The Annals of Statistics
2020, Vol. 48, No. 1, 560–583
https://doi.org/10.1214/19-AOS1823

ASYMPTOTIC GENEALOGIES OF INTERACTING PARTICLE SYSTEMS WITH AN APPLICATION TO SEQUENTIAL MONTE CARLO

BY JERE KOSKELA[1,*], PAUL A. JENKINS[2], ADAM M. JOHANSEN[1,**] AND DARIO SPANÒ[1,†]

[1]*Department of Statistics, University of Warwick,* [*]*j.koskela@warwick.ac.uk;* [**]*a.m.johansen@warwick.ac.uk;* [†]*d.spano@warwick.ac.uk*

[2]*Department of Statistics and Department of Computer Science, University of Warwick, p.jenkins@warwick.ac.uk*

We study weighted particle systems in which new generations are resampled from current particles with probabilities proportional to their weights. This covers a broad class of sequential Monte Carlo (SMC) methods, widely-used in applied statistics and cognate disciplines. We consider the genealogical tree embedded into such particle systems, and identify conditions, as well as an appropriate time-scaling, under which they converge to the Kingman n-coalescent in the infinite system size limit in the sense of finite-dimensional distributions. Thus, the tractable n-coalescent can be used to predict the shape and size of SMC genealogies, as we illustrate by characterising the limiting mean and variance of the tree height. SMC genealogies are known to be connected to algorithm performance, so that our results are likely to have applications in the design of new methods as well. Our conditions for convergence are strong, but we show by simulation that they do not appear to be necessary.

1. Introduction. Interacting particle systems (IPSs) are a broad class of stochastic models for phenomena in disciplines including physics, engineering, biology and finance. Prominent examples are particle filters, particle methods and sequential Monte Carlo (SMC), which feature prominently in numerical approximation schemes for nonlinear filtering, as well as mean field approximation of Feynman–Kac flows. For additional background, we direct readers to Del Moral (2004), and Doucet and Johansen (2011).

Central to these methods are discrete-time, evolving weighted particle systems. Correlations between particles arise out of *resampling*: a stochastic selection mechanism in which particles with high weight are typically replicated while those with low weight vanish, giving rise to an embedded genealogy. The contribution of this paper is to identify conditions under which these genealogies converge to the Kingman n-coalescent (Kingman (1982)) in the infinite system size limit under an appropriate rescaling of time, as well as to describe ways in which information about the limiting genealogy can be used to characterize particle systems in applications.

Consider a sequence of measurable spaces $(E_t, \mathcal{E}_t)_{t\in\mathbb{N}}$, each associated to a Markov transition kernel $K_{t+1} : E_t \times \mathcal{E}_{t+1} \mapsto (0, \infty)$, and a nonnegative potential $g_{t+1} : E_t \times E_{t+1} \mapsto (0, \infty)$. These correspond to state spaces, transition kernels and the importance weight function of our IPS, respectively.

Let $\zeta_t^{(N)} := \{(w_t^{(i)}, X_t^{(i)})\}_{i=1}^N$ be a weighted N-particle system at time $t \in \mathbb{N}$, where each $X_t^{(i)} \in E_t$, and the weights $w_t^{(i)}$ are nonnegative and satisfy $\sum_{i=1}^N w_t^{(i)} = 1$. Let S be a resampling operator which acts on $\zeta_t^{(N)}$ by assigning to each particle a random number of offspring. The total number of offspring is fixed at N, and the mean number of offspring

Received April 2018; revised January 2019.
MSC2010 subject classifications. Primary 60E15; secondary 60G99, 62E20.
Key words and phrases. Coalescent, finite-dimensional distributions, genealogy, interacting particle system, sequential Monte Carlo.

Algorithm 1 Simulation of an IPS

Require: Particle number N, Markov kernels K_t, potentials g_t, initial proposal distribution μ.

1: **for** $i \in \{1, \dots, N\}$ **do**
2: Sample $X_0^{(i)} \sim \mu$.
3: **for** $i \in \{1, \dots, N\}$ **do**
4: Set $w_0^{(i)} \leftarrow \frac{g_0(X_0^{(i)})}{g_0(X_0^{(1)})+\cdots+g_0(X_0^{(N)})}$.
5: **for** $t \in \{0, \dots, T-1\}$ **do**
6: Sample $(a_t^{(1)}, \dots, a_t^{(N)}) \sim \text{Resample}(w_t^{(1)}, \dots, w_t^{(N)})$.
7: **for** $i \in \{1, \dots, N\}$ **do**
8: Sample $X_{t+1}^{(i)} \sim K_{t+1}(X_t^{(a_t^{(i)})}, \cdot)$.
9: **for** $i \in \{1, \dots, N\}$ **do**
10: Set $w_{t+1}^{(i)} \leftarrow \frac{g_{t+1}(X_t^{(a_t^{(i)})}, X_{t+1}^{(i)})}{g_{t+1}(X_t^{(a_t^{(1)})}, X_{t+1}^{(1)})+\cdots+g_{t+1}(X_t^{(a_t^{(N)})}, X_{t+1}^{(N)})}$.

of particle $i \in \{1, \dots, N\}$ is $N w_t^{(i)}$. All offspring are assigned an equal weight $1/N$. More concretely,

$$S\zeta_t^{(N)} = \{(N^{-1}, X_t^{(a_t^{(i)})})\}_{i=1}^N,$$

where $a_t^{(i)} = j$ if j in $\zeta_t^{(N)}$ is the parent of i in $S\zeta_t^{(N)}$. Particles with low weight are randomly removed by having no offspring, while particles with high weight tend to have many offspring. Algorithm 2 gives an example.

The step from time t to time $t+1$ is completed by propagating each particle in $S\zeta_t^{(N)}$ independently through the transition kernel K_{t+1} to obtain particle locations $X_{t+1}^{(i)} \sim K_{t+1}(X_t^{(a_t^{(i)})}, \cdot)$. Finally, each particle $i \in \{1, \dots, N\}$ is assigned a weight proportional to the potential g_{t+1} evaluated at the locations of the particle and its parent so that the full update is

$$\zeta_t^{(N)} \overset{S}{\mapsto} \{N^{-1}, X_t^{(a_t^{(i)})}\}_{i=1}^N \overset{K_{t+1}}{\mapsto} \left\{ \frac{g_{t+1}(X_t^{(a_t^{(i)})}, X_{t+1}^{(i)})}{\sum_{j=1}^N g_{t+1}(X_t^{(a_t^{(j)})}, X_{t+1}^{(j)})}, X_{t+1}^{(i)} \right\}_{i=1}^N,$$

where $g_0(X_{-1}, X_0) \equiv g_0(X_0)$. A specification is given in Algorithm 1.

There are many options for the resampling step `Resample` in line 6 of Algorithm 1. The simplest is *multinomial resampling*, given in Algorithm 2. It is analytically tractable, but suboptimal in terms of Monte Carlo variance. Popular alternatives include *residual*, *stratified* and *systematic* resampling, which yield algorithms with lower variance (Douc, Cappé and Moulines (2005)). We prove our main theorem under multinomial resampling, and show by simulation that its conclusions also hold for these three alternatives, at least in our simple

Algorithm 2 Multinomial resampling

Require: Normalised particle weights $\{w_t^{(i)}\}_{i=1}^N$.

1: **for** $i \in \{1, \dots, N\}$ **do**
2: Sample $a_t^{(i)} \sim \text{Categorical}(w_t^{(1)}, \dots, w_t^{(N)})$.
3: **return** $(a_t^{(1)}, \dots, a_t^{(N)})$.

example. Other schemes which have been studied theoretically, but seem less widely used in applications, include the tree-based branching approximation of Crisan and Lyons (1997) and the McKean interpretation approach by Del Moral (2004), Section 2.5.3.

An example system of this form is particle filters: a class of IPSs for approximating integrals. For an overview see, for example, Doucet and Johansen (2011), and references therein. They are suited to settings in which expectations are computed with respect to the law of a latent, discrete-time Markov process conditioned on a sequence of observations, for example, state space models. A state space model is a Markov chain $(X_t, Y_t)_{t\geq 0}$ on $\boldsymbol{E}\times\boldsymbol{F}\subseteq\mathbb{R}^d\times\mathbb{R}^p$ with transition density $p(x,x')$ and emission density $\psi(x,y)$, both with respect to dominating measures abusively denoted dx' and dy, respectively. The spaces $(\boldsymbol{E},\boldsymbol{F})$ could be replaced with more general Polish spaces, but we focus on (subsets of) $\mathbb{R}^d\times\mathbb{R}^p$ for concreteness. Typically, X_t is an unobserved state at time t, and Y_t is a noisy or incomplete observation of the state.

Let $\boldsymbol{y}:=(y_0,\dots,y_T)$ be a vector of noisy, conditionally independent observations from the emission density $\psi(x,y)$ given an unobserved state trajectory $\boldsymbol{x}:=(x_0,\dots,x_T)$. Functionals of the *smoothing distribution*,

$$P(\boldsymbol{x}|\boldsymbol{y})\,d\boldsymbol{x}:=\mathbb{P}(\boldsymbol{X}\in d\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y})\propto\prod_{t=0}^{T}p(x_{t-1},x_t)\psi(x_t,y_t)\,dx_{0:T},$$

where we abuse notation and denote the law of X_0 by $p(x_{-1},x_0)\,dx_0$, are typically intractable, and often approximated by particle filters. The simplest is the algorithm introduced by Gordon, Salmond and Smith (1993), usually known as the bootstrap particle filter, although the term "(interacting) particle filter" seems first to have been used by Del Moral (1996). It fits into our IPS framework by taking $K_{t+1}(x_t,dx_{t+1})=p(x_t,x_{t+1})\,dx_{t+1}$ and $g_{t+1}(x_t,x_{t+1})=\psi(x_{t+1},y_{t+1})$ for a fixed observation sequence in Algorithm 1. Other particle filters can be considered by introducing a proposal density $q_{t+1}(x_t,x_{t+1})\,dx_{t+1}$ with $\operatorname{supp}(p(x,\cdot))\subseteq\operatorname{supp}(q_t(x,\cdot))$ for every x and t, and setting

$$K_{t+1}(x_t,dx_{t+1})=q_{t+1}(x_t,x_{t+1})\,dx_{t+1},$$

$$g_{t+1}(x_t,x_{t+1})=\frac{p(x_t,x_{t+1})\psi(x_{t+1},y_{t+1})}{q_{t+1}(x_t,x_{t+1})}.$$

Our IPS framework also includes more general SMC algorithms, such as SMC samplers (Del Moral, Doucet and Jasra (2006)) among others; see, for example, Del Moral (2004), Chapter 12.

There is a genealogy embedded in Algorithm 1. Consider $\zeta_t^{(N)}$ at a fixed time t. Tracing the ancestor indices $(a_t^{(1)},\dots,a_t^{(N)})$ backwards in time results in a coalescing forest of lineages. The forest forms a tree once the most recent common ancestor (MRCA) of all particles is reached, provided that happens before reaching the initial time 0. Our main result (Theorem 1) shows that, under certain conditions and an appropriate time-rescaling, functionals of this genealogy depending upon finite numbers of leaves converge to corresponding functionals of the Kingman n-coalescent (Kingman (1982)) as $N\to\infty$, in the sense of finite dimensional distributions. Hence, the tractable Kingman n-coalescent can be used to approximate functionals of SMC genealogies off-line, *before* the algorithm has been run. In particular, we show that the expected number of time steps from the leaves to the MRCA scales linearly in N for any finite number of leaves. This compares to an $N\log N$ upper bound of Jacob, Murray and Rubenthaler (2015) for the genealogy of all N leaves. We also provide scaling expressions for the variance of the number of time steps to the MRCA (see Corollary 2). Our result applies to the marginal genealogical tree structure, marginalised over locations and weights of particles.

The conditions for convergence to the Kingman n-coalescent can be satisfied by SMC algorithms under strong but standard conditions (18) and (19), used to control the fluctuations of family sizes. We expect that they can be relaxed, and simulations in Section 3 also suggest they are not necessary.

Mergers of lineages into common ancestors result in *path degeneracy* in SMC algorithms. Quantities of interest depending on times $t < T$ will be estimated using fewer than N particles, resulting in high variance. Path degeneracy can be reduced by resampling less frequently, but this increases variance across time. The extreme case of no resampling results in no path degeneracy, but yields estimators whose variances increase exponentially in T in most cases: a phenomenon known as *weight degeneracy* (Kong, Liu and Wong (1994)). Intermediate strategies balance these two difficulties (Liu and Chen (1995), Section 4.2). Our limiting genealogies are the only method of which we are aware that can provide a priori estimates of path degeneracy, and thus represent an important tool in minimising both degeneracies simultaneously. This is particularly important in the conditional SMC update in the particle Gibbs algorithm (Andrieu, Doucet and Holenstein (2010)), where retaining multiple trajectories across a fixed time window is essential.

Genealogies of SMC algorithms have found numerous applications, for example, Cérou, Del Moral and Guyader (2011), Chan and Lai (2013), Del Moral and Miclo (2001, 2007), Del Moral and Garnier (2005), Del Moral et al. (2010), Lee and Whiteley (2018), Olsson and Douc (2019); see also Del Moral (2004), Section 3.4 and Del Moral (2013), Chapter 15. Our explicit description of the limiting genealogical process is new for SMC, and has the potential to build on any of these works, as well as generate new ones in the IPS and SMC fields. We also demonstrate that non-Markovian prelimiting genealogies lie in the domain of attraction of Kingman's n-coalescent, in contrast with earlier results which assume Markovianity (Möhle (1998), Theorem 1). This extension is significant since reverse-time genealogies of Markov IPSs are not Markov processes in general.

The rest of this paper is structured as follows. In Section 2, we state and prove the convergence of genealogies of IPSs to the Kingman n-coalescent. Section 3 presents a simulation showing that the scalings predicted by our convergence result hold for an example outside our technical assumptions, and Section 4 concludes with a discussion. An Appendix contains some technical calculations. We conclude this section by summarising notation.

Let $(x)_b := x(x-1)\cdots(x-b+1)$ be the falling factorial. We adopt the conventions $\sum_{\varnothing} = 0$, $\prod_{\varnothing} = 1$. When a sum is written over a vector of indices, and that vector is of length 0, it should be interpreted as the identity operator; where this convention might hold, we emphasize it by writing $\sum^{\star}$. The statement $f(N) = O(g(N))$ (resp., $o(g(N))$) means $\limsup_{N\to\infty} |f(N)/g(N)| < \infty$, (resp., $= 0$), and thus corresponds to the usual Landau big-O (resp., little-o) notation. For an integer $n \in \mathbb{N}$, we define $[n] := \{1, \ldots, n\}$ with $[0] := \varnothing$, and for a finite set A, we let $\Pi_n(A)$ denote the set of partitions of A into at most n nonempty blocks, with $\Pi_n(\varnothing) := (\varnothing, \ldots, \varnothing)$. For a partition ξ, $|\xi|$ denotes the number of blocks in ξ, and $\boldsymbol{x} := (x_1, \ldots, x_N)$, where the length of the vector will be clear from context. For a vector $\boldsymbol{x}$, $|\boldsymbol{x}|$ denotes the L^1-norm.

2. The convergence theorem. It will be convenient to express our IPS in reverse time by denoting the initial time in Algorithm 1 by T and the terminal time by 0, and to describe the genealogy in terms of a partition-valued family of processes $(G_t^{(n,N)})_{t=0}^{T}$ indexed by $n \le N$, where n denotes the number of observed leaves (time 0 particles) in a system with N particles. The process $(G_t^{(n,N)})_{t=0}^{T}$ is defined in terms of the underlying IPS via its initial condition $G_0^{(n,N)} = \{\{1\}, \ldots, \{n\}\}$, and its dynamics, which are driven by the requirement that $i \neq j \in [n]$ belong to the same block in $G_t^{(n,N)}$ if leaves i and j have a common ancestor at time t. Boundary problems will be avoided by ensuring that $T \to \infty$ in our rescaled system as $N \to \infty$.

REMARK 1. Our genealogical process $(G_t^{(n,N)})_{t=0}^T$ evolves on a space which tracks the ancestral relationships of the observed particles but not their states. The process is a projection of the time reversal of the historical process of Del Moral and Miclo (2001), in which particle locations have been marginalised over. A consequence of this is that $(G_t^{(n,N)})_{t=0}^T$ is not Markovian in general. Indeed, Markovianity fails even for the forward-time evolution of lineages after removing locations. Genealogical processes typically do track locations in the SMC literature, whereas our marginal formulation is standard in genetics (Möhle (1998)). Our $(G_t^{(n,N)})_{t=0}^T$ also coincides with the random ancestral forest of Del Moral et al. (2010), who showed that the combinatorial structure of the reverse-time genealogy decouples from the particle locations in the case of *neutral* models ($g_t(x, y) \equiv 1$). Our contribution is to prove the same decoupling for suitably rescaled nonneutral models in the $N \to \infty$ limit, and to identify the limiting process.

Genealogical processes are a powerful tool in population genetics, where the genealogical tree is viewed as missing data to be imputed, or an object of inference in its own right. A common large population limit is given by the Kingman n-coalescent (Kingman (1982)), which is also a partition-valued stochastic process evolving in reverse time. The initial condition is the singleton partition $\{\{1\}, \ldots, \{n\}\}$. Each pair of blocks then merges to a common ancestor independently at rate 1, forming a death process on the total number of blocks with rate $\binom{k}{2}$ when there are k blocks. More formally, the generator of the Kingman n-coalescent, $Q = (q_{\xi\eta})_{\xi\eta}$, is the square matrix with a row and column corresponding to each partition of $[n]$, and

$$q_{\xi\eta} = \begin{cases} 1 & \text{if } \eta \text{ can be obtained from } \xi \text{ by merging two blocks,} \\ -\binom{|\xi|}{2} & \text{if } \eta = \xi, \end{cases}$$

and 0 otherwise. The dynamics terminate once the process hits the MRCA, that is, the trivial partition $\{\{1, \ldots, n\}\}$. See Wakeley (2009) for an introduction to coalescent processes and their use in population genetics.

Let $\nu_t^{(i)}$ denote the number of offspring that particle i at time t has at time $t-1$. The following standing assumption will be central to our results.

Standing assumption: The conditional distribution of parental indices given offspring counts, $\mathbf{a}_t | \boldsymbol{\nu}_t$, is uniform over all vectors which satisfy $\nu_t^{(i)} = \#\{j \in [N] : a_t^{(j)} = i\}$ for each $i \in [N]$.

REMARK 2. The standing assumption concerns the marginal distribution of parental assignments without particle locations. Conditionally uniform assignment given locations would be much stronger. A sufficient condition for the standing assumption is exchangeability of the `Resample` mechanism in line 6 of Algorithm 1. However, Möhle ((1998), page 446) provides an example of a nonexchangeable particle system which still satisfies the standing assumption. For SMC, multinomial and residual resampling can be implemented in ways which satisfy the standing assumption, and any resampling scheme can be made to satisfy it by applying a uniformly sampled permutation to each realisation of ancestor indices. This technique was suggested in Andrieu, Doucet and Holenstein (2010), page 290.

Let ξ and η be partitions of $[n]$ with blocks ordered by the least element in each block, and with η obtained from ξ by merging some subsets of blocks. For $i \in [|\eta|]$, let b_i be the number of blocks of ξ that have been merged to form block i in η, so that $b_1 + \cdots + b_{|\eta|} = |\xi|$, and

define

$$p_{\xi\eta}(t) := \frac{1}{(N)_{|\xi|}} \sum_{\substack{i_1 \neq \cdots \neq i_{|\eta|}=1 \\ \text{all distinct}}}^{N} \bigl(\nu_t^{(i_1)}\bigr)_{b_1} \cdots \bigl(\nu_t^{(i_{|\eta|})}\bigr)_{b_{|\eta|}} \tag{1}$$

as the conditional transition probability of the genealogical process from state ξ at time $t-1$ to state η at time t, given $\boldsymbol{\nu}_t$ (suppressed from the notation). The interpretation of (1) as a conditional transition probability of $G_t^{(n,N)}$ is justified by the standing assumption by associating offspring with balls, parents with boxes, and merger events with two or more balls occupying the same box. Finally, let $P_N(t) = (p_{\xi\eta}(t))_{\xi,\eta}$ denote the corresponding conditional transition probability matrix.

Now let the conditional probability (resp., upper bound on conditional probability) given $\boldsymbol{\nu}_t$ of two (resp., more than two) lineages at time $t-1$ coalescing at time t be denoted by

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

$$D_N(t) := \frac{1}{N(N)_2} \sum_{i=1}^{N} \bigl(\nu_t^{(i)}\bigr)_2 \biggl(\nu_t^{(i)} + \frac{1}{N} \sum_{j \neq i}^{N} \bigl(\nu_t^{(j)}\bigr)^2\biggr).$$

The interpretations as (upper bounds on) conditional probabilities are justified by the standing assumption, and Lemma 1 below. For $t > 0$, let $\tau_N(t)$ be the time change driven by the random offspring counts

$$\tau_N(t) := \min\biggl\{s \geq 1 : \sum_{r=1}^{s} c_N(r) \geq t\biggr\}.$$

REMARK 3. The quantity $c_N(t)$ will play the role of a merger rate between pairs of particles at time t. For multinomial resampling,

$$\mathbb{E}\bigl[c_N(t)|\mathbf{w}_t\bigr] = \sum_{i=1}^{N} \bigl(w_t^{(i)}\bigr)^2 = \frac{1}{\mathrm{ESS}(t)} \approx \frac{1}{N}\biggl(1 + \mathrm{Var}\biggl(\frac{g_t(X_{t+1}, X_t)}{\mathbb{E}[g_t(X_{t+1}, X_t)]}\biggr)\biggr),$$

where $\mathrm{ESS}(t)$ is the effective sample size of Kong, Liu and Wong (1994), who also justify the approximation of the random left-hand side by the deterministic right-hand side. The indices of X_{t+1} and X_t are flipped due to the time reversal described at the top of this section. Thus the coalescence rate is high when the unnormalised importance weights have high relative variance and vice versa.

LEMMA 1. *Suppose that the standing assumption holds.*
Case 1: *For any partition ξ we have the equality*

$$p_{\xi\xi}(t) = 1 - \binom{|\xi|}{2} \frac{(3|\xi| - 1)(|\xi| - 2)}{6N^2} + O\bigl(N^{-3}\bigr) - \binom{|\xi|}{2}\bigl\{1 + O\bigl(N^{-1}\bigr)\bigr\}c_N(t).$$

Case 2: *Let η be obtained from ξ by merging exactly two blocks. Then*

$$c_N(t) - \binom{|\xi| - 2}{2}\bigl\{1 + O\bigl(N^{-1}\bigr)\bigr\}D_N(t) \leq p_{\xi\eta}(t) \leq \bigl\{1 + O\bigl(N^{-1}\bigr)\bigr\}c_N(t).$$

Case 3: *For any η obtained from ξ by one or more mergers involving more than two blocks in total, we have*

$$p_{\xi\eta}(t) \leq \binom{|\xi| - 2}{2}\bigl\{1 + O\bigl(N^{-1}\bigr)\bigr\}D_N(t).$$

PROOF. We begin by proving Case 1 by setting $\xi = \eta$ in (1). A multinomial expansion in reverse yields

$$\begin{aligned} p_{\xi\xi}(t) &= \frac{1}{(N)_{|\xi|}} \left\{ \left(\sum_{i=1}^{N} \nu_t^{(i)} \right)^{|\xi|} - \binom{|\xi|}{2} \sum_{i=1}^{N} (\nu_t^{(i)})^2 \left(\sum_{j=1}^{N} \nu_t^{(j)} \right)^{|\xi|-2} \right\} \\ &= \frac{N^{|\xi|}}{(N)_{|\xi|}} \left[1 - \binom{|\xi|}{2} \frac{1}{N^2} \sum_{i=1}^{N} (\nu_t^{(i)})^2 \right] \\ &= 1 - \binom{|\xi|}{2} \frac{(3|\xi|-1)(|\xi|-2)}{6N^2} + O(N^{-3}) \\ &\quad - \binom{|\xi|}{2} (1 + O(N^{-1})) c_N(t), \end{aligned}$$

where the second and third equalities follow from $\sum_{i=1}^{N} \nu_t^{(i)} \equiv N$, and from

$$\frac{N^j}{(N)_j} = 1 + \binom{j}{2} \frac{1}{N} + \binom{j}{2} \frac{2j-1}{3N^2} + O(N^{-3}). \tag{2}$$

The proof of Case 2 is essentially identical to calculations in Möhle ((1998), pages 442–443) and is omitted.

For Case 3, we have

$$\begin{aligned} p_{\xi\eta}(t) &= \frac{1}{(N)_{|\xi|}} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ \left(\sum_{j=1}^{N} \nu_t^{(j)} \right)^{|\xi|-2} - \sum_{\substack{j_1 \neq \cdots \neq j_{|\xi|-2} \neq i \\ \text{all distinct}}}^{N} \nu_t^{(j_1)} \cdots \nu_t^{(j_{|\xi|-2})} \right\} \\ &= \frac{1}{(N)_{|\xi|}} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ N^{|\xi|-2} - \left(\sum_{j \neq i}^{N} \nu_t^{(j)} \right)^{|\xi|-2} \right. \\ &\quad \left. + \binom{|\xi|-2}{2} \sum_{j \neq i}^{N} (\nu_t^{(j)})^2 \left(\sum_{k \neq i} \nu_t^{(k)} \right)^{|\xi|-4} \right\} \\ &\leq \frac{1}{(N)_{|\xi|}} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ N^{|\xi|-2} - (N - \nu_t^{(i)})^{|\xi|-2} \right. \\ &\quad \left. + \binom{|\xi|-2}{2} \sum_{j \neq i}^{N} (\nu_t^{(j)})^2 N^{|\xi|-4} \right\} \\ &\leq \frac{1}{(N)_{|\xi|}} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ (|\xi|-2) \nu_t^{(i)} N^{|\xi|-3} + \binom{|\xi|-2}{2} \sum_{j \neq i}^{N} (\nu_t^{(j)})^2 N^{|\xi|-4} \right\}, \end{aligned}$$

where we have used the Bernoulli inequality in the last step. The result follows from (2). □

THEOREM 1. *Fix $n \leq N$ as the observed number of particles from the output of an IPS with N particles, and suppose that the standing assumption holds. Suppose also that for any $0 \leq s < t < \infty$, we have*

$$\lim_{N \to \infty} \mathbb{E} \left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r) \right] = 0, \tag{3}$$

$$\lim_{N\to\infty} \mathbb{E}\big[c_N(t)\big] = 0, \tag{4}$$

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r)^2\right] = 0 \quad \textit{and} \tag{5}$$

$$\mathbb{E}\big[\tau_N(t) - \tau_N(s)\big] \le C_{t,s} N, \tag{6}$$

for some constant $C_{t,s} > 0$ that is independent of N. Then $(G^{(n,N)}_{\tau_N(t)})_{t\ge 0}$ converges to the Kingman n-coalescent in the sense of finite-dimensional distributions as $N \to \infty$.

REMARK 4. While unbiasedness of resampling, that is, $\mathbb{E}[\nu_t^{(i)} | w_t^{(i)}] = N w_t^{(i)}$, is part of the definition of our IPS, it is not required for Theorem 1. The key ingredients are the symmetry in the standing assumption, and the control over moments of orders up to four implicit in (3)–(6). Loosely, condition (4) implies that $(G^{(n,N)}_{\tau_N(t)})_{t\ge 0}$ converges to a continuous-time process, and together with (5) ensures that binary mergers happen at the required unit rate. Condition (3) ensures that mergers involving more than two lineages happen on a slower timescale than binary mergers, and (6) controls the speed with which the convergence in (4) takes place.

PROOF OF THEOREM 1. For $k \ge 1$ and $0 \le t_1 < \cdots < t_k < \infty$, the finite-dimensional distributions of the process $(G^{(n,N)}_{\tau_N(t)})_{t\ge 0}$ have the form

$$\begin{aligned}
&\mathbb{P}\big(G^{(n,N)}_{\tau_N(t_1)} = \eta_1, \ldots, G^{(n,N)}_{\tau_N(t_k)} = \eta_k | G^{(n,N)}_{\tau_N(t_0)} = \eta_0\big) \\
&\quad = \mathbb{E}\left[\prod_{d=1}^{k} \left\{ \prod_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} P_N(t) \right\}_{\eta_{d-1},\eta_d}\right] =: \mathbb{E}\left[\prod_{d=1}^{k} \chi_d^*\right],
\end{aligned}$$

where $\eta_0, \eta_1, \ldots, \eta_k$ is a sequence of partitions in which the blocks of η_d are obtained by merging some subsets of blocks of η_{d-1}, or $\eta_d = \eta_{d-1}$, and where $P_N(r)$ is the conditional transition matrix given family sizes defined below (1). The probability associated with any other sequence of partitions is zero. We will prove the result by bounding these finite-dimensional distributions both above and below by those of a Kingman n-coalescent.

Consider a transition between η_{d-1} and η_d at respective times $\tau_N(t_{d-1})$ and $\tau_N(t_d)$. The corresponding conditional transition probability given offspring counts $\boldsymbol{\nu}_{\tau_N(t_{d-1})+1}, \ldots, \boldsymbol{\nu}_{\tau_N(t_d)}$ can be written

$$\chi_d^* = \sum_{\boldsymbol{\xi} \in \eta_{d-1} \rightsquigarrow \eta_d} \prod_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} p_{\xi_{t-1}\xi_t}(t),$$

where the sum on the right-hand side is over all paths from η_{d-1} to η_d of the requisite length, $\boldsymbol{\xi} = (\eta_{d-1}, \xi_{\tau_N(t_{d-1})+1}, \ldots, \xi_{\tau_N(t_d)-1}, \eta_d)$, where each successive element of $\boldsymbol{\xi}$ either equals its predecessor, or is obtained from its predecessor by merging some subsets of blocks. By decomposing based on $\alpha \ge 0$ (the number of times between η_{d-1} and η_d in which mergers occur), as well as based on whether each merger involves exactly two lineages or more than two lineages, we can use Lemma 1, Cases 2 and 3 to upper bound the conditional transition probability given family sizes by

$$\begin{aligned}
\chi_d^* \le \sum_{\alpha=1}^{|\eta_{d-1}|-|\eta_d|} &(1 + O(N^{-1}))^{\alpha} \sum_{(\lambda,\mu)\in\Pi_2([\alpha])} \\
&\sum_{s_1<\cdots<s_\alpha=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\star} \left\{\prod_{r\in\lambda} \binom{n}{2} c_N(s_r)\right\} \left\{\prod_{r\in\mu} n^n \binom{n-2}{2} D_N(s_r)\right\},
\end{aligned}$$

and $\chi_d^* \le 1$ if $\alpha = 0$, where n^n is an upper bound on the number of arrangements of at most n lineages into at most n mergers. A further expansion shows that the product of k transition probabilities has the following bound when $\alpha_d \ge 1$ for each $d \in [k]$:

$$
\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d^*\right] \\
&\quad \le n^{n^2 k} \binom{n}{2}^{nk} \sum_{\alpha_1=1}^{|\eta_0|-|\eta_1|} \cdots \sum_{\alpha_k=1}^{|\eta_{k-1}|-|\eta_k|} \sum_{(\lambda_1,\mu_1)\in\Pi_2([\alpha_1])} \\
&\qquad \cdots \sum_{(\lambda_k,\mu_k)\in\Pi_2([\alpha_k])} \lim_{N\to\infty} \mathbb{E}\Bigg[\sum_{s_1^{(1)}<\cdots<s_{\alpha_1}^{(1)}=\tau_N(t_0)+1}^{\tau_N(t_1)\,\star} \\
&\qquad \cdots \sum_{s_1^{(k)}<\cdots<s_{\alpha_k}^{(k)}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\,\star} \prod_{d=1}^{k} \Big\{ \prod_{r\in\lambda_d} c_N\big(s_r^{(d)}\big) \Big\} \prod_{r\in\mu_d} D_N\big(s_r^{(d)}\big) \Bigg];
\end{aligned} \tag{7}
$$

here, $n^{n^2k}\binom{n}{2}^{nk}$ bounds $\prod_{d=1}^{k} \binom{n}{2}^{|\lambda_d|} n^{|\mu_d| n} \binom{n-2}{2}^{|\mu_d|}$ using $|\lambda_d| + |\mu_d| \le n$. Transitions in which any $\alpha_d = 0$ result in a similar bound in which the corresponding factors on the right-hand side of (7) are replaced by 1. Next, a multinomial expansion in reverse gives

$$
\begin{aligned}
&\sum_{s_1<\cdots<s_\alpha=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\,\star} \prod_{r=1}^{\alpha} c_N(s_r) \\
&\quad = \frac{1}{\alpha!} \sum_{\substack{s_1\neq\cdots\neq s_\alpha=\tau_N(t_{d-1})+1 \\ \text{all distinct}}}^{\tau_N(t_d)\,\star} \prod_{r=1}^{\alpha} c_N(s_r) \\
&\quad = \frac{1}{\alpha!} \Bigg(\sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t) \Bigg)^{\alpha} - \frac{1}{\alpha!} \binom{\alpha}{2} \Bigg(\sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t)^2 \Bigg) \\
&\qquad \times \Bigg(\sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t) \Bigg)^{\alpha-2}
\end{aligned} \tag{8}
$$

(where we take $\binom{0}{2} = \binom{1}{2} = 0$), so that by definition of $\tau_N(t)$,

$$
\begin{aligned}
&\frac{1}{\alpha!} \sum_{\substack{s_1\neq\cdots\neq s_\alpha=\tau_N(t_{d-1})+1 \\ \text{all distinct}}}^{\tau_N(t_d)\,\star} \prod_{r=1}^{\alpha} c_N(s_r) \\
&\quad \le \frac{1}{\alpha!} \sum_{s_1,\ldots,s_\alpha=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\,\star} \prod_{r=1}^{\alpha} c_N(s_r) \\
&\quad \le \frac{[t_d - t_{d-1} + c_N(\tau_N(t_d))]^{\alpha}}{\alpha!} \le \frac{(t_d - t_{d-1} + 1)^{\alpha}}{\alpha!},
\end{aligned} \tag{9}
$$

because $c_N(t) \le 1$. Suppose that $|\mu_i| > 0$ in (7) for some $i \in [k]$. Using $D_N(s_r^{(d)}) \le c_N(s_r^{(d)})$, which is clear from

$$
\frac{1}{N}\left(\nu_t^{(i)} + \frac{1}{N} \sum_{j\neq i}^{N} (\nu_t^{(j)})^2 \right) \le 1,
$$

for all but one $D_N(s_r^{(d)})$-factor in (7), and substituting in (9), gives

$$\lim_{N\to\infty}\mathbb{E}\Biggl[\prod_{d=1}^{k}\chi_d^*\Biggr]$$
$$\le n^{n^2k}\binom{n}{2}^{nk}\sum_{\alpha_1=1}^{|\eta_0|-|\eta_1|}\cdots\sum_{\alpha_k=1}^{|\eta_{k-1}|-|\eta_k|}\sum_{(\lambda_1,\mu_1)\in\Pi_2([\alpha_1])}$$
$$\cdots\sum_{(\lambda_k,\mu_k)\in\Pi_2([\alpha_k])}\Biggl\{\prod_{d=1}^{k}\frac{(t_d-t_{d-1}+1)^{|\alpha_d|-\delta_{di}}}{(\alpha_d-\delta_{di})!}\Biggr\}\lim_{N\to\infty}\mathbb{E}\Biggl[\sum_{s=\tau_N(t_{i-1})+1}^{\tau_N(t_i)}D_N(s)\Biggr],$$

where $\delta_{di}=\mathbb{1}_{\{d=i\}}$ is the Kronecker delta. As for (7), cases for which some α_d are 0 can be handled by a straightforward modification. All sums outside the expectation consist of a bounded number of terms in N, so (3) guarantees that the contribution of paths with $\sum_{d=1}^{k}|\mu_d|>0$ vanishes in the limit and only isolated, binary mergers can take place.

To describe transitions in which the only mergers that occur are isolated, binary mergers, we define $\tilde{Q}=(\tilde{q}_{\xi\eta})_{\xi\eta}$ to be the matrix obtained from Q by setting its diagonal entries to 0. Note that $(\tilde{Q}^\alpha)_{\xi\eta}$ is precisely the number of ways of going from ξ to η in exactly α steps, where a step consists of merging a pair of blocks.

Now consider a transition from η_{d-1} to η_d at respective times $\tau_N(t_{d-1})$ and $\tau_N(t_d)$ via binary mergers, that is, with $\alpha=|\eta_{d-1}|-|\eta_d|$ and $\lambda=[\alpha]$. By Lemma 1, Cases 1 and 2, its conditional probability is bounded by

$$\chi_d\le(\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d}\bigl(1+O\bigl(N^{-1}\bigr)\bigr)^\alpha\sum_{s_1<\cdots<s_\alpha=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\star}\Biggl\{\prod_{r=1}^{\alpha}c_N(s_r)\Biggr\}$$
$$\times\prod_{\substack{r=\tau_N(t_{d-1})+1\\ r\neq s_1,\ldots,r\neq s_\alpha}}^{\tau_N(t_d)}\Biggl\{1-\binom{|\eta_{d-1}|-|\{i:s_i<r\}|}{2}\bigl\{1+O\bigl(N^{-1}\bigr)\bigr\}c_N(r)\Biggr\},$$

where χ_d is the restriction of χ_d^* to trajectories involving only isolated binary mergers. An expansion of the product on the second line gives

$$\prod_{\substack{r=\tau_N(t_{d-1})+1\\ r\neq s_1,\ldots,r\neq s_\alpha}}^{\tau_N(t_d)}\Biggl\{1-\binom{|\eta_{d-1}|-|\{i:s_i<r\}|}{2}\bigl\{1+O\bigl(N^{-1}\bigr)\bigr\}c_N(r)\Biggr\}$$
$$=\sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha}(-1)^\beta\bigl(1+O\bigl(N^{-1}\bigr)\bigr)^\beta$$
$$\times\sum_{\substack{r_1<\cdots<r_\beta=\tau_N(t_{d-1})+1\\ \forall i:r_i\neq s_1,\ldots,r_i\neq s_\alpha}}^{\tau_N(t_d)\star}\prod_{j=1}^{\beta}\binom{|\eta_{d-1}|-|\{i:s_i<r_j\}|}{2}c_N(r_j).$$

The product of binomial coefficients depends only on the pattern of orderings between times denoted by $\{s_i\}_{i\in[\alpha]}$ and $\{r_j\}_{j\in[\beta]}$, but is otherwise independent of the exact values of the time points. Hence we have the bound

$$\chi_d\le\sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha}\bigl(1+O\bigl(N^{-1}\bigr)\bigr)^{\alpha+\beta}(\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d}$$

$$\times \sum_{(\lambda,\mu)\in\Pi_2([\alpha+\beta]):|\lambda|=\alpha} \left\{ \prod_{r\in\mu} -\binom{|\eta_{d-1}| - |\{i\in\lambda : i<r\}|}{2} \right\} \times \sum_{s_1<\cdots<s_{\alpha+\beta}=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\,\star} \prod_{r=1}^{\alpha+\beta} c_N(s_r). \tag{10}$$

Now we can see that

$$(\tilde{Q}^{\alpha})_{\eta_{d-1}\eta_d} \sum_{(\lambda,\mu)\in\Pi_2([\alpha+\beta]):|\lambda|=\alpha} \left\{ \prod_{r\in\mu} -\binom{|\eta_{d-1}| - |\{i\in\lambda : i<r\}|}{2} \right\} = (Q^{\alpha+\beta})_{\eta_{d-1}\eta_d} \tag{11}$$

by noting that

$$\sum_{(\lambda,\mu)\in\Pi_2([\alpha+\beta]):|\lambda|=\alpha} (\tilde{Q}^{\alpha})_{\eta_{d-1}\eta_d} = [(\mathbb{I}+\tilde{Q})^{\alpha+\beta}]_{\eta_{d-1}\eta_d}$$

is the number of discrete time paths of length $\alpha+\beta$ from η_{d-1} to η_d using α binary mergers and β identity steps, where $\mathbb{I}$ is the identity matrix of the same size as $\tilde{Q}$. In (11), each identity step results in multiplication by the corresponding diagonal entry of Q, which justifies the equality. Substituting (11) into (10) gives

$$\chi_d \le \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (1+O(N^{-1}))^{\alpha+\beta} (Q^{\alpha+\beta})_{\eta_{d-1}\eta_d} \times \sum_{s_1<\cdots<s_{\alpha+\beta}=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\,\star} \prod_{r=1}^{\alpha+\beta} c_N(s_r).$$

Taking a product of k transition probabilities with $\alpha_d := |\eta_d| - |\eta_{d-1}|$ for $d\in[k]$, and expanding that product, gives the bound

$$\begin{aligned} &\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \\ &\le \lim_{N\to\infty} \mathbb{E}\Bigg[\sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} (1+O(N^{-1}))^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|} \\ &\quad \times \left\{ \prod_{d=1}^{k} (Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\ge\alpha_d+\beta_d\}} \right\} \\ &\quad \times \sum_{s_1^{(1)}<\cdots<s_{\alpha_1+\beta_1}^{(1)}=\tau_N(t_0)+1}^{\tau_N(t_1)\,\star} \cdots \sum_{s_1^{(k)}<\cdots<s_{\alpha_k+\beta_k}^{(k)}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\,\star} \prod_{d=1}^{k} \prod_{r=1}^{\alpha_d+\beta_d} c_N\big(s_r^{(d)}\big) \Bigg]. \end{aligned} \tag{12}$$

We show in the Appendix that the hypotheses of the Fubini and dominated convergence theorems are satisfied, so that the expectation and limit can be passed inside the k-fold infinite summation over $\boldsymbol{\beta}$. That leaves

$$\begin{aligned} &\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \\ &\le \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \left\{ \prod_{d=1}^{k} (Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d} \right\} \end{aligned} \tag{13}$$

$$\times \lim_{N\to\infty} \mathbb{E}\Bigg[\Bigg\{\prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg\} \sum_{s_1^{(1)}<\cdots<s_{\alpha_1+\beta_1}^{(1)}=\tau_N(t_0)+1}^{\tau_N(t_1)\star}$$

$$\cdots \sum_{s_1^{(k)}<\cdots<s_{\alpha_k+\beta_k}^{(k)}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\star} \prod_{d=1}^{k}\prod_{r=1}^{\alpha_d+\beta_d} c_N\big(s_r^{(d)}\big)\Bigg].$$

Consider now a generic term in (13) for which $\prod_{d=1}^{k}(Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d}$ is positive, which requires $|\boldsymbol{\beta}|$ to be even by (11). We use the penultimate inequality in (9), and expand the resulting product to bound such the expectations in (13) by

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\Bigg[\Bigg\{\prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg\} \sum_{s_1^{(1)}<\cdots<s_{\alpha_1+\beta_1}^{(1)}=\tau_N(t_0)+1}^{\tau_N(t_1)\star}\\
&\qquad \cdots \sum_{s_1^{(k)}<\cdots<s_{\alpha_k+\beta_k}^{(k)}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\star} \prod_{d=1}^{k}\prod_{r=1}^{\alpha_d+\beta_d} c_N\big(s_r^{(d)}\big)\Bigg]\\
&\leq \sum_{j_1=0}^{\alpha_1}\cdots\sum_{j_k=0}^{\alpha_k}\prod_{d=1}^{k}\binom{\alpha_d+\beta_d}{j_d}\frac{(t_d-t_{d-1})^{\alpha_d+\beta_d-j_d}}{(\alpha_d+\beta_d)!}\\
&\qquad \times \lim_{N\to 0}\mathbb{E}\big[c_N\big(\tau_N(t_d)\big)^{j_d}\big] = \prod_{d=1}^{k}\frac{(t_d-t_{d-1})^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!},
\end{aligned} \tag{14}$$

where the last step follows from $c_N(t)\leq 1$, and from (4). Expansion (8) also yields the lower bound

$$\begin{aligned}
&\sum_{s_1<\cdots<s_\alpha=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\star}\prod_{r=1}^{\alpha} c_N(s_r)\\
&\quad\geq \frac{\{t_d-t_{d-1}-c_N(\tau_N(t_{d-1}))\}^{\alpha}}{\alpha!}\\
&\qquad -\binom{\alpha}{2}\Bigg(\sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t)^2\Bigg)\frac{(t_d-t_{d-1}+1)^{\alpha-2}}{\alpha!},
\end{aligned}$$

meaning that expectations in (13) with odd $|\boldsymbol{\beta}|$ can be lower bounded by

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\Bigg[\Bigg\{\prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg\} \sum_{s_1^{(1)}<\cdots<s_{\alpha_1+\beta_1}^{(1)}=\tau_N(t_0)+1}^{\tau_N(t_1)\star}\\
&\qquad \cdots \sum_{s_1^{(k)}<\cdots<s_{\alpha_k+\beta_k}^{(k)}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\star} \prod_{d=1}^{k}\prod_{r=1}^{\alpha_d+\beta_d} c_N\big(s_r^{(d)}\big)\Bigg]\\
&\geq \sum_{(\lambda,\mu)\in\Pi_2([k])}\mathbb{E}\Bigg[\Bigg\{\prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg\}
\end{aligned}$$

$$\times \left\{ \prod_{d \in \lambda} \frac{\{t_d - t_{d-1} - c_N(\tau_N(t_{d-1}))\}^{\alpha_d + \beta_d}}{(\alpha_d + \beta_d)!} \right\}$$

$$\times \prod_{d \in \mu} -\binom{\alpha_d + \beta_d}{2} \left(\sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t)^2 \right) \frac{(t_d - t_{d-1} + 1)^{\alpha_d + \beta_d - 2}}{(\alpha_d + \beta_d)!} \Bigg].$$

Using the lower bound for terms with a positive sign and upper bound for terms with a negative sign from

$$0 \leq \sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t)^2 \leq \sum_{t=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} c_N(t) \leq (t_d - t_{d-1} + 1)$$

on all factors $d \in \mu$ apart from one, and then (5), shows that terms with $|\mu| > 0$ vanish in the limit, after which an expansion akin to (14) results in

$$\lim_{N \to \infty} \mathbb{E}\Bigg[\left\{ \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d) - \tau_N(t_{d-1}) \geq \alpha_d + \beta_d\}} \right\} \sum_{s_1^{(1)} < \cdots < s_{\alpha_1+\beta_1}^{(1)} = \tau_N(t_0)+1}^{\tau_N(t_1)\,\star}$$

$$\cdots \sum_{s_1^{(k)} < \cdots < s_{\alpha_k+\beta_k}^{(k)} = \tau_N(t_{k-1})+1}^{\tau_N(t_k)\,\star} \prod_{d=1}^{k} \prod_{r=1}^{\alpha_d+\beta_d} c_N\big(s_r^{(d)}\big) \Bigg] \tag{15}$$

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

Now, for $0 \leq s < t < \infty$ and $\alpha \in \mathbb{N}$, we have

$$\begin{aligned} \{\tau_N(t) - \tau_N(s) < \alpha\} &\subseteq \left\{ \min\left\{ p \geq 1 : \sum_{r=\tau_N(s)+1}^{\tau_N(s)+p} c_N(r) \geq t - s \right\} < \alpha \right\} \\ &= \left\{ \min\left\{ p \geq 1 : \sum_{r=1}^{p} c_N\big(\tau_N(s) + r\big) \geq t - s \right\} < \alpha \right\} \\ &\subseteq \left\{ \sum_{r=1}^{\alpha} c_N\big(\tau_N(s) + r\big) \geq t - s \right\}, \end{aligned}$$

and hence

$$\lim_{N \to \infty} \mathbb{E}\Bigg[\prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d) - \tau_N(t_{d-1}) \geq \alpha_d\}} \Bigg]$$

$$\geq 1 - \sum_{d=1}^{k} \lim_{N \to \infty} \mathbb{P}\left(\sum_{r=1}^{\alpha_d} c_N\big(\tau_N(t_{d-1}) + r\big) \geq t_d - t_{d-1} \right) \tag{16}$$

$$\geq 1 - \sum_{d=1}^{k} \frac{1}{t_d - t_{d-1}} \sum_{r=1}^{\alpha_d} \lim_{N \to \infty} \mathbb{E}\big[c_N\big(\tau_N(t_{d-1} + r)\big)\big] = 1,$$

where the last two steps follow by Markov's inequality, and (4). Substituting (14) for terms in (13) with $|\boldsymbol{\beta}|$ even, and (15) as well as (16) for terms in (13) with $|\boldsymbol{\beta}|$ odd, and using the

fact that $(Q^j)_{\eta_{d-1}\eta_d} = 0$ for $j < \alpha_d$ gives

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

where the required absolute convergence for the final rearrangement is guaranteed by the calculation in the Appendix used to bound (12).

In the other direction, Lemma 1, Cases 1 and 2, give the lower bound

$$\chi_d \ge \sum_{s_1<\cdots<s_\alpha=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\star} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d}$$
$$\times \left\{ \prod_{r=1}^{\alpha} \left(c_N(s_r) - \binom{n-2}{2} \{1 + O(N^{-1})\} D_N(s_r) \right) \right\}$$
$$\times \prod_{\substack{r=\tau_N(t_{d-1})+1 \\ r\neq s_1,\ldots,r\neq s_\alpha}}^{\tau_N(t_d)} \left\{ 1 - \binom{n}{2} \frac{(3n-1)(n-2)}{6N^2} \{1 + O(N^{-1})\} \right.$$
$$\left. - \binom{|\eta_{d-1}| - |\{i : s_i < r\}|}{2} \{1 + O(N^{-1})\} c_N(r) \right\}.$$

A multinomial expansion of the product spanning the last two lines yields

$$\chi_d \ge \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} \frac{(1 + O(N^{-1}))^\beta}{(1 + O(N^{-2}))^{\alpha+\beta}} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d}$$
$$\times \sum_{(\lambda,\mu)\in\Pi_2([\alpha+\beta]):|\lambda|=\alpha} \left\{ \prod_{r\in\mu} - \binom{|\eta_{d-1}| - |\{i \in \lambda : i < r\}|}{2} \right\}$$
$$\times \left(1 - \binom{n}{2}^2 \frac{1}{N^2}\right)^{\tau_N(t_d)-\tau_N(t_{d-1})} \sum_{s_1<\cdots<s_{\alpha+\beta}=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\star}$$
$$\left\{ \prod_{r\in\mu} c_N(s_r) \right\} \prod_{r\in\lambda} \left\{ c_N(s_r) - \binom{n-2}{2} \{1 + O(N^{-1})\} D_N(s_r) \right\},$$

for sufficiently large N; and expanding the product over λ gives

$$\chi_d \ge \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{(\lambda,\mu,\pi)\in\Pi_3([\alpha+\beta]):|\mu|=\beta} \left\{ -\binom{n-2}{2} \right\}^{|\pi|}$$
$$\times \frac{(1 + O(N^{-1}))^{\beta+|\pi|}}{(1 + O(N^{-2}))^{\alpha+\beta}} \left(1 - \binom{n}{2}^2 \frac{1}{N^2}\right)^{\tau_N(t_d)-\tau_N(t_{d-1})}$$

$$\times \left\{ \prod_{r\in\mu} -\binom{|\eta_{d-1}| - |\{i \in \lambda \cup \pi : i < r\}|}{2} \right\}$$

$$\times \sum_{s_1<\cdots<s_{\alpha+\beta}=\tau_N(t_{d-1})+1}^{\tau_N(t_d)\,\star} \left\{ \prod_{r\in\lambda\cup\mu} c_N(s_r) \right\} \prod_{r\in\pi} D_N(s_r).$$

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

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \\
&\quad \geq \lim_{N\to\infty} \mathbb{E}\Bigg[\left(1 - \binom{n}{2}^2 \frac{1}{N^2}\right)^{\tau_N(t_k)-\tau_N(t_0)} \sum_{\beta_1=0}^{\infty} \\
&\qquad \cdots \sum_{\beta_k=0}^{\infty} \sum_{(\lambda_1,\mu_1,\pi_1)\in\Pi_3([\alpha_1+\beta_1]):|\mu_1|=\beta_1} \cdots \sum_{(\lambda_k,\mu_k,\pi_k)\in\Pi_3([\alpha_k+\beta_k]):|\mu_k|=\beta_k} \\
&\qquad \times \left\{ -\binom{n-2}{2} \right\}^{\sum_{d=1}^{k}|\pi_d|} \frac{(1+O(N^{-1}))^{|\boldsymbol{\beta}|+\sum_{d=1}^{k}|\pi_d|}}{(1+O(N^{-2}))^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \\
&\qquad \times \left\{ \prod_{d=1}^{k} (\tilde{Q}^{\alpha_d})_{\eta_{d-1}\eta_d} \prod_{r\in\mu_d} -\binom{|\eta_{d-1}| - |\{i \in \lambda_d \cup \pi_d : i < r\}|}{2} \right\} \\
&\qquad \times \sum_{s_1^{(1)}<\cdots<s^{(1)}_{\alpha_1+\beta_1}=\tau_N(t_0)+1}^{\tau_N(t_1)\,\star} \cdots \sum_{s_1^{(k)}<\cdots<s^{(k)}_{\alpha_k+\beta_k}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\,\star} \\
&\qquad \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}} \left\{ \prod_{r\in\lambda_d\cup\mu_d} c_N\big(s_r^{(d)}\big) \right\} \prod_{r\in\pi_d} D_N\big(s_r^{(d)}\big) \Bigg].
\end{aligned} \tag{17}$$

As for the upper bound, we verify in the Appendix that passing the expectation and the limit through the infinite sums is justified, whereupon (3), and the argument used to show that mergers involving more than two lineages cannot happen in the limit, implies that the contribution of terms with $\sum_{d=1}^{k} |\pi_d| > 0$ vanishes in the limit. Applying (11) to the remaining terms gives

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \geq \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \left\{ \prod_{d=1}^{k} (Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d} \right\} \\
\times \lim_{N\to\infty} \mathbb{E}\Bigg[\left(1 - \binom{n}{2}^2 \frac{1}{N^2}\right)^{\tau_N(t_k)-\tau_N(t_0)} \\
\times \left\{ \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}} \right\} \sum_{s_1^{(1)}<\cdots<s^{(1)}_{\alpha_1+\beta_1}=\tau_N(t_0)+1}^{\tau_N(t_1)\,\star} \\
\cdots \sum_{s_1^{(k)}<\cdots<s^{(k)}_{\alpha_k+\beta_k}=\tau_N(t_{k-1})+1}^{\tau_N(t_k)\,\star} \prod_{d=1}^{k} \prod_{r=1}^{\alpha_d+\beta_d} c_N\big(s_r^{(d)}\big) \Bigg].
\end{aligned}$$

Likewise, the expansion (14) applied to terms with negative sign ($|\boldsymbol{\beta}|$ odd), and (15) applied to terms with a positive sign ($|\boldsymbol{\beta}|$ even), followed by (4) and (5) show that

$$\lim_{N\to\infty}\mathbb{E}\Bigg[\prod_{d=1}^{k}\chi_d\Bigg] \geq \sum_{\beta_1=0}^{\infty}\cdots\sum_{\beta_k=0}^{\infty}\Bigg\{\prod_{d=1}^{k}(Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d}\frac{(t_d-t_{d-1})^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!}\Bigg\}$$
$$\times \lim_{N\to\infty}\mathbb{E}\Bigg[\left(1-\binom{n}{2}^2\frac{1}{N^2}\right)^{\tau_N(t_k)-\tau_N(t_0)}$$
$$\times \prod_{d=1}^{k}\mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg].$$

For terms with even $|\boldsymbol{\beta}|$, the bound $1-x\leq e^{-x}$ gives the inequality

$$\mathbb{E}\Bigg[\left(1-\binom{n}{2}^2\frac{1}{N^2}\right)^{\tau_N(t_k)-\tau_N(t_0)}\prod_{d=1}^{k}\mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg]$$
$$\geq \mathbb{E}\Bigg[\Bigg\{\prod_{d=1}^{k}\mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg\}+\log\left(1-\binom{n}{2}^2\frac{1}{N^2}\right)$$
$$\times\big[\tau_N(t_k)-\tau_N(t_0)\big]\Bigg].$$

Using (16) on the first term and (6) on the second yields

$$\lim_{N\to\infty}\mathbb{E}\Bigg[\left(1-\binom{n}{2}^2\frac{1}{N^2}\right)^{\tau_N(t_k)-\tau_N(t_0)}\prod_{d=1}^{k}\mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\Bigg]$$
$$\geq 1+\lim_{N\to\infty}NC_{t_k,t_0}\log\left(1-\binom{n}{2}^2\frac{1}{N^2}\right)=1.$$

A corresponding upper bound of 1 for terms with odd $|\boldsymbol{\beta}|$ is immediate, resulting in the overall lower bound

$$\lim_{N\to\infty}\mathbb{E}\Bigg[\prod_{d=1}^{k}\chi_d\Bigg] \geq \sum_{\beta_1=0}^{\infty}\cdots\sum_{\beta_k=0}^{\infty}\Bigg\{\prod_{d=1}^{k}(Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d}\frac{(t_d-t_{d-1})^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!}\Bigg\}$$
$$\geq \prod_{d=1}^{k}(e^{Q(t_d-t_{d-1})})_{\eta_{d-1}\eta_d},$$

which again follows from the fact that $(Q^j)_{\eta_{d-1}\eta_d}=0$ for $j<\alpha_d$, and the bound obtained for (17) in the Appendix. □

Our next aim is to show that particle filters with multinomial resampling can satisfy (3)–(6). We define the filtration $\mathcal{F}_t:=\sigma(\boldsymbol{v}_s;1\leq s\leq t)$ and will make extensive use of Lemma 2, whose proof is given in the Appendix, and one further preparatory lemma.

LEMMA 2. *For any* $0\leq s\leq t<\infty$,

$$\mathbb{E}\Bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}c_N(r)\Bigg]=\mathbb{E}\Bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}\mathbb{E}\big[c_N(r)|\mathcal{F}_{r-1}\big]\Bigg].$$

REMARK 5. It will be clear from the proof that the choice of summand, $c_N(r)$, is not special in Lemma 2. Any family of bounded functions depending on family sizes at only a single time could be substituted, and we will also apply the same result to $c_N(r)^2$ and $D_N(r)$.

LEMMA 3. *Suppose that the kernels $K_t(x, dx')$ in Algorithm 1 have respective densities $q_t(x, x')\,dx'$, and that*

$$\frac{1}{a} \le g_t(x, x') \le a, \tag{18}$$

$$\varepsilon h(x') \le q_t(x, x') \le \frac{1}{\varepsilon} h(x'), \tag{19}$$

for some constants $0 < \varepsilon \le 1 \le a < \infty$, and probability density $h(x)$, uniformly in t as well as both arguments. Then SMC algorithms with multinomial resampling (i.e., Algorithm 1, for which the `Resample` *method on line* 6 *is Algorithm 2) satisfy*

$$\mathbb{E}\big[D_N(t)|\mathcal{F}_{t-1}\big] \le \frac{C_1}{N}\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big], \tag{20}$$

$$\mathbb{E}\big[c_N(t)^2|\mathcal{F}_{t-1}\big] \le \frac{C_2}{N}\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big], \tag{21}$$

for constants $C_1, C_2 > 0$ that are independent of N, and

$$\frac{\varepsilon^4}{Na^4} \le \mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big] \le \frac{a^4}{N\varepsilon^4}. \tag{22}$$

REMARK 6. Assumptions (18) and (19) are strong, and can only be expected to hold on compact state spaces. Many rigorous results about SMC require similarly strong assumptions, but are robust to violations of them in practice (Cérou, Del Moral and Guyader (2011), Chopin (2004), Del Moral and Guionnet (2001), Künsch (2005), Jacob, Murray and Rubenthaler (2015)).

PROOF OF LEMMA 3. Recall that our reverse-time perspective results in SMC algorithms whose time steps progress backwards through time points $t+1, t, t-1, \dots, 0$. Thus, for any integrable function $f(\mathbf{a}_t)$, the forwards-in-time Markov property of SMC algorithms gives

$$\begin{aligned}\mathbb{E}\big[f(\mathbf{a}_t)|\mathcal{F}_{t-1}\big] &= \mathbb{E}\big[\mathbb{E}\big[f(\mathbf{a}_t)|\mathbf{X}_{t-1}, \mathbf{w}_{t-1}\big]|\mathcal{F}_{t-1}\big] \\ &= \mathbb{E}\big[\mathbb{E}\big[f(\mathbf{a}_t)|\mathbf{a}_{t+1}, \mathbf{X}_{t+1}, \mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{w}_{t-1}\big]|\mathcal{F}_{t-1}\big].\end{aligned}$$

For multinomial resampling, the law with respect to which the inner conditional expectation is taken is

$$\begin{aligned}&\mathbb{P}(\mathbf{a}_t = \mathbf{a}|\mathbf{a}_{t+1}, \mathbf{X}_{t+1}, \mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{w}_{t-1}) \\ &\propto \prod_{i=1}^{N} g_t\big(X_{t+1}^{(a_{t+1}^{(a_i)})}, X_t^{(a_i)}\big) q_{t-1}\big(X_t^{(a_i)}, X_{t-1}^{(i)}\big),\end{aligned}$$

that is, the entries of $\mathbf{a}_t|\mathbf{a}_{t+1}, \mathbf{X}_{t+1}, \mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{w}_{t-1}$ are independent, with

$$\begin{aligned}&a_t^{(i)}|\mathbf{a}_{t+1}, \mathbf{X}_{t+1}, \mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{w}_{t-1} \\ &\sim \text{Categorical}\big(g_t\big(X_{t+1}^{(a_{t+1}^{(1)})}, X_t^{(1)}\big) q_{t-1}\big(X_t^{(1)}, X_{t-1}^{(i)}\big), \\ &\quad \cdots, g_t\big(X_{t+1}^{(a_{t+1}^{(N)})}, X_t^{(N)}\big) q_{t-1}\big(X_t^{(N)}, X_{t-1}^{(i)}\big)\big),\end{aligned}$$

where in this and subsequent uses, the probabilities parametrising categorical distributions are given up to a normalising constant.

We call a function f I-increasing if it is increasing in $\sum_{i\in I}|\{j\in[N]:a_t^{(j)}=i\}|$ for $I\subseteq[N]$. A balls-in-bins coupling shows that an I-increasing $\mathbb{E}[f(\mathbf{a}_t)|\mathbf{a}_{t+1},\mathbf{X}_{t+1},\mathbf{X}_t,\mathbf{X}_{t-1},\mathbf{w}_{t-1}]\leq\mathbb{E}[f(\tilde{\mathbf{a}}_t)]$, where $\tilde{\mathbf{a}}$ is independent of $\mathcal{F}_\infty$, the entries of $\tilde{\mathbf{a}}_t$ are independent of each other, and

$$\tilde{a}_t^{(j)}\sim\text{Categorical}\left(\left(\frac{a}{\varepsilon}\right)^{\mathbb{1}_{\{1\in I\}}-\mathbb{1}_{\{1\notin I\}}},\ldots,\left(\frac{a}{\varepsilon}\right)^{\mathbb{1}_{\{N\in I\}}-\mathbb{1}_{\{N\notin I\}}}\right),$$

which follows from substituting upper bounds from (18) and (19) for the probabilities corresponding to bins in I, corresponding lower bounds elsewhere, and canceling common factors. Writing

$$\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big]=\frac{1}{(N)_2}\sum_{i=1}^N\mathbb{E}\big[(\nu_t^{(i)})_2|\mathcal{F}_{t-1}\big]=:\frac{1}{(N)_2}\sum_{i=1}^N\mathbb{E}\big[f_i(\mathbf{a}_t)|\mathcal{F}_{t-1}\big],$$

noting that f_i is $\{i\}$-increasing, and using the binomial moment formula,

$$X\sim\text{Bin}(N;p)\quad\Rightarrow\quad\mathbb{E}\big[(X)_q\big]=(N)_q\,p^q,\tag{23}$$

which can be found, for example, in Mosimann (1962), applied to $\nu_t^{(i)}$, yields

$$\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big]\leq\frac{1}{(N)_2}\sum_{i=1}^N(N)_2\left(\frac{a/\varepsilon}{(N-1)\varepsilon/a+a/\varepsilon}\right)^2\leq\frac{a^4}{N\varepsilon^4}.$$

Flipping the upper and lower bounds in the argument establishes (22) via

$$\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big]\geq\frac{\varepsilon^4}{Na^4}.\tag{24}$$

To verify (21), we write

$$\begin{aligned}
&\mathbb{E}\big[c_N(t)^2|\mathcal{F}_{t-1}\big]\\
&\quad=\frac{1}{[(N)_2]^2}\mathbb{E}\left[\left(\sum_{i=1}^N(\nu_t^{(i)})_2\right)^2\Big|\mathcal{F}_{t-1}\right]\\
&\quad=\frac{1}{[(N)_2]^2}\left(\sum_{i=1}^N\mathbb{E}\big[[(\nu_t^{(i)})_2]^2|\mathcal{F}_{t-1}\big]+\sum_{i=1}^N\sum_{j\neq i}^N\mathbb{E}\big[(\nu_t^{(i)})_2(\nu_t^{(j)})_2|\mathcal{F}_{t-1}\big]\right)\qquad(25)\\
&\quad\leq\frac{1}{[(N)_2]^2}\sum_{i=1}^N\Bigg\{\mathbb{E}\big[(\nu_t^{(i)})_4+4(\nu_t^{(i)})_3+2(\nu_t^{(i)})_2|\mathcal{F}_{t-1}\big]\\
&\qquad+\sum_{j\neq i}^N\mathbb{E}\big[(\nu_t^{(i)}+\nu_t^{(j)})_4+4(\nu_t^{(i)}+\nu_t^{(j)})_3+2(\nu_t^{(i)}+\nu_t^{(j)})_2|\mathcal{F}_{t-1}\big]\Bigg\},
\end{aligned}$$

where the inequality uses $[(\nu)_2]^2=(\nu)_4+4(\nu)_3+2(\nu)_2$. The first expectation on the right-hand side is $\{i\}$-increasing, and so by (18), (19) and (23),

$$\mathbb{E}\big[(\nu_t^{(i)})_4+4(\nu_t^{(i)})_3+2(\nu_t^{(i)})_2|\mathcal{F}_{t-1}\big]\leq\frac{(N)_2a^4}{N^2\varepsilon^4}\left(\frac{a^4}{\varepsilon^4}+\frac{4a^2}{\varepsilon^2}+2\right).\tag{26}$$

The second is $\{i, j\}$-increasing, which gives

$$\begin{aligned}(27)\qquad &\mathbb{E}\big[(\nu_t^{(i)}+\nu_t^{(j)})_4+4(\nu_t^{(i)}+\nu_t^{(j)})_3+2(\nu_t^{(i)}+\nu_t^{(j)})_2|\mathcal{F}_{t-1}\big]\\ &\quad\le\frac{8(N)_2a^4}{N^2\varepsilon^2}\bigg(\frac{2a^4}{\varepsilon^4}+\frac{4a^2}{\varepsilon^2}+1\bigg).\end{aligned}$$

Substituting (26) and (27) into (25) yields

$$\mathbb{E}\big[c_N(t)^2|\mathcal{F}_{t-1}\big]\le\frac{16a^4}{(N)_2\varepsilon^4}\bigg(\frac{a^4}{\varepsilon^4}+\frac{2a^2}{\varepsilon^2}+1\bigg)\le\frac{C_2}{N}\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big],$$

where the last inequality follows from (24).

Finally, for (20) we write

$$\begin{aligned}&\mathbb{E}\big[D_N(t)|\mathcal{F}_{t-1}\big]\\ &\quad=\frac{1}{N(N)_2}\sum_{i=1}^N\bigg\{\mathbb{E}\big[(\nu_t^{(i)})_2(\nu_t^{(i)})|\mathcal{F}_{t-1}\big]+\frac{1}{N}\sum_{j\ne i}^N\mathbb{E}\big[(\nu_t^{(i)})_2(\nu_t^{(j)})^2|\mathcal{F}_{t-1}\big]\bigg\}\\ &\quad\le\frac{1}{N(N)_2}\sum_{i=1}^N\bigg\{\mathbb{E}\big[(\nu_t^{(i)})_3+2(\nu_t^{(i)})_2|\mathcal{F}_{t-1}\big]+\frac{1}{N}\sum_{j\ne i}^N\mathbb{E}\big[(\nu_t^{(i)}+\nu_t^{(j)})_4\\ &\qquad+5(\nu_t^{(i)}+\nu_t^{(j)})_3+4(\nu_t^{(i)}+\nu_t^{(j)})_2|\mathcal{F}_{t-1}\big]\bigg\},\end{aligned}$$

where the second line follows from $(\nu)_2\nu^2=(\nu)_4+5(\nu)_3+4(\nu)_2$. The expectations are $\{i\}$- and $\{i, j\}$-increasing, respectively, and so (23) gives

$$\begin{aligned}&\mathbb{E}\big[(\nu_t^{(i)})_3+2(\nu_t^{(i)})_2|\mathcal{F}_{t-1}\big]\le\frac{(N)_2a^4}{N^2\varepsilon^4}\bigg(\frac{a^2}{\varepsilon^2}+2\bigg),\\ &\mathbb{E}\big[(\nu_t^{(i)}+\nu_t^{(j)})_4+5(\nu_t^{(i)}+\nu_t^{(j)})_3+4(\nu_t^{(i)}+\nu_t^{(j)})_2|\mathcal{F}_{t-1}\big]\\ &\quad\le\frac{8(N)_2a^4}{N^2\varepsilon^4}\bigg(2+\frac{5a^2}{\varepsilon^2}+\frac{2a^4}{\varepsilon^4}\bigg).\end{aligned}$$

Using (24) then yields

$$\begin{aligned}\mathbb{E}\big[D_N(t)|\mathcal{F}_{t-1}\big]&\le\frac{1}{(N)_2}\bigg\{\frac{(N)_2a^4}{N^2\varepsilon^4}\bigg(\frac{a^2}{\varepsilon^2}+2\bigg)+\frac{8(N)_2a^4}{N^2\varepsilon^4}\bigg(2+\frac{5a^2}{\varepsilon^2}+\frac{2a^4}{\varepsilon^4}\bigg)\bigg\}\\ &\le\frac{C_1}{N}\mathbb{E}\big[c_N(t)|\mathcal{F}_{t-1}\big].\end{aligned}$$

□

COROLLARY 1. *Genealogies of n particles from SMC algorithms with multinomial resampling converge to the Kingman n-coalescent in the sense of finite-dimensional distributions under the time-scaling introduced in Theorem 1 under the conditions of Lemma 3.*

PROOF. The standing assumption holds by exchangeability of multinomial resampling. Condition (4) is immediate by taking expectations in (22).

To verify (5), we use Lemma 2 twice with (21) in between to obtain

$$\begin{aligned}\mathbb{E}\bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}c_N(r)^2\bigg]&=\mathbb{E}\bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}\mathbb{E}\big[c_N(r)^2|\mathcal{F}_{r-1}\big]\bigg]\\ &\le\frac{C_2}{N}\mathbb{E}\bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}c_N(r)\bigg]\le\frac{C_2(t-s+1)}{N}.\end{aligned}$$

Condition (3) can be checked using (20) via the same argument as for (5):

$$\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r)\right] \le \frac{C_1}{N}\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r)\right] \le \frac{C_1(t-s+1)}{N}.$$

Finally, for (6) we use (22) and Lemma 2 to obtain

$$\mathbb{E}\big[\tau_N(t)-\tau_N(s)\big] = \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} \frac{\mathbb{E}[c_N(r)|\mathcal{F}_{r-1}]}{\mathbb{E}[c_N(r)|\mathcal{F}_{r-1}]}\right] \le \frac{a^4}{\varepsilon^4}N(t-s+1).$$

□

We now demonstrate that while the mode of convergence in Theorem 1 is too weak for convergence of expectations of continuous, bounded test functions, useful information can still be obtained. For example, the time until the Kingman n-coalescent reaches its MRCA can be constructed as $T_n := \sum_{k=2}^{n} S_k$, where the $(S_2, \ldots, S_n)$ are independent, and $S_k \sim \mathrm{Exp}(\binom{k}{2})$. Moments of $\tau_N(T_n)$ give the time scale on which a SMC algorithm will reach its MRCA.

COROLLARY 2. *Under the assumptions of Corollary 1, the following bounds hold for any* $1 \le n \le N$, *and any coupling of* (τ_N, T_n):

$$\frac{2\varepsilon^4 N}{a^4}\big(1-n^{-1}\big) \le \mathbb{E}\big[\tau_N(T_n)\big] \le \frac{2a^4 N}{\varepsilon^4}\big(1-n^{-1}\big) + \frac{a^8}{\varepsilon^4},$$

$$\mathrm{Var}\big(\tau_N(T_n)\big) \le \frac{N^2 a^8}{\varepsilon^8}\left(\frac{4\pi^2}{3} - 12 + O\big(n^{-1}\big)\right) + O(N),$$

$$\mathrm{Var}\big(\tau_N(T_n)\big) \ge \frac{N^2 \varepsilon^8}{a^8}\left(\frac{4\pi^2}{3} - 12 + O\big(n^{-1}\big)\right).$$

PROOF. By (22), Lemma 2, the definition of τ_N, and the fact that

$$\mathbb{E}\big[c_N(t)\big] = \mathbb{E}\big[\mathbb{E}\big[c_N(t)|\mathbf{w}_t\big]\big] = \sum_{i=1}^{N}\mathbb{E}\big[\big(w_t^{(i)}\big)^2\big] \le \frac{a^4}{N},$$

we have

$$\begin{aligned}\mathbb{E}\big[\tau_N(t)\big] &= \mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} \frac{\mathbb{E}[c_N(s)|\mathcal{F}_{s-1}]}{\mathbb{E}[c_N(s)|\mathcal{F}_{s-1}]}\right] \le N\frac{a^4}{\varepsilon^4}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} \mathbb{E}\big[c_N(s)|\mathcal{F}_{s-1}\big]\right] \\ &= N\frac{a^4}{\varepsilon^4}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)\right] \le N\frac{a^4}{\varepsilon^4}\big\{t + \mathbb{E}\big[c_N\big(\tau_N(t)\big)\big]\big\} \le N\frac{a^4}{\varepsilon^4}\left(t + \frac{a^4}{N}\right).\end{aligned}$$

A corresponding lower bound of $Nt\varepsilon^4/a^4$ is obtained via the same argument. Conditioning on T_n in $\mathbb{E}[\tau_N(T_n)]$, and using $\mathbb{E}[T_n] = 2(1-n^{-1})$ (Wakeley (2009), page 76) establishes the claimed bounds on the mean.

A similar argument for the variance gives

$$\begin{aligned}\mathrm{Var}\big(\tau_N(t)\big) &= \mathrm{Var}\left(\sum_{s=1}^{\tau_N(t)} \frac{\mathbb{E}[c_N(s)|\mathcal{F}_{s-1}]}{\mathbb{E}[c_N(s)|\mathcal{F}_{s-1}]}\right) \le \mathrm{Var}\left(\frac{Na^4}{\varepsilon^4}\sum_{s=1}^{\tau_N(t)} \mathbb{E}\big[c_N(s)|\mathcal{F}_{s-1}\big]\right) \\ &= \frac{N^2 a^8}{\varepsilon^8}\left(\mathbb{E}\left[\sum_{s_1,s_2=1}^{\tau_N(t)} c_N(s_1)c_N(s_2)\right] - \mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)\right]^2\right) \\ &\le \frac{N^2 a^8}{\varepsilon^8}\big(\mathbb{E}\big[\big\{t + c_N\big(\tau_N(t)\big)\big\}^2\big] - t^2\big) \le \frac{2Na^{12}}{\varepsilon^8}t + O(1),\end{aligned}$$

because taking expectations in (21) shows that $\mathbb{E}[c_N(t)^2] = O(N^{-2})$. Thus, by the law of total variance,

$$\begin{aligned}\operatorname{Var}\big(\tau_N(T_n)\big) &= \operatorname{Var}\big(\mathbb{E}\big[\tau_N(T_n)|T_n\big]\big) + \mathbb{E}\big[\operatorname{Var}\big(\tau_N(T_n)|T_n\big)\big] \\ &\le \frac{N^2a^8}{\varepsilon^8}\operatorname{Var}(T_n) + \frac{2Na^{12}}{\varepsilon^8}\mathbb{E}[T_n] + O(1) \\ &= \frac{N^2a^8}{\varepsilon^8}\bigg(\frac{4\pi^2}{3} - 12 + O\big(n^{-1}\big)\bigg) + O(N),\end{aligned}$$

since $\operatorname{Var}(T_n) = 4\pi^2/3 - 12 + O(n^{-1})$ (Wakeley (2009), page 76). The other direction is much simpler since variance is nonnegative:

$$\begin{aligned}\operatorname{Var}\big(\tau_N(T_n)\big) &\ge \operatorname{Var}\big(\mathbb{E}\big[\tau_N(T_n)|T_n\big]\big) \ge \frac{N^2\varepsilon^8}{a^8}\operatorname{Var}(T_n) \\ &= \frac{N^2\varepsilon^8}{a^8}\bigg(\frac{4\pi^2}{3} - 12 + O\big(n^{-1}\big)\bigg). \qquad \square\end{aligned}$$

Simulations in the next section confirm that the scalings predicted by Corollary 2 hold for real algorithms. Strengthening the mode of convergence in Theorem 1 to obtain a wider class of bounds is a subject of ongoing work.

3. A numerical example. In this section, we study the robustness of Theorem 1 by demonstrating via simulation that the scalings of Corollary 2 hold nonasymptotically for a particle system for which (18) and (19) fail. We also show that the same scalings hold for popular alternatives to multinomial resampling which do not satisfy the standing assumption.

Let $(X_t, Y_t)_{t\ge 0}$ be the discretised Ornstein–Uhlenbeck process:

$$\begin{aligned}X_{t+1} &= (1-\Delta)X_t + \sqrt{\Delta}\xi_t, \\ X_0 &\sim N(0,1), \\ Y_t|X_t &\sim N\big(X_t, \sigma^2\big),\end{aligned}$$

where $\Delta > 0$ is the step size, σ^2 is the observation noise and $\boldsymbol{\xi} \overset{\text{iid}}{\sim} N(0,1)$.

We observe a realisation of the trajectory $\boldsymbol{Y} = \boldsymbol{y}$, but not $\boldsymbol{X}$, and specify a bootstrap particle filter targeting the smoothing distribution $P(\boldsymbol{x}|\boldsymbol{y})$ via

$$\begin{aligned}q_t(x,x') \equiv p(x,x') &:= (2\pi\Delta)^{-1/2}\exp\bigg(\frac{-(x'-(1-\Delta)x)^2}{2\Delta}\bigg), \\ g_t(x,x') \equiv \psi(x',y_t) &:= \big(2\pi\sigma^2\big)^{-1/2}\exp\bigg(\frac{-(y_t-x')^2}{2\sigma^2}\bigg), \\ p(x_{-1},x) \equiv \mu(x) &:= (2\pi)^{-1/2}\exp\bigg(\frac{-x^2}{2}\bigg),\end{aligned}$$

in Algorithm 1. We set $\Delta = \sigma = 0.1$, simulated an observed trajectory of length $T = 40{,}960$, and used it as the input for bootstrap particle filters with $N = 8192$, and recorded subtree heights for uniformly sampled subsets of leaves of size $n \in \{2, 4, 8, \dots, N\}$. The mean and variance of tree heights were then estimated from 1000 replicates for each of multinomial, residual, stratified and systematic resampling, with all four simulations run using the same observed data and the same random seed. Representative results are shown in Figure 1; results for filters as small as $N = 128$ were similar.

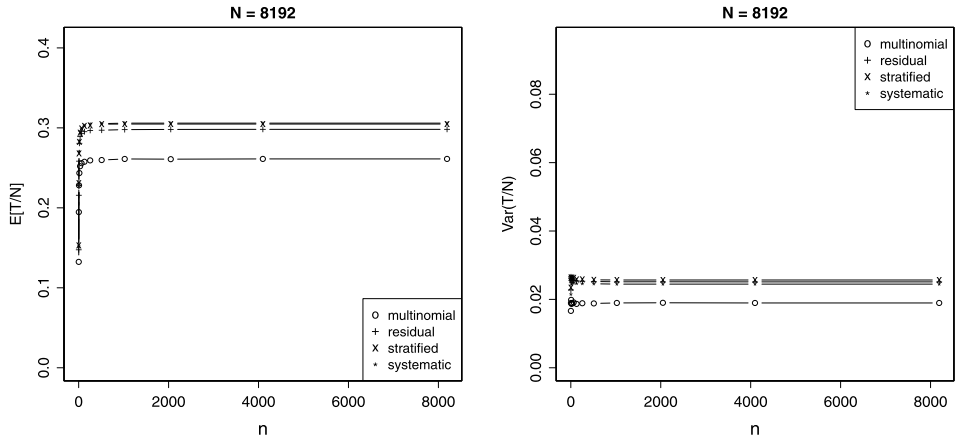

FIG. 1. *Mean (left) and variance (right) of tree heights for $n \leq N$.*

All depicted moment estimators remain bounded away from 0 and ∞ as predicted by Corollary 2, including in the $n \approx N$ regime to which Theorem 1 does not apply. As such, our simulations suggests that the $O(N \log N)$ bound of Jacob, Murray and Rubenthaler (2015) on the height of the genealogy of all N particles could be sharpened to an $O(N)$ bound.

4. Discussion. We have shown in Theorem 1 that genealogies of SMC algorithms converge, in the sense of finite-dimensional distributions, to the Kingman n-coalescent under a suitable rescaling of time, and certain assumptions. Thus, we are able to use the tractability of the Kingman n-coalescent to characterise SMC algorithms as well. We illustrated this in Corollary 2 by obtaining scaling results for the first two moments of the number of generations from the leaves to the MRCA. Asymptotic expressions for other quantities, for example, the probability of retaining at least two branches for a time window of a desired length can also be easily obtained. Our method is the only tool for a priori estimation of path degeneracy of which we are aware, and represents an important step towards practical guidelines for simultaneous minimisation of path and weight degeneracies. Strengthening the mode of convergence in Theorem 1, along the lines of Möhle (1999), Theorem 3.1, would be desirable to enable the analysis of larger classes of functionals, and is the subject of ongoing work.

Our assumptions essentially require a compact state space and multinomial resampling. However, the simulation study in Section 3 suggests that neither is necessary: our example did not satisfy (18) or (19), and a range of resampling methods fit the predictions of Corollary 2. We believe that our assumptions can be relaxed if subtler arguments are used to control the variability of family sizes. Our simulations also confirmed that the predicted scalings hold nonasymptotically, and hence are relevant for real algorithms.

Theorem 1 also demonstrates that the domain of attraction of the Kingman n-coalescent includes non-Markovian genealogies. Previous results have focused on Markovian genealogies, which are typical models for neutral genetic evolution (Möhle (1998), Theorem 1). Thus, our results improve the tractability of models with random but inherited fecundity.

APPENDIX

Here we verify that the Fubini and dominated convergence theorems apply to (12) and (17) are satisfied, and prove Lemma 2. Taking the modulus of summands on the right-hand side of (12) and using (9) yields

$$\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \leq \lim_{N\to\infty} \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \left(1 + O(N^{-1})\right)^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}$$

$$\times \prod_{d=1}^{k} \big(|Q|^{\alpha_d+\beta_d}\big)_{\eta_{d-1}\eta_d} \frac{(t_d - t_{d-1}+1)^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!}$$

$$= \prod_{d=1}^{k} \big(e^{|Q|(t_d-t_{d-1}+1)}\big)_{\eta_{d-1}\eta_d} < \infty,$$

where $|Q|$ is the matrix whose entries are the absolute values of those of Q, and the last equality holds because $(|Q|^j)_{\eta_{d-1}\eta_d} = 0$ for $j < \alpha_d$.

Similarly, for (17), using $D_N(t) \le c_N(t)$, (9) and (11) gives

$$\lim_{N\to\infty} \mathbb{E}\Bigg[\prod_{d=1}^{k} \chi_d\Bigg]$$

$$\le \lim_{N\to\infty} \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \binom{n-2}{2}^{|\boldsymbol{\alpha}|} \frac{(1+O(N^{-1}))^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}{(1+O(N^{-2}))^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}$$

$$\times \prod_{d=1}^{k} \big(|Q|^{\alpha_d+\beta_d}\big)_{\eta_{d-1}\eta_d} \frac{(t_d - t_{d-1}+1)^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!}$$

$$= \binom{n-2}{2}^{|\boldsymbol{\alpha}|} \prod_{d=1}^{k} \big(e^{|Q|(t_d-t_{d-1}+1)}\big)_{\eta_{d-1}\eta_d} < \infty.$$

PROOF OF LEMMA 2. Define $M_{s,t} := \sum_{r=s+1}^{t} c_N(r) - \mathbb{E}[c_N(r)|\mathcal{F}_{r-1}]$, and for fixed $K > 0$ note that $\tau_N(t) \wedge K$ is a bounded $\mathcal{F}_t$-stopping time. We have

$$\mathbb{E}[M_{s,\tau_N(t)\wedge K}]$$

$$= \sum_{r=s+1}^{K} \mathbb{E}\big[\big(c_N(r) - \mathbb{E}\big[c_N(r)|\mathcal{F}_{r-1}\big]\big)\mathbb{1}_{\{\tau_N(t)\wedge K \ge r\}}\big]$$

$$= \sum_{r=s+1}^{K} \mathbb{E}\big[\mathbb{1}_{\{\tau_N(t)\wedge K > r-1\}}\big(\mathbb{E}\big[c_N(r) - \mathbb{E}\big[c_N(r)|\mathcal{F}_{r-1}\big]\big)|\mathcal{F}_{r-1}\big]\big] = 0,$$

where the second line holds because $\mathbb{1}_{\{\tau_N(t)\wedge K \ge r\}} = \mathbb{1}_{\{\tau_N(t)\wedge K > r-1\}}$ is $\mathcal{F}_{r-1}$-measurable. Conditioning on $\tau_N(s)$ and using $\tau_N(s) \le \tau_N(t)$ yields

$$\mathbb{E}\Bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)\wedge K} c_N(r)\Bigg] = \mathbb{E}\Bigg[\sum_{r=\tau_N(s)+1}^{\tau_N(t)\wedge K} \mathbb{E}\big[c_N(r)|\mathcal{F}_{r-1}\big]\Bigg],$$

and the monotone convergence theorem concludes the proof. □

Acknowledgments. The first author was supported by EPSRC Grant EP/HO23364/1 as part of the MASDOC DTC at the University of Warwick, by DFG Grant BL 1105/3-2, and by EPSRC Grant EP/R044732/1.

The second author was supported in part by EPSRC Grant EP/L018497/1.

The third author was supported in part by funding from the Lloyd's Register Foundation—Alan Turing Institute Programme on Data-Centric Engineering.

REFERENCES

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

CÉROU, F., DEL MORAL, P. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068 https://doi.org/10.1214/10-AIHP358

CHAN, H. P. and LAI, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41** 2877–2904. MR3161451 https://doi.org/10.1214/13-AOS1172

CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989 https://doi.org/10.1214/009053604000000698

CRISAN, D. and LYONS, T. (1997). Nonlinear filtering and measure-valued processes. *Probab. Theory Related Fields* **109** 217–244. MR1477650 https://doi.org/10.1007/s004400050131

DEL MORAL, P. (1996). Nonlinear filtering: Interacting particle solution. *Markov Process. Related Fields* **2** 555–579. MR1431187

DEL MORAL, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration*. *Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333 https://doi.org/10.1111/j.1467-9868.2006.00553.x

DEL MORAL, P. and GARNIER, J. (2005). Genealogical particle analysis of rare events. *Ann. Appl. Probab.* **15** 2496–2534. MR2187302 https://doi.org/10.1214/105051605000000566

DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122 https://doi.org/10.1016/S0246-0203(00)01064-5

DEL MORAL, P. and MICLO, L. (2001). Genealogies and increasing propagation of chaos for Feynman–Kac and genetic models. *Ann. Appl. Probab.* **11** 1166–1198. MR1878294

DEL MORAL, P. and MICLO, L. (2007). Strong propagations of chaos in Moran's type particle interpretations of Feynman–Kac measures. *Stoch. Anal. Appl.* **25** 519–575. MR2321897 https://doi.org/10.1080/07362990701282666

DEL MORAL, P., MICLO, L., PATRAS, F. and RUBENTHALER, S. (2010). The convergence to equilibrium of neutral genetic models. *Stoch. Anal. Appl.* **28** 123–143. MR2597983 https://doi.org/10.1080/07362990903415833

DOUC, R., CAPPÉ, O. and MOULINES, E. (2005). Comparison of resampling schemes for particle filtering. In *Proceedings of the* 4*th International Symposium on Image and Signal Processing and Analysis* 64–69.

DOUCET, A. and JOHANSEN, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *The Oxford Handbook of Nonlinear Filtering* 656–704. Oxford Univ. Press, Oxford. MR2884612

GORDON, N. J., SALMOND, S. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140** 107–113.

JACOB, P. E., MURRAY, L. M. and RUBENTHALER, S. (2015). Path storage in the particle filter. *Stat. Comput.* **25** 487–496. MR3306720 https://doi.org/10.1007/s11222-013-9445-x

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034 https://doi.org/10.1016/0304-4149(82)90011-4

KONG, A., LIU, J. S. and WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89** 278–288.

KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077 https://doi.org/10.1214/009053605000000426

LEE, A. and WHITELEY, N. (2018). Variance estimation in the particle filter. *Biometrika* **105** 609–625. MR3842888 https://doi.org/10.1093/biomet/asy028

LIU, J. S. and CHEN, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.* **90** 567–576. MR3363399 https://doi.org/10.1080/01621459.1995.10476549

MÖHLE, M. (1998). Robustness results for the coalescent. *J. Appl. Probab.* **35** 438–447. MR1641829 https://doi.org/10.1239/jap/1032192859

MÖHLE, M. (1999). Weak convergence to the coalescent in neutral population models. *J. Appl. Probab.* **36** 446–460. MR1724816 https://doi.org/10.1017/s0021900200017241

MOSIMANN, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* **49** 65–82. MR0143299 https://doi.org/10.1093/biomet/49.1-2.65

OLSSON, J. and DOUC, R. (2019). Numerically stable online estimation of variance in particle filters. *Bernoulli* **25** 1504–1535. MR3920380 https://doi.org/10.3150/18-bej1028

WAKELEY, J. (2009). *Coalescent Theory*: *An Introduction*. Roberts & Co., Greenwood Village, CO.