J. Appl. Prob. **40,** 839–854 (2003)
Printed in Israel
© *Applied Probability Trust* 2003

CONVERGENCE TO THE COALESCENT WITH SIMULTANEOUS MULTIPLE MERGERS

SERIK SAGITOV,* *Chalmers University of Technology*

Abstract

The general coalescent process with simultaneous multiple mergers of ancestral lines was initially characterized by Möhle and Sagitov (2001) in terms of a sequence of measures defined on the finite-dimensional simplices. A more compact characterization of the general coalescent requiring a single probability measure Ξ defined on the infinite simplex Δ was suggested by Schweinsberg (2000). This paper presents a simple criterion of weak convergence to the Ξ-coalescent. In contrast to the earlier criterion of Möhle and Sagitov (2001) based on the moment conditions, the key condition here is expressed in terms of the joint distribution of the ranked offspring sizes. This criterion interprets a vector in Δ as the ranked fractions of the total population size assigned to sibling groups constituting a (rare) generation, where a merger might occur. An example of the general coalescent is developed on the basis of the Poisson–Dirichlet distribution. It suggests a simple algorithm of simulating the Kingman coalescent with occasional (simultaneous) multiple mergers.

Keywords: Ξ-coalescent; exchangeability; Poisson–Dirichlet distribution

AMS 2000 Subject Classification: Primary 92D25; 92D15; 60F17

1. Introduction

The well-known Kingman coalescent [7]–[9] is a continuous-time Markov chain $\{R_0(t)\}_{t\geq 0}$ describing a robust asymptotic pattern for the genealogical tree of a large population. The genealogical tree is built bottom up: starting with n leaves corresponding to a random sample of genes in the current generation and ending with the root of the tree corresponding to the most recent common ancestor of the sampled genes. The hold-and-jump picture of the Markov chain $\{R_0(t)\}_{t\geq 0}$ is very simple: if n is the current number of branches, wait an exponentially distributed time with mean $\binom{n}{2}$, and then reduce the number of branches to $n-1$ by joining a pair of branches picked uniformly at random from among the $\binom{n}{2}$ available pairs.

The first robustness result [7] concerning the Kingman coalescent is stated in the framework of the Cannings exchangeable population model [3] describing a haploid population of constant population size N with nonoverlapping generations. The assumptions of the Cannings model concern the vector of offspring sizes $(\nu_1, \ldots, \nu_N)$ for N individuals constituting one generation. The constant population size condition

$$\nu_1 + \cdots + \nu_N = N \tag{1.1}$$

requires that the total offspring number is equal to the number of parents. It is assumed also that there is no dependence between vectors $(\nu_1, \ldots, \nu_N)$ for different generations and that

Received 5 March 2003; revision received 24 April 2003.
* Postal address: School of Mathematical Sciences, Chalmers University of Technology, S-412 96 Göteborg, Sweden. Email address: serik@math.chalmers.se

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

all generations share the same reproduction law, that is, the joint distribution of $(\nu_1, \dots, \nu_N)$. Finally, the Cannings model assumes that

$$(\nu_1, \dots, \nu_N) \text{ has an exchangeable joint distribution.} \tag{1.2}$$

Note that under these assumptions all the offspring sizes ν_i within one generation are equally distributed and $\mathrm{E}(\nu_1) = 1$.

The genealogy of a population with nonoverlapping generations is defined by Kingman as a Markov chain $\{\mathcal{R}(k)\}_{k=0,1\dots}$ with a finite number of states which are different partitions of the set of labels $\{1, \dots, n\}$ representing n individuals sampled in generation $k = 0$. Theorem 1 in [7] states that the finite-dimensional distributions of $\{\mathcal{R}(\lfloor Nt \rfloor)\}_{t\geq 0}$ converge to those of $\{R_0(\sigma^2 t)\}_{t\geq 0}$ given that

$$\sigma_N^2 := \mathrm{var}(\nu_1) \to \sigma^2 \quad \text{as } N \to \infty, \text{ where } 0 < \sigma^2 < \infty \tag{1.3}$$

and $\sup_N \mathrm{E}(\nu_1^k) < \infty$ for $k \geq 3$. These conditions cover the famous Wright–Fisher population model, but fail to encompass the important Moran population model addressed by Theorem 3 in [7].

The cited convergence results by Kingman suggest the following general formula for the time scaling in the ancestral process:

$$T_N \sim \frac{N}{\sigma_N^2} \quad \text{as } N \to \infty. \tag{1.4}$$

This relation has a simple explanation in terms of the probability that a randomly chosen pair of individuals have the same parent:

$$c_N = \mathrm{E}\left(\frac{1}{N^{(2)}} \sum_{i=1}^{N} \nu_i^{(2)} \right) = \frac{\sigma_N^2}{N-1}. \tag{1.5}$$

Here and elsewhere, the notation $a^{(b)} = a(a-1)\cdots(a-b+1)$ is used for the descending factorial power. Since the Cannings model of nonoverlapping generations treats the reproduction process as a sequence of independent and identically distributed trials, the time until the most recent common ancestor for two individuals has a geometric distribution with mean $1/c_N \sim T_N$ as $N \to \infty$.

The minimal convergence conditions for the Kingman coalescent are established in [13] under the assumptions (1.1) and (1.2).

Proposition 1.1. (Möhle and Sagitov [13].) *Suppose that the conditions (1.1) and (1.2) hold. The time-scaled genealogical process $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ converges weakly to $\{R_0(t)\}_{t\geq 0}$ if and only if the time scale T_N is defined by (1.4),*

$$\sigma_N^2 = o(N) \quad \textit{as } N \to \infty, \tag{1.6}$$

and

$$\mathrm{E}((\nu_1 - 1)^3) = o(N\sigma_N^2) \quad \textit{as } N \to \infty. \tag{1.7}$$

From now on the time scale T_N is tacitly assumed to be defined by (1.4). In Section 4 we give an alternative proof of Proposition 1.1 as a corollary of the main result of this paper,

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Theorem 2.1. In particular, we show that under the conditions of Proposition 1.1 the asymptotic relation (1.7) is equivalent to

$$\mathrm{P}(\nu_1 > N\varepsilon) = o\left(\frac{\sigma_N^2}{N^2}\right) \quad \text{as } N \to \infty \text{ for } \varepsilon > 0. \tag{1.8}$$

The condition (1.6) stipulates that mergers of ancestor lines are rare events, which results in a continuous-time limit coalescent. Since (1.3) entails (1.6) and $T_N \sim N/\sigma^2$ as $N \to \infty$, Proposition 1.1 implies the following improvement of Theorem 1 in [7].

Corollary 1.1. *Under the assumptions (1.1), (1.2), and (1.3), the time-scaled genealogical process* $\{\mathcal{R}(\lfloor Nt \rfloor)\}_{t\geq 0}$ *weakly converges to* $\{R_0(t\sigma^2)\}_{t\geq 0}$ *if and only if* $\mathrm{E}(\nu_1^3) = o(N)$ *as* $N \to \infty$.

In contrast to the pairwise merger, a *multiple merger* is a coalescent event where three or more branches join at one point. The number of branches joining together will be called the *merger size*. The condition (1.7) effectively prohibits multiple mergers of ancestral lines in a large population, which is necessary for convergence to the Kingman coalescent process of pairwise mergers. A situation where the condition (1.7) does not hold was first considered in [15]. It was shown that in the case of increased offspring size variability a new coalescent pattern arises, which is described by a probability measure F on the unit interval $[0, 1]$. Given that (1.7) is not valid, the measure F has a positive mass away from zero and the coalescent with positive probability has multiple mergers (see Corollary 2.1). This coalescent model is called here an *asynchronous coalescent* $\{R_F(t)\}_{t\geq 0}$, since it prohibits simultaneous mergers of ancestral lines.

The measure F will be called the *coalescent rate measure*, since it determines the transition rates of the coalescent Markov chain $\{R_F(t)\}_{t\geq 0}$. The fact that F is a probability measure reflects the special choice (1.4) of the time scale T_N. Observe that a linear time acceleration by a factor λ transforms the coalescent rate measure into a finite measure $\Lambda(\mathrm{d}x_1) := \lambda F(\mathrm{d}x_1)$, so that $\{R_F(t/\lambda)\}_{t\geq 0} = \{R_\Lambda(t)\}_{t\geq 0}$. The Kingman coalescent (with unit coalescent rate) $\{R_0(t)\}_{t\geq 0}$ is a special asynchronous coalescent corresponding to $F = \delta_0$, the probability measure concentrated at zero.

The asynchronous Λ-coalescent was first introduced in [14] (the paper [15] is written later than though independently of [14]) in the spirit of the theory of exchangeable random partitions [9] using a Poisson process construction. Such a construction makes it possible to model a stochastic tree starting from infinitely many leaves. In [14] special attention is paid to the case of the uniform coalescent rate measure, first discussed in [2]. The important question of whether the asynchronous coalescent comes down from infinity (discussed earlier in [14] and [15]) is fully answered in [16]. An alternative way of introducing the asynchronous coalescent is suggested in Section 5.1 of [4].

The most general case of simultaneous multiple mergers was first addressed in [12], where the moment conditions of [15] concerning the joint distribution of offspring sizes are relaxed, allowing for stronger dependence between offspring sizes. The coalescent with simultaneous multiple mergers was further investigated in [17] by extending the Poisson process construction of [14]. It is shown that the general coalescent is characterized by a finite measure Ξ (which will also be called the coalescent rate measure) defined on the infinite simplex

$$\Delta = \left\{(x_1, x_2, \dots) : x_1 \geq x_2 \geq \cdots \geq 0,\ \sum_{i=1}^{\infty} x_i \leq 1\right\}.$$

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

In Section 2 we give necessary and sufficient conditions for convergence to the Ξ-coalescent with the coalescent rate measure Ξ being a probability measure. The main results of this paper are stated as Theorems 2.1 and 2.2. Theorem 2.1 is a general convergence result without the assumption (1.2) of exchangeability (cf. [11] dealing with the Kingman coalescent). It suggests that the components of a vector $(x_1, x_2, \dots) \in \Delta$ be interpreted as the population-size fractions assigned to the ranked offspring sizes in a rare generation, where a multiple merger has occurred. Theorem 2.2 is a counterpart of Theorem 2.1 concerning the discrete-time Ξ-coalescent. Theorem 2.1 is proved in Section 4, the proof of Theorem 2.2 is similar and is therefore omitted.

Section 3 is devoted to a simple example of an exchangeable population whose genealogical process is approximated by a Ξ-coalescent with Ξ being closely related to the Poisson–Dirichlet distribution. We call this particular coalescent process the Poisson–Dirichlet coalescent. The transition probabilities of the Poisson–Dirichlet coalescent are explicitly computed, and it is worth noticing that the Ewens distribution probabilities show up in the formula for the total collision rate. In Section 5 we return to the exchangeable case and state convergence results in terms of the raw moment conditions (in contrast to the central moment conditions of [15] and [12]).

2. Summary of convergence results

Consider a certain type of r simultaneous mergers transforming b ancestral lines to a smaller number $r + s$ of lines, where it is assumed that s lines remain unchanged and r new lines are formed by simultaneously coalescing groups of size $k_1, \dots, k_r \geq 2$, so that $b = k_1 + \cdots + k_r + s$. Following [17], we call this change in a tree a $(b; k_1, \dots, k_r; s)$-*collision*. Note that a permutation of indices $(k_1, \dots, k_r)$ does not change the collision type $(b; k_1, \dots, k_r; s)$, producing only another label for the same collision type. The full set of possible collision types is given by

$$K_b := \{(b; k_1, \dots, k_r; s) : k_1 + \cdots + k_r + s = b,\ k_1 \geq \cdots \geq k_r \geq 2,\ s \geq 0,\ r \geq 1\}.$$

Observe also that, given b distinct ancestral lines and a collision type $(b; k_1, \dots, k_r; s)$ with $r \geq 1$, there are several suitable partitions of the b lines into the $r + s$ clusters, and the number of possible $(b; k_1, \dots, k_r; s)$-collisions for

$$(k_1, \dots, k_r, s) = (\underbrace{b, \dots, b}_{l_b}, \underbrace{b-1, \dots, b-1}_{l_{b-1}}, \dots, \underbrace{2, \dots, 2}_{l_2}, l_1). \tag{2.1}$$

is

$$M_{b;k_1,\dots,k_r;s} = \binom{b}{k_1, \dots, k_r, s} \frac{1}{\prod_{j=2}^{b} l_j!}. \tag{2.2}$$

Let $P_{b;k_1,\dots,k_r;s}$ be the probability that a certain collision of type $(b; k_1, \dots, k_r; s)$ occurs, when b ancestral lines are traced one generation back. According to [7], given an exchangeable reproduction law $(\nu_1, \dots, \nu_N)$, this probability is

$$P_{b;k_1,\dots,k_r;s} = \frac{N^{(r+s)}}{N^{(b)}} \mathrm{E}(\nu_1^{(k_1)} \cdots \nu_r^{(k_r)} \nu_{r+1} \cdots \nu_{r+s}). \tag{2.3}$$

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Proposition 2.1. (Möhle and Sagitov [12].) *If (1.1), (1.2), and (1.6) hold, then the time-scaled genealogical process* $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ *converges weakly to some limit coalescent process* $\{R(t)\}_{t\geq 0}$ *if and only if there exist*

$$A_{k_1,\dots,k_r} = \lim_{N\to\infty} N^{r-b} T_N \,\mathrm{E}((\nu_1-1)^{k_1}\cdots(\nu_r-1)^{k_r}) \tag{2.4}$$

for all $k_1,\dots,k_r \geq 2$ *and* $r \geq 1$, *where* $b := k_1+\cdots+k_r$. *The* $(b; k_1,\dots,k_r; s)$*-collision rates of the Markov chain* $\{R(t)\}_{t\geq 0}$ *for* $s=0$ *are* $\lambda_{b;k_1,\dots,k_r;0} = A_{k_1,\dots,k_r}$ *and for* $s\geq 1$ *can be computed from the recursion*

$$\lambda_{b+1;k_1,\dots,k_r;s+1} = \lambda_{b;k_1,\dots,k_r;s} - \sum_{m=1}^{r} \lambda_{b+1;k_1,\dots,k_{m-1},k_m+1,k_{m+1},\dots,k_r;s} - s\lambda_{b+1;k_1,\dots,k_r,2;s-1}, \qquad k_1,\dots,k_r \geq 2,\ r\geq 1,\ s\geq 0. \tag{2.5}$$

The set of rates $\{\lambda_{b;k_1,\dots,k_r;s}\}$ provides the following hold-and-jump description of the general coalescent $\{R(t)\}_{t\geq 0}$. It holds on a state with b ancestral lines for an exponential time with mean $\lambda_b = \sum_{K_b} \lambda_{b;k_1,\dots,k_r;s} M_{b;k_1,\dots,k_r;s}$, where $M_{b;k_1,\dots,k_r;s}$ is given by (2.2). When the holding time is over, a collision of type $(b; k_1,\dots,k_r; s)$ (one among $M_{b;k_1,\dots,k_r;s}$ possible) occurs with probability $(1/\lambda_b)\lambda_{b;k_1,\dots,k_r;s}$.

It follows from [12] that, if $A_{k_1,\dots,k_r} = 0$ for some $k_1,\dots,k_r \geq 2$, then automatically $A_{k'_1,\dots,k'_{r'}} = 0$ for all $r'\geq r$ and $k'_1\geq k_1,\dots,k'_r\geq k_r, k'_{r+1}\geq 2,\dots,k'_{r'}\geq 2$. In particular, Proposition 2.1 entails the following improved version of the convergence result of [15].

Corollary 2.1. (Cf. [15].) *Suppose that the conditions (1.1), (1.2), and (1.6) hold while (1.7) does not. If*

$$\mathrm{E}((\nu_1-1)^2(\nu_2-1)^2) = o(N\sigma_N^2) \quad \text{as } N\to\infty, \tag{2.6}$$

then the time-scaled genealogical process $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ *converges weakly if and only if*

$$\mathrm{E}((\nu_1-1)^k) \sim A_k N^{k-2}\sigma_N^2 \quad \text{as } N\to\infty \text{ for } k\geq 3,$$

where $1\geq A_3\geq A_4\geq\cdots>0$. *In this case, the limit process is an asynchronous coalescent* $\{R_F(t)\}_{t\geq 0}$, *whose collision rates are specified in terms of a probability measure* F. *The coalescent rate measure* F *is uniquely defined on the unit interval* $[0,1]$ *via its moments* $\int_0^1 x^k F(\mathrm{d}x) = A_{k+2}$ *for* $k\geq 0$, *where* $A_2 := 1$. *The collision rates of* $\{R_F(t)\}_{t\geq 0}$ *satisfy the conditions*

$$\lambda_{b;k_1,\dots,k_r;s} = \mathbf{1}_{\{r=1\}} \int_0^1 x^{k_1-2}(1-x)^s F(\mathrm{d}x), \qquad r\geq 1,\ k_1,\dots k_r\geq 2,\ s\geq 0, \tag{2.7}$$

and

$$\lambda_b = \int_0^1 \frac{1-(1-x)^b - bx(1-x)^{b-1}}{x^2} F(\mathrm{d}x), \qquad b\geq 2.$$

This result excludes convergence to the Kingman coalescent and, by the condition (2.6), prohibits simultaneous mergers of ancestral lines. The asynchronous coalescent $\{R_F(t)\}_{t\geq 0}$ is characterized by positive tree collapse rates $A_k = \lambda_{k;k;0}$ for $k\geq 3$ (a tree collapse is a single merger involving all branches of the tree). It follows that the limit coalescent cannot have bounded merger sizes k_i: if $A_3>0$, then $A_k>0$ for all $k\geq 3$.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

According to [17] there is a unique probability measure Ξ defined on the simplex Δ determining the collision rates of the limit coalescent $\{R(t)\}_{t\geq 0}$ in Proposition 2.1 by

$$\lambda_{b;k_1,\dots,k_r;s} = a\,\mathbf{1}_{\{r=1,k_1=2\}} + \int_\Delta \left(\sum_{j=0}^{s} \sum_{i_1\neq\cdots\neq i_{r+j}} \binom{s}{j} x_{i_1}^{k_1}\cdots x_{i_r}^{k_r} x_{i_{r+1}}\cdots x_{i_{r+j}} x_0^{s-j} \right) \frac{\Xi_0(\mathrm{d}x)}{(x,x)}, \tag{2.8}$$

where $(x,x) := \sum_{i=1}^{\infty} x_i^2$, $x_0 := 1 - \sum_{i=1}^{\infty} x_i$, the nonnegative constant a is the mass assigned by Ξ to the zero point $(0,0,\dots)$ of the infinite simplex Δ, and Ξ_0 is Ξ with the atom at zero removed. Notice that in the last integral the integrand has no singularity at zero since

$$\begin{aligned} 0 &\leq \left(\sum_{j=0}^{s} \sum_{i_1\neq\cdots\neq i_{r+j}} \binom{s}{j} x_{i_1}^{k_1}\cdots x_{i_r}^{k_r} x_{i_{r+1}}\cdots x_{i_{r+j}} x_0^{s-j} \right) \frac{1}{(x,x)} \\ &\leq \left(\sum_{i_1=1}^{\infty} x_{i_1}^2 \sum_{j=0}^{s} \binom{s}{j} x_0^{s-j} \sum_{i_{r+1}=1}^{\infty} \cdots \sum_{i_{r+j}=1}^{\infty} x_{i_{r+1}}\cdots x_{i_{r+j}} \right) \frac{1}{(x,x)} = 1 \end{aligned}$$

for any nonzero $x \in \Delta$ given $r \geq 1$.

The set of rates

$$\lambda_{2;2;0} = 1, \qquad \lambda_{b;k_1,\dots,k_r;0} = \int_\Delta \left(\sum_{i_1\neq\cdots\neq i_r} x_{i_1}^{k_1}\cdots x_{i_r}^{k_r} \right) \frac{\Xi_0(\mathrm{d}x)}{(x,x)}$$

for all $k_1 \geq 3$ when $r = 1$ and for all $k_1,\dots,k_r \geq 2$ when $r \geq 2$, uniquely determines the coalescent rate measure Ξ. If Ξ is concentrated on the set $\{(x_1,0,0,\dots) : x_1 \in [0,1]\}$, then it is essentially a probability measure F on $[0,1]$, and the collision rates of the coalescent process $\{R_F(t)\}_{t\geq 0}$ obtained in Corollary 2.1 can be deduced from (2.8).

So far we have assumed that the joint distribution of the offspring sizes is exchangeable. This excludes for example the so-called generalized Wright–Fisher model [11] based on the not necessarily symmetric multinomial $\mathrm{Mn}(N; p_1,\dots,p_N)$ distribution for the vector of offspring sizes $(\nu_1,\dots,\nu_N)$. Next we present the main result of this paper, Theorem 2.1, which states necessary and sufficient conditions for weak convergence to the Ξ-coalescent in terms of the ordered offspring sizes $\nu_{(1)} \geq \cdots \geq \nu_{(N)}$ without the exchangeability assumption (1.2).

The proper time scale T_N in this more general case might again be defined by (1.4), with σ_N^2 now being the variance of the offspring size ν of a randomly chosen individual. This follows from the following counterpart of (1.5): since $\mathrm{E}(\nu) = 1$, the coalescence probability c_N in the nonexchangeable case is

$$c_N = \mathrm{E}\left(\frac{1}{N^{(2)}} \sum_{i=1}^{N} \nu_i^{(2)} \right) = \frac{1}{N-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathrm{E}(\nu_i^2) - 1 \right) = \frac{\mathrm{var}(\nu)}{N-1}.$$

Theorem 2.1. *Let $\Phi_N(\mathrm{d}x)$ be the joint distribution of the vector $(\nu_{(1)}/N,\dots,\nu_{(N)}/N)$ of the ranked relative frequencies for sizes of N sibling groups constituting a generation. If (1.1) and (1.6) hold, then the time-scaled genealogical process $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ converges weakly to the Ξ-coalescent if and only if the weak convergence condition*

$$T_N \Phi_N(\mathrm{d}x) \to \frac{\Xi(\mathrm{d}x)}{(x,x)} \quad \text{as } N \to \infty \tag{2.9}$$

holds on $\Delta \cap \{x_1 > \varepsilon\}$ for all $\varepsilon > 0$, where Ξ is some probability measure on Δ.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Following Section 4 of [1], we treat Δ as a subset of the metric space $\mathbb{R}^\infty$ with the topology of coordinatewise convergence. The weak convergence (2.9) holds if and only if the weak convergence of the tail distribution functions

$$T_N \, \mathrm{P}(\nu_{(1)} > Nx_1, \ldots, \nu_{(r)} > Nx_r) \to \int_\Delta \mathbf{1}_{\{y_1 > x_1, \ldots, y_r > x_r\}} \frac{\Xi(\mathrm{d}y)}{(y, y)} \quad \text{as } N \to \infty \tag{2.10}$$

holds over $x_1 \geq \cdots \geq x_r > 0$ for all $r \geq 1$.

The next two corollaries of Theorem 2.1 address the cases of the asynchronous coalescent and the Kingman coalescent.

Corollary 2.2. *Under the assumptions (1.1) and (1.6), the process* $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t \geq 0}$ *converges weakly to the asynchronous* F*-coalescent if and only if*

$$T_N \, \mathrm{P}(\nu_{(1)} > Nx) \to \int_x^1 y^{-2} F(\mathrm{d}y) \quad \textit{as } N \to \infty \tag{2.11}$$

for all points $x > 0$ *of continuity of a probability measure* F *defined on the unit interval* $[0, 1]$ *and*

$$T_N \, \mathrm{P}(\nu_{(2)} > N\varepsilon) \to 0 \quad \textit{as } N \to \infty \textit{ for } \varepsilon > 0. \tag{2.12}$$

The condition (2.12) excludes simultaneous mergers by prohibiting two large sibling groups (of size no less than $N\varepsilon$) in the same generation among T_N consecutive generations. The convergence (2.11) illuminates the role of the largest offspring size in determining the merger size distribution (2.7). The result (2.7) implies in turn that, if F is concentrated at zero, then only pairwise mergers are possible.

Corollary 2.3. *Suppose that (1.1) and (1.6) hold. The condition that*

$$T_N \, \mathrm{P}(\nu_{(1)} > N\varepsilon) \to 0 \quad \textit{as } N \to \infty \textit{ for } \varepsilon > 0 \tag{2.13}$$

is necessary and sufficient for the weak convergence of $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t \geq 0}$ *to the Kingman coalescent.*

If instead of (1.6) it is assumed that

$$\sigma_N^2 \sim Nc \quad \text{as } N \to \infty \text{ for } 0 < c \leq 1 \tag{2.14}$$

(the constant c cannot exceed 1 since $\sigma_N^2 \leq N - 1$), then there might exist a discrete-time limit for the unscaled genealogical process. The condition (2.14) means that, with probability c, a pair of randomly chosen individuals in a large population are siblings.

Proposition 2.2. (Möhle and Sagitov [12].) *If (1.1), (1.2), and (2.14) hold, then the discrete-time genealogical process* $\{\mathcal{R}(k)\}_{k \geq 0}$ *converges weakly to some discrete-time Markov chain* $\{R(k)\}_{k \geq 0}$ *if and only if there exist*

$$A_{k_1, \ldots, k_r} = \lim_{N \to \infty} N^{r-b} \, \mathrm{E}((\nu_1 - 1)^{k_1} \cdots (\nu_r - 1)^{k_r})$$

for all $k_1, \ldots, k_r \geq 2$ *and* $r \geq 1$*, where* $b = k_1 + \cdots + k_r$.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

The one-step transition probabilities of the discrete-time coalescent $\{R(k)\}_{k\geq 0}$ *corresponding to* $(b; k_1, \dots, k_r; s)$*-collisions are computed from the recursion*

$$p_{b+1;k_1,\dots,k_r;s+1} = p_{b;k_1,\dots,k_r;s} - \sum_{m=1}^{r} p_{b+1;k_1,\dots,k_{m-1},k_m+1,k_{m+1},\dots,k_r;s} - s p_{b+1;k_1,\dots,k_r,2;s-1}, \qquad k_1, \dots, k_r \geq 2,\ r \geq 1,\ s \geq 0,$$

where $p_{b;k_1,\dots,k_r;0} = A_{k_1,\dots,k_r}$.

A characterization of the discrete-time coalescent in terms of a single measure defined on the infinite simplex Δ is proposed in [17]. Let Φ be an arbitrary probability measure on the simplex Δ. The Markov chain with transition probabilities

$$p_{b;k_1,\dots,k_r;s} = \int_{\Delta} \left(\sum_{j=0}^{s} \sum_{i_1 \neq \dots \neq i_{r+j}} \binom{s}{j} x_{i_1}^{k_1} \cdots x_{i_r}^{k_r} x_{i_{r+1}} \cdots x_{i_{r+j}} x_0^{s-j} \right) \Phi(\mathrm{d}x)$$

for $r \geq 1$, $k_1 \geq 2, \dots, k_r \geq 2$, $s \geq 0$, and $b = k_1 + \cdots + k_r + s$ is called here a *discrete-time* Φ*-coalescent*. What follows is a discrete-time counterpart of Theorem 2.1.

Theorem 2.2. *If (1.1) and (2.14) hold, then the discrete-time genealogical process* $\{\mathcal{R}(k)\}_{k\geq 0}$ *converges weakly to the discrete-time* Φ*-coalescent if and only if* Φ *is the weak limit distribution of the vector of ordered offspring proportions* $(\nu_{(1)}/N, \dots, \nu_{(N)}/N)$.

Remark 2.1. The definition of the discrete-time Φ-coalescent given above slightly modifies Definition 36 of [17]. The limit process of Theorem 2.2 is the discrete-time Ξ-coalescent in the strict sense of Definition 36 from [17] if the limit distribution of $(\nu_{(1)}/N, \dots, \nu_{(N)}/N)$ is of the form

$$\Phi(\mathrm{d}x) = \frac{\Xi(\mathrm{d}x)}{(x,x)} \mathbf{1}_{\{x_1>0\}} + a_0 \delta_0(\mathrm{d}x)\, \mathbf{1}_{\{x_1=0\}},$$

where δ_0 is the Dirac measure on Δ concentrated at zero and a_0 is a nonnegative constant such that $\Phi(\Delta) = 1$.

3. Poisson–Dirichlet coalescent

The best known example of a nontrivial probability measure on an infinite simplex is the Poisson–Dirichlet distribution $\Pi_\theta(\mathrm{d}x)$ with a positive parameter θ defined on the infinite simplex $\Delta^* := \{x \in \Delta : \sum_{i=1}^{\infty} x_i = 1\}$. The calculations of Section 9.5 of [10] show that

$$\int_{\Delta^*} \left(\sum_{i_1 \neq \dots \neq i_{r+s}} x_{i_1}^{k_1} \cdots x_{i_r}^{k_r} x_{i_{r+1}} \cdots x_{i_{r+s}} \right) \Pi_\theta(\mathrm{d}x) = \frac{\theta^{r+s}}{\theta^{[b]}} \prod_{i=1}^{r} (k_i - 1)!$$

for all $r \geq 1$, $k_1, \dots, k_r \geq 2$, and $s \geq 0$, where as usual $b = k_1 + \cdots + k_r + s$ and $\theta^{[b]} := \theta(\theta+1)\cdots(\theta+b-1)$ is the ascending factorial power. This formula suggests a natural example of the discrete-time Π_θ-coalescent with explicit transition probabilities

$$p_{b;k_1,\dots,k_r;s} = \frac{\theta^{r+s}}{\theta^{[b]}} \prod_{i=1}^{r} (k_i - 1)!. \tag{3.1}$$

This will be called the Poisson–Dirichlet coalescent with parameter θ.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Using the representation (2.1), put $p(l_1, \dots, l_b) := p_{b;k_1,\dots,k_r;s}$ and $l := l_1 + \cdots + l_b$. To compute the total probability p_b of all mergers in the Poisson–Dirichlet coalescent with b ancestral lines, rewrite (3.1) as

$$p(l_1, \dots, l_b) = \frac{\theta^l}{\theta^{[b]}} \prod_{j=1}^{b} ((j-1)!)^{l_j}.$$

Observe that, due to (2.2),

$$p_b = \sum_{l_1,\dots,l_b} p(l_1, \dots, l_b) \frac{b!}{l_1! \cdots l_b!\, (2!)^{l_2} \cdots (b!)^{l_b}} = \sum_{l_1,\dots,l_b} \frac{\theta^l}{\theta^{[b]}} \frac{b!}{l_1! \cdots l_b!\, 2^{l_2} \cdots b^{l_b}},$$

where the sum is taken over all nonnegative integers $l_1, \dots, l_b$ satisfying $\sum_{j=1}^{b} j l_j = b$ and $l_1 < b$. Since the last sum contains all but one (that with $l_1 = b$) of the probabilities constituting the Ewens distribution (see [6] or [10]), we obtain $p_b = 1 - \theta^b/\theta^{[b]}$. This simple formula clearly shows how the coalescent slows down as the number of branches in a tree decreases. It reveals also the influence of the parameter θ on the overall speed of the coalescent: larger value of θ brings lower speed. Notice that $1 - \theta^b/\theta^{[b]} \sim \theta^{-1}\binom{b}{2}$ as $\theta \to \infty$, implying the Kingman coalescent approximation for the Poisson–Dirichlet coalescent.

We further enhance the example of the Poisson–Dirichlet coalescent by deducing it as the limit process for an exchangeable reproduction model described by a compound multinomial distribution. Recall first that the Wright–Fisher population model uses the symmetric multinomial $\mathrm{Mn}(N; p_1, \dots, p_N)$ with $p_1 = \cdots = p_N = 1/N$ as the joint distribution of offspring sizes $(\nu_1, \dots, \nu_N)$. To extend this model, we assume that the multinomial parameters $(p_1, \dots, p_N)$ are random and have a symmetric Dirichlet distribution $\mathrm{D}(\alpha, \dots, \alpha)$, so that the resulting distribution of $(\nu_1, \dots, \nu_N)$ is the so-called Dirichlet compound multinomial (cf. [6]):

$$\mathrm{P}(\nu_1 = n_1, \dots, \nu_N = n_N) = \binom{N}{n_1, \dots, n_k} \frac{\alpha^{[n_1]} \cdots \alpha^{[n_N]}}{(N\alpha)^{[N]}}.$$

This model is particularly convenient for the coalescent calculations in view of the following formula for the joint moments:

$$\mathrm{E}(\nu_1^{(k_1)} \cdots \nu_r^{(k_r)} \nu_{r+1} \cdots \nu_{r+s}) = \frac{N^{(b)}}{(N\alpha)^{[b]}} \alpha^{[k_1]} \cdots \alpha^{[k_r]} \alpha^s. \tag{3.2}$$

This formula implies that $\mathrm{E}(\nu_1^{(n)}) = N^{(n)}\alpha^{[n]}/(N\alpha)^{[n]}$, so that $\sigma_N^2 = (N-1)(\alpha+1)/N\alpha + 1$: the smaller α is, the more variable the offspring sizes become, and the larger α is, the closer the model is to the Wright–Fisher model in which $\sigma_N^2 = (N-1)/N$. If α does not depend on N, then the asymptotic genealogy of the Dirichlet compound Wright–Fisher model is the Kingman coalescent. If $\alpha = \theta/N$, it is easy to verify that (2.3) and (3.2) imply convergence to the Poisson–Dirichlet coalescent. This derivation of the Poisson–Dirichlet coalescent illuminates the meaning of the parameter θ: the smaller θ is, the more variable the offspring sizes are and therefore the faster coalescing proceeds.

It is a less straightforward exercise to obtain a coalescent with continuous time from the Dirichlet compound Wright–Fisher model. It requires a randomization of the parameter α in

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

such a way that most of the time α takes a value A which does not depend on N, and with a small probability α is inversely proportional to N:

$$\mathrm{P}(\alpha = A) = 1 - \frac{p}{N}, \qquad \mathrm{P}\left(\alpha = \frac{\theta}{N}\right) = \frac{p}{N}. \tag{3.3}$$

In this case, the condition (1.3) holds with the limit variance $\sigma^2 = (A+1)/A + p/(1+\theta)$. Moreover, it follows that the genealogical process $\{\mathcal{R}(\lfloor Nt \rfloor)\}_{t\geq 0}$ converges weakly to the Ξ-coalescent, where $\Xi := \Xi(A, p, \theta)$ has an atom at zero of size $a := (A+1)/A$, which is the first term in the formula for σ^2, and away from zero the measure $\Xi(A, p, \theta)$ is closely related to the Poisson–Dirichlet distribution.

Observe first that (3.2) and (3.3) imply that

$$\mathrm{E}(\nu_1^{(k_1)} \cdots \nu_r^{(k_r)} \nu_{r+1} \cdots \nu_{r+s}) \sim pN^{b-r-1} \frac{\theta^{r+s}}{\theta^{[b]}} \prod_{i=1}^{r} (k_i - 1)! + a\, \mathbf{1}_{\{r=1, k_1=2\}},$$

which according to Proposition 2.1 and (2.3) entails that $\{\mathcal{R}(\lfloor Nt \rfloor)\}_{t\geq 0}$ converges to a coalescent with the collision rates

$$\lambda_{b;k_1,\ldots,k_r;s} = p \frac{\theta^{r+s}}{\theta^{[b]}} \prod_{i=1}^{r} (k_i - 1)! + a\, \mathbf{1}_{\{r=1, k_1=2\}}.$$

The total rate of all mergers for b ancestral lines in the limit coalescent is

$$\lambda_b = a\binom{b}{2} + p\left(1 - \frac{\theta^b}{\theta^{[b]}}\right).$$

Notice that when the parameter p is relatively small, we have a model of the Kingman coalescent of asynchronous pairwise mergers with occasional (simultaneous) multiple mergers.

4. Proof of Theorem 2.1

Given a probability measure Ξ on the infinite simplex Δ and a positive integer r, we can define a symmetric measure on the r-dimensional simplex

$$\Delta_r := \{(x_1, \ldots, x_r) : x_1, \ldots, x_r \geq 0,\ x_1 + \cdots + x_r \leq 1\}$$

by

$$\Psi_r(S) := \int_\Delta \left(\sum_{i_1 \neq \cdots \neq i_r} \mathbf{1}_{\{(y_{i_1}, \ldots, y_{i_r}) \in S\}} \right) \frac{\Xi(\mathrm{d}y)}{(y, y)} \tag{4.1}$$

for all measurable S. These measures are infinite: $\Psi_r(\Delta_r \cap \{x_1, \ldots, x_{r-1} > \varepsilon,\ x_r > 0\}) = \infty$, but on the other hand, for all $\varepsilon > 0$ and $r \geq 1$,

$$\Psi_r(\Delta_{r,\varepsilon}) \leq \int_\Delta \left(\sum_{i_1 \neq \cdots \neq i_r} \left(\frac{y_{i_1}}{\varepsilon}\right)^2 \cdots \left(\frac{y_{i_r}}{\varepsilon}\right)^2 \mathbf{1}_{\{y_{i_1}, \ldots, y_{i_r} > \varepsilon\}} \right) \frac{\Xi(\mathrm{d}y)}{(y, y)} \leq \varepsilon^{-2r},$$

where $\Delta_{r,\varepsilon} := \Delta_r \cap \{x_1, \ldots, x_r > \varepsilon\}$. It turns out that the probability measure Ξ is uniquely determined by the sequence of measures $\{\Psi_r\}_{r\geq 1}$.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Lemma 4.1. *Assume that (1.1), (1.2), and (1.6) hold and let $\Psi_{r,N}$ denote the symmetric measure on Δ_r giving the joint distribution of the vector $(\nu_1/N, \dots, \nu_r/N)$ of normalized offspring sizes. The time-scaled genealogical process $(\mathcal{R}_{\lfloor T_N t \rfloor})_{t\geq 0}$ converges weakly to the Ξ-coalescent if and only if the weak convergence*

$$N^r T_N \Psi_{r,N} \to \Psi_r \quad \text{as } N \to \infty \text{ for } r \geq 1 \tag{4.2}$$

holds over $\Delta_{r,\varepsilon}$ for any $\varepsilon > 0$, where Ψ_r is connected to Ξ via (4.1).

Proof. The condition (4.2) is similar to the following weak convergence condition given in [12]. It is shown there that, under (1.1), (1.2), and (1.6), the weak convergence of $(\mathcal{R}_{\lfloor T_N t \rfloor})_{t\geq 0}$ takes place if and only if, for all $r \geq 1$, there exists a symmetric measure F_r defined on Δ_r such that the weak convergence

$$N^r T_N \Psi_{r,N}(\mathrm{d}x_1, \dots, \mathrm{d}x_r) \to \frac{F_r(\mathrm{d}x_1, \dots, \mathrm{d}x_r)}{x_1^2 \cdots x_r^2} \quad \text{as } N \to \infty \tag{4.3}$$

holds for $(x_1, \dots, x_r) \in \Delta_{r,\varepsilon}$ and any given $\varepsilon > 0$, and moreover

$$N^{-r} T_N \,\mathrm{E}((\nu_1 - 1)^2 \cdots (\nu_r - 1)^2) \to F_r(\Delta_r) \quad \text{as } N \to \infty. \tag{4.4}$$

Furthermore, it is known that the total masses of F_r form a monotone sequence $F_1(\Delta_1) \geq F_2(\Delta_2) \geq \cdots$. Notice that, for $r = 1$, the convergence (4.4) follows automatically from (1.4), the definition of the time scale T_N. Therefore, our task reduces to verifying that the limits in (4.2) and (4.3) coincide, and that (4.4) for $r \geq 2$ follows from (1.1), (1.2), and (1.6).

The equality

$$F_r(\mathrm{d}x_1, \dots, \mathrm{d}x_r) = x_1^2 \cdots x_r^2 \Psi_r(\mathrm{d}x_1, \dots, \mathrm{d}x_r), \qquad (x_1, \dots, x_r) \in \Delta_{r,\varepsilon},\ \varepsilon > 0$$

follows immediately from the representation

$$F_r(S) = \int_\Delta \sum_{i_1 \neq \cdots \neq i_r} y_{i_1}^2 \cdots y_{i_r}^2 \, \mathbf{1}_{\{(y_{i_1}, \dots y_{i_r}) \in S\}} \frac{\Xi_0(\mathrm{d}y)}{(y,\, y)} + a\, \mathbf{1}_{\{r=1, (0,\dots,0) \in S\}},$$

which according to Proposition 11 in [17] is valid for any subset $S \subset \Delta_r$. The same representation implies that, for any $\varepsilon > 0$ and $r \geq 2$,

$$F_r(\Delta_r \cap \{x_r \leq \varepsilon\}) \leq \varepsilon^2 \int_\Delta \sum_{i_1=1}^{\infty} y_{i_1}^2 \frac{\Xi(\mathrm{d}y)}{(y,\, y)} = \varepsilon^2,$$

which together with (4.3) entails

$$\lim_{\varepsilon \to 0+} \limsup_{N \to \infty} |N^{-r} T_N \,\mathrm{E}((\nu_1 - 1)^2 \cdots (\nu_r - 1)^2 \, \mathbf{1}_{\{\nu_1, \dots, \nu_r > N\varepsilon\}}) - F_r(\Delta_r)| = 0.$$

To deduce (4.4) from the last convergence, it suffices to prove that

$$N^{-r} T_N \,\mathrm{E}((\nu_1 - 1)^2 \cdots (\nu_r - 1)^2 (1 - \mathbf{1}_{\{\nu_1, \dots, \nu_r > N\varepsilon\}})) \to 0$$

as first $N \to \infty$ and then $\varepsilon \to 0$. Therefore, due to the upper bound $1 - \mathbf{1}_{\{\nu_1 > N\varepsilon, \dots, \nu_r > N\varepsilon\}} \leq \sum_{i=1}^r \mathbf{1}_{\{\nu_i \leq N\varepsilon\}}$, it is enough to show that

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} N^{-r} T_N \,\mathrm{E}((\nu_1 - 1)^2 \cdots (\nu_r - 1)^2 \, \mathbf{1}_{\{\nu_r \leq N\varepsilon\}}) = 0. \tag{4.5}$$

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Observe that $\sum_{i=2}^{N} \nu_i^2 \mathbf{1}_{\{\nu_i \le N\varepsilon\}} \le N\varepsilon \sum_{i=2}^{N} \nu_i \le N^2\varepsilon$ (where the second inequality follows from (1.1)) implies that

$$\begin{aligned}(N-1)\,\mathrm{E}((\nu_1-1)^2(\nu_2-1)^2\,\mathbf{1}_{\{\nu_2\le N\varepsilon\}}) &\le (N-1)\,\mathrm{E}((\nu_1-1)^2\nu_2^2\,\mathbf{1}_{\{\nu_2\le N\varepsilon\}}) + (N-1)\sigma_N^2\\ &= \sum_{i=2}^{N}\mathrm{E}((\nu_1-1)^2\nu_i^2\,\mathbf{1}_{\{\nu_i\le N\varepsilon\}}) + (N-1)^2 c_N\\ &\le N^3\varepsilon c_N + N^2 c_N\end{aligned}$$

(where the equality follows from (1.2)). Dividing the leftmost and the rightmost sides of this inequality by $N^3 c_N$, we obtain an upper bound confirming that (4.5) is valid for $r = 2$. Finally, the validity of (4.5) for $r \ge 3$ follows by induction over r due to the inequality

$$\begin{aligned}&(N-r+1)\,\mathrm{E}((\nu_1-1)^2\cdots(\nu_r-1)^2\,\mathbf{1}_{\{\nu_r\le N\varepsilon\}})\\ &\quad\le \sum_{i=r}^{N}\mathrm{E}((\nu_1-1)^2\cdots(\nu_{r-1}-1)^2\nu_i^2\,\mathbf{1}_{\{\nu_i\le N\varepsilon\}}) + N\,\mathrm{E}((\nu_1-1)^2\cdots(\nu_{r-1}-1)^2)\\ &\quad\le (N^2\varepsilon+N)\,\mathrm{E}((\nu_1-1)^2\cdots(\nu_{r-1}-1)^2).\end{aligned}$$

4.1. Proof of Theorem 2.1

We first prove this theorem under the exchangeable reproduction assumptions (1.1), (1.2), and (1.6).

To prove the 'if' part of Theorem 2.1 under these assumptions, fix an arbitrary ε with $0 < \varepsilon < 1$, and let $(x_1, \dots, x_r) \in \Delta_{r,\varepsilon}$. Due to the obvious equality

$$\sum_{i_1\ne\cdots\ne i_r} \mathbf{1}_{\{\nu_{i_1}>Nx_1,\dots,\nu_{i_r}>Nx_r\}} = \sum_{i_1\ne\cdots\ne i_r} \mathbf{1}_{\{\nu_{(i_1)}>Nx_1,\dots,\nu_{(i_r)}>Nx_r\}},$$

we have

$$N^{(r)}\,\mathrm{P}(\nu_1 > Nx_1, \dots, \nu_r > Nx_r) = \sum_{1\le i_1\ne\cdots\ne i_r\le i_\varepsilon} \mathrm{P}(\nu_{(i_1)} > Nx_1, \dots, \nu_{(i_r)} > Nx_r).$$

In the last sum there is no need to consider $\nu_{(i)}$ for i above $i_\varepsilon := \lfloor \varepsilon^{-1} \rfloor$, since $i\nu_{(i)} \le \nu_{(i)} + \cdots + \nu_{(1)} \le N$ and $\nu_{(i)} \le N\varepsilon$ for $i \ge \varepsilon^{-1}$. Thus, (2.9) implies (4.2) via (2.10) and according to Lemma 4.1 we have established the asserted convergence to the Ξ-coalescent.

We now prove the 'only if' part of Theorem 2.1 under the exchangeable reproduction assumptions. If the convergence to the Ξ-coalescent takes place, it follows from Lemma 4.1 that $N^r T_N\,\mathrm{P}(\nu_1 > Nx_1, \dots, \nu_r > Nx_r)$ converges weakly for $(x_1, \dots, x_r) \in \Delta_{r,\varepsilon}$ given an arbitrary ε with $0 < \varepsilon < 1$. We will use this convergence to show that the weak convergence

$$T_N\,\mathrm{P}(\nu_{(1)} > Nx_1, \dots, \nu_{(r)} > Nx_r) \to \int_\Delta \mathbf{1}_{\{y_1\ge x_1,\dots,y_r\ge x_r\}}\,\Phi(\mathrm{d}y) \quad \text{as } N\to\infty \tag{4.6}$$

holds over $x_1 \ge \cdots \ge x_r \ge \varepsilon$ for all $r \ge 1$.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

For the purpose of illustration we start with the simplest case of (4.6) with $r = 1$ and $i_1 = 1$. The representation $\{\nu_{(1)} > x\} = \{\nu_1 > x\} \cup \cdots \cup \{\nu_N > x\}$ and the inclusion–exclusion method imply that

$$\begin{aligned} \mathrm{P}(\nu_{(1)} > Nx_1) = N\,\mathrm{P}(\nu_1 > Nx_1) &- \binom{N}{2}\mathrm{P}(\nu_1 > Nx_1, \nu_2 > Nx_1) \\ &+ \binom{N}{3}\mathrm{P}(\nu_1 > Nx_1, \nu_2 > Nx_1, \nu_3 > Nx_1) - \cdots \\ &+ (-1)^{i_\varepsilon}\binom{N}{i_\varepsilon}\mathrm{P}(\nu_1 > Nx_1, \ldots, \nu_{i_\varepsilon} > Nx_1). \end{aligned} \tag{4.7}$$

It follows that (4.2) does entail (4.6) and that even in this simple case the emerging relation between Φ and $\{\Psi_r\}_{r\geq 1}$ is rather complex. Turning to the general case, we see how overwhelming the full set of formulae expressing Φ in terms of $\{\Psi_r\}_{r\geq 1}$ becomes.

Giving up the task of finding the exact relationship, we just verify that (4.6) holds for some finite measure Φ on $\Delta \cap \{x_1 \geq \varepsilon\}$. The idea is to apply the inclusion–exclusion method to the representation

$$\{\nu_{(1)} > Nx_1, \ldots, \nu_{(r)} > Nx_r\} = \bigcup_{i_1 \neq \cdots \neq i_r} \{\nu_{i_1} > Nx_1, \ldots, \nu_{i_r} > Nx_r\},$$

where $r \geq 1$, $x_1 \geq \cdots \geq x_r \geq \varepsilon$. In view of Lemma 4.1, it is enough to show that the resulting formula for $T_N\,\mathrm{P}(\nu_{(i_1)} > Nx_1, \ldots, \nu_{(i_r)} > Nx_r)$ is a sum of the products

$$\begin{aligned} c(m_1, \ldots, m_r)N^{m_1+\cdots+m_r}\,T_N\,\mathrm{P}(&\nu_1 > Nx_1, \ldots, \nu_{m_1} > Nx_1, \\ &\nu_{m_1+1} > Nx_2, \ldots, \nu_{m_1+m_2} > Nx_2, \ldots, \\ &\nu_{m_1+\cdots+m_{r-1}+1} > Nx_r, \ldots, \nu_{m_1+\cdots+m_r} > Nx_r), \end{aligned} \tag{4.8}$$

where $m_1, \ldots, m_r \geq 1$, $m_1 + \cdots + m_r \leq i_\varepsilon$ and the $c(m_1, \ldots, m_r)$ are some negative or positive finite coefficients asymptotically independent of N.

Consider an intersection of m distinct sets $\{\nu_{i_{11}} > Nx_1, \ldots, \nu_{i_{r1}} > Nx_r\}, \ldots, \{\nu_{i_{1m}} > Nx_1, \ldots, \nu_{i_{rm}} > Nx_r\}$ and let

M_r be the set of distinct i_{rv}, $1 \leq v \leq m$,

M_{r-1} be the set of distinct $i_{r-1,v}$ such that $i_{r-1,v} \notin M_r$, $1 \leq v \leq m$,

$\vdots$

M_1 be the set of distinct $i_{1,v}$ such that $i_{1v} \notin M_2 \cup \cdots \cup M_r$, $1 \leq v \leq m$.

We will classify such intersections using labels $(m_1, \ldots, m_r)$, where m_i is the number of elements in M_i. Observe that an $(m_1, \ldots, m_r)$-intersection contributes at the mth round of the inclusion–exclusion procedure of the formula for $\mathrm{P}(\nu_{(i_1)} > Nx_1, \ldots, \nu_{(i_r)} > Nx_r)$ with the probability term in (4.8). Since the total number of $(m_1, \ldots, m_r)$-intersections is of the form

$$c_1(m_1, \ldots, m_r)\binom{N}{m_1, \ldots, m_r, N - m_1 - \cdots - m_r},$$

it follows that the combined contribution of the $(m_1, \ldots, m_r)$-intersections is of the form (4.8).

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Now, when (4.6) is established for some measure Φ, let us check the inequality

$$\int_{\Delta}(x,x)\Phi(\mathrm{d}x)\le 1, \tag{4.9}$$

assuming that $(x,x)\Phi(\mathrm{d}x)$ has no atom at zero. Since $\sum_{i=1}^{N}\nu_i^2\,\mathbf{1}_{\{\nu_i\ge N\varepsilon\}}=\sum_{i=1}^{i_\varepsilon}\nu_{(i)}^2\,\mathbf{1}_{\{\nu_{(i)}\ge N\varepsilon\}}$, we have

$$NT_N\int_{\varepsilon}^{1}x_1^2\Psi_{1,N}(\mathrm{d}x_1)=T_N\int_{\Delta}\Bigl(\sum_{i=1}^{i_\varepsilon}x_i^2\,\mathbf{1}_{\{x_i\ge\varepsilon\}}\Bigr)\Phi_N(\mathrm{d}x).$$

Applying Lemma 4.1 and (4.6), we obtain at

$$\int_{\Delta}\Bigl(\sum_{i=1}^{i_\varepsilon}x_i^2\,\mathbf{1}_{\{x_i\ge\varepsilon\}}\Bigr)\Phi(\mathrm{d}x)\le 1$$

and letting $\varepsilon\to 0$ we arrive at (4.9).

It follows from (4.9) that we can define a probability measure on Δ by

$$\Xi'(\mathrm{d}x):=(x,x)\Phi(\mathrm{d}x)+a'\delta_0(\mathrm{d}x),$$

where $a':=1-\int_{\Delta}(x,x)\Phi(\mathrm{d}x)$. It remains to apply (4.6) together with the 'if' part of the theorem established above and conclude that Ξ' must coincide with the coalescent rate measure Ξ.

We now prove Theorem 2.1 under the general reproduction assumptions (1.1) and (1.6).

This case follows from the exchangeable reproduction case above, due to the invariance of the condition (2.9) with respect to permutation of the offspring sizes $(\nu_1,\dots,\nu_N)$. Namely, if the joint distribution of $(\nu_1,\dots,\nu_N)$ is not exchangeable, we can switch to an exchangeable version $(\nu_1',\dots,\nu_N')$ keeping unchanged the distribution of the ordered offspring sizes $(\nu_{(1)},\dots,\nu_{(N)})$. The symmetrized vector $(\nu_1',\dots,\nu_N')$ can be viewed as a random permutation of $(\nu_1,\dots,\nu_N)$, which is equal to $(\nu_{\pi 1},\dots,\nu_{\pi N})$ with probability $1/N!$, where $(\pi 1,\dots,\pi N)$ is an arbitrary permutation of $(1,\dots,N)$.

4.2. Proof of Proposition 1.1

We have to verify that, under the conditions of Proposition 1.1, the asymptotic relations (2.13) and (1.7) are equivalent. Let ε be an arbitrary positive number. Using (4.7) we can replace (2.13) with (1.8). Since $\mathrm{E}((\nu_1-1)^3)=\mathrm{E}([\nu_1-1]_+^3)-\mathrm{P}(\nu_1=0)$, where $[c]_+:=c\vee 0$, it follows from the inequality $\mathrm{P}(\nu_1=0)\le\sigma_N^2$ that (1.7) is equivalent to

$$\mathrm{E}([\nu_1-1]_+^3)=o(N\sigma_N^2)\quad\text{as } N\to\infty. \tag{4.10}$$

Thus, it remains to show that (4.10) is equivalent to (1.8).

The relation (1.8) immediately follows from (4.10) due to the Markov inequality. On the other hand, (1.8) implies (4.10) because

$$\begin{aligned}\mathrm{E}([\nu_1-1]_+^3)&\le\mathrm{E}((\nu_1-1)^2\nu_1\,\mathbf{1}_{\{\nu_1\le N\varepsilon\}})+\mathrm{E}(\nu_1^3\,\mathbf{1}_{\{\nu_1>N\varepsilon\}})\\&\le N\varepsilon\sigma_N^2+N^3\,\mathrm{P}(\nu_1>N\varepsilon).\end{aligned}$$

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

5. Convergence conditions involving the raw moments

Convergence conditions in Proposition 2.1 are stated in terms of the central moments of the reproduction law. Here we restate these conditions putting emphasis on the raw moments (moments around zero). We start with two corollaries of Proposition 2.1 where all the moment conditions (except the condition on the variance σ_N^2) are given in terms of the raw moments. The third corollary deals with very slow reproduction models similar to the Moran model with $\mathrm{P}(\nu_{(1)} = 2, \nu_{(2)} = \nu_{(N-1)} = 1, \nu_{(N)} = 0) = 1$ and $\sigma_N^2 = 2/N$.

Corollary 5.1. *Under the assumptions (1.1), (1.2), and (1.6), letting $N\sigma_N^2 \to \infty$ the time-scaled genealogical process $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ converges weakly to some limit process $\{R(t)\}_{t\geq 0}$ if and only if there exist*

$$A_{k_1,\dots,k_r} = \lim_{N\to\infty} N^{r-b} T_N \,\mathrm{E}(\nu_1^{k_1} \cdots \nu_r^{k_r}) \tag{5.1}$$

for all $k_1 \geq 3$ when $r = 1$ and all $k_1, \dots, k_r \geq 2$ when $r \geq 2$, where $b = k_1 + \cdots + k_r$. The $(b; k_1, \dots, k_r; s)$-collision rates of the coalescent $\{R(t)\}_{t\geq 0}$ for $s = 0$ are $\lambda_{b;k_1,\dots,k_r;0} = A_{k_1,\dots,k_r}$ and for $s \geq 1$ can be computed from the recursion (2.5).

Corollary 5.2. *Under the assumptions (1.1), (1.2), and (1.6), letting $N\sigma_N^2 \to \infty$ the time-scaled genealogical process $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ converges weakly to the asynchronous coalescent $\{R_F(t)\}_{t\geq 0}$ if and only if $\mathrm{E}(\nu_1^2 \nu_2^2) = o(N\sigma_N^2)$ and there exist $A_k = \lim_{N\to\infty} N^{1-k} T_N \,\mathrm{E}(\nu_1^k)$ for all $k \geq 3$. The measure F and the collision rates of $\{R_F(t)\}_{t\geq 0}$ are determined from $\{A_k\}_{k\geq 3}$ in the same way as described in Corollary 2.1.*

The replacement of the central moments in (2.4) by the raw moments is justified by two mutual expansions

$$\mathrm{E}(\nu_1^{k_1} \cdots \nu_r^{k_r}) = \sum_{i_1=0}^{k_1} \cdots \sum_{i_r=0}^{k_r} \binom{k_1}{i_1} \cdots \binom{k_r}{i_r} \mathrm{E}((\nu_1 - 1)^{i_1} \cdots (\nu_r - 1)^{i_r}),$$

$$\mathrm{E}((\nu_1 - 1)^{k_1} \cdots (\nu_r - 1)^{k_r}) = \sum_{i_1=0}^{k_1} \cdots \sum_{i_r=0}^{k_r} (-1)^{b-b_i} \binom{k_1}{i_1} \cdots \binom{k_r}{i_r} \mathrm{E}(\nu_1^{i_1} \cdots \nu_r^{i_r}),$$

where $b = k_1 + \cdots + k_r$ and $b_i = i_1 + \cdots + i_r$. According to (2.4) and (5.1), all the terms except $i_1 = \cdots = i_r = 0$ and $i_1 = k_1, \dots, i_r = k_r$ are negligible in the limit. Thus, given that $N\sigma_N^2 \to \infty$, we have

$$N^{r-b} T_N \,\mathrm{E}(\nu_1^{k_1} \cdots \nu_r^{k_r}) \sim N^{r-b} T_N \,\mathrm{E}((\nu_1 - 1)^{k_1} \cdots (\nu_r - 1)^{k_r}) \quad \text{as } N \to \infty$$

since the absolute value of the terms $i_1 = \cdots = i_r = 0$ in both mutual expansions is $1 = o(N\sigma_N^2) = o(N^{b-r} T_N^{-1})$ for all $k_1 \geq 3$ when $r = 1$ and all $k_1, \dots, k_r \geq 2$ when $r \geq 2$.

Corollary 5.3. *Let (1.1), (1.2), and (1.6) hold, and let*

$$k_\sigma := \min\Big\{\text{integer } k \geq 1 : \lim_{N\to\infty} N^k \sigma_N^2 = \infty\Big\}.$$

The time-scaled genealogical process $\{\mathcal{R}(\lfloor T_N t \rfloor)\}_{t\geq 0}$ converges weakly if and only if there exist limits

(a) *for the central moments (2.4) with $k_1, \dots, k_r \geq 2$, $r \geq 1$, and $\sum_{i=1}^r (k_i - 1) \leq k_\sigma$, and*

(b) *for the raw moments (5.1) with $k_1, \dots, k_r \geq 2$, $r \geq 1$, and $\sum_{i=1}^r (k_i - 1) \geq k_\sigma + 1$.*

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press

Due to (1.4) we have an upper bound $T_N = o(N^{k_\sigma+1})$ on the time scale, and a lower bound $N^{k_\sigma} = O(T_N)$. The parameter k_σ can be estimated in terms of the probability $p_N := \mathrm{P}(D_N)$ of the event $D_N := \{(\nu_1, \ldots, \nu_N) \neq (1, \ldots, 1)\}$ that in a given generation at least one parent has no children (for example, in the Moran model case $p_N = 1$). Using the inequalities

$$\sigma_N^2 = \sum_{i=0}^{N} (i-1)^2 \,\mathrm{P}(\nu_1 = i) \geq \mathrm{P}(\nu_1 \neq 1)$$

and

$$p_N = \mathrm{P}(\{\nu_1 \neq 1\} \cup \cdots \cup \{\nu_N \neq 1\}) \leq N\,\mathrm{P}(\nu_1 \neq 1)$$

we obtain $N\sigma_N^2 \geq p_N$ (recall that in the Moran model case $N\sigma_N^2 = 2p_N$). In particular, if it is known that $N^u p_N \to \infty$, we have $k_\sigma \leq u + 1$.

Acknowledgements

The author thanks Professor Lars Holst for a quick introduction to the theory of the Poisson–Dirichlet distribution based on [5], as well as an anonymous referee for valuable comments on and corrections to an earlier version of the manuscript. This work was supported by the Knut and Alice Wallenberg Foundation.

References

[1] Billingsley, P. (1999). *Weak Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
[2] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197,** 247–276.
[3] Cannings, C. (1974). The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Adv. Appl. Prob.* **6,** 260–290.
[4] Donnelly, P. and Kurtz, T. G. (1999). Particle representations for measure-valued population models. *Ann. Prob.* **27,** 166–205.
[5] Holst, L. (2001). The Poisson–Dirichlet distribution and its relatives revisited. Preprint, Department of Mathematics, Royal Institute of Technology, Stockholm. Available at http://www2.math.kth.se/~gunnare/lasse_e.htm.
[6] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley, New York.
[7] Kingman, J. F. C. (1982). Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics*, eds G. Koch and F. Spizzichino, North-Holland, Amsterdam, pp. 97–112.
[8] Kingman, J. F. C. (1982). On the genealogy of large populations. In *Essays in Statistical Science* (J. Appl. Prob. Spec. Vol. **19A**), eds J. Gani and E. J. Hannan, Applied Probability Trust, Sheffield, pp. 27–43.
[9] Kingman, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13,** 235–248.
[10] Kingman, J. F. C. (1993). *Poisson Processes*. Oxford University Press.
[11] Möhle, M. (1998). Robustness results for the coalescent. *J. Appl. Prob.* **35,** 438–447.
[12] Möhle, M. and Sagitov, S. (2001). A classification of coalescent processes for haploid exchangeable population models. *Ann. Prob.* **29,** 1547–1562.
[13] Möhle, M. and Sagitov, S. (2003). Coalescent patterns in exchangeable diploid population models. To appear in *J. Math. Biol.*
[14] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Prob.* **27,** 1870–1902.
[15] Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Prob.* **36,** 1116–1125.
[16] Schweinsberg, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Prob.* **5,** No. 1.
[17] Schweinsberg, J. (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Prob.* **5,** No. 12.

https://doi.org/10.1239/jap/1067436085 Published online by Cambridge University Press