Fifth Colloquium on Mathematics and Computer Science DMTCS proc. **AI**, 2008, 311–324

Convergence to the coalescent and its relation to the time back to the most recent common ancestor

Martin Möhle

Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany
E-mail address: moehle@math.uni-duesseldorf.de

For the class of haploid exchangeable population models with non-overlapping generations and population size N it is shown that, as N tends to infinity, convergence of the time-scaled ancestral process to Kingman's coalescent and convergence in distribution of the scaled times back to the most recent common ancestor (MRCA) to the corresponding times back to the MRCA of the Kingman coalescent are equivalent.

Extensions of this equivalence are derived for exchangeable population models being in the domain of attraction of a coalescent process with multiple collisions. The proofs are based on the property that the total rates of a coalescent with multiple collisions already determine the distribution of the coalescent.

It is finally shown that similar results cannot be obtained for the full class of exchangeable coalescents allowing for simultaneous multiple collisions of ancestral lineages, essentially because the total rates do not determine the distribution of a general exchangeable coalescent.

Keywords: absorption time, ancestral process, coalescent, exchangeability, most recent common ancestor, simultaneous multiple collisions

1 Introduction

Coalescent theory has been proven to be a powerful tool to analyse the ancestry of a sample of n individuals (genes, particles, DNA-sequences) taken from a large (haploid) population. In the literature a variety of convergence results can be found ([13], [17], [18]) which ensure convergence to a coalescent process, in particular, to the Kingman coalescent ([5] - [8], [9]).

In its kernel coalescent theory boils down to a collection of methods, formulas, and convergence results for the class of exchangeable population models with non-overlapping generations and fixed population size $N \in \mathbb{N} := \{1, 2, \ldots\}$ introduced by Cannings ([1], [2]). These models assume that each individual $i \in \{1, \ldots, N\}$ in generation r produces a random number $\nu_i^{(r)}$ of offspring and that the offspring of all the N individuals of generation r form the following $(r+1)$-th generation. Since the total population size is assumed to be fixed ($= N$), the relation $\nu_1^{(r)} + \cdots + \nu_N^{(r)} = N$ has to be satisfied for each generation r. Therefore, for fixed r the offspring variables cannot be stochastically independent, except for the trivial

case when $\nu_i^{(r)} \equiv 1$ for all $i \in \{1,\dots,N\}$. It is therefore assumed that, for each fixed generation r, the offspring sizes $\nu_1^{(r)},\dots,\nu_N^{(r)}$ are exchangeable, i.e. the joint distribution of $(\nu_{\pi 1}^{(r)},\dots,\nu_{\pi N}^{(r)})$ does not depend on the permutation π of the indices $1,\dots,N$. It is furthermore assumed that the random vectors $\nu^{(r)} := (\nu_1^{(r)},\dots,\nu_N^{(r)})$, $r \in \mathbb{Z}$, are independent and identically distributed, which makes the model homogeneous over time. We will often write ν_i instead of $\nu_i^{(0)}$ for convenience.

In recent years the literature on coalescent processes increased rapidly. More and more problems around these processes have been addressed and got solved. However, the number of open problems increased as well. In this paper it is discussed how convergence results for ancestral processes and convergence results for times back to the most recent common ancestor (MRCA) are related.

The paper is structured as follows. In Sections 2 and 3 we briefly recall the well-known theory on time-discrete ancestral processes and their convergence to time-continuous limiting coalescent processes as the total population size N tends to infinity. In Section 4 we show that convergence of the time-scaled ancestral process always implies the convergence of the corresponding times back to the MRCA (Proposition 4.1). More delicate is the converse question, i.e. whether the convergence of the times back to the MRCA already ensures the convergence of the time-scaled ancestral processes to some limiting ancestral process. The answer to this question depends on the type of the limiting coalescent process. In Section 5 we give a positive answer to this question for the situation when the population model is in the domain of attraction of the Kingman coalescent (Proposition 5.1). In Section 6 (see Theorem 6.4) the result is extended to the class of coalescent processes allowing for multiple collisions. For more information on such processes we refer to Pitman [15] and Sagitov [16]. In Section 7 we finally verify that results of this type cannot be extended to the full class of exchangeable coalescent processes ([13], [17], [18]) allowing for simultaneous multiple mergers of ancestral lineages.

2 Time-discrete coalescent processes

Consider an exchangeable population model with total population size N as introduced in Section 1. Fix $n \in \{1,\dots,N\}$ and sample n individuals from the current generation. For $r \in \mathbb{N}_0 := \{0,1,2,\dots\}$ let $\mathcal{R}_r := \mathcal{R}_r^{(n)}(N)$ denote the random equivalence relation on $\{1,\dots,n\}$ containing (i,j) if and only if the individuals i and j have a common ancestor r generations backward in time. It is well known (Kingman [6]) that $\mathcal{R} := (\mathcal{R}_r)_{r\in\mathbb{N}_0}$ is a time-homogeneous Markov chain with state space $\mathcal{E}_n$, the set of all equivalence relations on $\{1,\dots,n\}$, and initial state $\mathcal{R}_0 \equiv \Delta_n := \{(i,i) \,|\, 1 \le i \le n\}$, the diagonal relation. The transition probabilities $p_{\xi\eta} = p_{\xi\eta}(N) := P(\mathcal{R}_r = \eta \,|\, \mathcal{R}_{r-1} = \xi)$, $\xi,\eta \in \mathcal{E}_n$, are equal to zero for $\xi \not\subseteq \eta$ and equal to

$$p_{\xi\eta} \;=\; \frac{(N)_a}{(N)_b}\mathrm{E}((\nu_1)_{b_1}\cdots(\nu_a)_{b_a}) \;=:\; \Phi_a^{(N)}(b_1,\dots,b_a), \tag{1}$$

for $\xi \subseteq \eta$, where $(x)_0 := 1$ and $(x)_k := x(x-1)\cdots(x-k+1)$ for $x \in \mathbb{R}$ and $k \in \mathbb{N}$. Here $a = |\eta|$ denotes the number of equivalence classes (blocks) of η, $b = |\xi|$ the number of classes of ξ, and $b_1,\dots,b_a$ are the group sizes of merging classes of ξ ($\Rightarrow b_1 + \cdots + b_a = b$). The process $\mathcal{R}$ is called a discrete ancestral process or a discrete n-coalescent.

The functions $\Phi_j := \Phi_j^{(N)}$, $j \in \{1,\dots,N\}$, are consistent (see, for example, Eq. (3) of [10]) in the

sense that

$$\Phi_{a+1}(b_1,\dots,b_a,1) \;=\; \Phi_a(b_1,\dots,b_a) - \sum_{i=1}^{a} \Phi_a(b_1,\dots,b_{i-1},b_i+1,b_{i+1},\dots,b_a) \tag{2}$$

for $a, b_1, \dots, b_a \in \mathbb{N}$ with $b := b_1 + \cdots + b_a < N$. In the appendix (Lemma 8.1) it is shown that the consistency (2) implies that the functions Φ_j, $j \in \{1,\dots,N\}$, are monotone in the sense that

$$\Phi_j(k_1,\dots,k_j) \;\le\; \Phi_l(m_1,\dots,m_l) \tag{3}$$

for $1 \le l \le j \le N$ and $k_1,\dots,k_j,m_1,\dots,m_l \in \mathbb{N}$ with $k_1 \ge m_1,\dots,k_l \ge m_l$ and $k_1+\cdots+k_j \le N$.

Instead of considering the random relation $\mathcal{R}_r$ it is often convenient to record in a sample of size $n \in \{1,\dots,N\}$ only the number $\mathcal{D}_r = \mathcal{D}_r^{(n)}(N) := |\mathcal{R}_r^{(n)}(N)|$ of ancestors r generations backward in time. It is well known and follows easily from (1) that $\mathcal{D} := (\mathcal{D}_r)_{r\in\mathbb{N}_0}$ is a Markov chain with state space $\{1,\dots,n\}$, initial state $\mathcal{D}_0 \equiv n$, and transition probabilities $p_{ij} = p_{ij}(N) := P(\mathcal{D}_r = j \,|\, \mathcal{D}_{r-1} = i)$ given by (see Kingman [5] - [7])

$$\begin{aligned} p_{ij} \;&=\; \frac{i!}{j!} \sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ i_1+\cdots+i_j=i}} \frac{\Phi_j(i_1,\dots,i_j)}{i_1!\cdots i_j!} \\ &=\; \frac{\binom{N}{j}}{\binom{N}{i}} \sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ i_1+\cdots+i_j=i}} \mathrm{E}\left(\binom{\nu_1}{i_1}\cdots\binom{\nu_j}{i_j}\right), \qquad i,j \in \{1,\dots,n\}. \end{aligned} \tag{4}$$

Note that $p_{ij} = 0$ for $i < j$ and, therefore, the transition matrix $(p_{ij})_{i,j\in\{1,\dots,n\}}$ has eigenvalues

$$\lambda_i \;=\; \lambda_i(N) \;:=\; p_{ii} \;=\; \Phi_i(1,\dots,1) \;=\; \mathrm{E}(\nu_1\cdots\nu_i), \quad i \in \{1,\dots,n\}. \tag{5}$$

Obviously, the consistency relation (2) puts certain constraints on the transition probabilities p_{ij}. It does not seem to be straightforward to describe these constraints directly in terms of the p_{ij}. The process $\mathcal{D}$ is called the block-counting process or class-counting process.

3 Convergence to time-continuous coalescent processes

The consistency (2) and the monotonicity (3) of the functions Φ_j, $j \in \{1,\dots,N\}$, are two important properties needed in order to prove convergence-to-the-coalescent results. All these convergence-to-the-coalescent results have of course been stimulated by the seminal work of Kingman [5] and finally led to a full classification of all possible limiting processes for the case of exchangeable reproduction. For detailed information on convergence-to-the-coalescent results we refer to [13], [17] and [18]. In order to state such results we need to introduce for $N \in \mathbb{N}\setminus\{1\}$ the so-called *coalescence probability*

$$c_N \;:=\; p_{21}(N) \;=\; \Phi_1^{(N)}(2) \;=\; \frac{\mathrm{E}(\nu_1(\nu_1-1))}{N-1} \;=\; \frac{\mathrm{Var}(\nu_1)}{N-1},$$

i.e., the probability that two individuals, randomly chosen from some generation, are descended from the same parent. Note that $c_N = 0$ if and only if $P(\nu_1 = 1) = 1$. The coalescence probability c_N (see,

for example, [9], [13]) is a crucial quantity in coalescent theory. Roughly speaking, the time has to be measured in units of $\lfloor t/c_N \rfloor$ generations in order to achieve convergence to the coalescent as the total population size N tends to infinity. The following convergence result is essentially Theorem 2.1 of [13], however we state this theorem in the notation used in [12].

Theorem 3.1 *Assume that $c_N > 0$ for all sufficiently large N, that $\lim_{N\to\infty} c_N = 0$, and that the limits*

$$\phi_j(k_1,\ldots,k_j) := \lim_{N\to\infty} \frac{\Phi_j^{(N)}(k_1,\ldots,k_j)}{c_N}, \tag{6}$$

exist for all $j \in \mathbb{N}$ and $k_1 \geq \cdots \geq k_j \geq 2$. Then, the limits (6) exist for the wider range of parameters $j, k_1, \ldots, k_j \in \mathbb{N}$ satisfying $k_1 + \cdots + k_j > j$ and also the limits

$$\phi_j(1,\ldots,1) := \lim_{N\to\infty} \frac{\Phi_j^{(N)}(1,\ldots,1) - 1}{c_N}$$

exist for all $j \in \mathbb{N}$. Moreover, for each sample size $n \in \mathbb{N}$, the time-scaled process $(\mathcal{R}^{(n)}_{\lfloor t/c_N \rfloor}(N))_{t\geq 0}$ converges in the Skorohod topology as N tends to infinity to a time-continuous Markov chain $(R_t^{(n)})_{t\geq 0}$, with transition matrices e^{tQ}, $t \geq 0$, and generator $Q = (q_{\xi\eta})_{\xi,\eta\in\mathcal{E}_n}$ with entries

$$q_{\xi\eta} := \begin{cases} \phi_a(b_1,\ldots,b_a) & \text{if } \xi \subseteq \eta, \\ 0 & \text{otherwise,} \end{cases}$$

where $a := |\eta|$ and $b_1, \ldots, b_a \in \mathbb{N}$ are the group sizes of merging classes of ξ.

Note that, in general, the limiting process $(R_t^{(n)})_{t\geq 0}$ allows for simultaneous multiple collisions of ancestral lineages. It is therefore called an n-coalescent with simultaneous multiple collisions.

Theorem 3.1 remains valid, if the population size N is replaced by N_l, where $(N_l)_{l\in\mathbb{N}}$ is a subsequence of population sizes satisfying $\lim_{l\to\infty} N_l = \infty$. In that case of course all limits '$N \to \infty$' appearing in Theorem 3.1 have to be interpreted as '$l \to \infty$'.

Convergence to Kingman's n-coalescent $(R_t^{(n)})_{t\geq 0}$, which allows only for binary mergers of ancestral lineages, occurs if all the limits in (6) are zero except for $\phi_1(2)$. The following convergence theorem is one of the fundamental results of coalescent theory.

Theorem 3.2 *Assume that $c_N > 0$ for all sufficiently large N. Then, the following conditions are equivalent.*

(i) $\phi_1(3) := \lim_{N\to\infty} \Phi_1^{(N)}(3)/c_N = 0$, *i.e., triple mergers of ancestral lineages are asymptotically negligible in comparison with binary mergers.*

(ii) For each sample size $n \in \mathbb{N}$, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_N \rfloor})_{t\geq 0}$ converges in the Skorohod topology to Kingman's n-coalescent $(R_t^{(n)})_{t\geq 0}$ as $N \to \infty$.

Remark: Note (see Theorem 6.1 of [10]) that Theorem 3.2 remains valid if in (ii) the symbol $\mathcal{R}$ is replaced by $\mathcal{D}$ and $R_t^{(n)}$ by $D_t^{(n)} := |R_t^{(n)}|$. There is also an equivalent formulation of (i) in terms of the eigenvalues λ_2 and λ_3 introduced in (5), namely

(iii) $\lim_{N\to\infty} \dfrac{1-\lambda_3(N)}{1-\lambda_2(N)} \;=\; 3.$

The equivalence of (i) and (iii) is seen as follows. The consistency (2) implies that $\lambda_2 = \Phi_2(1,1) = \Phi_1(1) - \Phi_1(2) = 1 - c_N$ and that $\lambda_3 = \Phi_3(1,1,1) = \Phi_2(1,1) - 2\Phi_2(2,1) = 1 - c_N - 2\Phi_2(2,1)$, i.e. $\Phi_2(2,1) = (1 - c_N - \lambda_3)/2$. Therefore, again using (2),

$$\Phi_1(3) \;=\; \Phi_1(2) - \Phi_2(2,1) \;=\; c_N - \frac{1-c_N-\lambda_3}{2} \;=\; \frac{3}{2}c_N - \frac{1-\lambda_3}{2}$$

or, equivalently, $2\Phi_1(3)/c_N = 3 - (1-\lambda_3)/c_N$, from which the equivalence of (i) and (iii) follows immediately.

At the end of Section 5 further equivalent conditions are formulated involving times back to the most recent common ancestor.

4 Time back to the most recent common ancestor

For $N \in \mathbb{N}$ and $n \in \{1,\ldots,N\}$ let $T_{n,N} := \inf\{r \in \mathbb{N}_0 \,:\, \mathcal{D}_r^{(n)}(N) = 1\}$ denote the number of generations back to the most recent common ancestor (MRCA) of a sample of size n.

Proposition 4.1 *Assume that the conditions of Theorem 3.1 or Theorem 3.2 are satisfied. Then, for each sample size $n \in \mathbb{N}$, as $N \to \infty$, the scaled time $c_N\, T_{n,N}$ back to the MRCA converges in distribution to*

$$T_n \;:=\; \inf\{t > 0 : D_t^{(n)} = 1\},$$

where $D_t^{(n)} := |R_t^{(n)}|$, $t \geq 0$, and $(R_t^{(n)})_{t\geq 0}$ is the limiting n-coalescent with simultaneous multiple collisions appearing in Theorem 3.1 or Kingman's n-coalescent appearing in Theorem 3.2.

Proof: Fix $n \in \mathbb{N}$ and $t \geq 0$. By Theorem 3.1 or Theorem 3.2, $\mathcal{R}^{(n)}_{\lfloor t/c_N\rfloor} \xrightarrow{d} R_t^{(n)}$ as $N \to \infty$. As the function $\mathcal{E}_n \ni \xi \mapsto |\xi| \in \{1,\ldots,n\}$ is continuous, it follows that $\mathcal{D}^{(n)}_{\lfloor t/c_N\rfloor} \xrightarrow{d} D_t^{(n)}$ as $N \to \infty$. Thus, for $t \geq 0$,

$$P(c_N T_{n,N} > t) \;=\; P(T_{n,N} > \lfloor t/c_N\rfloor) \;=\; P(\mathcal{D}^{(n)}_{\lfloor t/c_N\rfloor} > 1) \;\to\; P(D_t^{(n)} > 1) \;=\; P(T_n > t),$$

which implies that $c_N T_{n,N} \xrightarrow{d} T_n$ as $N \to \infty$. □

Remark: If the eigenvalues $\lambda_i := \lambda_i(N)$, $i \in \{1,\ldots,N\}$, are distinct, then there exists the following alternative proof of Proposition 4.1 based on an explicit formula for the distribution function of $T_{n,N}$. Fix $n \in \mathbb{N}$ and $t \geq 0$. By Lemma 3.1. of [11], for $N \geq n$,

$$P(c_N T_{n,N} > t) \;=\; \sum_{k=1}^{n-1} \sum_{i_0,\ldots,i_k} \frac{p_{i_k,i_{k-1}}}{1-\lambda_{i_k}} \cdots \frac{p_{i_1,i_0}}{1-\lambda_{i_1}} \sum_{m=1}^{k} \lambda_{i_m}^{\lfloor t/c_N\rfloor} \prod_{\substack{j=1\\ j\neq m}}^{k} \frac{1-\lambda_{i_j}}{\lambda_{i_m}-\lambda_{i_j}}, \tag{7}$$

where the second sum extends over all integers $i_0,\ldots,i_k$ satisfying $1 = i_0 < i_1 < \cdots < i_{k-1} < i_k = n$. Under the conditions of Theorem 3.1 or Theorem 3.2, the asymptotic formula

$$p_{ij}(N) \;=\; \delta_{ij} + c_N g_{ij} + o(c_N), \qquad i,j \in \{1,\ldots,n\} \tag{8}$$

holds, where δ_{ij} denotes the Kronecker symbol and $G = (g_{ij})_{i,j\in\{1,\ldots,n\}}$ is the generator of the process $(D_t^{(n)})_{t\geq 0}$ with entries

$$g_{ij} \;=\; \frac{i!}{j!} \sum_{\substack{i_1,\ldots,i_j\in\mathbb{N}\\ i_1+\cdots+i_j=i}} \frac{\phi_j(i_1,\ldots,i_j)}{i_1!\cdots i_j!}, \tag{9}$$

$i,j\in\{1,\ldots,n\}$. Now, (8) ensures that, for $i,j\in\{1,\ldots,n\}$ with $i\neq j$,

$$\lambda_i^{\lfloor t/c_N\rfloor}\to e^{-g_i t}, \quad \frac{1-\lambda_j}{\lambda_i-\lambda_j}\to\frac{g_j}{g_j-g_i}, \quad \text{and} \quad \frac{p_{ij}}{1-\lambda_i}\to\frac{g_{ij}}{g_i} \;=:\; r_{ij}$$

with $g_i := -g_{ii}$, $i\in\mathbb{N}$. From (7) it follows that

$$\lim_{N\to\infty} P(c_N T_{n,N}>t) \;=\; \sum_{k=1}^{n-1}\sum_{i_0,\ldots,i_k} r_{i_k,i_{k-1}}\cdots r_{i_1,i_0}\sum_{m=1}^{k} e^{-g_{i_m}t}\prod_{\substack{j=1\\ j\neq m}}^{k}\frac{g_{i_j}}{g_{i_j}-g_{i_m}} \;=\; P(T_n>t). \tag{10}$$

The alternative proof of Proposition 4.1 is complete.

Remark: Note that the alternative proof above is valid as long as the eigenvalues $\lambda_1(N),\ldots,\lambda_N(N)$ are distinct for all sufficiently large N.

The sequence $(T_n)_{n\in\mathbb{N}}$ satisfies (see, for example, Eq. (2) of [4]) the distributional recursion $T_1=0$ and $T_n\stackrel{d}{=}\tau_n+T_{I_n}$, $n\in\{2,3,\ldots\}$, where τ_n is exponentially distributed with parameter g_n and I_n is a random variable independent of $T_2,\ldots,T_{n-1},\tau_n$ with distribution $P(I_n=k)=r_{nk}=g_{nk}/g_n$, $k\in\{1,\ldots,n-1\}$.

Under the conditions of Theorem 3.1 or Theorem 3.2 we have, for each $n\in\mathbb{N}$, pointwise convergence of the corresponding transforms $\lim_{N\to\infty}\mathrm{E}(e^{zc_N T_{n,N}})=\mathrm{E}(e^{zT_n})$, $z\in\mathbb{C}$, $\mathrm{Re}(z)<g_2$.

Under the conditions of Theorem 3.2 (Kingman case), $T_n\stackrel{d}{=}\tau_2+\cdots+\tau_n$, $n\in\mathbb{N}$, where $\tau_2,\tau_3,\ldots$ are independent random variables and τ_i is exponentially distributed with parameter $g_i=i(i-1)/2$, $i\in\{2,3,\ldots\}$. In particular,

$$\mathrm{E}(e^{zT_n}) \;=\; \prod_{i=2}^{n}\mathrm{E}(e^{z\tau_i}) \;=\; \prod_{i=2}^{n}\frac{g_i}{g_i-z}, \quad z\in\mathbb{C},\mathrm{Re}(z)<g_2.$$

The rest of this section deals with a result similar to that provided in Proposition 4.1, but for the situation when the sample size $n=n_N$ depends on the total population size N.

Theorem 4.2 *Suppose that the conditions of Theorem 3.1 or Theorem 3.2 are satisfied. Then, for any sequence $(n_N)_{N\in\mathbb{N}}$ satisfying $n_N\in\{1,\ldots,N\}$ and $\lim_{N\to\infty} n_N=\infty$, the process $(\mathcal{D}^{(n_N)}_{\lfloor t/c_N\rfloor})_{t\geq 0}$ converges in the Skorohod topology to $(D_t)_{t\geq 0}$ as $N\to\infty$, where $(D_t)_{t\geq 0}$ is a non-increasing Markov process with state space $\mathbb{N}$ and generator $G=(g_{ij})_{i,j\in\mathbb{N}}$ with entries (9), $i,j\in\mathbb{N}$.*

Proof: The proof is identical to that of Theorem 3 in [3]. Note that the formula for the rates g_{ij} can be easily read off from (4) or (9). □

Corollary 4.3 *Suppose that the conditions of Theorem 3.1 or 3.2 are satisfied. Then, for any sequence* $(n_N)_{N\in\mathbb{N}}$ *satisfying* $n_N \in \{1,\dots,N\}$ *and* $\lim_{N\to\infty} n_N = \infty$*, the scaled time* $c_N T_{n_N,N}$ *back to the MRCA converges in distribution to* $T := \inf\{t > 0 : D_t = 1\}$.

Proof: By Theorem 4.2, for $t \geq 0$, $P(c_N T_{n_N,N} > t) = P(\mathcal{D}^{(n_N)}_{\lfloor t/c_N\rfloor} > 1) \to P(D_t > 1) = P(T > t)$. □ The proof of Proposition 4.1 shows that, whenever the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_N\rfloor})_{t\geq 0}$ converges to an exchangeable n-coalescent, then the convergence $c_N T_{n,N} \xrightarrow{d} T_n$ holds for the corresponding times $T_{n,N}$ and T_n back to the MRCA. In the following we are interested in the converse of Proposition 4.1. Is the convergence $c_N\, T_{n,N} \xrightarrow{d} T_n$, $n \in \mathbb{N}$, where $(T_n)_{n\in\mathbb{N}}$ is some sequence of random variables, sufficient to ensure convergence in the Skorohod topology of the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_N\rfloor})_{t\geq 0}$ to some n-coalescent process $(R^{(n)}_t)_{t\geq 0}$, $n \in \mathbb{N}$? The answer to this questions is less obvious and more subtile as it looks at a first glance. We therefore approach this question in three steps. First we consider the Kingman case (Section 5). Afterwards, the results are extended to coalescents with multiple collisions (Section 6). Finally we address this problem for the full class of exchangeable coalescents allowing for simultaneous multiple collisions of ancestral lineages (Section 7).

5 The Kingman case

The following proposition is a kind of converse of Proposition 4.1 for the Kingman case.

Proposition 5.1 *Assume that, for* $n \in \{2,3\}$, $c_N T_{n,N} \xrightarrow{d} T_n := \sum_{i=2}^n \tau_i$ *as* $N \to \infty$*, where* τ_2, τ_3 *are independent and* τ_i *is exponentially distributed with parameter* $g_i := i(i-1)/2$, $i \in \{2,3\}$*. Then, the condition (i) (and hence (ii)) of Theorem 3.2 holds.*

Proof: Define $\mu_i := 1 - \lambda_i$ for convenience. Obviously, $c_N = p_{21} = 1 - p_{22} = 1 - \lambda_2 = \mu_2$. For $n = 2$, Eq. (7) reduces to

$$P(c_N T_{2,N} > t) \;=\; \frac{p_{21}}{1-\lambda_2}\lambda_2^{\lfloor t/c_N\rfloor} \;=\; (1-c_N)^{\lfloor t/c_N\rfloor}.$$

Now let $N \to \infty$ to conclude that $\lim_{N\to\infty}(1-c_N)^{\lfloor t/c_N\rfloor} = P(T_2 > t) = e^{-g_2 t} = e^{-t}$. Taking the logarithm yields $\lim_{N\to\infty}\lfloor t/c_N\rfloor \log(1-c_N) = -t$. Thus, $c_N \to 0$ as $N \to \infty$. Next, consider (7) for $n = 3$, i.e.,

$$\begin{aligned} P(c_N T_{3,N} > t) &= \frac{p_{31}}{\mu_3}\lambda_3^{\lfloor t/c_N\rfloor} + \frac{p_{32}}{\mu_3}\frac{p_{21}}{\mu_2}\Big(\lambda_2^{\lfloor t/c_N\rfloor}\frac{\mu_3}{\mu_3-\mu_2} + \lambda_3^{\lfloor t/c_N\rfloor}\frac{\mu_2}{\mu_2-\mu_3}\Big) \\ &= \frac{p_{31}}{\mu_3}\lambda_3^{\lfloor t/c_N\rfloor} + \Big(1-\frac{p_{31}}{\mu_3}\Big)\Big(\lambda_2^{\lfloor t/c_N\rfloor}\frac{\mu_3}{\mu_3-\mu_2} + \lambda_3^{\lfloor t/c_N\rfloor}\frac{\mu_2}{\mu_2-\mu_3}\Big), \end{aligned}$$

as $p_{21} = \mu_2$ and $p_{32} = \mu_3 - p_{31}$. Define $x := x_N := \mu_2/\mu_3$ and $y := y_N := p_{31}/p_{21} = p_{31}/\mu_2$. From the remark after Theorem 3.2 it is known that x and y are related via

$$y \;=\; \frac{3}{2} - \frac{1}{2x}.$$

We have

$$\begin{aligned} P(c_N T_{3,N} > t) &= xy\lambda_3^{\lfloor t/c_N \rfloor} + (1-xy)\Big(\lambda_2^{\lfloor t/c_N \rfloor}\frac{1}{1-x} + \lambda_3^{\lfloor t/c_N \rfloor}\frac{x}{x-1}\Big) \\ &= \frac{1-xy}{1-x}\lambda_2^{\lfloor t/c_N \rfloor} + \frac{x(1-y)}{x-1}\lambda_3^{\lfloor t/c_N \rfloor} = \frac{3}{2}\lambda_2^{\lfloor t/c_N \rfloor} - \frac{1}{2}\lambda_3^{\lfloor t/c_N \rfloor}, \qquad (11) \end{aligned}$$

as $xy = (3x-1)/2$ and hence $(1-xy)/(1-x) = 3/2$. By assumption, as $N \to \infty$, (11) converges to

$$P(T_3 > t) = \frac{g_3}{g_3 - g_2}e^{-g_2 t} + \frac{g_2}{g_2 - g_3}e^{-g_3 t} = \frac{3}{2}e^{-t} - \frac{1}{2}e^{-3t}.$$

As $\lambda_2^{\lfloor t/c_N \rfloor} \to e^{-t}$, the convergence

$$\lim_{N\to\infty} \lambda_3^{\lfloor t/c_N \rfloor} = e^{-3t}$$

must hold. Thus, $[t/c_N]\log(1-\mu_3) \to -3t$, and, hence, $\log(1-\mu_3)/\mu_2 \to -3$, which is only possible, if $\mu_3/\mu_2 \to 3$. Thus, $x_N \to 1/3$, or, equivalently, $y_N \to 0$, and Proposition 5.1 is established. □

Remark: For $n \in \mathbb{N}$ let $T_n := \sum_{i=2}^n \tau_i$, where $\tau_2, \tau_3, \ldots$ are independent random variables and τ_i is exponentially distributed with parameter $g_i = i(i-1)/2$, $i \in \mathbb{N} \setminus \{1\}$. Proposition 5.1 together with Proposition 4.1 show that condition (i) of Theorem 3.2 is equivalent to

(iv) $c_N\, T_{n,N} \xrightarrow{d} T_n$ as $N \to \infty$ for $n \in \{2,3\}$.

Since condition (i) of Theorem 3.2 implies (ii) of the same theorem, an application of Theorem 4.1 shows that (iv) is also equivalent to

(v) $c_N\, T_{n,N} \xrightarrow{d} T_n$ as $N \to \infty$ for all $n \in \mathbb{N}$.

6 Multiple collisions

We now generalize Proposition 5.1 to coalescents with multiple collisions. Roughly speaking, we will verify that convergence of all the times back to the most recent common ancestor already implies convergence of the time-scaled ancestral processes.

Exchangeable coalescents are time-continuous $\mathcal{E}$-valued Markov processes, where $\mathcal{E}$ denotes the set of all equivalence relations on $\mathbb{N}$. During each transition equivalence classes are allowed to merge together. Pitman [15] and Sagitov [16] independently introduced the class of coalescents allowing for multiple collisions, also called Λ-coalescents as they can be characterized by a finite measure Λ on the unit interval $[0,1]$. The full class of exchangeable coalescent processes allowing for simultaneous multiple mergers of ancestral lineages has been studied by Möhle and Sagitov [13] and Schweinsberg [18]. For $n \in \mathbb{N}$ let $\varrho_n : \mathcal{E} \to \mathcal{E}_n$ denote the natural restriction to the set $\mathcal{E}_n$ of all equivalence relations on $\{1, \ldots, n\}$.

Lemma 6.1 *Assume that* $\lim_{N\to\infty} c_N = 0$. *Then there exists a subsequence* $(N_l)_{l\in\mathbb{N}}$ *of positive integers with* $\lim_{l\to\infty} N_l = \infty$ *and an exchangeable coalescent* $R = (R_t)_{t\geq 0}$ *such that, for each sample size* $n \in \mathbb{N}$, *the time-scaled ancestral process* $(\mathcal{R}^{(n)}_{\lfloor t/c_{N_l} \rfloor})_{t\geq 0}$ *converges in the Skorohod topology to* $(\varrho_n R_t)_{t\geq 0}$ *as* $l \to \infty$.

Proof: Let $k_1, \ldots, k_j \in \mathbb{N}$ with $k_1 + \cdots + k_j > j$. The monotonicity (3) ensures that $\Phi_j^{(N)}(k_1, \ldots, k_j) \leq \Phi_1(2) = c_N$. Thus, the sequence $(\Phi_j^{(N)}(k_1, \ldots, k_j)/c_N)_{N \in \mathbb{N}}$ is bounded. Because these are countable many such sequences (for each j there are '$|\mathbb{N}^j|-1$' such sequences), there exists a (diagonal) subsequence $(N_l)_{l \in \mathbb{N}}$ with $\lim_{l \to \infty} N_l = \infty$ such that the limit $\phi_j(k_1, \ldots, k_j) := \lim_{l \to \infty} \Phi_j^{(N_l)}(k_1, \ldots, k_j)/c_{N_l}$ exists for all $j, k_1, \ldots, k_j \in \mathbb{N}$ with $k_1 + \cdots + k_j > j$. Now apply Theorem 3.1 (with N replaced by N_l) to conclude that, for each sample size n, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_{N_l} \rfloor})_{t \geq 0}$ converges in the Skorohod topology to $R^{(n)}$ as $l \to \infty$, where $R^{(n)} = (R_t^{(n)})_{t \geq 0}$ is some n-coalescent allowing for simultaneous multiple collisions. The sequence $(R^{(n)})_{n \in \mathbb{N}}$ of n-coalescents is consistent in the sense that for sample sizes $m, n \in \mathbb{N}$ with $m \leq n$ the process $(\varrho_m R_t^{(n)})_{t \geq 0}$ is distributional equal to $R^{(m)} = (R_t^{(m)})_{t \geq 0}$. Therefore, by an application of Kolmogoroff's extension theorem, there exists an exchangeable coalescent process $R = (R_t)_{t \geq 0}$ such that the distribution of $(\varrho_n R_t)_{t \geq 0}$ coincides with that of $R^{(n)}$ for all $n \in \mathbb{N}$. □

Lemma 6.2 *Assume that* $\lim_{N \to \infty} c_N = 0$ *and that, for each sample size* $n \in \mathbb{N}$*, the time-scaled number of generations back to the MRCA* $c_N T_{n,N}$ *converges in distribution to some random variable* W_n *as* $N \to \infty$*. Then, there exists an exchangeable coalescent process* $R = (R_t)_{t \geq 0}$ *such that* $W_n \stackrel{d}{=} T_n := \inf\{t > 0 : |\varrho_n R_t| = 1\}$ *for all* $n \in \mathbb{N}$*.*

Proof: By Lemma 6.1 there exists a subsequence $(N_l)_{l \in \mathbb{N}}$ of positive integers with $\lim_{l \to \infty} N_l = \infty$ and some exchangeable coalescent process $R = (R_t)_{t \geq 0}$ such that, for each $n \in \mathbb{N}$, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_{N_l} \rfloor})_{t \geq 0}$ converges in the Skorohod topology to $(\varrho_n R_t)_{t \geq 0}$ as $l \to \infty$. In particular, for all sample sizes $n \in \mathbb{N}$ and all times $t \geq 0$,

$$P(c_{N_l} T_{n,N_l} > t) \;=\; P(|\mathcal{R}^{(n)}_{\lfloor t/c_{N_l} \rfloor}| > 1) \;\to\; P(|\varrho_n R_t| > 1) \;=\; P(T_n > t),$$

i.e., $c_{N_l} T_{n,N_l} \stackrel{d}{\to} T_n$ as $l \to \infty$ for all $n \in \mathbb{N}$. By assumption, $c_N T_{n,N} \stackrel{d}{\to} W_n$ for all $n \in \mathbb{N}$. Therefore, $W_n \stackrel{d}{=} T_n$ for all $n \in \mathbb{N}$. □

Remark: Let R be an exchangeable coalescent process. In general (see the following section for more details), the marginal distributions of the times $T_n := \inf\{t > 0 : |\varrho_n R_t| = 1\}$, $n \in \{2, 3, \ldots\}$, do not determine the distribution of R completely. In particular, the uniqueness of the (distribution of the) coalescent R in Lemma 6.2 is in general not guaranteed. However, if it is in addition know that the coalescent process R allows only for multiple mergers of ancestral lineages, i.e. if R is a Λ-coalescent with Λ some finite measure on the unit interval $[0, 1]$, then (see the following Lemma 6.3) the distribution of R is already uniquely determined by the marginal distributions of the times T_n, $n \in \{2, 3, \ldots\}$.

Lemma 6.3 *The measure* Λ *of a* Λ*-coalescent* $R = (R_t)_{t \geq 0}$ *is uniquely determined by the marginal distributions of the times* $T_n := \inf\{t > 0 : |\varrho_n R_t| = 1\}$*,* $n \in \{2, 3, \ldots\}$*.*

Proof: The times T_n can take the value ∞ with positive probability only if $\Lambda \equiv 0$. Thus, without loss of generality we can and do assume that $\Lambda \neq 0$, in which case all the random variables T_n, $n \in \mathbb{N}$, are

almost surely real valued. Assume for a moment that Λ is not concentrated in 1. Then, the basic formula

$$g_{n+1} - g_n \;=\; n \int_{[0,1]} (1-x)^{n-1}\, \Lambda(dx), \qquad n \in \mathbb{N}, \tag{12}$$

(which follows easily from (14)) implies that the sequence $(g_n)_{n\in\mathbb{N}}$ of total rates is strictly increasing. In particular, the total rates are pairwise distinct, and, therefore, as in (10), for $n \in \{2,3,\ldots\}$ and $t \geq 0$,

$$P(T_n > t) \;=\; \sum_{k=1}^{n-1} \sum_{i_0,\ldots,i_k} r_{i_k,i_{k-1}} \cdots r_{i_1,i_0} \sum_{m=1}^{k} e^{-g_{i_m} t} \prod_{\substack{j=1 \\ j\neq m}}^{k} \frac{g_{i_j}}{g_{i_j} - g_{i_m}},$$

where the probabilities $r_{ij} := g_{ij}/g_i$, $1 \leq j < i \leq n$, are the transition probabilities of the jump process of $(|\varrho_n R_t|)_{t\geq 0}$ and the second sum extends over all integers $i_0, \ldots, i_k$ satisfying $1 = i_0 < i_1 < \cdots < i_{k-1} < i_k = n$. The above formula shows that

$$P(T_n > t) \;=\; \sum_{k=2}^{n} a_{nk} e^{-g_k t}, \qquad t \geq 0, \tag{13}$$

where the a_{nk} are some non-zero coefficients which do not depend on t (we do not need to know these coefficients in detail). If Λ is concentrated in 1, then $g_k = 1$ for all $k \in \{2,3,\ldots\}$ and (13) holds as well, if we choose $a_{nk} := 1/(n-1)$ for all $k \in \{2,\ldots,n\}$. From (13) it follows that the total rates g_k, $k \in \{2,\ldots,n\}$, are uniquely determined by the distribution of T_n. Thus, the total rates g_n, $n \in \mathbb{N}$, are uniquely determined by the marginal distributions of the times T_n, $n \in \mathbb{N}$. From (12) it follows that the moments of Λ are uniquely determined by the sequence $(g_n)_{n\in\mathbb{N}}$. As the measure Λ is concentrated on $[0,1]$, it is uniquely determined by its moments which completes the proof. □ If the given sequence of Cannings models is in the domain of attraction of a coalescent with multiple collisions (Λ-coalescent), i.e., if $\phi_2(2,2) := \lim_{N\to\infty} \Phi_2^{(N)}(2,2)/c_N = 0$, then Lemma 6.2 can be strengthened as follows.

Theorem 6.4 *Suppose that $c_N > 0$ for sufficiently large N, that $\lim_{N\to\infty} c_N = 0$, that $\phi_2(2,2) := \lim_{N\to\infty} \Phi_2^{(N)}(2,2)/c_N = 0$, and that, for each sample size $n \in \mathbb{N}$, the time-scaled number of generations $c_N T_{n,N}$ back to the MRCA converges in distribution to some random variable W_n as $N \to \infty$. Then, there exists a Λ-coalescent $R = (R_t)_{t\geq 0}$ such that*

(i) $W_n \stackrel{d}{=} T_n := \inf\{t > 0 : |\varrho_n R_t| = 1\}$ for all $n \in \mathbb{N}$ and

(ii) for each sample size $n \in \mathbb{N}$, as N tends to infinity, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_N \rfloor})_{t\geq 0}$ converges in the Skorohod topology to $(\varrho_n R_t)_{t\geq 0}$.

The measure Λ is uniquely determined by (i).

Proof: Let $(N_l^{(0)})_{l\in\mathbb{N}}$ be some arbitrary subsequence of $\mathbb{N}$. As in the proof of Lemma 6.2, but starting with the arbitrary subsequence $(N_l^{(0)})_{l\in\mathbb{N}}$, it follows that there exists a subsequence $(N_l)_{l\in\mathbb{N}}$ of $(N_l^{(0)})_{l\in\mathbb{N}}$ such that, for each sample size $n \in \mathbb{N}$, as $l \to \infty$, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{\lfloor t/c_{N_l} \rfloor})_{t\geq 0}$ converges in the Skorohod topology to $(\varrho_n R_t)_{t\geq 0}$, where $R = (R_t)_{t\geq 0}$ is some exchangeable coalescent process. From

the assumption $\phi_2(2,2) = 0$ it follows that R cannot have simultaneous multiple collisions, i.e. R must be a Λ-coalescent. The convergence in the Skorohod topology in particular implies that $c_{N_l} T_{n,N_l} \stackrel{d}{\to} T_n$, $n \in \mathbb{N}$, and from the assumption $c_N T_{n,N} \stackrel{d}{\to} W_n$ we conclude that $W_n \stackrel{d}{=} T_n$ for all $n \in \mathbb{N}$, which proves (i). In order to verify (ii) it remains, by the criterion of subsequences, to show that the distribution of R does not depend on the particular subsequence $(N_l^{(0)})_{l\in\mathbb{N}}$. By Lemma 6.3, the distribution of R is uniquely determined by the sequence $(W_n)_{n\in\mathbb{N}}$. In particular, the distribution of R does not depend on the particular subsequence $(N_l^{(0)})_{l\in\mathbb{N}}$. □

7 Simultaneous multiple collisions

The results presented so far are essentially based on the property that the total rates

$$g_n := \lim_{t\searrow 0} \frac{P(|\varrho_n R_t| < n)}{t} = \int_{[0,1]} \frac{1-(1-x)^{n-1}(1-x+nx)}{x^2}\,\Lambda(dx), \quad n \in \mathbb{N}, \tag{14}$$

of a Λ-coalescent $R = (R_t)_{t\geq 0}$, i.e. a coalescent with multiple collisions, already determine the measure Λ completely. This property follows simply from (12) and the fact that - since the measure Λ is concentrated on the unit interval - the moments of Λ uniquely determine the measure Λ. Note that there is even a formula available (see Eq. (16) of [12]) which expresses the rates

$$\begin{aligned} g_{nk} &= \lim_{t\searrow 0} \frac{P(|\varrho_n R_t| = k)}{t} = \binom{n}{k-1} \phi_k(n-k+1,1,\ldots,1) \\ &= \binom{n}{k-1} \int_{[0,1]} x^{n-k-1}(1-x)^{k-1}\,\Lambda(dx), \end{aligned}$$

$n, k \in \mathbb{N}$ with $k < n$, directly in terms of the total rates g_n, $n \in \mathbb{N}$.

It is natural to ask whether the above mentioned property carries over to the wider class of exchangeable coalescent processes with simultaneous multiple collisions. In the following it is shown that, in general, the total rates do not determine the distribution of an exchangeable coalescent.

Schweinsberg [18] characterizes exchangeable coalescents via a finite measure Ξ on the infinite simplex $\Delta := \{x = (x_1, x_2, \ldots) : x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i=1}^\infty x_i \leq 1\}$. Decomposing $\Xi = \Xi_0 + a\delta_0$ with $0 \leq a < \infty$ and Ξ_0 having no atom at zero, each $(k_1,\ldots,k_j)$-collision ($k_1,\ldots,k_j \in \mathbb{N}$ with $k_1 \geq \cdots \geq k_j$ and $k_1 \geq 2$) occurs at the rate (see [18, Eq. (11)])

$$\phi_j(k_1,\ldots,k_j) = a\,1_{\{r=1,k_1=2\}} + \int_\Delta \sum_{l=0}^{s} \binom{s}{l} (1-|x|)^{s-l} \sum_{\substack{i_1,\ldots,i_{r+l}\in\mathbb{N} \\ \text{all distinct}}} x_{i_1}^{k_1}\cdots x_{i_{r+l}}^{k_{r+l}} \frac{\Xi_0(dx)}{(x,x)}, \tag{15}$$

where $s := |\{1 \leq i \leq j : k_i = 1\}|$ denotes the number of singletons, $r := j - s$, $|x| := \sum_{i=1}^\infty x_i$ and $(x,x) := \sum_{i=1}^\infty x_i^2$. In particular,

$$\phi_j(2,1,\ldots,1) = a + \int_\Delta \sum_{l=0}^{j-1} \binom{j-1}{l} (1-|x|)^{j-1-l} \sum_{\substack{i_1,\ldots,i_{l+1}\in\mathbb{N} \\ \text{all distinct}}} x_{i_1}^2 x_{i_2}\cdots x_{i_{l+1}} \frac{\Xi_0(dx)}{(x,x)}. \tag{16}$$

If the measure Ξ is concentrated on the subset of points $x \in \Delta$ satisfying $x_i = 0$ for $i \geq 3$, then only for $l \in \{0, 1\}$ the last sum in (16) is non-zero and it follows that

$$\phi_j(2, 1, \ldots, 1) \;=\; a + \int_\Delta (1 - x_1 - x_2)^{j-1} \Xi_0(dx) \\ + (j-1) \int_\Delta (1 - x_1 - x_2)^{j-2} x_1 x_2 (x_1 + x_2) \frac{\Xi_0(dx)}{x_1^2 + x_2^2}. \tag{17}$$

In the following we construct a class of Ξ-coalescents which all have the same total rates. Fix a constant $c \in (0, 1)$. If the measure Ξ is concentrated on the subset of points $x \in \Delta$ satisfying $x_i = 0$ for $i \geq 3$ and $x_1 + x_2 = c$, then (17) reduces to

$$\phi_j(2, 1, \ldots, 1) \;=\; (1-c)^{j-1} \Xi(\Delta) + (j-1) c (1-c)^{j-2} \int_\Delta x_1 x_2 \, \frac{\Xi(dx)}{x_1^2 + x_2^2}.$$

Let X be a random variable taking values in the interval $[c/2, c]$ and let Ξ denote the distribution of $(X, c - X, 0, 0, \ldots)$. Then, we have

$$\phi_j(2, 1, \ldots, 1) \;=\; (1-c)^{j-1} + (j-1) c (1-c)^{j-2} \mathrm{E}\Big(\frac{X(c-X)}{X^2 + (c-X)^2} \Big). \tag{18}$$

It is straightforward (Take, for example, $X \equiv 3c/4$ and Y such that $P(Y = c/2) = 3/5$ and $P(Y = c) = 2/5$.) to construct two (even infinitely many) random variables X and Y both taking values in $[c/2, c]$, such that $P_X \neq P_Y$ but

$$\mathrm{E}\Big(\frac{X(c-X)}{X^2 + (c-X)^2} \Big) \;=\; \mathrm{E}\Big(\frac{Y(c-Y)}{Y^2 + (c-Y)^2} \Big).$$

Let Ξ and Ξ' denote the distribution of $(X, c - X, 0, 0, \ldots)$ and $(Y, c - Y, 0, 0, \ldots)$ respectively. Then $\Xi \neq \Xi'$. Let R denote a standard Ξ-coalescent and R' denote a standard Ξ'-coalescent. The processes R and R' do not have the same distribution. However, from (18) and the general formula (see, for example, p. 1556 of [13])

$$g_n \;=\; \sum_{j=1}^{n-1} j \phi_j(2, 1, \ldots, 1), \qquad n \in \mathbb{N}$$

for the total rates $g_1, g_2, \ldots$ of exchangeable coalescent processes it follows that the total rates of R and R' coincide. Thus, in general the total rates do not determine the distribution of an exchangeable coalescent completely.

As a consequence, the marginal distributions of the times T_n, $n \in \{2, 3, \ldots\}$, do not determine the measure Ξ of the coalescent. Therefore, Lemma 6.3 cannot be extended to the full class of all exchangeable coalescents and, as a consequence, without the assumption $\phi_2(2, 2) = 0$ in Theorem 6.4, a coalescent R satisfying the condition (i) of Theorem 6.4 is in general not uniquely determined such that condition (ii) of Theorem 6.4 cannot hold in general.

8 Appendix

Lemma 8.1 *The functions Φ_j, $j \in \{1, \ldots, N\}$, defined via (1), are monotone in the sense that*

$$\Phi_j(k_1, \ldots, k_j) \;\leq\; \Phi_l(m_1, \ldots, m_l) \tag{19}$$

for $1 \le l \le j \le N$ *and* $k_1, \dots, k_j, m_1, \dots, m_l \in \mathbb{N}$ *with* $k_1 \ge m_1, \dots, k_l \ge m_l$ *and* $k_1 + \cdots + k_j \le N$.

Remark: The following proof of Lemma 8.1 uses only the consistency relation (2) and is hence shorter and more transparent than earlier proofs mentioned in [14] and going back to [13].

Proof: We prove (19) inductively on the difference $d := j - l \in \{0, \dots, N-1\}$. By (2),

$$\Phi_l(m_1, \dots, m_l) \;\ge\; \Phi_l(m_1, \dots, m_{i-1}, m_i + 1, m_{i+1}, \dots, m_l)$$

for $i \in \{1, \dots, l\}$. Thus, iteratively, (19) holds for $j = l$, i.e. for $d = 0$. Again, from (2) and (19) (for $d = 0$), it follows that

$$\Phi_l(m_1, \dots, m_l) \;\ge\; \Phi_{l+1}(m_1, \dots, m_l, 1) \;\ge\; \Phi_{l+1}(k_1, \dots, k_{l+1}),$$

i.e., (19) holds for $j = l + 1$, i.e. for $d = 1$. Now apply (19) (with $d = 1$) $(j - l)$-times to conclude that

$$\begin{aligned} \Phi_l(m_1, \dots, m_l) \;&\ge\; \Phi_{l+1}(k_1, \dots, k_{l+1}) \;\ge\; \Phi_{l+2}(k_1, \dots, k_{l+2}) \\ &\ge\; \cdots \;\ge\; \Phi_{j-1}(k_1, \dots, k_{j-1}) \;\ge\; \Phi_j(k_1, \dots, k_j), \end{aligned}$$

and (19) is established. □

References

[1] CANNINGS, C. (1974) The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. *Adv. Appl. Probab.* **6**, 260–290.

[2] CANNINGS, C. (1975) The latent roots of certain Markov chains arising in genetics: a new approach, II. Further haploid models. *Adv. Appl. Probab.* **7**, 264–282.

[3] DONNELLY, P. (1991) Weak convergence to a Markov chain with an entrance boundary: ancestral processes in population genetics. *Ann. Probab.* **19**, 1102–1117.

[4] FREUND, F. AND MÖHLE, M. (2007) On the time back to the most recent common ancestor and the external branch length of the Bolthausen-Sznitman coalescent. Preprint.

[5] 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.

[6] KINGMAN, J.F.C. (1982) On the genealogy of large populations. *J. Appl. Probab.* **19A**, 27–43.

[7] KINGMAN, J.F.C. (1982) The coalescent. *Stoch. Process. Appl.* **13**, 235–248.

[8] KINGMAN, J.F.C. (2000) Origins of the coalescent: 1974–1982. *Genetics* **156**, 1461–1463.

[9] MÖHLE, M. (1998) Robustness results for the coalescent. *J. Appl. Probab.* **35**, 438–447.

[10] MÖHLE, M. (2000) Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.* **32**, 983–993.

[11] MÖHLE, M. (2004) The time back to the most recent common ancestor in exchangeable population models. *J. Appl. Probab.* **36**, 78–97.

[12] MÖHLE, M. (2006) On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12**, 35–53.

[13] MÖHLE, M. AND SAGITOV, S. (2001) A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562.

[14] MÖHLE, M. AND SAGITOV, S. (2003) Coalescent patterns in diploid exchangeable population models. *J. Math. Biol.* **47**, 337–352.

[15] PITMAN, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870–1902.

[16] SAGITOV, S. (1999) The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, 1116–1125.

[17] SAGITOV, S. (2003) Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40**, 839–854.

[18] SCHWEINSBERG, J. (2000) Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, 1–50.