha+\beta])} B_n^{\| \sigma \|} \left( \frac{n-2}{2} \right) \| \pi | \| \sigma \|
\times \{ 1 + O(N^{-1}) \}^\beta \| \pi \|
\times \prod_{r \in \lambda} \left\{ - \left( \frac{|\eta_d-1| - |\{ i \in \lambda \cup \pi : i < r \}|}{2} \right) c_N(s_r) - B_n D_N(s_r) \right\}
\times \prod_{s_1 < \ldots < s_{n+\beta} = \tau_N(t_{d-1})+1} \prod_{r \in \lambda \cup \mu} c_N(s_r) \prod_{r \in \pi \cup \sigma} D_N(s_r).
$$

Via a further multinomial expansion, the lower bound for the $k$-step transition probability
can be written as

\[
\lim_{N \to \infty} \mathbb{E} \left[ \prod_{d=1}^{k} \chi_d \right] \geq \lim_{N \to \infty} \mathbb{E} \left[ \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \sum_{(\lambda_1, \mu_1, \tau_1, \sigma_1) \in \mathcal{E}_1([\alpha_1 + \beta_1])} \cdots \sum_{(\lambda_k, \mu_k, \tau_k, \sigma_k) \in \mathcal{E}_k([\alpha_k + \beta_k])} \right.
\]

\[
B_N^{\sum_{d=1}^{k} \left| \sigma_d \right|} \left( n - 2 \right) \frac{\sum_{d=1}^{k} \left| \sigma_d \right|}{2} - \left( 1 + O(N^{-1}) \right) 2 \sum_{d=1}^{k} \left| \sigma_d \right| \left| \sigma_d \right| \left( \{1 \in \lambda_d \cup \pi_d : i < r\} \right) \right]
\]

\[
\times \prod_{d=1}^{k} \mathbb{I}_{\{\tau_N(t_d) - \tau_N(t_{d-1}) \geq \alpha_d + \beta_d\}} \left\{ \prod_{r \in \lambda_d \cup \mu_d} c_N(s_r^{(d)}) \right\} \prod_{r \in \pi_d \cup \sigma_d} D_N(s_r^{(d)}) \right].
\]

An argument completely analogous to that in [20, Appendix] shows that passing the expectation and the limit through the infinite sums is justified, whereupon the contribution of terms with \( \sum_{d=1}^{k} (|\sigma_d| + |\sigma_d|) > 0 \) vanishes. To see why, follow the argument used to show that the contribution of multiple merger trajectories vanishes in the corresponding upper bound in [20]. That leaves

\[
\lim_{N \to \infty} \mathbb{E} \left[ \prod_{d=1}^{k} \chi_d \right] \geq \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \sum_{(\lambda_1, \mu_1, \tau_1, \sigma_1) \in \mathcal{E}_1([\alpha_1 + \beta_1])} \cdots \sum_{(\lambda_k, \mu_k, \tau_k, \sigma_k) \in \mathcal{E}_k([\alpha_k + \beta_k])} \left( \prod_{d=1}^{k} \mathbb{I}_{\{\tau_N(t_d) - \tau_N(t_{d-1}) \geq \alpha_d + \beta_d\}} \right)
\]

\[
\times \lim_{N \to \infty} \mathbb{E} \left[ \prod_{r \in \mu} \mathbb{I}_{\{\tau_N(t_r) \geq \alpha_d + \beta_d\}} \left\{ \prod_{r \in \lambda_d \cup \mu_d} c_N(s_r^{(d)}) \right\} \right].
\]

Recall [20, Eq. (11)]:

\[
\sum_{(\lambda, \mu) \in \Pi_2(\alpha + \beta)} \mathbb{I}_{|\mu| = \beta} (\tilde{Q}^\alpha)_{\eta_d-1, \eta_d} \prod_{r \in \mu} \left( \left| \eta_d - 1 \right| - \left| \{i \in \lambda \cup \pi : i < r\} \right| \right) = (Q^{\alpha + \beta})_{\eta_d-1, \eta_d}.
\]

Applying this \( k \) times in (3.15) yields

\[
\lim_{N \to \infty} \mathbb{E} \left[ \prod_{d=1}^{k} \chi_d \right] \geq \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \left\{ \prod_{d=1}^{k} (Q^{\alpha_d + \beta_d})_{\eta_d-1, \eta_d} \right\}
\]

\[
\times \lim_{N \to \infty} \mathbb{E} \left[ \prod_{d=1}^{k} \mathbb{I}_{\{\tau_N(t_d) - \tau_N(t_{d-1}) \geq \alpha_d + \beta_d\}} \right. \sum_{s_1^{(1)} < \cdots < s_k^{(k-1)} = \tau_N(t_0) + 1} \prod_{d=1}^{k} \prod_{r \in \lambda_d \cup \mu_d} c_N(s_r^{(d)}) \right].
\]
We now apply equations (14) and (15), respectively, of [20], to those terms with a negative (|β| odd) and positive (|β| even) sign, respectively, and obtain

\[
\lim_{N \to \infty} \mathbb{E}\left[ \prod_{d=1}^{k} \chi_d \right] \geq \sum_{\beta_1 = 0}^{\infty} \cdots \sum_{\beta_k = 0}^{\infty} \left\{ \prod_{d=1}^{k} (Q^{\alpha_d + \beta_d})_{t_d - t_{d-1}} \frac{(t_d - t_{d-1})^{\alpha_d + \beta_d}}{\alpha_d + \beta_d} \right\}
\]

\[
\times \lim_{N \to \infty} \mathbb{E}\left[ \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d) - \tau_N(t_{d-1}) \geq \alpha_d + \beta_d\}} \right].
\]

An invocation of [20, Eq. (16)] concludes the proof. \[\square\]

4 Illustrative applications

4.1 Resampling with stochastic rounding

Definition 4.1. Let \(X = (X_1, \ldots, X_N)\) be a \(\mathbb{R}^N\)-valued random variable. Then \(Y = (Y_1, \ldots, Y_N) \in \mathbb{N}^N\) is a stochastic rounding of \(X\) if each element \(Y_i\) takes values

\[
Y_i \mid X = \begin{cases} |X_i| & \text{with probability } 1 - X_i + |X_i| \\ |X_i| + 1 & \text{with probability } X_i - |X_i|. \end{cases}
\]

By construction, \(\mathbb{E}[Y_i \mid X] = X_i\) for each \(i\). Taking \(X\) to be \(N\) times the vector of particle weights, we can therefore use stochastic rounding for the RESAMPLE procedure in Algorithm 1, under the further constraint that \(Y_1 + \cdots + Y_N = N\). Several ways to enforce this constraint have been proposed, including systematic resampling [3, 31], residual resampling with stratified or systematic residuals [31], the branching system of [5], and the Srinivasan sampling process resampling introduced in [14].

Corollary 4.2. Consider an SMC algorithm using any stochastic rounding as its resampling scheme, such that the standing assumption is satisfied. Assume that there exists a constant \(a \in [1, \infty)\) such that for all \(x, x', t, 1 / a \leq g(x, x') \leq a. \) (4.1)

Assume that \(\mathbb{P}[\tau_N(t) = \infty] = 0\) for all finite \(t\). Let \((G^{(n,N)}_{\tau_N(t)})_{t \geq 0}\) denote the genealogy of a random sample of \(n\) terminal particles from the output of the algorithm when the total number of particles used is \(N\). Then, for any fixed \(n\), the time-scaled genealogy \((G^{(n,N)}_{\tau_N(t)})_{t \geq 0}\) converges to Kingman’s \(n\)-coalescent as \(N \to \infty\), in the sense of finite-dimensional distributions.

Condition (4.1) is strong, but is widespread in the literature. By contrast, the assumption that the time-scale does not explode is a technicality, which holds under an easier-to-verify condition when the transition densities are bounded above and below (see Lemma B.1). We conjecture that in general bounded transition densities are not necessary in the case of stochastic rounding.

Remark 4.3. In a similar vein to [20, Remark 3], if we consider the weight vector as fixed, the time scale induced by stochastic rounding is slower than that induced by multinomial resampling. Details are given in Appendix A.

Remark 4.4. Since every stochastic rounding has the same marginal distributions and the first moment of (3.1) depends only on marginal family size distributions, the expected coalescence rate is the same whichever stochastic rounding is used for resampling. Thus the time-scale on which the \(n\)-coalescent is recovered is equal in expectation for every such scheme.
We also have the lower bounds which follow from (4.1) along with the form of the weights in Algorithm 1:

\[ \mathcal{H}_t := \sigma(X_{i-1}^{(1:N)}, X_i^{(1:N)}, w_{i-1}^{(1:N)}, w_i^{(1:N)}) \]  

(4.2)

such that \( \nu_t^{(1:N)} \) is conditionally independent of the filtration \( F_{t-1} \) given \( \mathcal{H}_t \). The full D-separation argument is presented in Appendix C.

Defining the family sizes \( \nu_t^{(i)} = \{ \{ j : a_j^{(i)} = i \} \} \) as functions of \( a_t^{(1:N)} \), we have the almost sure constraint \( \nu_t^{(i)} \in \{ [N w_t^{(i)}], [N w_t^{(i)}] + 1 \} \). Denote \( p_0^{(i)} := \mathbb{P}[\nu_t^{(i)} = [N w_t^{(i)}] | \mathcal{H}_t] \) and \( p_1^{(i)} := \mathbb{P}[\nu_t^{(i)} = [N w_t^{(i)}] + 1 | \mathcal{H}_t] = 1 - p_0^{(i)} \).

We obtain the following upper bounds, using the almost sure bounds \( w_t^{(i)} \leq a^2/N \) which follow from (4.1) along with the form of the weights in Algorithm 1:

\[
\mathbb{E}[\nu_t^{(i)}] | \mathcal{H}_t] = p_0^{(i)}([N w_t^{(i)}] + 1) + p_0^{(i)}([N w_t^{(i)}] + 1)
\]

\[
= [N w_t^{(i)}]([N w_t^{(i)}] + 1) + p_0^{(i)}([N w_t^{(i)}] + 1)
\]

\[
= [N w_t^{(i)}]([N w_t^{(i)}] + 1)(p_0^{(i)} + p_1^{(i)}) - 2p_0^{(i)} + p_1^{(i)}
\]

\[
\leq a^2(1 - 1)(a^2 - 0 + 1)I_{\{[N w_t^{(i)}] \geq 2\}}
\]

\[
\leq (a^2 + 1)^3I_{\{[N w_t^{(i)}] \geq 2\}}.
\]

We also have the lower bounds

\[
\mathbb{E}[\nu_t^{(i)}] | \mathcal{H}_t] = p_0^{(i)}([N w_t^{(i)}]) + p_1^{(i)}([N w_t^{(i)}] + 1)
\]

\[
= [N w_t^{(i)}]([N w_t^{(i)}] - 1) + p_0^{(i)}([N w_t^{(i)}] + 1)
\]

\[
= [N w_t^{(i)}]([N w_t^{(i)}])(p_0^{(i)} + p_1^{(i)}) - p_0^{(i)} + p_1^{(i)}
\]

\[
\geq 2(2 - 1 + 0)I_{\{[N w_t^{(i)}] \geq 2\}} = 2I_{\{[N w_t^{(i)}] \geq 2\}}.
\]

Applying the tower property and conditional independence,

\[
\frac{1}{(N)^2} \sum_{i=1}^{N} \mathbb{E}[\nu_t^{(i)}] = \frac{1}{(N)^2} \mathbb{E}_t \left[ \sum_{i=1}^{N} \mathbb{E}_t \left[ (\nu_t^{(i)})_2 | \mathcal{H}_t, F_{t-1} \right] \right]
\]

\[
= \frac{1}{(N)^2} \mathbb{E}_t \left[ \sum_{i=1}^{N} \mathbb{E}_t \left[ (\nu_t^{(i)})_2 | \mathcal{H}_t \right] \right] \geq \frac{1}{(N)^2} 2 \mathbb{E}_t \left[ \{ i: [N w_t^{(i)}] \geq 2 \} \right]
\]

and similarly

\[
\frac{1}{(N)^3} \sum_{i=1}^{N} \mathbb{E}_t[\nu_t^{(i)}] \leq \frac{1}{(N)^3} (a^2 + 1)^3 \mathbb{E}_t \left[ \{ i: [N w_t^{(i)}] \geq 2 \} \right]
\]

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

where

\[
b_N := \frac{1}{N - 2} \frac{(a^2 + 1)^3}{2} \rightarrow 0 \quad N \rightarrow \infty
\]

is independent of \( F_\infty \), satisfying (3.2). The result follows by applying Theorem 3.2. \( \square \)
4.2 Conditional sequential Monte Carlo updates

Conditional SMC differs from Algorithm 1 in that one predetermined trajectory is conditioned to survive all of the resampling steps. We refer to this sequence of positions as the immortal trajectory, and the immortal particle is the particle in a particular generation that is part of the immortal trajectory. Conditional SMC was introduced as a component of the particle Gibbs algorithm [1] but has found somewhat wider application in fields as diverse as smoothing [17, 29] and optimization [12, Chapter 6].

In particle Gibbs, the immortal trajectory \( x_{0:T}^* \) at each time step is sampled from the output of the previous conditional SMC run. It is therefore important that with high probability at least two distinct lineages survive each run so that the immortal trajectory can be updated. A single run of conditional SMC with multinomial resampling is presented in Algorithm 2.

![Algorithm 2](https://www.imstat.org/ejp)

Although it is also possible to construct a conditional SMC algorithm using a low-variance resampling scheme, here we treat only the case of multinomial resampling. We believe that the result can be extended to other resampling schemes by similar arguments to those of Corollary 4.2.

**Corollary 4.5.** Consider a conditional SMC algorithm using multinomial resampling, such that the standing assumption is satisfied. Assume there exist constants \( \varepsilon \in (0, 1], a \in [1, \infty) \) and probability density \( h \) such that for all \( x, x', t \),

\[
\frac{1}{a} \leq g_t(x, x') \leq a, \quad \varepsilon h(x') \leq q_t(x, x') \leq \frac{1}{\varepsilon} h(x'). \tag{4.3}
\]

Let \( (G_t^{(n,N)})_{t \geq 0} \) denote the genealogy of a random sample of \( n \) terminal particles from the output of the algorithm when the total number of particles used is \( N \). Then, for any fixed \( n \), the time-scaled genealogy \( (G_t^{(n,N)})_{t \geq 0} \) converges to Kingman’s \( n \)-coalescent as \( N \to \infty \), in the sense of finite-dimensional distributions.

Condition (4.3) is widespread in the SMC literature, where it is known as the *strong mixing condition* [6, Section 3.5.2]; it further strengthens the condition (4.1) used in the case of stochastic rounding. It is to be expected that some additional control over transition densities is required in the case of multinomial resampling compared to stochastic rounding.

**Proof.** Define the conditioning \( \sigma \)-algebra \( \mathcal{H}_t \) as in (4.2). We assume without loss of generality that the immortal particle takes index 1 in each generation. This significantly
simplifies the notation, but the same argument holds if the immortal indices are taken to be $a_{(0:T)}$ rather than $(1, \ldots, 1)$.

The parental indices are conditionally independent, as in standard SMC with multinomial resampling, but we have to treat $i = 1$ as a special case. We have the following conditional law on parental indices

$$P\left[a_i^{(i)} = a_i \mid \mathcal{H}_t\right] \propto \begin{cases} \mathbb{I}_{\{a_i = 1\}} w_i^{(a_i)} q_t^{-1}(X_{t_i}^{(a_i)}, X_{t-1}^{(i)}) & i = 1 \\ w_i^{(a_i)} q_t^{-1}(X_{t_i}^{(a_i)}, X_{t-1}^{(i)}) & i = 2, \ldots, N. \end{cases}$$

The joint conditional law is therefore

$$P\left[a_1^{(1:N)} = a_1; N \mid \mathcal{H}_t\right] \propto \prod_{i=2}^N w_i^{(a_i)} q_t^{-1}(X_{t_i}^{(a_i)}, X_{t-1}^{(i)}).$$

First we make the following observation, which follows from a balls-in-bins coupling. Assume (4.3). Then for any function $f : \{1, \ldots, N\}^N \rightarrow \mathbb{R}$ such that (for a fixed $i$) $f(a_1^{(1:N)}) \geq f(a_i^{(1:N)})$ whenever $|\{j : a_i^{(j)} = i\}| \geq |\{j : a_t^{(j)} = i\}|$,

$$E[f(a_1^{(1:N)})] \leq E[f(a_t^{(1:N)}) \mid \mathcal{H}_t] \leq E[f(A_{ch}^{(1:N)})] \tag{4.4}$$

where the elements of $A_1^{(1:N)}, A_2^{(1:N)}$ are all mutually independent and independent of $\mathcal{F}_\infty$, and distributed according to

$$A_1^{(j)} \sim \begin{cases} \delta_1 & j = 1 \\ \text{Categorical} \left(\frac{\varepsilon}{a}, \frac{\varepsilon}{a} + (N-1)(a/\varepsilon)\right) & j \neq 1 \end{cases},$$

$$A_2^{(j)} \sim \begin{cases} \delta_1 & j = 1 \\ \text{Categorical} \left(\frac{a}{\varepsilon}, \frac{a}{\varepsilon} + (N-1)(\varepsilon/a)\right) & j \neq 1 \end{cases}.$$

where the vector of probabilities is given up to a constant in the argument of Categorical distributions. We use these random vectors to construct bounds that are independent of $\mathcal{F}_\infty$. Also define the corresponding offspring counts $V_1^{(i)} = |\{j : A_1^{(j)} = i\}|$, $V_2^{(i)} = |\{j : A_2^{(j)} = i\}|$, for $i = 1, \ldots, N$, which have marginal distributions

$$V_1^{(i)} \overset{d}{=} \mathbb{I}_{\{i=1\}} + \text{Binomial} \left(N - 1, \frac{\varepsilon}{a} + \left(N - 1\right)(a/\varepsilon)\right),$$

$$V_2^{(i)} \overset{d}{=} \mathbb{I}_{\{i=1\}} + \text{Binomial} \left(N - 1, \frac{a}{\varepsilon} + \left(N - 1\right)(\varepsilon/a)\right),$$

Now consider the function $f_i(a_1^{(1:N)}) := (V_1^{(i)})_2$. We can apply (4.4) to obtain the lower bound

$$\frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}[V_1^{(i)}]_2 \mid \mathcal{H}_t] \geq \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}[V_1^{(i)}]_2 = \frac{1}{(N)_2} \left[ \mathbb{E}[V_1^{(i)}]_2 + \sum_{i=2}^N \mathbb{E}[V_1^{(i)}]_2 \right]$$

$$= \frac{1}{(N)_2} \left[ \frac{(N - 1)_2(\varepsilon/a)^2}{(\varepsilon/a) + (N - 1)(a/\varepsilon)} + \frac{2(N - 1)(\varepsilon/a)}{\varepsilon/a + (N - 1)(a/\varepsilon)} + \frac{N(N - 1)_2(\varepsilon/a)^2}{(\varepsilon/a) + (N - 1)(a/\varepsilon)} \right]$$

$$= \frac{1}{(N)_2} \left[ \frac{2(N - 1)(\varepsilon/a)}{\varepsilon/a + (N - 1)(a/\varepsilon)} + \sum_{i=1}^N \frac{(N - 1)_2(\varepsilon/a)^2}{(\varepsilon/a) + (N - 1)(a/\varepsilon)} \right]$$

EJP 26 (2021), paper 1. https://www.imstat.org/ejp
using the moments of the Binomial distribution (see [26] for example) along with the identity \((X + 1)^2 \equiv 2(X)_1 + (X)_2\). This is further bounded by

\[
\frac{1}{N} \sum_{i=1}^{N} E[(\nu_t^{(i)}_{t+1} \mid H_t)] \geq \frac{1}{N} \left\{ \frac{2(N - 1)(\varepsilon/a)}{N(a/\varepsilon)} + \frac{(N)_3(\varepsilon/a)^2}{N^2(a/\varepsilon)^2} \right\} \\
= \frac{1}{N^2} \left\{ \frac{2\varepsilon^2}{a^2} + \frac{(N - 2)\varepsilon^4}{a^4} \right\}. \tag{4.5}
\]

Similarly, we derive an upper bound on \(f_t(a_t^{1:N}) := (\nu_t^{(i)})_{1} \), this time using the identity \((X + 1)^3 \equiv 3(X)_2 + (X)_3^2\);

\[
\frac{1}{N} \sum_{i=1}^{N} E[(\nu_t^{(i)}_{t+1} \mid H_t)] \\
\leq \frac{1}{N} \left[ E[(V_{2}^{(1)}_{t+1})] + \sum_{i=1}^{N} E[(V_{2}^{(i)}_{t+1})] \right] \\
\leq \frac{1}{N} \left[ \frac{3(N - 1)^2(a/\varepsilon)^2}{(a/\varepsilon) + (N - 1)(\varepsilon/a))^3} + \sum_{i=1}^{N} \frac{(N - 1)(\varepsilon/a)^3}{((a/\varepsilon) + (N - 1)(\varepsilon/a))^3} \right] \\
\leq \frac{1}{N} \left[ \frac{3(N - 1)^2(a/\varepsilon)^2}{N^2(a/\varepsilon)^2} + \frac{(N)_4(a/\varepsilon)^3}{N^3(\varepsilon/a)^3} \right] \\
= \frac{1}{N} \left\{ \frac{3a^4}{\varepsilon^4} + \frac{(N - 3)a^6}{\varepsilon^6} \right\} \\
= \frac{1}{N^3} \left\{ \frac{3a^4}{\varepsilon^4} + \frac{(N - 3)a^6}{\varepsilon^6} \right\}
\]

We apply the tower property and conditional independence as in Corollary 4.2, upper bounding the ratio by

\[
\frac{1}{N^3} \sum_{i=1}^{N} E_t[(\nu_t^{(i)}_{t+1})] \leq \frac{N^2}{N^3} \left\{ \frac{3a^4}{\varepsilon^4} + \frac{(N - 3)a^6}{\varepsilon^6} \right\} \leq \frac{1}{N} \left\{ \frac{3a^4}{2\varepsilon^4} + \frac{(N - 3)a^6}{2\varepsilon^6} \right\} \\
\leq \frac{1}{N} \left\{ \frac{3a^4}{2\varepsilon^4} + \frac{(N - 3)a^6}{2\varepsilon^6} \right\} \leq \frac{1}{N} \left\{ \frac{3a^4}{2\varepsilon^4} + \frac{a^{10}}{2\varepsilon^{10}} \right\} =: b_N \xrightarrow{N \to \infty} 0.
\]

Thus (3.2) is satisfied. It remains to show that, for \(N\) sufficiently large, \(P[\tau_N(t) = \infty] = 0\) for all finite \(t\), a technicality which is proved in Lemma B.2 in Appendix B. Applying Theorem 3.2 gives the result.

\[\square\]

A Comparison of time scales for multinomial resampling and stochastic rounding

Fix \(N \geq 1\). Suppose we are given a fixed vector of weights \(w_t^{(1:N)}\). Let \(E^M\) and \(E^{SR}\) denote expectations with respect to the laws of resampling steps using multinomial resampling and stochastic rounding respectively. Let \(\nu_t^{(1:N)}\) be the resulting offspring counts.

**Proposition A.1.** For any \(i \in \{1, \ldots, N\}\), \(E^M[w_t^{1:N} \mid c_N(t)] \geq E^{SR}[w_t^{1:N}].\)
Simple conditions for convergence of SMC genealogies

Proof. It is sufficient to show that \( E^M[(v_t^{(i)})_2 \mid w_t^{(1:N)}] \geq \) \( E^{SR}[(v_t^{(i)})_2 \mid w_t^{(1:N)}] \) for each \( i \in \{1, \ldots, N\} \). Using properties of the Multinomial distribution [26], we have

\[
E^M[(v_t^{(i)})_2 \mid w_t^{(1:N)}] = N(N-1)(w_t^{(i)})^2.
\]

Directly from Definition 4.1, we calculate the corresponding quantity in the case of stochastic rounding to be

\[
E^{SR}[(v_t^{(i)})_2 \mid w_t^{(1:N)}] = \lfloor w_t^{(i)} \rfloor (\lfloor w_t^{(i)} \rfloor - 1)(1 - w_t^{(i)}) + \lfloor w_t^{(i)} \rfloor \lfloor w_t^{(i)} \rfloor - w_t^{(i)} - \lfloor w_t^{(i)} \rfloor - 1.
\]

We define the difference \( \Delta_i := E^M[(v_t^{(i)})_2 \mid w_t^{(1:N)}] - E^{SR}[(v_t^{(i)})_2 \mid w_t^{(i)}] \), and show that \( \Delta_i \geq 0 \) for all \( 0 \leq w_t^{(i)} \leq 1 \). Partition the interval \([0, 1]\) into the half-open intervals \([k/N, (k+1)/N)\) for each \( k \in \{0, \ldots, N-1\} \), plus the singleton \( (1) \).

If \( w_t^{(i)} = 1 \), it follows easily that \( E^M[(v_t^{(i)})_2 \mid w_t^{(1:N)}] = E^{SR}[(v_t^{(i)})_2 \mid w_t^{(1:N)}] = N(N-1) \).

For the other cases, suppose \( k/N \leq w_t^{(i)} < (k+1)/N \) for some \( k \in \{0, \ldots, N-1\} \). Then \( \lfloor w_t^{(i)} \rfloor = k \), and

\[
\Delta_i = (\lfloor w_t^{(i)} \rfloor - k)^2 - \lfloor w_t^{(i)} \rfloor N(N-1)\]
\[
= N(N-1)\left(\frac{w_t^{(i)} - \frac{k}{N}}{N-1}\right)^2 - \frac{k^2}{N-1} + \frac{k(k+1)}{N(N-1)}\]
\[
= N(N-1)\left(\frac{w_t^{(i)} - \frac{k}{N}}{N-1}\right)^2 + k\left(1 - \frac{k}{N-1}\right)\]
\[
\geq k\left(1 - \frac{k}{N-1}\right) > 0.
\]

For each \( N \geq 2 \), any \( w_t^{(i)} \in [0, 1] \) falls into one of the above cases, so for any fixed vector \( w_t^{(1:N)} \) of weights, we have that \( \Delta_i \geq 0 \) for all \( i \). For \( N = 1 \) the result is immediate. This concludes the proof. \( \square \)

B Proof of finite time-scale condition for corollaries

Lemma B.1. Consider an SMC algorithm using any stochastic rounding as its resampling scheme. Suppose that \( \varepsilon \leq q_t(x, x') \leq \varepsilon^{-1} \) uniformly for some \( \varepsilon \in (0, 1] \), and that there exist \( \zeta > 0 \) and \( \delta \in (0, 1) \) such that \( P[\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N \mid F_{t-1}] \geq \zeta \) for infinitely many \( t \). Then, for all \( N > 1 \), \( P[\tau_N(t) = \infty] = 0 \) for all finite \( t \).

Proof. Let \( \mathcal{H}_t \) be defined as in (4.2). The first step is to show that whenever \( \max_i w_t^{(i)} \geq (1 + \delta)/N, P[\varepsilon_N(t) > 2N^2/|\mathcal{H}_t| = P[\varepsilon_N(t) \neq 0/|\mathcal{H}_t|] \) is bounded below uniformly in \( t \). For this purpose we need consider only weight vectors such that \( w_t^{(i)} \in (0, 2/N) \) for all \( i \); otherwise \( P[\varepsilon_N(t) \neq 0/|\mathcal{H}_t|] = 1 \) by the definition of stochastic rounding.

Denote \( S_{N-1}^\delta = \{w_t^{(1:N)} \in S_{N-1} : \forall i, 0 < w_t^{(i)} < 2/N; \max_i w_t^{(i)} \geq (1 + \delta)/N\} \) for any \( \delta \in (0, 1) \), where \( S_d \) denotes the \( d \)-dimensional probability simplex. Fix arbitrary \( w_t^{(1:N)} \in S_{N-1}^\delta \). Set \( i^* = \arg \max_i w_t^{(i)} \) and denote \( I = \{i \in \{1, \ldots, N\} : w_t^{(i)} > 1/N\} \).

Since all weights are in \((0, 2/N)\), \( i^* \in I, v_t^{(i)} \in \{1, 2\} \) and \( i \notin I, v_t^{(i)} \in \{0, 1\}; \) and
since the offspring counts must sum to \( N \), we can write

\[
P[c_N(t) \leq 2/N^2|H_t] = \mathbb{P}[\nu_t(i) = 1 \forall i \in \{1, \ldots, N\}|H_t]
\]

\[
= \prod_{i \in I} \mathbb{P}[\nu_t(i) = 1|\nu_t(j) = 1 \forall j \in I: j < i; H_t]
\]

\[
= \mathbb{P}[\nu_t(i') = 1|H_t] \prod_{i \in I} \mathbb{P}[\nu_t(i) = 1|\nu_t(j') = 1; \nu_t(i) = 1 \forall j \in I: j < i; H_t]
\]

\[
\leq \mathbb{P}[\nu_t(i') = 1|H_t].
\]

The final inequality holds with equality when \( |I| = 1 \), i.e. the only weight larger than \( 1/N \) is \( w_t(i') \). Thus \( \mathbb{P}[c_N(t) > 2/N^2|H_t] \) is minimised on \( S_{N-1}^k \) when only one weight is larger than \( 1/N \), in which case the values of the other weights do not affect this probability.

Define \( w_N = \{(1, \ldots, 1) + \delta^* e_i - \delta^* e_j\}/N \) for fixed \( i^* \neq j^* \) and \( \delta^* \in (0, 1) \), where \( e_i \) denotes the \( i \)th canonical basis vector in \( \mathbb{R}^N \). As in the proof of Corollary 4.2, define \( p^{(i)}_0 = \mathbb{P}[^{i}\nu_t(i) = |Nw_t(i')|, H_t \text{ and } p^{(i)}_1 = \mathbb{P}[\nu_t(i) = |Nw_t(i')| + 1 | H_t \}. \] From (B.1) we have

\[
P[c_N(t) > 2/N^2 | H_t, w_t(1:N) = w_N] = 1 - \mathbb{P}[\nu_t(i') = 1 | H_t, w_t(1:N) = w_N] = p^{(i*)}_1,
\]
evaluated on \( w_N \). We will need a lower bound on \( p^{(i*)}_1 \) when \( w_t(1:N) = w_N \). We first derive expressions for \( p^{(i)}_0 \) and \( p^{(i)}_1 \) up to a constant, then use \( p^{(i)}_0 + p^{(i)}_1 = 1 \) to get a normalised bound. We have

\[
p^{(i)}_0 = C(1 - Nw_t(i') + |Nw_t(i')|) \times \sum_{\alpha_1, \ldots, \alpha_N: \{j : \alpha_j = i\}|=|Nw_t(i')|} N_{q_{t-1}(X^{a(k)}_t, X^{(k)}_{t-1})}
\]

\[
p^{(i)}_1 = C(Nw_t(i') - |Nw_t(i')|) \times \sum_{\alpha_1, \ldots, \alpha_N: \{j : \alpha_j = i\}|=|Nw_t(i')|+1} N_{q_{t-1}(X^{a(k)}_t, X^{(k)}_{t-1})}
\]

Applying the bounds on \( q_t \), we have

\[
C(1 - Nw_t(i') + |Nw_t(i')|)\varepsilon_N \leq p^{(i)}_0 \leq C(1 - Nw_t(i') + |Nw_t(i')|)\varepsilon^{-N},
\]

\[
C(Nw_t(i') - |Nw_t(i')|)\varepsilon_N \leq p^{(i)}_1 \leq C(Nw_t(i') - |Nw_t(i')|)\varepsilon^{-N}
\]

from which we construct the normalised bound

\[
p^{(i)}_1 \geq \frac{(Nw_t(i') - |Nw_t(i')|)\varepsilon_N}{(Nw_t(i') - |Nw_t(i')|)\varepsilon^{-N} + (1 - Nw_t(i') + |Nw_t(i')|)\varepsilon^{-N}} = (Nw_t(i') - |Nw_t(i')|)\varepsilon^{2N}.
\]

When \( w_t(1:N) = w_N \), we have \( w_t(i') = (1 + \delta')/N \), so \( p^{(i*)}_1 \geq \delta'\varepsilon^{2N} \), which is increasing in \( \delta' \). We conclude that \( \mathbb{P}[c_N(t) > 2/N^2|H_t, \max w_t(i) \geq (1 + \delta')/N] \geq \min_{\delta' > 0} \delta'\varepsilon^{2N} = \delta\varepsilon^{2N} \).

A slight modification of this argument yields \( \mathbb{P}[c_N(t) > 2/N^2|H_t, \min w_t(i) \leq (1 - \delta')/N] \geq \delta\varepsilon^{2N} \). Whenever \( \max w_t(i) - \min w_t(i) \geq 2\delta/N \), either \( \max w_t(i) \geq (1 + \delta)/N \) or
Simple conditions for convergence of SMC genealogies

\[ \min_i w_i^{(t)} \leq (1 - \delta)/N, \] so we have \( \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{H}_t, \max_i w_i^{(t)} - \min_i w_i^{(t)} \geq 2\delta/N] \geq \delta e^{2N}. \] Thus

\[ \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{H}_t] \geq \delta e^{2N} \mathbb{P}[\max_i w_i^{(t)} - \min_i w_i^{(t)} \geq 2\delta/N]. \]

Using the D-separation established in Appendix C combined with the tower property, we have

\[ \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{F}_{t-1}] = \mathbb{E}_t \left[ \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{H}_t, \mathcal{F}_{t-1}] \right] = \mathbb{E}_t \left[ \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{H}_t] \right] \geq \delta e^{2N} \mathbb{P}[\max_i w_i^{(t)} - \min_i w_i^{(t)} \geq 2\delta/N | \mathcal{F}_{t-1}], \]

which is bounded below by \( \zeta \delta e^{2N} \) for infinitely many \( t \). Hence,

\[ \sum_{t=0}^{\infty} \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{F}_{t-1}] = \infty. \]

By a filtered version of the second Borel–Cantelli lemma (see for example [10, Theorem 4.3.4]), this implies that \( \epsilon_N(t) > 2/N^2 \) for infinitely many \( t \), almost surely. This ensures, for all \( t < \infty \), that \( \mathbb{P} \left[ \exists s < \infty : \sum_{i=1}^{s} \epsilon_N(r) \geq t \right] = 1 \), which by definition of \( \tau_N(t) \) is equivalent to \( \mathbb{P}[\tau_N(t) = \infty] = 0 \).

**Lemma B.2.** Consider a conditional SMC algorithm using multinomial resampling, satisfying the standing assumption and (4.3). Then, for all \( N > 2 \), \( \mathbb{P}[\tau_N(t) = \infty] = 0 \) for all finite \( t \).

**Proof.** Since \( \epsilon_N(t) \in [0, 1] \) almost surely and has strictly positive expectation, for any fixed \( N \) the distribution of \( \epsilon_N(t) \) with given expectation that maximises \( \mathbb{P}[\epsilon_N(t) = 0 | \mathcal{F}_{t-1}] \) is two atoms, at 0 and 1 respectively. To ensure the correct expectation, the atom at 1 should have mass \( \mathbb{P}[\epsilon_N(t) = 1 | \mathcal{F}_{t-1}] = \mathbb{E}_t[\epsilon_N(t)] \), which is bounded below by (4.5). If \( \epsilon_N(t) > 0 \) then \( \epsilon_N(t) \geq 2/(N) > 2/N^2 \). Hence, in general \( \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{F}_{t-1}] \geq \mathbb{E}_t[\epsilon_N(t)] \). Applying (4.5), we have for any finite \( N \),

\[ \sum_{t=0}^{\infty} \mathbb{P}[\epsilon_N(t) > 2/N^2 | \mathcal{F}_{t-1}] \geq \sum_{t=0}^{\infty} \mathbb{E}_t[\epsilon_N(t)] \geq \sum_{t=0}^{\infty} \frac{1}{N^2} \left\{ \frac{2\delta^2}{a^2} + \frac{(N-2)\delta^4}{a^4} \right\} = \infty. \]

By an argument analogous to the conclusion of Lemma B.1, \( \mathbb{P}[\tau_N(t) = \infty] = 0 \) for all \( t < \infty \).

**C D-separation argument to establish conditional independence of \( \nu_t^{(1:N)} \) and \( \mathcal{F}_{t-1} \) given \( \mathcal{H}_t \)**

Figure 1 shows part of the conditional dependence graph implied by Algorithm 1. Our aim is to find a \( \sigma \)-algebra \( \mathcal{H}_t \) at each time \( t \) that separates the ancestral process (encoded by \( \nu_t^{(1:N)} \)) from the filtration \( \mathcal{F}_{t-1} \). That is, \( \nu_t^{(1:N)} \) is conditionally independent of \( \mathcal{F}_{t-1} \) given \( \mathcal{H}_t \). By a D-separation argument (see [30]), the nodes highlighted in grey suffice as the generator of \( \mathcal{H}_t \). That is, for each \( t \), we take

\[ \mathcal{H}_t = \sigma(\chi_t^{(1:N)}, \nu_t^{(1:N)}, w_{t-1}^{(1:N)}, w_t^{(1:N)}). \]

Notice that \( \nu_t^{(1:N)} \) can be expressed as a non-injective function of \( \nu_t^{(1:N)} \), and as such carries less information.
Simple conditions for convergence of SMC genealogies

Figure 1: Part of the conditional dependence graph implied by Algorithm 1. The direction of time is from left to right. The reverse-time filtration is indicated by the dashed areas. The nodes highlighted in grey generate the separatrix $H_t$ between $a_{t-1}^{1:N}$ and $F_{t-1}$.

References

[1] C. Andrieu, A. Doucet, and R. Holenstein, Particle Markov chain Monte Carlo methods, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72 (2010), no. 3, 269–342. MR-2758115
[2] M. Briers, A. Doucet, and S. Maskell, Smoothing algorithms for state space models, Annals of the Institute of Statistical Mathematics 62 (2010), no. 1, 61–89. MR-2577439
[3] J. Carpenter, P. Clifford, and P. Fearnhead, Improved particle filter for nonlinear problems, IEE Proceedings – Radar, Sonar and Navigation 146 (1999), no. 1, 2–7.
[4] N. Chopin and O. Papaspiliopoulos, An Introduction to Sequential Monte Carlo, Springer, 2020.
[5] D. Crisan and T. Lyons, Nonlinear filtering and measure-valued processes, Probability Theory and Related Fields 109 (1997), no. 2, 217–244. MR-1477650
[6] P. Del Moral, Feynman-Kac formulae: Genealogical and Interacting Particle Systems with Applications, Springer, 2004. MR-2044973
[7] P. Del Moral, A. Doucet, and A. Jasra, Sequential Monte Carlo samplers, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (2006), no. 3, 411–436. MR-2278333
[8] A. Doucet and A. M. Johansen, A tutorial on particle filtering and smoothing: Fifteen years later, Handbook of Nonlinear Filtering, OUP, 2011, pp. 656–704. MR-2884612
[9] R. Durrett, Probability Models for DNA Sequence Evolution, Springer Science & Business Media, 2008. MR-2439767
[10] R. Durrett, Probability: Theory and Examples, Cambridge University Press, 2019. MR-3930614
[11] P. Fearnhead and H. R. Künsch, Particle filters and data assimilation, Annual Review of Statistics and Its Application 5 (2018), no. 1, 421–449. MR-3774754
[12] A. Finke, On extended state-space constructions for Monte Carlo methods, Ph.D. thesis, University of Warwick, 2015.
[13] Y.-X. Fu, Exact coalescent for the Wright-Fisher model, Theoretical Population Biology 69 (2006), 385–394.
Simple conditions for convergence of SMC genealogies

[14] M. Gerber, N. Chopin, and N. Whiteley, Negative association, ordering and convergence of resampling methods, Annals of Statistics 47 (2019), no. 4, 2236–2260. MR-3953450

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proceedings F (Radar and Signal Processing) 140 (1993), no. 2, 107–113.

[16] G. H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, Proceedings of the London Mathematical Society s2-17 (1918), no. 1, 75–115. MR-1575586

[17] P. E. Jacob, F. Lindsten, and T. B. Schön, Smoothing with couplings of conditional particle filters, Journal of the American Statistical Association 115 (2020), no. 530, 721–729. MR-4107675

[18] P. E. Jacob, L. M. Murray, and S. Rubenthaler, Path storage in the particle filter, Statistics and Computing 25 (2015), no. 2, 487–496. MR-3306720

[19] J. F. C. Kingman, The coalescent, Stochastic Processes and Their Applications 13 (1982), no. 3, 235–248. MR-0671034

[20] J. Koskela, P. A. Jenkins, A. M. Johansen, and D. Spanò, Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo, Mathematics e-print arXiv:1804.01811v6, 2018. MR-4065174

[21] J. Koskela, P. A. Jenkins, A. M. Johansen, and D. Spanò, Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo, Annals of Statistics 48 (2020), no. 1, 560–583. MR-4065174

[22] A. Lee and N. Whiteley, Variance estimation in the particle filter, Biometrika 105 (2018), no. 3, 609–625. MR-3842888

[23] T. M. Liggett, Interacting Particle Systems, Springer Berlin/Heidelberg, 2005. MR-2108619

[24] M. Möhle and S. Sagitov, Coalescent patterns in exchangeable diploid population models, Journal of Mathematical Biology 47 (2003), 337–352. MR-2024501

[25] M. Möhle, Robustness results for the coalescent, Journal of Applied Probability 35 (1998), no. 2, 438–447. MR-1641829

[26] J. E. Mosimann, On the compound multinomial distribution, the multivariate \( \beta \)-distribution, and correlations among proportions, Biometrika 49 (1962), no. 1/2, 65–82. MR-0143299

[27] C. A. Naesseth, F. Lindsten, and T. B. Schön, Elements of sequential Monte Carlo, Foundations and Trends® in Machine Learning 12 (2019), no. 3, 307–392.

[28] J. Olsson and R. Douc, Numerically stable online estimation of variance in particle filters, Bernoulli 25 (2019), no. 2, 1504–1535. MR-3920380

[29] A. Shestopaloff and A. Doucet, Replica conditional sequential Monte Carlo, in: Proceedings of the 36th International Conference on Machine Learning (Long Beach, California, USA) (Kamalika Chaudhuri and Ruslan Salakhutdinov, eds.), Proceedings of Machine Learning Research, vol. 97, PMLR, 09–15 Jun 2019, pp. 5749–5757.

[30] T. Verma and J. Pearl, Causal networks: Semantics and expressiveness, in: Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence (Minneapolis, MN, Mountain View, CA), 1988, pp. 352–359. MR-1166827

[31] D. Whitley, A genetic algorithm tutorial, Statistics and Computing 4 (1994), no. 2, 65–85.

Acknowledgments. This work was supported by The Engineering and Physical Sciences Council, The Medical Research Council, The Alan Turing Institute and the Alan Turing Institute — Lloyd’s Register Foundation Programme on Data-centric Engineering, under grant numbers EP/L016710/1, EP/N510129/1, EP/R034710/1, EP/R044732/1 and EP/T004134/1.