

# Convergence to the coalescent with simultaneous multiple mergers\*

Serik Sagitov

School of Mathematical Sciences  
Chalmers University of Technology  
S-412 96 Göteborg, Sweden  
serik@math.chalmers.se

October 21, 2002

## Abstract

The general coalescent process with simultaneous multiple mergers of ancestral lines was initially characterized in [13] 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  $\Xi$  defined on the infinite simplex  $\Delta$  was suggested in [17].

This paper presents a simple criterium of weak convergence to the  $\Xi$ -coalescent. In contrast to the earlier criterium of [13] based on the moment conditions, the key condition here is expressed in terms of the joint distribution of the ranked offspring sizes. This criterium interprets a vector  $x \in \Delta$  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.

AMS 1991 subject classifications: 92D25, 92D15, 60F17.

Keywords:  $\Xi$ -coalescent, exchangeability, Poisson-Dirichlet distribution

## 1 Introduction

The well-known Kingman coalescent ([7], [8], [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

---

\*This work has been supported by the Knut and Alice Wallenberg Foundation

the Markov chain  $\{R_0(t)\}_{t \geq 0}$  is very simple: if  $n$  is the current number of branches, wait a period proportional to  $\binom{n}{2}$ , and then reduce the number of branches to  $(n - 1)$  by joining a pair of branches picked up uniformly at random among  $\binom{n}{2}$  available pairs.

The first robustness result [8] 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 non-overlapping generations. The assumptions of Cannings model concern the vector of offspring sizes  $(\nu_1, \dots, \nu_N)$  for  $N$  individuals constituting one generation. The constant population size condition

$$\nu_1 + \dots + \nu_N = N \tag{1}$$

requires that the total offspring number equals to the number of parents. It is assumed also that there is no dependence between vectors  $(\nu_1, \dots, \nu_N)$  for different generations and that all generations share the same reproduction law, that is the joint distribution of  $(\nu_1, \dots, \nu_N)$ . Finally, the Cannings model assumes that

$$(\nu_1, \dots, \nu_N) \text{ has an exchangeable joint distribution.} \tag{2}$$

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

The genealogy of a population with non-overlapping 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  $k = 0$ . The initial state  $\mathcal{R}(0) = \{(1); \dots; (n)\}$  says that the genealogical tree construction starts with  $n$  leaves. For  $k \geq 1$  the partition  $\mathcal{R}(k)$  describes the tree state  $k$  generations back. The number of parts forming the partition  $\mathcal{R}(k)$  gives the number of branches in the tree at the backward time  $k$ . Different parts of the partition  $\mathcal{R}(k)$  label the branches in an informative way showing the subsets of leaves stemming from different branches. For instance, if  $n = 5$ , then partition  $\mathcal{R}(6) = \{(1, 3); (2, 4, 5)\}$  says that six generations ago there were two ancestors to the five sampled individuals and that individuals (1,3) stem from one the ancestors while individuals (4,5,6) stem from the other.

It turns out that the coalescent rate for the Cannings population model depends on the asymptotics of the offspring size variance  $\sigma_N^2 := \text{Var}(\nu_1)$ .

**Proposition 1.1** (from [8]) *Suppose conditions (1) and (2) hold and*

$$\sigma_N^2 \rightarrow \sigma^2, \quad N \rightarrow \infty, \quad 0 < \sigma^2 < \infty. \tag{3}$$

*If furthermore,*

$$\sup_N E(\nu_1^k) < \infty, \quad k \geq 3, \tag{4}$$

*then the finite-dimensional distributions of  $\{\mathcal{R}([Nt])\}_{t \geq 0}$  converge to those of  $\{R_0(\sigma^2 t)\}_{t \geq 0}$ .*

Conditions (3) and (4) introduce a wide class of reproduction laws falling into the domain of attraction of the Kingman coalescent. A key example here is the Wright-Fisher

population model with the symmetric multinomial  $\text{Mn}(N; \frac{1}{N}, \dots, \frac{1}{N})$  reproduction law obviously satisfying (2). In this case the marginal distribution of the offspring size is binomial  $\nu_1 \in \text{Bin}(N, \frac{1}{N})$  with variance  $\sigma_N^2 = 1 - \frac{1}{N}$  and descending factorial moments

$$\mathbb{E} \left( \nu_1^{(k)} \right) = \frac{N^{(k)}}{N^k}, \quad N^{(k)} := N(N-1)\dots(N-k+1).$$

Thus for the Wright-Fisher model conditions (3) and (4) are satisfied with  $\sigma^2 = 1$ .

However, this class fails to encompass the important Moran population model with the ordered offspring sizes being equal  $\nu_{(1)} = 2, \nu_{(2)} = \nu_{(N-1)} = 1, \nu_{(N)} = 0$  with probability one. Due to the exchangeability assumption the marginal offspring distribution for the Moran population is

$$\mathbb{P}(\nu_1 = 1) = 1 - \frac{2}{N}, \quad \mathbb{P}(\nu_1 = 0) = \mathbb{P}(\nu_1 = 2) = \frac{1}{N},$$

which implies that  $\sigma_N^2 = \frac{2}{N}$  converges to zero as  $N \rightarrow \infty$ , so that (3) does not hold. Nevertheless, as it was shown in [7], the Moran population model does belong to the domain of attraction of the Kingman coalescent, though under a different time scaling  $T_N = N^2/2$ , so that  $\{\mathcal{R}([N^2t/2])\}_{t \geq 0}$  converges to  $\{R_0(t)\}_{t \geq 0}$  as  $N \rightarrow \infty$ .

The cited convergence results by Kingman suggest a general formula

$$T_N = N/\sigma_N^2 \tag{5}$$

for the time scaling in the ancestral process. This relation has a simple explanation in terms of

$$c_N = \mathbb{E} \left( \frac{1}{N^{(2)}} \sum_{i=1}^N \nu_i^{(2)} \right) = \frac{\sigma_N^2}{N-1}, \tag{6}$$

the probability that a randomly chosen pair of individuals have the same parent. Since the Cannings model of non-overlapping generations treats the reproduction process as a sequence of i.i.d. 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  $N \rightarrow \infty$ .

The minimal convergence conditions for the Kingman coalescent are established in [12] under assumptions (1) and (2).

**Proposition 1.2** (from [12]) *Suppose conditions (1) and (2) hold. The time-scaled genealogical process  $\{\mathcal{R}([T_N t])\}_{t \geq 0}$  weakly converges to  $\{R_0(t)\}_{t \geq 0}$  if and only if the time scale  $T_N$  is defined by (5),*

$$\sigma_N^2 = o(N), \quad N \rightarrow \infty, \tag{7}$$

and

$$\mathbb{E}((\nu_1 - 1)^3) = o(N\sigma_N^2), \quad N \rightarrow \infty. \tag{8}$$

From now on the time scale  $T_N$  is tacitly assumed to be defined by (5). In Section 4 we give an alternative proof of Proposition 1.2 as a corollary of the main result of this paper Theorem 2.1. In particular, we show that under conditions of Proposition 1.2 the asymptotic relation (8) is equivalent to

$$\mathbb{P}(\nu_1 > N\epsilon) = o\left(\frac{\sigma_N^2}{N^2}\right), \quad N \rightarrow \infty, \quad \epsilon > 0. \tag{9}$$

Condition (7) stipulates that the mergers of ancestor lines are rare events, which results in a continuous time limit coalescent. Since (3) entails (7) and  $T_N \sim \frac{N}{\sigma^2}$  as  $N \rightarrow \infty$ , Proposition 1.2 implies the following improvement of Proposition 1.1.

**Corollary 1.1** *Under assumptions (1), (2), and (3) the time-scaled genealogical process  $\{\mathcal{R}([Nt])\}_{t \geq 0}$  weakly converges to  $\{R_0(t\sigma^2)\}_{t \geq 0}$  if and only if  $E(\nu_1^3) = o(N)$  as  $N \rightarrow \infty$ .*

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*. Condition (8) 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 condition (8) 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 (8) is not valid, the measure  $F$  has a positive mass outside 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 (5) of the time scale  $T_N$ . Observe that a linear time acceleration by factor  $\lambda$  transforms the coalescent rate measure to a finite measure  $\Lambda(dx_1) := \lambda F(dx_1)$ , 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  $\Lambda$ -coalescent was first introduced in [14] (paper [15] is written later though independently from [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 [13], 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  $\Xi$  (which will be also called the coalescent rate measure) defined on the infinite simplex

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

In Section 2 we give necessary and sufficient conditions for convergence to the  $\Xi$ -coalescent with the coalescent rate measure  $\Xi$  being a probability measure. The main results of this

paper are stated as Theorem 2.1 and Theorem 2.2. Theorem 2.1 is a general convergence result without the assumption (2) of exchangeability (cf. [11] dealing with the Kingman coalescent). It suggests to interpret the components of a vector  $(x_1, x_2, \dots) \in \Delta$  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  $\Xi$ -coalescent. Theorem 2.1 is proven in Section 4, the proof of Theorem 2.2 is similar and therefore omitted.

Section 3 is devoted to a simple example of an exchangeable population, whose genealogical process is approximated by a  $\Xi$ -coalescent with  $\Xi$  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 to notice 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 [13]).

## 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  $b = k_1 + \dots + k_r + s$ . 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

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

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 ways of 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

$$(k_1, \dots, k_r, s) = (\underbrace{b, \dots, b}_{l_b}, \underbrace{b-1, \dots, b-1}_{l_{b-1}}, \dots, \underbrace{2, \dots, 2}_{l_2}, l_1). \quad (10)$$

is

$$M_{b; k_1, \dots, k_r; s} = \binom{b}{k_1, \dots, k_r, s} \frac{1}{\prod_{j=2}^b l_j!}. \quad (11)$$

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 [8] given an exchangeable reproduction law  $(\nu_1, \dots, \nu_N)$  this probability equals

$$P_{b; k_1, \dots, k_r; s} = \frac{N^{(r+s)}}{N^{(b)}} \mathbb{E} \left( \nu_1^{(k_1)} \dots \nu_r^{(k_r)} \nu_{r+1} \dots \nu_{r+s} \right). \quad (12)$$

**Proposition 2.1** (from [13]) *If (1), (2), and (7) hold, the time-scaled genealogical process  $\{\mathcal{R}([T_N t])\}_{t \geq 0}$  weakly converges to some limit coalescent process  $\{R(t)\}_{t \geq 0}$  if and only if there exist*

$$A_{k_1, \dots, k_r} = \lim_{N \rightarrow \infty} N^{r-b} T_N \mathbb{E}((\nu_1 - 1)^{k_1} \dots (\nu_r - 1)^{k_r}) \quad (13)$$

for all  $k_1, \dots, k_r \geq 2$  and  $r \geq 1$ , where  $b := k_1 + \dots + k_r$ . 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

$$\begin{aligned} \lambda_{b+1; k_1, \dots, k_r; s+1} &= \lambda_{b; k_1, \dots, k_r; s} - \sum_{m=1}^r \lambda_{b+1; k_1, \dots, k_{m-1}, k_m+1, k_{m+1}, \dots, k_r; s} \\ &\quad - s \lambda_{b+1; k_1, \dots, k_r, 2; s-1}, \quad k_1, \dots, k_r \geq 2, \quad r \geq 1, \quad s \geq 0. \end{aligned} \quad (14)$$

The set of rates  $\{\lambda_{b; k_1, \dots, k_r; s}\}$  provides with 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 (11). 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  $\frac{1}{\lambda_b} \lambda_{b; k_1, \dots, k_r; s}$ .

It follows from [13] 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 conditions (1), (2), (7) hold while condition (8) does not. If*

$$\mathbb{E}((\nu_1 - 1)^2 (\nu_2 - 1)^2) = o(N \sigma_N^2), \quad N \rightarrow \infty, \quad (15)$$

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

$$\mathbb{E}((\nu_1 - 1)^k) \sim A_k N^{k-2} \sigma_N^2, \quad N \rightarrow \infty, \quad k \geq 3, \quad (16)$$

where  $1 \geq A_3 \geq A_4 \geq \dots > 0$ . 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  $\int_0^1 x^k F(dx) = A_{k+2}$ ,  $k \geq 0$ , where  $A_2 := 1$ . The collision rates of  $\{R_F(t)\}_{t \geq 0}$  satisfy

$$\lambda_{b; k_1, \dots, k_r; s} = 1_{\{r=1\}} \int_0^1 x^{k_1-2} (1-x)^s F(dx), \quad r \geq 1, \quad k_1, \dots, k_r \geq 2, \quad s \geq 0, \quad (17)$$

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

This result excludes convergence to the Kingman coalescent and by condition (15) prohibits simultaneous mergers of ancestral lines. The asynchronous coalescent  $\{R_F(t)\}_{t \geq 0}$  is characterised by positive tree collapse rates  $A_k = \lambda_{k; k; 0}$ ,  $k \geq 3$  (a tree collapse is a single merger involving all branches of the tree). It follows that the limit coalescent can not have bounded merger sizes  $k_i$ : if  $A_3 > 0$ , then  $A_k > 0$  for all  $k \geq 3$ .

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

$$\lambda_{b; k_1, \dots, k_r; s} = a \mathbf{1}_{\{r=1, k_1=2\}} + \int_{\Delta} \left( \sum_{j=0}^s \sum_{i_1 \neq \dots \neq i_{r+j}} \binom{s}{j} x_{i_1}^{k_1} \dots x_{i_r}^{k_r} x_{i_{r+1}} \dots x_{i_{r+j}} x_0^{s-j} \right) \Xi(dx)/(x, x), \quad (18)$$

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

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

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

The set of rates

$$\lambda_{2; 2; 0} = 1; \quad \lambda_{b; k_1, \dots, k_r; 0} = \int_{\Delta} \left( \sum_{i_1 \neq \dots \neq i_r} x_{i_1}^{k_1} \dots x_{i_r}^{k_r} \right) \Xi(dx)/(x, x)$$

for all  $k_1 \geq 3$ ,  $r = 1$  and  $k_1, \dots, k_r \geq 2$ ,  $r \geq 2$ , uniquely determines the coalescent rate measure  $\Xi$ . If  $\Xi$  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 (18).

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  $\text{Mn}(N; p_1, \dots, p_N)$  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 the weak convergence to the  $\Xi$ -coalescent in terms of the ordered offspring sizes  $\nu_{(1)} \geq \dots \geq \nu_{(N)}$  without the exchangeability assumption (2).

The proper time scale  $T_N$  in this more general case might be again defined by (5), now with  $\sigma_N^2$  being the variance of the offspring size  $\nu$  of a randomly chosen individual. This follows from the next counterpart of relation (6): since  $\mathbb{E}(\nu) = 1$ , the coalescence probability  $c_N$  in the non-exchangeable case equals

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

**Theorem 2.1** *Let  $\Phi_N(dx)$  be the joint distribution of the vector  $(\frac{\nu_{(1)}}{N}, \dots, \frac{\nu_{(N)}}{N})$  of the ranked relative frequencies for sizes of  $N$  sibling groups constituting a generation. If (1) and (7) hold, then the time-scaled genealogical process  $\{\mathcal{R}([T_N t])\}_{t \geq 0}$  weakly converges to the  $\Xi$ -coalescent if and only if the weak convergence condition*

$$T_N \Phi_N(dx) \rightarrow \Xi(dx)/(x, x), \quad N \rightarrow \infty \quad (19)$$

holds on  $\Delta \cap \{x_1 > \epsilon\}$  for any  $\epsilon > 0$ , where  $\Xi$  is some probability measure on  $\Delta$ .

Following Section 4 of [1] we treat  $\Delta$  as a subset of the metric space  $R^\infty$  with the topology of coordinatewise convergence. The weak convergence (19) holds if and only if the weak convergence of the tail distribution functions

$$T_N \mathbb{P}(\nu_{(1)} > Nx_1, \dots, \nu_{(r)} > Nx_r) \rightarrow \int_{\Delta} 1_{\{y_1 > x_1, \dots, y_r > x_r\}} \Xi(dy)/(y, y), \quad N \rightarrow \infty \quad (20)$$

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

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

**Corollary 2.2** *Under assumptions (1) and (7) the process  $\{\mathcal{R}([T_N t])\}_{t \geq 0}$  weakly converges to the asynchronous  $F$ -coalescent if and only if*

$$T_N \mathbb{P}(\nu_{(1)} > Nx) \rightarrow \int_x^1 y^{-2} F(dy), \quad N \rightarrow \infty \quad (21)$$

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

$$T_N \mathbb{P}(\nu_{(2)} > N\epsilon) \rightarrow 0, \quad N \rightarrow \infty, \quad \epsilon > 0. \quad (22)$$

Condition (22) excludes simultaneous mergers by prohibiting two large sibling groups (of size  $\geq N\epsilon$ ) in the same generation among  $T_N$  consecutive generations. Relation (21) illuminates the role of the largest offspring size in determining the merger size distribution (17). This result implies in turn that if  $F$  is concentrated at zero, then only pairwise mergers are possible.

**Corollary 2.3** *Suppose that (1) and (7) hold. Condition*

$$T_N \mathbb{P}(\nu_{(1)} > N\epsilon) \rightarrow 0, \quad N \rightarrow \infty, \quad \epsilon > 0 \quad (23)$$

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

If instead of (7) it is assumed that

$$\sigma_N^2 \sim Nc, \quad N \rightarrow \infty, \quad 0 < c \leq 1 \quad (24)$$

(constant  $c$  can not exceed one since  $\sigma_N^2 \leq N - 1$ ), then there might exist a discrete-time limit for the unscaled genealogical process. Condition (24) means that with probability  $c$  a pair of randomly chosen individuals in a large population are siblings.

**Proposition 2.2** (from [13]) *If (1), (2), and (24) hold, the discrete-time genealogical process  $\{\mathcal{R}(k)\}_{k \geq 0}$  weakly converges to some discrete-time Markov chain  $\{R(k)\}_{k \geq 0}$  if and only if there exist*

$$A_{k_1, \dots, k_r} = \lim_{N \rightarrow \infty} N^{r-b} \mathbb{E}((\nu_1 - 1)^{k_1} \dots (\nu_r - 1)^{k_r})$$

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



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

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

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  $\Delta$  is proposed in [17]. Let  $\Phi$  be an arbitrary probability measure on the simplex  $\Delta$ . The Markov chain with transition probabilities

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

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

**Theorem 2.2** *If (1) and (24) hold, the discrete-time genealogical process  $\{\mathcal{R}(k)\}_{k \geq 0}$  weakly converges to the discrete-time  $\Phi$ -coalescent if and only if  $\Phi$  is the weak limit distribution of the vector of ordered offspring proportions  $(\frac{\nu(1)}{N}, \dots, \frac{\nu(N)}{N})$ .*

**Remark.** The definition of the discrete-time  $\Phi$ -coalescent given above slightly modifies the Definition 36 of [17]. The limit process of Theorem 2.2 is the discrete-time  $\Xi$ -coalescent in the strict sense of Definition 36 from [17], if the limit distribution of  $(\frac{\nu(1)}{N}, \dots, \frac{\nu(N)}{N})$  is of the form  $\Phi(dx) = \frac{\Xi(dx)}{(x, x)} \mathbf{1}_{\{x_1 > 0\}} + a_0 \delta_0(dx) \mathbf{1}_{\{x_1 = 0\}}$ , where  $\delta_0$  is the Dirac measure on  $\Delta$  concentrated at zero and a non-negative constant  $a_0$  is such that  $\Phi(\Delta) = 1$ .

### 3 Poisson-Dirichlet coalescent

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

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

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

$$p_{b; k_1, \dots, k_r; s} = \frac{\theta^{r+s}}{\theta^{[b]}} \prod_{i=1}^r (k_i - 1)! \quad (25)$$

This will be called the Poisson-Dirichlet coalescent with parameter  $\theta$ .

Using representation (10) put  $p(l_1, \dots, l_b) := p_{b; k_1, \dots, k_r; s}$  and  $l := l_1 + \dots + l_b$ . To compute the total probability  $p_b$  of all mergers in the Poisson-Dirichlet coalescent with  $b$  ancestral lines rewrite (25) as  $p(l_1, \dots, l_b) = \frac{\theta^l}{\theta^{[b]}} \prod_{j=1}^b ((j-1)!)^{l_j}$ . Observe that due to (11)

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

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 except one (that with  $l_1 = b$ ) 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  $\theta$  on the overall speed of the coalescent: larger value of  $\theta$  brings lower speed. Recall that with the Kingman coalescent  $p_b = \binom{b}{2}$ , and notice that  $1 - \theta^b / \theta^{[b]} \sim \theta^{-1} \binom{b}{2}$  as  $\theta \rightarrow \infty$ . Thus, if  $\theta \rightarrow \infty$  the Poisson-Dirichlet coalescent can be approximated by the Kingman coalescent, as the multiple mergers become less and less frequent.

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  $\text{Mn}(N; p_1, \dots, p_N)$  with  $p_1 = \dots = p_N = \frac{1}{N}$  as the joint distribution of offspring sizes  $(\nu_1, \dots, \nu_N)$ . To extent this model we assume that the multinomial parameters  $(p_1, \dots, p_N)$  are random and have a symmetric Dirichlet distribution  $D(\alpha, \dots, \alpha)$ , so that the resulting distribution of  $(\nu_1, \dots, \nu_N)$  is the so-called Dirichlet compound multinomial (cf. [6]):

$$P(\nu_1 = n_1, \dots, \nu_N = n_N) = \binom{N}{n_1, \dots, n_N} \frac{\alpha^{[n_1]} \dots \alpha^{[n_N]}}{(N\alpha)^{[N]}}. \quad (26)$$

This model is particularly convenient for the coalescent calculations in view of the next formular for the joint moments:

$$E \left( \nu_1^{(k_1)} \dots \nu_r^{(k_r)} \nu_{r+1} \dots \nu_{r+s} \right) = \frac{N^{(b)}}{(N\alpha)^{[b]}} \alpha^{[k_1]} \dots \alpha^{[k_r]} \alpha^s. \quad (27)$$

Equality (27) implies  $E \left( \nu_1^{(n)} \right) = \frac{N^{(n)} \alpha^{[n]}}{(N\alpha)^{[n]}}$ , so that  $\sigma_N^2 = \frac{(N-1)(\alpha+1)}{N\alpha+1}$ : the smaller is  $\alpha$  the more variable become the offspring sizes, and the larger is  $\alpha$  the closer is the model to the Wright-Fisher model in which  $\sigma_N^2 = \frac{N-1}{N}$ . If  $\alpha$  does not depend on  $N$  the asymptotic genealogy of the Dirichlet compound Wright-Fisher model is the Kingman coalescent. If  $\alpha = \frac{\theta}{N}$  it is easy to verify that (12) and (27) imply the convergence to the Poisson-Dirichlet coalescent. This derivation of the Poisson-Dirichlet coalescent illuminates the meaning of the parameter  $\theta$ : the smaller is  $\theta$  the more variable are offspring sizes and therefore the faster goes coalescing.

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

$$P(\alpha = A) = 1 - \frac{p}{N}, \quad P(\alpha = \frac{\theta}{N}) = \frac{p}{N}. \quad (28)$$

In this case condition (3) holds with the limit variance  $\sigma^2 = \frac{A+1}{A} + \frac{p}{1+\theta}$ . Moreover, it follows that the genealogical process  $\{\mathcal{R}([Nt])\}_{t \geq 0}$  weakly converges to the  $\Xi$ -coalescent, where  $\Xi := \Xi(A, p, \theta)$  has an atom at zero of size  $a := \frac{A+1}{A}$ , which is the first term in the formula for  $\sigma^2$ , and outside zero the measure  $\Xi(A, p, \theta)$  is closely related to the Poisson-Dirichlet distribution.

Observe first that (27) and (28) imply

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

which according to Proposition 2.1 and (12) entails that  $\{\mathcal{R}([Nt])\}_{t \geq 0}$  converges to a coalescent with the collision rates

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

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 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  $\Xi$  on the infinite simplex  $\Delta$  and a positive integer  $r$  we can introduce a symmetric measure on the  $r$ -dimensional simplex

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

by

$$\Psi_r(S) := \int_{\Delta} \left( \sum_{i_1 \neq \dots \neq i_r} 1_{\{(y_{i_1}, \dots, y_{i_r}) \in S\}} \right) \Xi(dy)/(y, y) \quad (29)$$

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

$$\Psi_r(\Delta_{r, \epsilon}) \leq \int_{\Delta} \left( \sum_{i_1 \neq \dots \neq i_r} \left(\frac{y_{i_1}}{\epsilon}\right)^2 \cdots \left(\frac{y_{i_r}}{\epsilon}\right)^2 1_{\{y_{i_1}, \dots, y_{i_r} > \epsilon\}} \right) \Xi(dy)/(y, y) \leq \epsilon^{-2r},$$

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

**Lemma 4.1** *Assume (1), (2), (7) and let  $\Psi_{r, N}$  stand for the symmetric measure on  $\Delta_r$  giving the joint distribution of the vector  $(\frac{\nu_1}{N}, \dots, \frac{\nu_r}{N})$  of normalized offspring sizes. The time-scaled genealogical process  $(\mathcal{R}_{[TNt]})_{t \geq 0}$  weakly convergence to the  $\Xi$ -coalescent if and only if the weak convergence*

$$N^r T_N \Psi_{r, N} \rightarrow \Psi_r, \quad N \rightarrow \infty, \quad r \geq 1 \quad (30)$$

holds over  $\Delta_{r, \epsilon}$  for any  $\epsilon > 0$ , where  $\Psi_r$  is connected to  $\Xi$  via (29).

PROOF OF LEMMA 4.1

Condition (30) is similar to the following weak convergence condition given in [13]. It is shown there that under (1), (2), and (7) the weak convergence of  $(\mathcal{R}_{[T_N t]})_{t \geq 0}$  takes place if and only if for all  $r \geq 1$  there exist a symmetric measure  $F_r$  defined on  $\Delta_r$  such that the weak convergence

$$N^r T_N \Psi_{r,N}(dx_1, \dots, dx_r) \rightarrow \frac{F_r(dx_1, \dots, dx_r)}{x_1^2 \dots x_r^2}, \quad N \rightarrow \infty \quad (31)$$

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

$$N^{-r} T_N \mathbb{E}((\nu_1 - 1)^2 \dots (\nu_r - 1)^2) \rightarrow F_r(\Delta_r), \quad N \rightarrow \infty. \quad (32)$$

Furthermore, it is known that the total masses of  $F_r$  form a monotone sequence  $F_1(\Delta_1) \geq F_2(\Delta_2) \geq \dots$ . Notice that for  $r = 1$  convergence (32) follows automatically from (5), the definition of the time scale  $T_N$ . Therefore, our task reduces to verifying that the limits in (30) and (31) coincide, and that (32) for  $r \geq 2$  is always true under conditions of the lemma.

The equality

$$F_r(dx_1, \dots, dx_r) = x_1^2 \dots x_r^2 \Psi_r(dx_1, \dots, dx_r), \quad (x_1, \dots, x_r) \in \Delta_{r,\epsilon}, \quad \epsilon > 0$$

immediately follows from the representation

$$F_r(S) = \int_{\Delta} \sum_{i_1 \neq \dots \neq i_r} y_{i_1}^2 \dots y_{i_r}^2 \mathbf{1}_{\{(y_{i_1}, \dots, y_{i_r}) \in S\}} \Xi(dy)/(y, y) + a \mathbf{1}_{\{r=1, (0, \dots, 0) \in S\}},$$

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

$$F_r(\Delta_r \cap \{x_r \leq \epsilon\}) \leq \epsilon^2 \int_{\Delta} \sum_{i_1=1}^{\infty} y_{i_1}^2 \Xi(dy)/(y, y) = \epsilon^2,$$

which together with (31) entail

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

To deduce (32) from the last convergence it suffices to prove that

$$N^{-r} T_N \mathbb{E}((\nu_1 - 1)^2 \dots (\nu_r - 1)^2 (1 - \mathbf{1}_{\{\nu_1, \dots, \nu_r > N\epsilon\}})) \rightarrow 0$$

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

$$\lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} N^{-r} T_N \mathbb{E}((\nu_1 - 1)^2 \dots (\nu_r - 1)^2 \mathbf{1}_{\{\nu_r \leq N\epsilon\}}) = 0. \quad (33)$$

Observe that  $\sum_{i=2}^N \nu_i^2 \mathbf{1}_{\{\nu_i \leq N\epsilon\}} \leq N\epsilon \sum_{i=2}^N \nu_i \stackrel{(1)}{\leq} N^2\epsilon$  implies

$$\begin{aligned} & (N-1) \mathbb{E}((\nu_1 - 1)^2 (\nu_2 - 1)^2 \mathbf{1}_{\{\nu_2 \leq N\epsilon\}}) \\ & \leq (N-1) \mathbb{E}((\nu_1 - 1)^2 \nu_2^2 \mathbf{1}_{\{\nu_2 \leq N\epsilon\}}) + (N-1) \sigma_N^2 \\ & \stackrel{(2)}{=} \sum_{i=2}^N \mathbb{E}((\nu_1 - 1)^2 \nu_i^2 \mathbf{1}_{\{\nu_i \leq N\epsilon\}}) + (N-1)^2 c_N \leq N^3 \epsilon c_N + N^2 c_N. \end{aligned}$$

Dividing the leftmost and the rightmost sides by  $N^3 c_N$  we obtain an upper bound confirming that (33) is valid for  $r = 2$ . Finally, validity of (33) for  $r \geq 3$  follows by induction over  $r$  due to the inequality

$$\begin{aligned} & (N - r + 1) \mathbb{E} \left( (\nu_1 - 1)^2 \cdots (\nu_r - 1)^2 \mathbf{1}_{\{\nu_r \leq N\epsilon\}} \right) \\ & \leq \sum_{i=r}^N \mathbb{E} \left( (\nu_1 - 1)^2 \cdots (\nu_{r-1} - 1)^2 \nu_i^2 \mathbf{1}_{\{\nu_i \leq N\epsilon\}} \right) + N \mathbb{E} \left( (\nu_1 - 1)^2 \cdots (\nu_r - 1)^2 \right) \\ & \leq (N^2 \epsilon + N) \mathbb{E} \left( (\nu_1 - 1)^2 \cdots (\nu_r - 1)^2 \right). \end{aligned}$$

PROOF OF THEOREM 2.1

A. Exchangeable reproduction assumptions: (1), (2), (7).

A1. The “if” part of Theorem 2.1. Fix an arbitrary  $0 < \epsilon < 1$ , and let  $(x_1, \dots, x_r) \in \Delta_{r,\epsilon}$ . Due to the obvious equality

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

we have

$$(N)_r \mathbb{P}(\nu_1 > Nx_1, \dots, \nu_r > Nx_r) = \sum_{1 \leq i_1 \neq \dots \neq i_r \leq i_\epsilon} \mathbb{P}(\nu_{(i_1)} > Nx_1, \dots, \nu_{(i_r)} > Nx_r).$$

In the last sum there is no need to consider  $\nu_{(i)}$  for  $i$  above  $i_\epsilon := \lceil \epsilon^{-1} \rceil$ , since  $i\nu_{(i)} \leq \nu_{(i)} + \dots + \nu_{(1)} \leq N$  and  $\nu_{(i)} \leq N\epsilon$  for  $i \geq \epsilon^{-1}$ . Thus, (19) via (20) implies (30) and according to Lemma 4.1 we have established the asserted convergence to the  $\Xi$ -coalescent.

A2. The “only if” part of Theorem 2.1. If the convergence to the  $\Xi$ -coalescent takes place, it follows from Lemma 4.1 that  $N^r T_N \mathbb{P}(\nu_1 > Nx_1, \dots, \nu_r > Nx_r)$  weakly converges for  $(x_1, \dots, x_r) \in \Delta_{r,\epsilon}$  given an arbitrary  $0 < \epsilon < 1$ . We will use this convergence to show that the weak convergence

$$T_N \mathbb{P}(\nu_{(i_1)} > Nx_1, \dots, \nu_{(i_r)} > Nx_r) \rightarrow \int_{\Delta} \mathbf{1}_{\{y_1 \geq x_1, \dots, y_r \geq x_r\}} \Phi(dy), \quad N \rightarrow \infty \quad (34)$$

holds over  $x_1 \geq \dots \geq x_r \geq \epsilon$  for all  $r \geq 1$ .

For the purpose of illustration we start with the simplest case of (34) with  $r = 1$  and  $i_1 = 1$ . The representation  $\{\nu_{(1)} > x\} = \{\nu_1 > x\} \cup \dots \cup \{\nu_N > x\}$  and the inclusion-exclusion method imply

$$\begin{aligned} \mathbb{P}(\nu_{(1)} > Nx_1) &= N \mathbb{P}(\nu_1 > Nx_1) - \binom{N}{2} \mathbb{P}(\nu_1 > Nx_1, \nu_2 > Nx_1) \\ &+ \binom{N}{3} \mathbb{P}(\nu_1 > Nx_1, \nu_2 > Nx_1, \nu_3 > Nx_1) \\ &- \dots + (-1)^{i_\epsilon} \binom{N}{i_\epsilon} \mathbb{P}(\nu_1 > Nx_1, \dots, \nu_{i_\epsilon} > Nx_1). \end{aligned} \quad (35)$$

It follows that (30) does entail (34) and that even in this simple case the emerging relation between  $\Phi$  and  $\{\Psi_r\}_{r \geq 1}$  is rather complex. Turning to the general case we see how overwhelming becomes the full set of formulae expressing  $\Phi$  in terms of  $\{\Psi_r\}_{r \geq 1}$ .

Giving up the task of finding the exact relationship we just verify that (34) holds for some finite measure  $\Phi$  on  $\Delta \cap \{x_1 \geq \epsilon\}$ . The idea is to apply the inclusion-exclusion method to the representation

$$\{\nu_{(1)} > Nx_1, \dots, \nu_{(r)} > Nx_r\} = \cup_{i_1 \neq \dots \neq i_r} \{\nu_{i_1} > Nx_1, \dots, \nu_{i_r} > Nx_r\},$$

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

$$\begin{aligned} & c(m_1, \dots, m_r) N^{m_1 + \dots + m_r} T_N \\ & \times P(\nu_1 > Nx_1, \dots, \nu_{m_1} > Nx_1, \nu_{m_1+1} > Nx_2, \dots, \nu_{m_1+m_2} > Nx_2, \\ & \dots, \nu_{m_1+\dots+m_{r-1}+1} > Nx_r, \dots, \nu_{m_1+\dots+m_r} > Nx_r), \end{aligned} \quad (36)$$

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

Consider an intersection of  $m$  distinct sets  $\{\nu_{i_{11}} > Nx_1, \dots, \nu_{i_{r1}} > Nx_r\}, \dots, \{\nu_{i_{1m}} > Nx_1, \dots, \nu_{i_{rm}} > Nx_r\}$  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  $i_{r-1,v} \notin M_r$ ,  $1 \leq v \leq m, \dots$ ,

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

We will classify such intersections using labels  $(m_1, \dots, m_r)$ , where  $m_i$  is the number of elements in  $M_i$ . Observe that a  $(m_1, \dots, m_r)$ -intersection contributes at the  $m$ -th round of the inclusion-exclusion procedure to the formula for  $P(\nu_{(i_1)} > Nx_1, \dots, \nu_{(i_r)} > Nx_r)$  with the probability mentioned in (36). Since the total number of the  $(m_1, \dots, m_r)$ -intersections is of the form  $c_1(m_1, \dots, m_r) \binom{N}{m_1, \dots, m_r, N-m_1-\dots-m_r}$ , it follows that the combined contribution of the  $(m_1, \dots, m_r)$ -intersections is of the form (36).

Now, when (34) is established for some measure  $\Phi$ , let us check the inequality

$$\int_{\Delta} (x, x) \Phi(dx) \leq 1, \quad (37)$$

assuming that  $(x, x) \Phi(dx)$  has no atom at zero. Since  $\sum_{i=1}^N \nu_i^2 1_{\{\nu_i \geq N\epsilon\}} = \sum_{i=1}^{i_\epsilon} \nu_{(i)}^2 1_{\{\nu_{(i)} \geq N\epsilon\}}$ , we have

$$NT_N \int_{\epsilon}^1 x_1^2 \Psi_{1,N}(dx_1) = T_N \int_{\Delta} \left( \sum_{i=1}^{i_\epsilon} x_i^2 1_{\{x_i \geq \epsilon\}} \right) \Phi_N(dx).$$

Applying Lemma 4.1 and (34) we obtain  $\int_{\Delta} \left( \sum_{i=1}^{i_\epsilon} x_i^2 1_{\{x_i \geq \epsilon\}} \right) \Phi(dx) \leq 1$  and letting  $\epsilon \rightarrow 0$  we arrive at (37).

It follows from (37) that we can introduce a probability measure on  $\Delta$  by  $\Xi'(dx) := (x, x) \Phi(dx) + a' \delta_0(dx)$ , where  $a' := 1 - \int_{\Delta} (x, x) \Phi(dx)$ . It remains to apply (34) together with the already established “if” part of the theorem and conclude that  $\Xi'$  must coincide with the coalescent rate measure  $\Xi$ .

B. General reproduction assumptions: (1), (7).

The general reproduction case follows from the exchangeable reproduction case above, due to the invariance of the condition (19) 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  $(1, \dots, N)$ .

PROOF OF PROPOSITION 1.2

We have to verify that under the conditions of Proposition 1.2 the asymptotic relations (23) and (8) are equivalent. Let  $\epsilon$  be an arbitrary positive number. Using (35) we can replace (23) with (9). Since  $\mathbb{E}((\nu_1 - 1)^3) = \mathbb{E}([\nu_1 - 1]_+^3) - \mathbb{P}(\nu_1 = 0)$ , where  $[c]_+ := c \vee 0$ , it follows from  $\mathbb{P}(\nu_1 = 0) \leq \sigma_N^2$  that (8) is equivalent to

$$\mathbb{E}([\nu_1 - 1]_+^3) = o(N\sigma_N^2), \quad N \rightarrow \infty. \quad (38)$$

Thus it remains to show that (38) is equivalent to (9).

Relation (9) immediately follows from (38) due to the Markov inequality. On the other hand, (9) implies (38) because

$$\mathbb{E}([\nu_1 - 1]_+^3) \leq \mathbb{E}((\nu_1 - 1)^2 \nu_1 \mathbf{1}_{\{\nu_1 \leq N\epsilon\}}) + \mathbb{E}(\nu_1^3 \mathbf{1}_{\{\nu_1 > N\epsilon\}}) \leq N\epsilon\sigma_N^2 + N^3\mathbb{P}(\nu_1 > N\epsilon).$$

## 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  $\sigma_N^2$ ) are given in terms of the raw moments. Notice that it does not cover the Moran model in which  $\sigma_N^2 = 2/N$ .

**Corollary 5.1** *Under assumptions (1), (2), (7), and  $N\sigma_N^2 \rightarrow \infty$  the time-scaled genealogical process  $\{\mathcal{R}([T_N t])\}_{t \geq 0}$  weakly converges to some limit process  $\{R(t)\}_{t \geq 0}$  if and only if there exist*

$$A_{k_1, \dots, k_r} = \lim_{N \rightarrow \infty} N^{r-b} T_N \mathbb{E}(\nu_1^{k_1} \dots \nu_r^{k_r}) \quad (39)$$

for all  $k_1 \geq 3$ ,  $r = 1$  and  $k_1, \dots, k_r \geq 2$ ,  $r \geq 2$ , where  $b = k_1 + \dots + k_r$ . The  $(b; k_1, \dots, k_r; s)$ -collision rates of the coalescent  $\{R(t)\}_{t \geq 0}$  are 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 (14).

**Corollary 5.2** *Under assumptions (1), (2), (7), and  $N\sigma_N^2 \rightarrow \infty$  the time-scaled genealogical process  $\{\mathcal{R}([T_N t])\}_{t \geq 0}$  weakly converges to the asynchronous coalescent  $\{R_F(t)\}_{t \geq 0}$  if and only if  $\mathbb{E}(\nu_1^2 \nu_2^2) = o(N\sigma_N^2)$  and there exist  $A_k = \lim_{N \rightarrow \infty} N^{1-k} T_N \mathbb{E}(\nu_1^k)$  for all  $k \geq 3$ . 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 (13) by the raw moments is justified by two mutual expansions

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

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

where  $b = k_1 + \dots + k_r$  and  $b_i = i_1 + \dots + i_r$ . All except the free terms in these sums are regulated by conditions (13) and (39). The common absolute value of the free terms is 1, and in the case  $N\sigma_N^2 \rightarrow \infty$  we can conclude that

$$N^{r-b} T_N \mathbb{E}(\nu_1^{k_1} \dots \nu_r^{k_r}) \sim N^{r-b} T_N \mathbb{E}((\nu_1 - 1)^{k_1} \dots (\nu_r - 1)^{k_r}), \quad N \rightarrow \infty \quad (40)$$

since the free terms from the mutual expansions make negligible contribution of size  $N^{r-b} T_N = o(N^{r+2-b}) = o(1)$  for all  $k_1 \geq 3$ ,  $r = 1$  and  $k_1, \dots, k_r \geq 2$ ,  $r \geq 2$ . In general, the condition  $T_N = o(N^{b-r})$  is sufficient to justify the replacement (40).

**Corollary 5.3** *Let (1), (2), (7) hold, and let*

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

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

- a) for the central moments (13) with  $k_1, \dots, k_r \geq 2$ ,  $r \geq 1$ ,  $\sum_{i=1}^r (k_i - 1) \leq k_\sigma$ ,  
and  
b) for the raw moments (39) with  $k_1, \dots, k_r \geq 2$ ,  $r \geq 1$ ,  $\sum_{i=1}^r (k_i - 1) \geq k_\sigma + 1$ .*

Due to (5) we have an upper bound on the time scale  $T_N = o(N^{k_\sigma+1})$ , and a lower bound  $N^{k_\sigma} = O(T_N)$ . Parameter  $k_\sigma$  can be estimated in terms of the probability  $p_N := \mathbb{P}(D_N)$  of the event  $D_N := \{(\nu_1, \dots, \nu_N) \neq (1, \dots, 1)\}$  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

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

and

$$p_N = \mathbb{P}(\{\nu_1 \neq 1\} \cup \dots \cup \{\nu_N \neq 1\}) \leq N \mathbb{P}(\nu_1 \neq 1)$$

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  $N^u p_N \rightarrow \infty$ , we have  $k_\sigma \leq u - 1$ .

**Acknowledgement.** The author thanks Prof. Lars Holst for a quick introduction to the theory of the Poisson-Dirichlet distribution based on [5].

## References

- [1] BILLINGSLEY, P. (1999) Weak Convergence of Probability Measures, 2nd edition. 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. Probab.* 27, 166 – 205.
- [5] HOLST, L. (2001) The Poisson-Dirichlet distribution and its relatives revisited. Preprint: Dept. Math., Royal Inst. Tech., Stockholm. Downloadable at <http://www2.math.kth.se/~gunnare/lasse.e.htm>
- [6] JOHNSON, N.L., KOTZ, S. AND BALAKRISHNAN, N. (1997) *Discrete Multivariate Distributions*. Wiley, New York.
- [7] KINGMAN, J.F.C. (1982a) On the Genealogy of Large Populations. *J. Appl. Prob.* 19A, 27–43.
- [8] KINGMAN, J.F.C. (1982b) Exchangeability and the Evolution of Large Populations. in: KOCH, G. AND SPIZZICHINO, F.: *Exchangeability in Probability and Statistics*, North-Holland Publishing Company, pp. 97–112.
- [9] KINGMAN, J.F.C. (1982c) 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. (1999) Coalescent Patterns in Exchangeable Diploid Population Models. Preprint: *Berichte zur Stochastik und verwandten Gebieten*, Johannes Gutenberg-Universität Mainz, July 1999. Downloadable from <http://www.math.chalmers.se/~serik/two-sex.html>.
- [13] MÖHLE, M. AND SAGITOV, S. (2001) A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* 29, 1547 – 1562.
- [14] PITMAN, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* 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  $\Lambda$ -coalescent to come down from infinity, *Electron. Comm. Probab.* 5, 1–11 (electronic).
- [17] SCHWEINSBERG, J. (2000) Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* 5, 1 – 50 (electronic).