ON THE GENEALOGY OF MULTI-TYPE CANNINGS MODELS AND THEIR LIMITING EXCHANGEABLE COALESCENTS

Maximilian Flamm and Martin Möhle[1]

Date: April 14, 2025

Abstract

We study the multi-type Cannings population model. Each individual has a type belonging to a given at most countable type space E. The population is hence divided into $|E|$ subpopulations. The subpopulation sizes are assumed to be constant over the generations, whereas the number of offspring of type $\ell \in E$ of all individuals of type $k \in E$ is allowed to be random. Under a joint exchangeability assumption on the offspring numbers, the transition probabilities of the ancestral process of a sample of individuals satisfy a multi-type consistency property, paving a way to prove in the limit for large subpopulation sizes the existence of multi-type exchangeable coalescent processes via Kolmogorov's extension theorem. Integral representations for the infinitesimal rates of these multi-type exchangeable coalescents and some of their properties are studied. Examples are provided, among them multi-type Wright–Fisher models and multi-type pure mutation models. The results contribute to the foundations of multi-type coalescent theory and provide new insights into (the existence of) multi-type exchangeable coalescents.

Keywords: Consistency; exchangeability; integral representation; multi-type Cannings model; multi-type coalescent

2020 Mathematics Subject Classification: Primary 60J90; 60J10 Secondary 92D15; 92D25

1 Introduction and model definition

Multi-type population models in the spirit of Cannings [6, 7, 8] are studied, where each individual is equipped with a certain type taken from a given type space E. It is assumed that E is at most countable. The population is hence divided into $|E|$ subpopulations. It is furthermore assumed that the size $N_k \in \mathbb{N} := \{1, 2, \ldots\}$ of each subpopulation $k \in E$ is deterministic and constant over the generations. We denote by $N := (N_k)_{k\in E} \in \mathbb{N}^E$ the vector of all subpopulation sizes. In the following the dynamics of the model is first described for a single generation step. For $k, \ell \in E$ and $i \in [N_k] := \{1, \ldots, N_k\}$ let $\nu_{k,\ell,i}$ denote the (random) number of offspring of type $\ell \in E$ of the i-th individual of type $k \in E$. Clearly, $N_{k,\ell} := \sum_{i\in[N_k]} \nu_{k,\ell,i}$ is the number of offspring of type $\ell \in E$ of all individuals of type $k \in E$. Note that $\sum_{k\in E} N_{k,\ell}$ is the total number of offspring of type $\ell \in E$, whereas $\sum_{\ell\in E} N_{k,\ell}$ is the number of offspring of all individuals of type $k \in E$. The assumption that all subpopulation

[1]Fachbereich Mathematik, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany
E-mail addresses: maximilian.flamm@uni-tuebingen.de, martin.moehle@uni-tuebingen.de

sizes are constant over time puts the restrictions

$$\sum_{k\in E} N_{k,\ell} \;=\; N_\ell, \qquad \ell \in E, \tag{1}$$

on the random variables $N_{k,\ell}$, $k,\ell \in E$. Eq. (1) is the analog of Cannings [7, Eq. (30)] and Hössjer [13, Eq. (2.1)]. The subclass of models satisfying $N_{k,k} = N_k$ for all $k \in E$ or, equivalently, $N_{k,\ell} = 0$ for all $k,\ell \in E$ with $k \neq \ell$, is studied in [22]. Our analysis is based on the assumption (compare also with [7, p. 266]) that the offspring sizes are jointly exchangeable, i.e.,

(A) for all permutations $\pi_{k,\ell}$ of $[N_k]$, $k,\ell \in E$, $(\nu_{k,\ell,\pi_{k,\ell}(i)})_{k,\ell\in E, i\in[N_k]}$ has the same distribution as $(\nu_{k,\ell,i})_{k,\ell\in E, i\in[N_k]}$.

Less general, as in [13], one may assume that the offspring sizes are

(A1) exchangeable in each subpopulation, i.e., for each $k \in E$ the random vectors $\nu_{k,i} := (\nu_{k,\ell,i})_{\ell\in E}$, $i \in [N_k]$, are exchangeable and

(A2) independent in different subpopulations, i.e., the $(\nu_{k,i})_{i\in[N_k]}$, $k \in E$, are independent.

Clearly, (A) implies (A1). If (A1) and (A2) are satisfied, then (A) holds as well, since independence implies exchangeability. If (A2) is satisfied, then, for each $\ell \in E$, the $N_{k,\ell}$, $k \in E$, are independent, which together with (1) implies that the $N_{k,\ell}$ are deterministic almost surely, which is the reason why in [13] it is assumed that the $N_{k,\ell}$ are deterministic. Our results however hold under the more general assumption (A). In particular, the $N_{k,\ell}$ are allowed to be random and all results in this article hold for random $N_{k,\ell}$ satisfying (1).

Up to now the model is defined for a single generation step. The description of the model over the generations is rather simple. Offspring sizes in different generations, labeled with $r \in \mathbb{Z} := \{\ldots, -1, 0, 1, \ldots\}$, are simply assumed to be independent and identically distributed (iid).

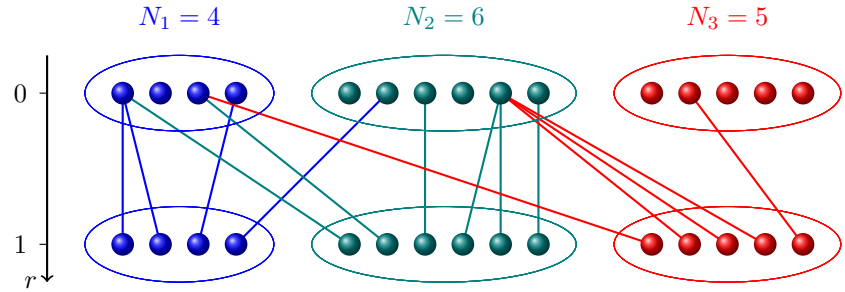
Figure 1 illustrates a realization of two consecutive generations of the model with type space $E = \{1, 2, 3\}$, subpopulation sizes $N = (N_1, N_2, N_3) = (4, 6, 5)$ and with the realization

$$(N_{k,\ell}(\omega))_{k,\ell\in E} = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

The offspring numbers can be read off from Figure 1. For example, the first individual of the first (blue) subpopulation has $\nu_{1,1,1}(\omega) = 2$ offspring of type 1 (blue), $\nu_{1,2,1}(\omega) = 1$ offspring of type 2 (green) and $\nu_{1,3,1}(\omega) = 0$ offspring of type 3 (red). For $|E| = 1$ and under (A), this model reduces to the classical single-type exchangeable population model introduced by Cannings [6, 7].

The paper is organized as follows. In Section 2 basic properties on the ancestry of multi-type Cannings models are derived. Section 3 provides new multi-type consistency equations and discusses symmetry and exchangeability issues of multi-type Cannings models. Convergence results as all subpopulation sizes become large are provided in Section 4. Sections 5 and 6 deal with (the existence of) multi-type exchangeable coalescent processes and their integral representations. Two examples, the multi-type Wright–Fisher model and a multi-type pure

Figure 1: An illustration of two consecutive generations with three subpopulations



mutation model, are extensively analysed in Section 7. The proofs are provided in Sections 8 except for the proof of the integral representation (Theorem 3), which is – because of its length – provided in the separate Section 9. The article finishes with short appendix collecting some basic bounds, definitions, illustrations and results used throughout the article. The results extend those obtained in [22] and, hence, contribute to the foundations of multi-type coalescent theory.

It is almost impossible to mention all relevant literature here. Precise citations are made at the appropriate places throughout the article. For recent related works on multi-type Λ-coalescents we refer the reader exemplary to González et al. [10] and Johnston, Kyprianou and Rogers [15].

2 Ancestral structure

Before we will define the multi-type ancestral process, let us recall some basics on set partitions having labeled blocks, also called labeled set partitions. We shall also introduce some notations for random labeled partitions.

2.1 Block labeled set partitions

The literature on labeled set partitions, also called colored, marked or typed set partitions, is sparse compared to the immense literature on standard set partitions. We refer the reader to Kallenberg [17] for probabilistic theory on marked partitions, to Goyt and Pudwell [11, 12] for some literature on the combinatorics of colored set partitions and to Alberti [1], where a labeled partitioning process is studied. For $n \in \mathbb{N}$ let $\mathcal{P}_n$ denote the space of partitions of $[n]$. Note that $|\mathcal{P}_n| = \sum_{j=1}^{n} S(n,j)$, where the $S(.,.)$ denote the Stirling numbers of the second kind, i.e., $S(n,j)$ is the number of partitions of $[n]$ having j blocks. Any partition $\pi \in \mathcal{P}_n$ can be written as $\pi = \{B_1, \ldots, B_j\}$, where $B_1, \ldots, B_j$ are the (non-empty) blocks of π. The order of the blocks is not relevant, but usually the blocks are listed in order of appearance, i.e. $1 \in B_1$, $\min([n] \setminus B_1) \in B_2$ and so on. Given a partition $\{B_1, \ldots, B_j\} \in \mathcal{P}_n$ having j blocks $B_1, \ldots, B_j$, one may equip each block B_i with a type $k_i \in E$ leading to the partition $\{(B_1, k_1), \ldots, (B_j, k_j)\}$ having labeled (colored) blocks. Such a partition is called a block-labeled partition or simply a labeled (or colored, marked or typed) partition. Let $\mathcal{P}_{n,E}$ denotes the space of labeled partitions of $[n]$. If $d := |E| < \infty$, then $|\mathcal{P}_{n,E}| = \sum_{j=1}^{n} d^j S(n,j) < \infty$.

Alternatively, by Dobiński's formula (see, for example, Eq. (21) of Hsu and Shiue [14] or Mansour [19, p. 384, Example 8.12]), $|\mathcal{P}_{n,E}| = e^{-d}\sum_{j\geq 0} d^j j^n/j!$ if $d < \infty$. For $\pi \in \mathcal{P}_{n,E}$ any block $(B,k) \in \pi$ is called a k-block of π.

It is sometimes useful to delete all labels of a labeled partition. Formally, for $n \in \mathbb{N}$, the function $r_n : \mathcal{P}_{n,E} \to \mathcal{P}_n$, defined via $r_n(\pi) := \{B_1, \ldots, B_j\}$ for all $\pi = \{(B_1,k_1), \ldots, (B_j,k_j)\} \in \mathcal{P}_{n,E}$, is called the ($n$-th) label removal function. The function r_n simply maps each labeled partition to its non-labeled counterpart by removing all labels.

For a permutation $\sigma \in S_n$ and a block $B \subseteq [n]$ define $\sigma(B) := \{\sigma(i) : i \in B\}$ and for $\pi = \{(B_1,k_1), \ldots, (B_j,k_j)\} \in \mathcal{P}_{n,E}$ define $\sigma(\pi) := \{(\sigma(B_1),k_1), \ldots, (\sigma(B_j),k_j)\} \in \mathcal{P}_{n,E}$.

2.2 Random labeled partitions

A random labeled partition of $[n]$ (with label space E) is a random variable Π taking values in $\mathcal{P}_{n,E}$. The following notions of exchangeability are in the spirit of Aldous [2], Kallenberg [17, p. 343] and Pitman [25]. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $A \in \mathcal{F}$ let $\mathbb{P}_A$ denote the restriction of $\mathbb{P}$ to A defined via $\mathbb{P}_A(B) := \mathbb{P}(B \cap A)$ for all $B \in \mathcal{F}$.

Definition 1 (Partial exchangeability)**.** *Let $n \in \mathbb{N}$, $G \subseteq S_n$ and $A \in \mathcal{F}$. A random labeled partition Π of $[n]$ is called partially exchangeable with respect to G, if $\sigma(\Pi) \stackrel{d}{=} \Pi$ for all $\sigma \in G$, i.e., $\mathbb{P}(\sigma(\Pi) = \pi) = \mathbb{P}(\Pi = \pi)$ for all $\sigma \in G$ and all $\pi \in \mathcal{P}_{n,E}$. A random labeled partition Π is called partially exchangeable with respect to G on the event A, if $\mathbb{P}_A(\sigma(\Pi) = \pi) = \mathbb{P}_A(\Pi = \pi)$ for all $\sigma \in G$ and all $\pi \in \mathcal{P}_{n,E}$, where $\mathbb{P}_A$ denotes the restriction of $\mathbb{P}$ to A.*

Remark. For $A = \Omega$, partial exchangeability with respect to G on A is the same as partial exchangeability with respect to G. If Π is partially exchangeable with respect to G, then Π is even partially exchangeable with respect to the group $\langle G \rangle$ generated by G. Partial exchangeability with respect to the full group $G = S_n$ of permutations of $[n]$ is simply called exchangeability. We will come back to exchangeability issues at the end of Section 3.

2.3 Multi-type ancestral process

Suppose that one has sampled $n \in [\sum_{k\in E} N_k]$ individuals from the current generation 0. One may order these n individuals (in some arbitrary way) and let $k_1, \ldots, k_n \in E$ denote their types. Looking $r \in \mathbb{N}_0 := \{0, 1, \ldots\}$ generations backward in time, the ancestry of the sample can be captured by defining a random labeled partition $\mathcal{A}_r = \mathcal{A}_r^{(n,N)}$ of $[n]$ such that (by definition) $i, j \in [n]$ belong to the same k-block of $\mathcal{A}_r$ if and only if the individuals i and j have a common ancestor r generations backward in time and this ancestor has type k. Note that $\mathcal{A}_r$ not only depends on the sample size n and on the subpopulation sizes N_k, $k \in E$, but also on the distribution of the offspring numbers $\nu_{k,\ell,i}$, but this dependence is suppressed in our notation for simplicity. The process $\mathcal{A} := (\mathcal{A}_r)_{r\in\mathbb{N}_0}$ is called a *multi-type ancestral process* or, alternatively, a *multi-type backward process* or a *multi-type discrete n-coalescent process*. The assumption that offspring sizes in different generations are iid ensures that $\mathcal{A}$ is a homogeneous Markov chain (HMC) with state space $\mathcal{P}_{n,E}$ and initial state $\mathcal{A}_0 = \{(\{1\},k_1), \ldots, (\{n\},k_n)\}$. In this case the initial state $\mathcal{A}_0$ is a deterministic labeled partition of $[n]$ into singletons. Note however, that we could have sampled from generation 0 according to some different (even random) scheme, for example in such a way that $\mathcal{A}_0$ is an exchangeable

random labeled partition. We will come back to the impact of the distribution of $\mathcal{A}_0$ at the end of Section 3. Let

$$p_{\pi,\pi'} \;:=\; \mathbb{P}(\mathcal{A}_r = \pi' \,|\, \mathcal{A}_{r-1} = \pi), \qquad \pi, \pi' \in \mathcal{P}_{n,E}, r \in \mathbb{N}, \tag{2}$$

denote the transition probabilities of $\mathcal{A}$. Note that $p_{\pi,\pi'} = p^{(N)}_{\pi,\pi'}$ depends on the subpopulation sizes $N = (N_k)_{k\in E}$, but this dependence is often suppressed in our notation. Transitions from $\pi \in \mathcal{P}_{n,E}$ to $\pi' \in \mathcal{P}_{n,E}$ are only possible if each block of π' is a union of some blocks of π (types of the blocks disregarded here). We write $\pi \subseteq \pi'$ in this case. For $x \in \mathbb{R}$ and $n \in \mathbb{N}_0$ let $(x)_n := \prod_{i=0}^{n-1}(x-i)$ denote the descending factorials. Since, in each subpopulation, offspring to parents are randomly assigned (random assignment condition), it follows under (A) that the transition probability (2) can be expressed in terms of the offspring variables $\nu_{k,\ell,s}$, $k, \ell \in E$, $s \in [j_k]$, via

$$p_{\pi,\pi'} \;=\; \frac{\prod_{k\in E}(N_k)_{j_k}}{\prod_{\ell\in E}(N_\ell)_{i_\ell}} \mathbb{E}\Bigg(\prod_{k,\ell\in E} \prod_{s=1}^{j_k} (\nu_{k,\ell,s})_{i_{k,\ell,s}} \Bigg), \qquad \pi, \pi' \in \mathcal{P}_{n,E}, \pi \subseteq \pi', \tag{3}$$

where i_ℓ and j_k are the number of ℓ-blocks of π and k-blocks of π' respectively and $i_{k,\ell,s}$, $k, \ell \in E$, $s \in [j_k]$, are the group sizes of ℓ-blocks of π merging to the s-th k-block of π'. A proof of (3) is provided in Section 8. The structure of the transition matrix P is illustrated exemplary for sample size $n = 2$ in Section 10.2 in the appendix. Note that $\sum_{k\in E}\sum_{s=1}^{j_k} i_{k,\ell,s} = i_\ell$, $\ell \in E$. In particular, the transition matrix $P := (p_{\pi,\pi'})_{\pi,\pi'\in\mathcal{P}_{n,E}}$ has diagonal entries

$$p_{\pi,\pi} \;=\; \mathbb{E}\Bigg(\prod_{k\in E} \prod_{s=1}^{i_k} \nu_{k,k,s} \Bigg), \qquad \pi \in \mathcal{P}_{n,E}. \tag{4}$$

Consistency properties of the transition probabilities (3) and symmetry properties of the ancestral process are deferred to Section 3. If $N_{k,\ell} = 0$ for all $k \neq \ell$, then (3) reduces to the transition probabilities [22, Eq. (9)] for multi-type Cannings models, where each offspring has the same type as its parent. For $E = \{1\}$, Eq. (3) reduces to the well-known formula for the transition probabilities of the discrete coalescent for single-type Cannings models (see, for example, [23, Eq. (3)])

$$p_{\pi,\pi'} \;=\; \frac{(N)_j}{(N)_i} \mathbb{E}\big((\nu_1)_{i_1} \cdots (\nu_j)_{i_j} \big), \tag{5}$$

where N ($:= N_1$) is the total population size, i and j are the number of blocks of π and π' respectively, $\nu_s := \nu_{1,1,s}$ for $s \in [N]$ and $i_1, \ldots, i_j$ are the group sizes of merging blocks of π. Note that $i_1 + \cdots + i_j = i$.

The coalescence probability that two individuals of the same type $\ell \in E$ share a common parent of type $k \in E$ one generation backward in time is

$$\mathbb{E}\Bigg(\sum_{i=1}^{N_k} \frac{(\nu_{k,\ell,i})_2}{(N_\ell)_2} \Bigg) \;=\; \frac{N_k}{(N_\ell)_2} \mathbb{E}((\nu_{k,\ell,1})_2) \;=\; \frac{N_k}{(N_\ell)_2} \mathbb{E}(\nu^2_{k,\ell,1}) - \frac{\mathbb{E}(N_{k,\ell})}{(N_\ell)_2} \;=:\; c_{k,\ell}(N_k, N_\ell), \tag{6}$$

provided that $N_\ell > 1$, in agreement with (3) for $i_\ell := 2$, $i_s := 0$ for $s \in E \setminus \{\ell\}$, $j_k := 1$ and $j_s := 0$ for $s \in E \setminus \{k\}$.

Similarly, for $k, \ell_1, \ell_2 \in E$ with $\ell_1 \neq \ell_2$, the coalescence probability that two individuals of different types ℓ_1 and ℓ_2 respectively share a common parent of type k one generation backward in time is

$$\mathbb{E}\bigg(\sum_{i=1}^{N_k} \frac{\nu_{k,\ell_1,i}}{N_{\ell_1}} \frac{\nu_{k,\ell_2,i}}{N_{\ell_2}}\bigg) \;=\; \frac{N_k}{N_{\ell_1} N_{\ell_2}} \mathbb{E}(\nu_{k,\ell_1,1}\nu_{k,\ell_2,1}) \;=:\; c_{k,\ell_1,\ell_2}(N_k, N_{\ell_1}, N_{\ell_2}), \tag{7}$$

again in agreement with (3) for $i_{\ell_1} := i_{\ell_2} := 1$, $i_s := 0$ for $s \in E \setminus \{\ell_1, \ell_2\}$, $j_k := 1$ and $j_s := 0$ for $s \in E \setminus \{k\}$. We call $c_{k,k}(N_k, N_k)$, $k \in E$, the *diagonal coalescence probabilities* and all the other coalescence probabilities the *off-diagonal coalescence probabilities*. There are $d := |E|$ diagonal coalescence probabilities and $(d)_2 + d(d)_2 = d^3 - d$ off-diagonal coalescence probabilities, thus, altogether d^3 coalescence probabilities.

For $r \in \mathbb{N}_0$ and $k \in E$ let $Y_{r,k}$ denote the number of k-blocks of $\mathcal{A}_r$ and define $Y_r := (Y_{r,k})_{k\in E}$. Using an argument in the spirit of Burke and Rosenblatt [5] it is readily seen that $Y := (Y_r)_{r\in\mathbb{N}_0}$, called the (multi-type) block counting process of $\mathcal{A}$, is a HMC with (partially ordered) state space $\mathbb{N}_0^E$. Let $p_{i,j} := \mathbb{P}(Y_r = j \,|\, Y_{r-1} = i)$, $i, j \in \mathbb{N}_0^E$, $r \in \mathbb{N}$, denote the transition probabilities of Y. From (3) it follows that, for all $i, j \in \mathbb{N}_0^E$ with $i \geq j$ (componentwise),

$$p_{i,j} \;=\; \frac{\prod_{k\in E}\binom{N_k}{j_k}}{\prod_{\ell\in E}\binom{N_\ell}{i_\ell}} \sum \mathbb{E}\bigg(\prod_{k,\ell\in E}\prod_{s=1}^{j_k}\binom{\nu_{k,\ell,s}}{i_{k,\ell,s}}\bigg),$$

where the sum $\sum$ extents over all $i_{k,\ell,s} \in \mathbb{N}_0$, $k, \ell \in E$, $s \in [j_k]$, with $\sum_{k\in E}\sum_{s=1}^{j_k} i_{k,\ell,s} = i_\ell$ for all $\ell \in E$. It is often useful to study the $\mathbb{N}_0^E$-valued process Y before considering the full partition-valued process Π.

3 Consistency and symmetry

Single-type Cannings models satisfy a fundamental consistency property (see Eq. (11) below) being crucial for the analysis (of the ancestral structure) of these models. Proposition 1 below shows that multi-type Cannings models satisfy a similar but more involved multi-type consistency property. In order to state the result, it is useful to introduce matrices of the form $T := (t_{k,\ell})_{k,\ell\in E}$, where each entry $t_{k,\ell}$ of T is a (possibly empty) vector of the form $t_{k,\ell} := (i_{k,\ell,s})_{s\in[j_k]}$ with $j_k \in \mathbb{N}_0$ for all $k \in E$ and $i_{k,\ell,s} \in \mathbb{N}_0$ for all $k, \ell \in E$ and all $s \in [j_k]$ satisfying $i_\ell := \sum_{k\in E}\sum_{s=1}^{j_k} i_{k,\ell,s} \leq N_\ell$ for all $\ell \in E$. If $j_k = 0$ then $t_{k,\ell} = () =: \mathbf{0} \in \mathbb{R}^0$ is the empty vector (neutral and only element of $\mathbb{R}^0$) for every $\ell \in E$. The symbol T is used since T is a tensor. With this notation, the transition probability (3) is of the form

$$p_{\pi,\pi'} \;=\; \Phi_j(T) \quad \text{with} \quad \Phi_j(T) \;:=\; \Phi_j^{(N)}(T) \;:=\; \frac{\prod_{k\in E}(N_k)_{j_k}}{\prod_{\ell\in E}(N_\ell)_{i_\ell}} \mathbb{E}\bigg(\prod_{k,\ell\in E}\prod_{s=1}^{j_k}(\nu_{k,\ell,s})_{i_{k,\ell,s}}\bigg), \tag{8}$$

where $j := (j_k)_{k\in E}$ and T is the tensor defined above. Note that, if $j = e_k$ is the k-th unit vector in $\mathbb{R}^E$, then

$$\Phi_{e_k}(T) \;=\; \frac{N_k}{\prod_{\ell\in E}(N_\ell)_{i_{k,\ell,1}}} \mathbb{E}\bigg(\prod_{\ell\in E}(\nu_{k,\ell,1})_{i_{k,\ell,1}}\bigg), \qquad k \in E.$$

In particular, the coalescence probability (6) has the form $c_{k,\ell}(N_k, N_\ell) = \Phi_{e_k}(T)$, where $T := (t_{k',\ell'})_{k',\ell' \in E}$ is the tensor with entries $t_{k',\ell'} := (2)$ if $k' = k$ and $\ell' = \ell$, $t_{k',\ell'} := (0) = \mathbf{0} \in \mathbb{R}^1$ if $k' = k$ and $\ell' \neq \ell$ and $t_{k',\ell'} := () = \mathbf{0} \in \mathbb{R}^0$ otherwise. Similarly, the coalescence probability (7) has the form $c_{k,\ell_1,\ell_2}(N_k, N_{\ell_1}, N_{\ell_2}) = \Phi_{e_k}(T)$, where $T := (t_{k',\ell'})_{k',\ell' \in E}$ is the tensor with entries $t_{k',\ell'} := (1)$ if $k' = k$ and $\ell' \in \{\ell_1, \ell_2\}$, $t_{k',\ell'} := (0) = \mathbf{0} \in \mathbb{R}^1$ if $k' = k$ and $\ell' \notin \{\ell_1, \ell_2\}$ and $t_{k',\ell'} := () = \mathbf{0} \in \mathbb{R}^0$ otherwise.

In the following, for given $j = (j_k)_{k \in E} \in \mathbb{N}_0^E$, the space of all tensors $T = (t_{k,\ell})_{k,\ell \in E}$ with $t_{k,\ell} = (i_{k,\ell,s})_{s \in [j_k]}$, where the entries $i_{k,\ell,s} \in \mathbb{N}_0$ satisfy $i_\ell := \sum_{k \in E} \sum_{s=1}^{j_k} i_{k,\ell,s} \leq N_\ell$ for all $\ell \in E$, is denoted by $\mathcal{T}_j$. Note that $\mathcal{T}_j = \mathcal{T}_j^{(N)}$ depends on the subpopulation sizes $N := (N_k)_{k \in E}$ and that $\Phi_j(T)$ is well defined for all $T \in \mathcal{T}_j$. Thus, Φ_j is a function from $\mathcal{T}_j$ to $[0, 1]$. The following proposition provides full information on the consistency property of multi-type Cannings models. Despite the fact that its proof, provided in Section 8, is rather short, this consistency property turns out to be crucial for essentially all what follows.

Proposition 1 (Multi-type consistency)**.** *Under (A), the functions $\Phi_j : \mathcal{T}_j \to [0, 1]$, $j := (j_k)_{k \in E} \in \mathbb{N}_0^E$, are consistent in the following sense. For all $j = (j_k)_{k \in E} \in \mathbb{N}_0^E$ and all tensors $T \in \mathcal{T}_j$, the equality*

$$\Phi_j(T) \;=\; \sum_{k \in E} \Phi_{j+e_k}\big(T(k,\ell)\big) + \sum_{k \in E} \sum_{s=1}^{j_k} \Phi_j\big(T(k,\ell,s)\big) \tag{9}$$

holds for each $\ell \in E$ with $i_\ell := \sum_{k \in E} \sum_{s \in [j_k]} i_{k,\ell,s} < N_\ell$, where e_k denotes the k-th unit vector in $\mathbb{R}^E$, the tensor $T(k,\ell)$ is obtained from T by replacing the (possibly empty) vector $t_{k,\ell} = (i_{k,\ell,1}, \ldots, i_{k,\ell,j_k})$ by $(i_{k,\ell,1}, \ldots, i_{k,\ell,j_k}, 1)$ and the (possibly empty) vector $t_{k,\ell'} = (i_{k,\ell',1}, \ldots, i_{k,\ell',j_k})$ by $(i_{k,\ell',1}, \ldots, i_{k,\ell',j_k}, 0)$ for all $\ell' \neq \ell$, and the tensor $T(k,\ell,s)$ is obtained from T by replacing the single entry $i_{k,\ell,s}$ by $i_{k,\ell,s} + 1$.

Remark. In particular, the right-hand side of (9) takes the same value for each $\ell \in E$ with $i_\ell < N_\ell$, which is a-priori not obvious. For $j = (0)_{k \in E}$ (null vector), Eq. (9) reduces to the normalizing condition

$$1 \;=\; \sum_{k \in E} \Phi_{e_k}\big(T(k,\ell)\big), \qquad \ell \in E, \tag{10}$$

where $T(k,\ell)$ is the tensor with entries $t_{k,\ell} := (1)$, $t_{k,\ell'} := (0)$ for $\ell' \neq \ell$ and $t_{k',\ell'} := ()$ (empty vector) otherwise. Eq. (10) is easily seen as follows. For all $k, \ell \in E$ we have $\Phi_{e_k}(T(k,\ell)) = (N_k/N_\ell)\mathbb{E}(\nu_{k,\ell,1}) = \mathbb{E}(N_{k,\ell})/N_\ell$. Summing over all $k \in E$ and taking (1) into account yields (10).

For $E = \{1\}$ and $j := j_1 \in \mathbb{N}$, Eq. (9) reduces to the consistency equation for single-type Cannings models (see, for example, [21, Eq. (3) and (4)])

$$\Phi_j(i_1, \ldots, i_j) \;=\; \Phi_{j+1}(i_1, \ldots, i_j, 1) + \sum_{s=1}^{j} \Phi_j(i_1, \ldots, i_{s-1}, i_s + 1, i_{s+1}, \ldots, i_j), \tag{11}$$

$i_1, \ldots, i_j \in \mathbb{N}$ with $i_1 + \cdots + i_j < N$ $(:= N_1)$. The multi-type consistency property provided in Proposition 1 is more involved than its single-type counterpart (11) in the sense that it cannot be derived from (11) alone. For $E = \{1\}$, the normalizing condition (10) reduces to

$\Phi_1(1) = 1$. Note that the same consistency relation (11) holds for all $i_1, \ldots, i_j \in \mathbb{N}$ if Φ_j is the exchangeable partition probability function (EPPF) of an infinite exchangeable random partition (see, for example, Pitman [27, Eq. (2.9)]).

As in the single-type case, multi-type consistency has fundamental consequences, two of the probably most important of them provided in Corollary 1 and Corollary 2 below.

Let $j, j' \in \mathbb{N}_0^k$ with $j \leq j'$ (componentwise) and $T \in \mathcal{T}_j$, $T' \in \mathcal{T}_{j'}$ be two tensors. Recall that $T = (t_{k,\ell})_{k,\ell \in E}$ with $t_{k,\ell} = (i_{k,\ell,s})_{s \in [j_k]}$ and, similarly, $T' = (t'_{k,\ell})_{k,\ell \in E}$ with $t'_{k,\ell} = (i'_{k,\ell,s})_{s \in [j'_k]}$. We say that $T \leq T'$ if $i_{k,\ell,s} \leq i'_{k,\ell,s}$ for all $k, \ell \in E$ and all $s \in [j_k]$. The following monotonicity property is a direct consequence of the consistency. Again, its proof is provided in Section 8.

Corollary 1 (Monotonicity)**.** *Under (A), the functions $\Phi_j : \mathcal{T}_j \to [0,1]$, $j \in \mathbb{N}_0^E$, are monotone in the sense that*

$$\Phi_{j'}(T') \;\leq\; \Phi_j(T) \tag{12}$$

for all $j, j' \in \mathbb{N}_0^E$ with $j \leq j'$ and all tensors $T \in \mathcal{T}_j$ and $T' \in \mathcal{T}_{j'}$ with $T \leq T'$.

For $m, n \in \mathbb{N}$ with $m \leq n$ let $\varrho_{n,m} : \mathcal{P}_{n,E} \to \mathcal{P}_{m,E}$ denote the natural restriction from $\mathcal{P}_{n,E}$ to $\mathcal{P}_{m,E}$ defined via

$$\varrho_{n,m}(\pi) \;:=\; \{(B_i \cap [m], k_i) : 1 \leq i \leq j, B_i \cap [m] \neq \emptyset\} \tag{13}$$

for all $\pi = \{(B_1, k_1), \ldots, (B_j, k_j)\} \in \mathcal{P}_{n,E}$. The following corollary shows that multi-type Cannings models satisfy the natural coupling property. For general information on the natural coupling property of the Kingman n-coalescent we refer the reader to Section 7 of [18].

Corollary 2 (Natural coupling)**.** *Under (A), for all $m, n \in \mathbb{N}$ with $m \leq n$, the restricted process $(\varrho_{n,m} \circ \mathcal{A}_r^{(n)})_{r \in \mathbb{N}_0}$ has the same distribution as the multi-type ancestral process $(\mathcal{A}_r^{(m)})_{r \in \mathbb{N}_0}$.*

We now turn to symmetry and exchangeability properties of multi-type Cannings models. For $n \in \mathbb{N}$ let S_n be the set of permutations of $[n]$. Each permutation $\sigma \in S_n$ acts on $\mathbb{R}^n$ via $\sigma x := (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. This definition is naturally extended to $n = 0$ by assuming that S_0 contains only one particular permutation σ_0 acting on the empty vector $() \in \mathbb{R}^0$ via $\sigma_0() := ()$.

Let $j = (j_k)_{k \in E} \in \mathbb{N}_0^E$. The definition of the function Φ_j (see (8)) and (A) imply that Φ_j is symmetric in the following sense. For all tensors $T = (t_{k,\ell})_{k,\ell \in E} \in \mathcal{T}_j$ and all permutations $\sigma_{k,\ell} \in S_{j_k}$, $k, \ell \in E$,

$$\Phi_j(T) \;=\; \Phi_j(\sigma(T)), \tag{14}$$

where $\sigma := (\sigma_{k,\ell})_{k,\ell \in E}$ and the tensor $\sigma(T) \in \mathcal{T}_j$ is defined via $\sigma(T) := (\sigma_{k,\ell} t_{k,\ell})_{k,\ell \in E}$.

The distribution of the ancestral process $(\mathcal{A}_r)_{r \in \mathbb{N}_0}$ clearly does not only depend on the transition probabilities (8) but also on the distribution of $\mathcal{A}_0$. Lemma 1 below shows that the ancestral process is exchangeable, if one assumes that the individuals from generation 0 are sampled in such a way that $\mathcal{A}_0$ is an exchangeable random labeled partition. For example, one may sample all n individuals from a single subpopulation or one may sample the n individuals in an exchangeable manner such that their random types are exchangeable E-valued random variables. The proof of Lemma 1 is provided on Section 8.

Lemma 1 (Exchangeability)**.** *If $\mathcal{A}_0$ is exchangeable then $\mathcal{A}_r$ is exchangeable for all $r \in \mathbb{N}_0$.*

Remark. If $\mathcal{A}_0$ is not exchangeable, then all the $\mathcal{A}_r$, $r \in \mathbb{N}$, are in general not exchangeable. Still, the ancestral process shares some 'reduced' symmetry properties. To explain this it turns out to be useful to introduce certain subgroups of the permutation group S_n as follows.

Let $k_1, \ldots, k_n \in E$. For $\ell \in E$ define $A_\ell := \{i \in [n] : k_i = \ell\}$. Note that $A_\ell = A_\ell^{(k_1,\ldots,k_n)}$ depends on $k_1, \ldots, k_n$ and that $\sum_{\ell \in E} |A_\ell| = n$. For $\ell \in E$ let σ_ℓ be a permutation of A_ℓ and for $i \in A_\ell$ define $\sigma(i) := \sigma_\ell(i)$. Use the symbol $S_{k_1,\ldots,k_n}$ for the set of all such permutations σ. Note that $S_{k_1,\ldots,k_n}$ is a subgroup of S_n.

Now, let $\pi \in \mathcal{P}_{n,E}$ and let $k_i = k_i(\pi)$ denote the type of individual $i \in [n]$. Then, for every generation $r \in \mathbb{N}$, the random labeled partition $\mathcal{A}_r$ is partially exchangeable with respect to $S_{k_1,\ldots,k_n}$ on the event $\{\mathcal{A}_{r-1} = \pi\}$ in the sense of Definition 1, i.e.,

$$\mathbb{P}(\sigma(\mathcal{A}_r) = \pi', \mathcal{A}_{r-1} = \pi) \;=\; \mathbb{P}(\mathcal{A}_r = \pi', \mathcal{A}_{r-1} = \pi)$$

for all $\sigma \in S_{k_1,\ldots,k_n}$ and $\pi' \in \mathcal{P}_{n,E}$.

For the single-type situation $|E| = 1$, all the $k_1, \ldots, k_n$ are equal and, hence, the subgroups $S_{k_1,\ldots,k_n}$ coincide with the full permutation group S_n. Thus, in this case, the equality $\mathbb{P}(\sigma(\mathcal{A}_r) = \pi', \mathcal{A}_{r-1} = \pi) = \mathbb{P}(\mathcal{A}_r = \pi', \mathcal{A}_{r-1} = \pi)$ holds for all $\sigma \in S_n$ and $\pi, \pi' \in \mathcal{P}_n$. Summing over all π shows that $\sigma(\mathcal{A}_r) \stackrel{d}{=} \mathcal{A}_r$ for all $\sigma \in S_n$. Thus, $\mathcal{A}_r$ is exchangeable. Similar issues concerning d-type exchangeability for infinite labeled partitions and d-type coalescents are addressed in [15, Section 3.1].

4 Limiting results

It is natural to investigate the behaviour of the ancestral process $\mathcal{A}^{(n,N)} = (\mathcal{A}_r^{(n,N)})_{r \in \mathbb{N}_0}$ for large subpopulation sizes N_k, $k \in E$, i.e., when the minimal subpopulation size

$$N_{\min} \;:=\; \min_{k \in E} N_k \tag{15}$$

tends to infinity. In the following, $P_N = (p_{\pi,\pi'}^{(N)})_{\pi,\pi' \in \mathcal{P}_{n,E}}$ denotes the transition matrix of the ancestral process $(\mathcal{A}_r^{(n,N)})_{r \in \mathbb{N}_0}$ and I the identity matrix (of the same size as P_N).

Lemma 2. *Assume that $n \geq 2$. Then the convergence $P_N \to I$ as $N_{\min} \to \infty$ holds if and only if $\mathbb{E}(\nu_{k,k,1}) \to 1$ and $\mathbb{E}(\nu_{k,k,1}\nu_{k,k,2}) \to 1$ as $N_{\min} \to \infty$ for all $k \in E$. In this case, $(N_k/N_\ell)^2 \mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2}) \to \delta_{k,\ell}$ (Kronecker symbol) as $N_{\min} \to \infty$ for all $k, \ell \in E$.*

Let us first focus on a situation, where it will turn out that the ancestral process is in the domain of attraction of a discrete-time limiting multi-type process. Let $j = (j_k)_{k \in E} \in \mathbb{N}_0^E$. A tensor $T = (t_{k,\ell})_{k,\ell \in E} \in \mathcal{T}_j$ is called a *diagonal tensor*, if $t_{k,\ell} = \mathbf{0}$ ($\in \mathbb{R}^{j_k}$) for all $k, \ell \in E$ with $k \neq \ell$. The particular diagonal tensor $(t_{k,\ell})_{k,\ell \in E} \in \mathcal{T}_j$ with diagonal entries $t_{k,k} := (1, \ldots, 1) \in \mathbb{R}^{j_k}$ for all $k \in E$ is denoted by $\mathbf{1}_j$.

Theorem 1. *(Convergence of the ancestral process, discrete-time limit)*
Assume that for all $j = (j_k)_{k \in E} \in \mathbb{N}_0^E$ and all tensors $T = (t_{k,\ell})_{k,\ell \in E} \in \mathcal{T}_j \setminus \{\mathbf{1}_j\}$, the convergence $\Phi_j^{(N)}(T) \to \phi_j(T)$ as $N_{\min} \to \infty$ holds for some constant $\phi_j(T) \in [0,1]$. Let

$n \in \mathbb{N}$. If $\mathcal{A}_0^{(n,N)} \to \Pi_0^{(n)}$ in distribution as $N_{\min} \to \infty$ for some random labeled partition $\Pi_0^{(n)}$ of $[n]$, then the multi-type ancestral process $(\mathcal{A}_r^{(n,N)})_{r\in\mathbb{N}_0}$ converges in $D_{\mathcal{P}_{n,E}}(\mathbb{N}_0)$ as $N_{\min} \to \infty$ to a discrete-time limiting Markov chain $\Pi^{(n)} = (\Pi_r^{(n)})_{r\in\mathbb{N}_0}$ with state space $\mathcal{P}_{n,E}$ and transition matrix $A := (a_{\pi,\pi'})_{\pi,\pi'\in\mathcal{P}_{n,E}}$ having non-diagonal entries $a_{\pi,\pi'} := \phi_j(T)$ if $\pi \subset \pi'$ and $a_{\pi,\pi'} := 0$ otherwise, and diagonal entries $a_{\pi,\pi} := 1 - \sum_{\pi'\neq\pi'} a_{\pi,\pi'}$.

For many multi-type Cannings models (see Lemma 2), the transition matrix P_N of the ancestral process satisfies $P_N \to I$ as $N_{\min} \to \infty$. This holds for example under a sort of weak mutation assumption for the multi-type Wright–Fisher model discussed in Section 7.1. In this situation, the limits $\phi_j(T)$ in Theorem 1 are all equal to zero and, hence, Theorem 1 is not useful, since the limiting Markov chain $\Pi^{(n)}$ satisfies $\Pi_r^{(n)} = \Pi_0^{(n)}$ almost surely for all $r \in \mathbb{N}_0$. Thus, other assumptions are required in this situation to obtain convergence of the ancestral process to a non-degenerate limiting process. The following result covers such cases and it turns out that the limiting process is a continuous-time multi-type process.

Theorem 2. *(Convergence of the ancestral process, continuous-time limit)*
Suppose that, for every $N = (N_k)_{k\in E}$, there exists $c_N > 0$ such that $c_N \to 0$ as $N_{\min} \to \infty$ and such that, for all $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ and all tensors $T = (t_{k,\ell})_{k,\ell\in E} \in \mathcal{T}_j \setminus \{\mathbf{1}_j\}$, the convergence $\Phi_j^{(N)}(T)/c_N \to \phi_j(T)$ as $N_{\min} \to \infty$ holds for some real constant $\phi_j(T)$. Let $n \in \mathbb{N}$. If $\mathcal{A}_0^{(n,N)} \to \Pi_0^{(n)}$ in distribution as $N_{\min} \to \infty$ for some random labeled partition $\Pi_0^{(n)}$ of $[n]$, then the time-scaled multi-type ancestral process $(\mathcal{A}_{\lfloor t/c_N\rfloor}^{(n,N)})_{t\geq 0}$ converges in $D_{\mathcal{P}_{n,E}}([0,\infty))$ as $N_{\min} \to \infty$ to a continuous-time limiting Markov process $\Pi^{(n)} = (\Pi_t^{(n)})_{t\geq 0}$ with state space $\mathcal{P}_{n,E}$ and infinitesimal generator $Q := (q_{\pi,\pi'})_{\pi,\pi'\in\mathcal{P}_{n,E}}$ having non-diagonal entries $q_{\pi,\pi'} := \phi_j(T)$ if $\pi \subset \pi'$ and $q_{\pi,\pi'} := 0$ otherwise, and diagonal entries $q_{\pi,\pi} := -\sum_{\pi'\neq\pi} q_{\pi,\pi'}$, $\pi \in \mathcal{P}_{n,E}$.

Remark. The standard choice for the time-scaling is the supremum over all coalescence probabilities

$$c_N \;:=\; \sup\{c_{k,\ell}(N_k, N_\ell), c_{k,\ell_1,\ell_2}(N_k, N_{\ell_1}, N_{\ell_2}) : k, \ell, \ell_1, \ell_2 \in E, \ell_1 \neq \ell_2\}, \tag{16}$$

where $c_{k,\ell}(N_k, N_\ell)$ and $c_{k,\ell_1,\ell_2}(N_k, N_{\ell_1}, N_{\ell_2})$ are the coalescence probabilities defined via (6) and (7), in agreement with the standard choice (see [23]) $c_N = \Phi_1(2)$ for the 1-type case. Intuitively, if time is measured in units of $\lfloor 1/c_N \rfloor$ generations, a possible limiting process of the time-scaled ancestral process as $N_{\min} \to \infty$ should capture the genealogy for all parts of the ancestral process emerging from maximal coalescence probabilities, i.e., for all parts having the fastest evolution backward in time. This intuition is made rigorous by Theorem 2 above. If $c_N \to c$ as $N_{\min} \to \infty$ for some constant $c > 0$, then we are back in the situation of Theorem 1, where c_N is not needed for the formulation of the convergence result. In Section 7.1, Theorem 2 is applied with $c_N := 1/N_{\min}$ to the multi-type Wright–Fisher model leading in the limit to the multi-type Kingman n-coalescent.

5 Multi-type exchangeable coalescents

In the following we restrict our attention to the continuous-time setting in the spirit of Theorem 2. Similar results hold for the discrete-time setting considered in Theorem 1, but are omitted here for the sake of shortness and simplicity.

For $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ let $\mathcal{T}_j$ denote the set of all tensors $T = (t_{k,\ell})_{k,\ell\in E}$ with vector entries $t_{k,\ell} \in \mathbb{N}_0^{j_k}$ for all $k, \ell \in E$. As in Theorem 2 it is from now on assumed that for every $N = (N_k)_{k\in E}$ there exists a constant $c_N > 0$ such that the limits

$$\phi_j(T) \;:=\; \lim_{N_{\min}\to\infty} \frac{\Phi_j^{(N)}(T) - \delta_{T,\mathbf{1}_j}}{c_N} \tag{17}$$

exist for all $j \in \mathbb{N}_0^E$ and all tensors $T \in \mathcal{T}_j$. Note that $\phi_j(\mathbf{1}_j) \le 0$ for all $j \in \mathbb{N}_0^E$. In the following it is shown that the consistency equation (10) of the functions $\Phi_j^{(N)}$ carries over to the limits (17). More precisely, for all $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ and all $T = (t_{k,\ell})_{k,\ell\in E} \in \mathcal{T}_j$, the equality

$$\phi_j(T) \;=\; \sum_{k\in E} \phi_{j+e_k}(T(k,\ell)) + \sum_{k\in E}\sum_{s=1}^{j_k} \phi_j(T(k,\ell,s)) \tag{18}$$

holds for all $\ell \in E$, where e_k denotes the k-th unit vector in $\mathbb{R}^E$, the tensor $T(k,\ell)$ is obtained from T by replacing the vector $t_{k,\ell} = (i_{k,\ell,1}, \ldots, i_{k,\ell,j_k})$ by $(i_{k,\ell,1}, \ldots, i_{k,\ell,j_k}, 1)$ and the (possibly empty) vector $t_{k,\ell'} = (i_{k,\ell',1}, \ldots, i_{k,\ell',j_k})$ by $(i_{k,\ell',1}, \ldots, i_{k,\ell',j_k}, 0)$ for all $\ell' \ne \ell$, and the tensor $T(k,\ell,s)$ is obtained from T by replacing the single entry $i_{k,\ell,s}$ by $i_{k,\ell,s} + 1$. For $T \in \mathcal{T}_j \setminus \{\mathbf{1}_j\}$, Eq. (18) follows by dividing Eq. (9) by c_N and taking the limit $N_{\min} \to \infty$. For the particular tensor $T = \mathbf{1}_j$, we conclude from (9) that, for all $\ell \in E$,

$$\begin{aligned}
\frac{\Phi_j^{(N)}(\mathbf{1}_j) - 1}{c_N} &= \sum_{k\in E} \frac{\Phi_{j+e_k}^{(N)}(\mathbf{1}_j(k,\ell))}{c_N} + \sum_{k\in E}\sum_{s=1}^{j_k} \frac{\Phi_j^{(N)}(\mathbf{1}_j(k,\ell,s))}{c_N} - \frac{1}{c_N} \\
&= \frac{\Phi_{j+e_\ell}^{(N)}(\mathbf{1}_{j+e_\ell}) - 1}{c_N} + \sum_{k\ne\ell} \frac{\Phi_{j+e_k}^{(N)}(\mathbf{1}_j(k,\ell))}{c_N} + \sum_{k\in E}\sum_{s=1}^{j_k} \frac{\Phi_j^{(N)}(\mathbf{1}_j(k,\ell,s))}{c_N}.
\end{aligned}$$

Taking the limit $N_{\min} \to \infty$ yields

$$\begin{aligned}
\phi_j(\mathbf{1}_j) &= \phi_{j+e_\ell}(1_{j+e_\ell}) + \sum_{k\ne\ell} \phi_{j+e_k}(\mathbf{1}_j(k,\ell)) + \sum_{k\in E}\sum_{s=1}^{j_k} \phi_j(\mathbf{1}_j(k,\ell,s)) \\
&= \sum_{k\in E} \phi_{j+e_k}(\mathbf{1}_j(k,\ell)) + \sum_{k\in E}\sum_{s=1}^{j_k} \phi_j(\mathbf{1}_j(k,\ell,s)),
\end{aligned}$$

which shows that (18) holds as well for $T = \mathbf{1}_j$.

From Corollary 1 it follows that the functions ϕ_j, $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$, are monotone in the sense that

$$0 \;\le\; \phi_{j'}(T') \;\le\; \phi_j(T) \tag{19}$$

for all $j, j' \in \mathbb{N}_0^E$ with $j \le j'$ and all tensors $T \in \mathcal{T}_j \setminus \{\mathbf{1}_j\}$ and $T' \in \mathcal{T}_{j'} \setminus \{\mathbf{1}_{j'}\}$ with $T \le T'$. Furthermore, $\phi_{j'}(\mathbf{1}_{j'}) \le \phi_j(\mathbf{1}_j) \le 0$ for all $j, j' \in \mathbb{N}_0^E$ with $j \le j'$.

Following the proof of Corollary 2, but with conditional probabilities replaced by infinitesimal rates (see also Burke and Rosenblatt [5]), it follows by exploiting the consistency equations (18) that the limiting processes $\Pi^{(n)}$, $n \in \mathbb{N}$, satisfy the natural coupling property,

i.e., for all $m, n \in \mathbb{N}$ with $m \leq n$, the process $(\varrho_{n,m} \circ \Pi_t^{(n)})_{t\geq 0}$ has the same distribution as $\Pi^{(m)}$. The Daniell–Kolmogorov extension theorem ensures that there exists a process Π taking values in the space $\mathcal{P}_{\infty,E}$ of labeled partitions of $\mathbb{N}$ such that, for each $n \in \mathbb{N}$, the restricted process $(\varrho_n \circ \Pi_t)_{t\geq 0}$ has the same distribution as $\Pi^{(n)}$. The process Π is called a *multi-type coalescent*. Its distribution is determined by the rate functions ϕ_j, $j \in \mathbb{N}_0^E$. We have therefore verified the existence of a general class of multi-type coalescent processes allowing for simultaneous multiple collisions of ancestral lineages. Note that if $\mathcal{A}_0$ is exchangeable, then, by Lemma 1, the ancestral process $(\mathcal{A}_r^{(n,N)})_{r\in\mathbb{N}_0}$ is exchangeable and this exchangeability carries over to the limiting multi-type coalescent Π. As we shall see in the following section, the structure and analysis of these multi-type exchangeable coalescent processes is more involved as it seems to be at a first glance and leads to further open questions.

6 Integral representations

It is well-known (Pitman [26], Sagitov [28]) that (single-type) coalescents with multiple collisions can be characterized by a finite measure Λ on the unit interval (Λ-coalescent). Similarly, single-type exchangeable coalescents (allowing for simultaneous multiple mergers of ancestral lineages) can be either characterized (see, for example, M. and Sagitov [23]) by a certain sequence $(Q_j)_{j\in\mathbb{N}}$ of finite measures Q_j on the finite simplex

$$\Delta_j \;:=\; \{(x_1, \ldots, x_j) \in [0,1]^j : \sum_{i=1}^{j} x_i \leq 1\}, \tag{20}$$

$j \in \mathbb{N}$, or (Schweinsberg [31]; see also Sagitov [29]) by a single finite measure Ξ on the infinite simplex

$$\Delta \;:=\; \{x = (x_i)_{i\in\mathbb{N}} : x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i\in\mathbb{N}} x_i \leq 1\} \tag{21}$$

(Ξ-coalescent). Note that Δ is a compact Hausdorff Polish space. In common for these characterizations is the fact that the infinitesimal rates of the coalescent are expressed in terms of certain integrals over the corresponding measure Λ, the sequence of measures Q_j, $j \in \mathbb{N}$, or the measure Ξ, respectively. In this section we aim for similar characterizations for multi-type exchangeable coalescents.

In the following we focus on the particular subclass of multi-type exchangeable coalescents having the additional property that the support of the rate functions ϕ_j is concentrated on diagonal tensors $T \in \mathcal{T}_j$ as already defined at the beginning of Section 4. Having a sort of asymptotically weak mutation regime in mind, this is a natural condition. In this case the general consistency equation (18) considerably simplifies to

$$\phi_j(T) \;=\; \phi_{j+e_\ell}(T(\ell,\ell)) + \sum_{s=1}^{j_\ell} \phi_j(T(\ell,\ell,s)), \qquad \ell \in E. \tag{22}$$

Compared to the rather involved general equations (18), the advantage of (22) is the fact that the values $\phi_j(T)$ of the family of functions ϕ_j with diagonal tensors T having only entries greater than or equal to 2 fully determine the functions ϕ_j. This crucial fact is readily verified using (22) by induction on the number of ones in the tensor T, and paves the way

to obtain integral representations in the spirit of M. and Sagitov [23], Schweinsberg [31] and Sagitov [29] for rate functions ϕ_j, $j \in \mathbb{N}_0^E$, with support concentrated on diagonal tensors.

Integral representations for rate functions ϕ_j, $j \in \mathbb{N}_0^E$, satisfying the general consistency equations (18), however, remain unknown. The general consistency equations (18) also motivate to study a new class of multi-type exchangeable partition probability functions (M-EPPF) as defined in Section 10.3 in the appendix. These stimulating open problems are left for future work.

In the remaining part of this section it is always assumed that the support of each function ϕ_j is concentrated on diagonal tensors. We now turn to the desired integral representations.

For $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ let Δ_j denote the simplex of all $x := (x_{k,s})_{k\in E, s\in[j_k]}$ satisfying $x_{k,s} \geq 0$ for all $k \in E$ and all $s \in [j_k]$ and $\sum_{k\in E}\sum_{s=1}^{j_k} x_{k,s} \leq 1$. For $|E| = 1$, this definition of Δ_j is in agreement with (20). The following integral representation is the natural extension of [23, Lemma 3.1].

Proposition 2. *If the limits $\phi_j(T)$ exist for all $j \in \mathbb{N}_0^E$ and all diagonal tensors $T = (t_{k,\ell})_{k,\ell\in E} \in \mathcal{T}_j$ with diagonal entries $t_{k,k} \in \{2,3,\ldots\}^{j_k}$ for all $k,\ell \in E$, then there exists for every $j \in \mathbb{N}_0^E$ a finite measure Q_j on Δ_j uniquely determined by its moments*

$$\int_{\Delta_j} \prod_{k\in E}\prod_{s=1}^{j_k} x_{k,s}^{i_{k,s}-2}\, Q_j\big(\mathrm{d}((x_{k,s})_{k\in E, s\in[j_k]})\big) \;=\; \phi_j(T), \tag{23}$$

$i_{k,s} \in \{2,3,\ldots\}$, $k \in E$, $s \in [j_k]$, where $T := (t_{k,\ell})_{k,\ell\in E}$ is the diagonal tensor with diagonal entries $t_{k,k} := (i_{k,s})_{s\in[j_k]}$, $k \in E$. Moreover, the total masses of the family of measures Q_j, $j \in \mathbb{N}_0^E$, satisfy $Q_j(\Delta_j) \geq Q_{j'}(\Delta_{j'})$ for all $j, j' \in \mathbb{N}_0^E$ with $j \leq j'$.

Remark. If all the measures Q_j, $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ with $\sum_{k\in E} j_k > 1$, are equal to the zero measure, then the corresponding coalescent Π is a multi-type Λ-coalescent characterized by the family $\Lambda := (\Lambda_k)_{k\in E}$ of measures Λ_k defined via $\Lambda_k := Q_{e_k}$ for all $k \in E$. The measure Λ_k characterizes the rates of multiple collisions of k-blocks to a single k-block. In particular, simultaneous multiple collisions of ancestral lineages are impossible and as well multiple collisions of k-blocks to an ℓ-block with $\ell \neq k$ are impossible for this coalescent.

The following theorem extends the 'only if' part of Schweinsberg [31, Theorem 2]. It characterizes the rates of an exchangeable multi-type coalescent Π in terms of a sequence $(a_k)_{k\in E} \in [0,\infty)^E$ and a single finite measure Ξ on $(\Delta \setminus \{0\}) \times E^{\mathbb{N}}$. Because of its length the proof of Theorem 3 is provided in the separate Section 9. For $x = (x_i)_{i\in\mathbb{N}} \in \Delta$ define $(x,x) := \sum_{i\in\mathbb{N}} x_i^2$.

Theorem 3. *Suppose that the rate functions ϕ_j of the multi-type coalescent Π are concentrated on diagonal tensors. Then there exists a sequence $(a_k)_{k\in E} \in [0,\infty)^E$ and a finite measure Ξ on $(\Delta \setminus \{0\}) \times E^{\mathbb{N}}$ such that for all $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ and all diagonal tensors $T = (t_{k,\ell})_{k,\ell\in E}$ with diagonal entries $t_{k,k} := (i_{k,s})_{s\in[j_k]} \in \{2,3,\ldots\}^{j_k}$, the representation*

$$\phi_j(T) \;=\; \sum_{k\in E} a_k 1_{\{j=\mathbf{e}_k, i_{k,1}=2\}} + \int_{(\Delta\setminus\{\mathbf{0}\})\times E^{\mathbb{N}}} \sum_{m_{k,s}} \prod_{k\in E}\prod_{s=1}^{j_k} x_{m_{k,s}}^{i_{k,s}}\, 1_{\{k\}}(y_{m_{k,s}})\, \frac{\Xi(\mathrm{d}(x,y))}{(x,x)} \tag{24}$$

holds, where the sum $\sum_{m_{k,s}}$ extents over all pairwise distinct $m_{k,s} \in \mathbb{N}$, $k \in E$, $s \in [j_k]$.

Remark. For the single-type case, (24) reduces to the well-known representation of the rates in terms of a finite measure Ξ on Δ via

$$\phi_j(i_1,\ldots,i_j) \;=\; a1_{\{j=1,i_1=2\}} + \int_{\Delta\setminus\{\mathbf{0}\}} \sum_{\substack{m_1,\ldots,m_j\in\mathbb{N}\\ \text{all distinct}}} \prod_{s=1}^{j} x_{m_s}^{i_s} \frac{\Xi(\mathrm{d}x)}{(x,x)}$$

for all $j\in\mathbb{N}$ and $i_1,\ldots,i_j\in\{2,3,\ldots\}$, where $a:=\Xi(\{\mathbf{0}\})$.

Remark. Although rather different in detail, the representation (24) shares some structural similarities with the Lévy-Khintchine formula for infinitely divisible distributions on multi-dimensional spaces. We refer the reader exemplary to Applebaum [3] for some background on Lévy processes and infinite divisibility. The first part $\sum_{k\in E} a_k 1_{\{j=e_k,i_{k,1}=2\}}$ in (24), called the Kingman part, characterizing the binary merging rates of the multi-type exchangeable coalescent. This part can in some sense be viewed as a 'normal' or 'Brownian' part reflecting the generally accepted statement that the Kingman coalescent is the 'normal distribution' of ancestral population genetics. The second (integral) part in (24) characterises the simultaneous multiple merger rates (resulting from large offspring sizes). This part can be viewed as a sort of jump Lévy part.

7 Examples

In the following two examples the numbers $N_{k,\ell}$, $k,\ell\in E$, are (mainly) deterministic. For simplicity it is also assumed that the type space E is finite, i.e., $d:=|E|\in\mathbb{N}$.

7.1 Multi-type Wright–Fisher model

Let $N_{k,\ell}\in\mathbb{N}_0$, $k,\ell\in E$, be given non-negative integer constants with (see (1)) $N_\ell := \sum_{k\in E} N_{k,\ell}\in\mathbb{N}$ for all $\ell\in E$. Assume that for every $k,\ell\in E$ the vector $\nu_{k,\ell}:=(\nu_{k,\ell,i})_{i\in[N_k]}$ has a symmetric multinomial distribution with parameters $N_{k,\ell}$ and $(1/N_k)_{i\in[N_k]}$. Assume furthermore that the $\nu_{k,\ell}$, $k,\ell\in E$, are independent. We call this model the *multi-type Wright–Fisher model*. In this case the joint distribution of the offspring sizes is given by

$$\mathbb{P}\Big(\bigcap_{k,\ell\in E}\bigcap_{i=1}^{N_k}\{\nu_{k,\ell,i}=n_{k,\ell,i}\}\Big) \;=\; \prod_{k,\ell\in E}\frac{N_{k,\ell}!}{\prod_{i=1}^{N_k} n_{k,\ell,i}!}\Big(\frac{1}{N_k}\Big)^{N_{k,\ell}} \tag{25}$$

for all $n_{k,\ell,i}\in\mathbb{N}_0$ with $\sum_{i=1}^{N_k} n_{k,\ell,i}=N_{k,\ell}$ for all $k,\ell\in E$. It is readily seen from (25) that (A), (A1) and (A2) hold.

Any multinomial random variable $X:=(X_s)_{s\in S}$ with at most countable set S and parameters $n\in\mathbb{N}_0$ and $(p_s)_{s\in S}$ has joint descending factorial moments $\mathbb{E}(\prod_{s\in S}(X_s)_{i_s}) = (n)_{\sum_{s\in S} i_s}\prod_s p_s^{i_s}$ for all $(i_s)_{s\in S}\in\mathbb{N}_0^S$. Thus, the transition probability (3) of the ancestral process takes the form

$$\begin{aligned} p_{\pi,\pi'} \;&=\; \prod_{k\in E}\Big(\frac{(N_k)_{j_k}}{(N_k)_{i_k}}\prod_{\ell\in E}\mathbb{E}\Big(\prod_{s=1}^{j_k}(\nu_{k,\ell,s})_{i_{k,\ell,s}}\Big)\Big)\\ &=\; \prod_{k\in E}\Big(\frac{(N_k)_{j_k}}{(N_k)_{i_k}}\prod_{\ell\in E}\frac{(N_{k,\ell})_{i_{k,\ell}}}{N_k^{i_{k,\ell}}}\Big) \;=:\; \Phi_j(T), \end{aligned} \tag{26}$$

where $i_{k,\ell} := \sum_{s=1}^{j_k} i_{k,\ell,s}$ for all $k,\ell \in E$. In particular, the diagonal entries are

$$p_{\pi,\pi} \;=\; \prod_{k\in E} \frac{(N_{k,k})_{i_k}}{N_k^{i_k}}, \qquad \pi \in \mathcal{P}_{n,E}, \tag{27}$$

since for $\pi = \pi'$, $i_{k,\ell,s} = \delta_{k,\ell}$ (Kronecker symbol) and, hence, $i_{k,\ell} = j_k\delta_{k,\ell}$, $k,\ell \in E$. For the single-type case $E = \{1\}$, (26) reduces to the well-known expression $p_{\pi,\pi'} = (N)_j/N^i$ for the classical (single-type) Wright–Fisher model with $N := N_1$, $i := i_1$ and $j := j_1$. For the two-type case $E = \{1,2\}$ and sample size $n = 2$, the state space $\mathcal{P}_{n,E} = \mathcal{P}_{2,\{1,2\}}$ consists of the six states

$$\begin{array}{lll}
\pi_1 := \{(\{1,2\},1)\}, & \pi_2 := \{(\{1,2\},2)\}, & \pi_3 := \{(\{1\},1),(\{2\},1)\}, \\
\pi_4 := \{(\{1\},1),(\{2\},2)\}, & \pi_5 := \{(\{1\},2),(\{2\},1)\}, & \pi_6 := \{(\{1\},2),(\{2\},2)\}.
\end{array}$$

In this case, by (26), the transition matrix $P = (p_{\pi_i,\pi_j})_{1\le i,j\le 6}$ of the ancestral process is given by

$$P \;=\; \left(\begin{array}{cc|cccc}
m_{11} & m_{12} & 0 & 0 & 0 & 0 \\
m_{21} & m_{22} & 0 & 0 & 0 & 0 \\
\hline
\frac{(N_{1,1})_2}{N_1(N_1)_2} & \frac{(N_{2,1})_2}{(N_1)_2N_2} & \frac{(N_{1,1})_2}{N_1^2} & \frac{N_{1,1}N_{2,1}}{(N_1)_2} & \frac{N_{1,1}N_{2,1}}{(N_1)_2} & \frac{(N_2)_2(N_{2,1})_2}{(N_1)_2N_2^2} \\
\frac{N_{1,1}N_{1,2}}{N_1^2N_2} & \frac{N_{2,1}N_{2,2}}{N_1N_2^2} & \frac{(N_1)_2N_{1,1}N_{1,2}}{N_1^3N_2} & \frac{N_{1,1}N_{2,2}}{N_1N_2} & \frac{N_{1,2}N_{2,1}}{N_1N_2} & \frac{(N_2)_2N_{2,1}N_{2,2}}{N_1N_2^3} \\
\frac{N_{1,1}N_{1,2}}{N_1^2N_2} & \frac{N_{2,1}N_{2,2}}{N_1N_2^2} & \frac{(N_1)_2N_{1,1}N_{1,2}}{N_1^3N_2} & \frac{N_{1,2}N_{2,1}}{N_1N_2} & \frac{N_{1,1}N_{2,2}}{N_1N_2} & \frac{(N_2)_2N_{2,1}N_{2,2}}{N_1N_2^3} \\
\frac{(N_{1,2})_2}{N_1(N_2)_2} & \frac{(N_{2,2})_2}{N_2(N_2)_2} & \frac{(N_1)_2(N_{1,2})_2}{N_1^2(N_2)_2} & \frac{N_{1,2}N_{2,2}}{(N_2)_2} & \frac{N_{1,2}N_{2,2}}{(N_2)_2} & \frac{(N_{2,2})_2}{N_2^2}
\end{array}\right),$$

where $m_{k,\ell} := N_{\ell,k}/N_k$, $k,\ell \in E$, denote the backward mutation probabilities.

The coalescence probabilities (6) and (7) reduce to

$$c_{k,\ell}(N_k,N_\ell) \;=\; \frac{N_k}{(N_\ell)_2}\mathbb{E}((\nu_{k,\ell,1})_2) \;=\; \frac{(N_{k,\ell})_2}{N_k(N_\ell)_2}, \qquad k,\ell \in E, \tag{28}$$

and, for $k,\ell_1,\ell_2 \in E$ with $\ell_1 \neq \ell_2$,

$$\begin{aligned}
c_{k,\ell_1,\ell_2}(N_k,N_{\ell_1},N_{\ell_2}) \;&=\; \frac{N_k}{N_{\ell_1}N_{\ell_2}}\mathbb{E}(\nu_{k,\ell_1,1})\mathbb{E}(\nu_{k,\ell_2,1}) \\
&=\; \frac{N_k}{N_{\ell_1}N_{\ell_2}}\frac{N_{k,\ell_1}}{N_k}\frac{N_{k,\ell_2}}{N_k} \;=\; \frac{N_{k,\ell_1}N_{k,\ell_2}}{N_kN_{\ell_1}N_{\ell_2}}.
\end{aligned} \tag{29}$$

Clearly, the consistency equation (9) holds by Proposition 1. Alternatively, (9) is verified directly as follows. From (26) it is readily seen that

$$\Phi_{j+e_k}(T(k,\ell)) \;=\; \Phi_j(T)\frac{N_k-j_k}{N_\ell-i_\ell}\frac{N_{k,\ell}-i_{k,\ell}}{N_k} \;=\; \frac{\Phi_j(T)}{N_\ell-i_\ell}\left(1-\frac{j_k}{N_k}\right)(N_{k,\ell}-i_{k,\ell})$$

and

$$\Phi_j(T(k,\ell,s)) \;=\; \Phi_j(T)\frac{1}{N_\ell-i_\ell}\frac{N_{k,\ell}-i_{k,\ell}}{N_k} \;=\; \frac{\Phi_j(T)}{N_\ell-i_\ell}\frac{1}{N_k}(N_{k,\ell}-i_{k,\ell}).$$

In particular, the latter expression does not depend on s. Therefore,

$$\begin{aligned}
&\sum_{k\in E}\Phi_{j+e_k}(T(k,\ell)) + \sum_{k\in E}\sum_{s=1}^{j_k}\Phi_j(T(k,\ell,s)) \\
&\quad = \frac{\Phi_j(T)}{N_\ell - i_\ell}\sum_{k\in E}\left(1-\frac{j_k}{N_k}\right)(N_{k,\ell}-i_{k,\ell}) + \frac{\Phi_j(T)}{N_\ell - i_\ell}\sum_{k\in E}\frac{j_k}{N_k}(N_{k,\ell}-i_{k,\ell}) \\
&\quad = \frac{\Phi_j(T)}{N_\ell - i_\ell}\sum_{k\in E}(N_{k,\ell}-i_{k,\ell}) \;=\; \frac{\Phi_j(T)}{N_\ell - i_\ell}(N_\ell - i_\ell) \;=\; \Phi_j(T),
\end{aligned}$$

which is (9).

A weak mutation example

In the following all convergence statements and all asymptotic relations are meant as $N_{\min} := \min_{k\in E} N_k \to \infty$. Assume that $N_{k,k}/N_k \to 1$ as $N_{\min}\to\infty$ for all $k\in E$ or, equivalently, that $N_{k,\ell}/N_\ell \to 0$ as $N_{\min}\to\infty$ for all $k,\ell\in E$ with $k\neq\ell$. Note that this corresponds to a weak mutation regime. From $\mathbb{E}(\nu_{k,k,1}) = N_{k,k}/N_k \to 1$ and $\mathbb{E}(\nu_{k,k,1}\nu_{k,k,2}) = (N_{k,k})_2/N_k^2 \to 1$ for all $k\in E$ we conclude (Lemma 2) that $P_N \to I$. Let us also assume that the subpopulation sizes N_k satisfy the calibration assumption

$$\frac{N_{\min}}{N_k} \;\to\; a_k \tag{30}$$

as $N_{\min}\to\infty$ for some given real constants $a_k \geq 0$. If $a_k > 0$ for all $k\in E$ then (30) means that all subpopulation sizes are asymptotically of the same order. From (28) and (29) it follows for all $k,\ell\in E$ that $N_k c_{k,\ell}(N_k,N_\ell)\to\delta_{k,\ell}$ as $N_{\min}\to\infty$ and for all $k,\ell_1,\ell_2\in E$ with $\ell_1\neq\ell_2$ that $N_k c_{k,\ell_1,\ell_2}(N_k,N_{\ell_1},N_{\ell_2})\to 0$ as $N_{\min}\to\infty$. Recall that $N := (N_k)_{k\in E}$. The natural choice for the time scaling c_N is

$$c_N \;:=\; \frac{1}{N_{\min}} \;=\; \frac{1}{\min_{k\in E} N_k}. \tag{31}$$

Comparing the coalescence probability (28) with c_N we have

$$\frac{c_{k,\ell}(N_k,N_\ell)}{c_N} \;=\; N_{\min}c_{k,\ell}(N_k,N_\ell) \;\sim\; a_k N_k c_{k,\ell}(N_k,N_\ell) \;\to\; a_k\delta_{k,\ell},$$

which shows that, in the limit, binary mergers of two ℓ-individuals to a k-individual are only possible for $k=\ell$ (at the rate a_k). Similarly, for $k,\ell_1,\ell_2\in E$ with $\ell_1\neq\ell_2$,

$$\frac{c_{k,\ell_1,\ell_2}(N_k,N_{\ell_1},N_{\ell_2})}{c_N} \;=\; a_k N_k c_{k,\ell_1,\ell_2}(N_k,N_{\ell_1},N_{\ell_2}) \;\to\; 0,$$

showing that binary mergers of individuals of different types are impossible in the limit. Having the monotonicity (Corollary 1)) in mind, it follows that the assumptions of Theorem 2 hold with all the limits $\phi_j(T)$ equal to zero except for $\phi_{e_k}(T) = a_k$ in case of a binary merging of two k-individuals to a single k-individual. Thus, Theorem 2 is applicable. The

limiting Markov chain $\Pi^{(n)}$ in Theorem 2 is a multi-type Kingman n-coalescent in the sense that binary mergers of two ancestral lineages of the same type k to a single k-individual occur at the rate a_k. In particular, during any binary merging event a change of type is impossible. The measures Q_j in Proposition 2 are all equal to the zero measure except for Q_{e_k} being a Dirac measure on Δ_{e_k} assigning mass a_k to the single point $0 \in \Delta_{e_k}$, $k \in E$. The measure Ξ in Theorem 3 is the zero measure on $(\Delta \setminus \{0\}) \times E^{\mathbb{N}}$. If $a_k = 0$, then there is no activity in subpopulation k, which could be interpreted as a sort of seed bank. For general information on dormancy and seed bank models we refer the reader to the survey article [4] and the references therein. Similar multi-type Kingman (n-)coalescents, even with additional mutation forces, occur for example in [15, p. 4219] and [22, p. 107].

A strong mutation example

Let $M \in \mathbb{N}$ and assume that $N_{k,\ell} = M$ for all $k, \ell \in E$, corresponding to strong mutation. Then, $N_k = dM$, $k \in E$ and, hence, $N_{\min} := \min_{k \in E} N_k = dM$. Both coalescence probabilities, (28) and (29), are asymptotically equal to $1/(dM^3) = 1/(d^2 N_{\min})$. Therefore, a reasonable time-scaling is $c_N := 1/N_{\min}$. From (26) it follows for all $\pi, \pi' \in \mathcal{P}_{n,E}$ with $\pi \subseteq \pi'$ that

$$\begin{aligned} p_{\pi,\pi'} &= \frac{\prod_{k \in E} (N_{\min})_{j_k}}{\prod_{\ell \in E} (N_{\min})_{i_\ell}} \frac{\prod_{k,\ell \in E} (N_{\min}/d)_{i_{k,\ell}}}{\prod_{\ell \in E} N_{\min}^{i_\ell}} \\ &= \frac{N_{\min}^{|\pi'|}\Big(1 - c_N \sum_{k \in E} \binom{j_k}{2} + o(c_N)\Big)}{N^{|\pi|}\Big(1 - c_N \sum_{\ell \in E} \binom{i_\ell}{2} + o(c_N)\Big)} \frac{(\frac{N_{\min}}{d})^{|\pi|}\Big(1 - \frac{d}{N} \sum_{k,\ell \in E} \binom{i_{k,\ell}}{2} + o(c_N)\Big)}{N_{\min}^{|\pi|}} \\ &= \frac{N_{\min}^{|\pi'|-|\pi|}}{d^{|\pi|}} \big(1 + c_N \kappa(\pi, \pi') + o(c_N)\big) \end{aligned}$$

with $\kappa(\pi, \pi') := \sum_{\ell \in E} \binom{i_\ell}{2} - \sum_{k \in E} \binom{j_k}{2} - d \sum_{k,\ell \in E} \binom{i_{k,\ell}}{2}$. Thus, the transition matrix P_N of the ancestral process satisfies $P_N = A + c_N B + o(c_N)$ as $N_{\min} \to \infty$, where the matrices $A := (a_{\pi,\pi'})_{\pi,\pi' \in \mathcal{P}_{n,E}}$ and $B := (b_{\pi,\pi'})_{\pi,\pi' \in \mathcal{P}_{n,E}}$ have entries

$$a_{\pi,\pi'} := \begin{cases} d^{-|\pi|} & \text{if } r_n(\pi) = r_n(\pi'), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$b_{\pi,\pi'} := \begin{cases} d^{-|\pi|}\kappa(\pi, \pi') & \text{if } r_n(\pi) = r_n(\pi'), \\ d^{-|\pi|} & \text{if } r_n(\pi) \subset r_n(\pi') \text{ and } |\pi'| = |\pi| - 1, \\ 0 & \text{otherwise.} \end{cases}$$

The matrix A is stochastic and a projection ($A^2 = A$). By Theorem 1, the ancestral process $(\mathcal{A}_r^{(n,N)})_{r \in \mathbb{N}_0}$ converges in $D_{\mathcal{P}_{n,E}}(\mathbb{N}_0)$ as $N_{\min} \to \infty$ to a limiting discrete-time Markov chain $(\Pi_r^{(n)})_{r \in \mathbb{N}_0}$ with transition matrix A. No merging events occur in this limit. In each time step only the types of the blocks may change. More precisely, in each time step the resulting type of every block is chosen at random and independently of everything else. In this case the support of the limiting function ϕ_j in Theorem 1 is concentrated on tensors $T := (t_{k,\ell})_{k,\ell \in E}$,

where the entries $t_{k,\ell} := (i_{k,\ell,s})_{s\in[j_k]}$ satisfy the constrains

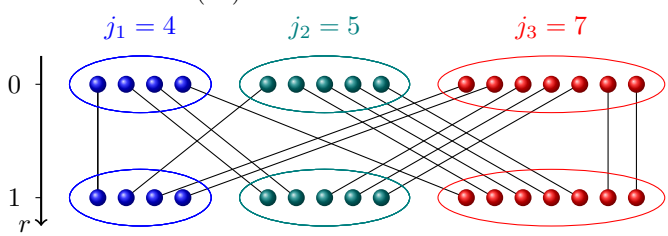
$$\begin{aligned}
&\bullet\ i_{k,\ell,s} \in \{0,1\} \text{ for all } k,\ell \in E \text{ and } s \in [j_k],\\
&\bullet\ \textstyle\sum_{k\in E}\sum_{s=1}^{j_k} i_{k,\ell,s} = j_\ell \text{ for all } \ell \in E \text{ and}\\
&\bullet\ \textstyle\sum_{\ell\in E}\sum_{s=1}^{j_\ell} i_{k,\ell,s} = j_k \text{ for all } k \in E.
\end{aligned} \tag{32}$$

In particular, the support of ϕ_j is not concentrated on diagonal tensors. For $E = \{1,2,3\}$, an example of a transition step with $j = (j_1, j_2, j_3) = (4,5,7)$ and corresponding tensor

$$T = \begin{pmatrix} (1,0,0,0) & (0,1,1,0) & (0,0,0,1) \\ (1,0,0,0,0) & (0,0,0,0,0) & (0,1,1,1,1) \\ (1,1,0,0,0,0,0) & (0,0,1,1,1,0,0) & (0,0,0,0,0,1,1) \end{pmatrix} \tag{33}$$

is illustrated in Figure 2.

Figure 2: A transition step of the limiting process Π for $E = \{1,2,3\}$ and $j = (4,5,7)$ corresponding to the tensor T in (33)



Having the expansion $P_N = A + c_N B + o(c_N)$ in mind, the ancestral structure can be analysed in more detail as follows. By [20, Theorem 1], the finite-dimensional distributions of the time-scaled ancestral process $(\mathcal{A}^{(n,N)}_{\lfloor t/c_N \rfloor})_{t\geq 0}$ converge as $N_{\min} \to \infty$ to those of a continuous-time limiting Markov process $(\widetilde{\Pi}^{(n)}_t)_{t\geq 0}$ with state space $\mathcal{P}_{n,E}$ and transition matrix Ae^{tG}, $t > 0$, with infinitesimal generator $G := ABA$. For example, if $n = 2$ and $d = 2$, then $N_{\min} = 2M$,

$$P_N = \left(\begin{array}{cc|cccc}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\hline
\frac{M-1}{4M(2M-1)} & \frac{M-1}{4M(2M-1)} & \frac{M-1}{4M} & \frac{M}{2(2M-1)} & \frac{M}{2(2M-1)} & \frac{M-1}{4M} \\
\frac{1}{8M} & \frac{1}{8M} & \frac{2M-1}{8M} & \frac{1}{4} & \frac{1}{4} & \frac{2M-1}{8M} \\
\frac{1}{8M} & \frac{1}{8M} & \frac{2M-1}{8M} & \frac{1}{4} & \frac{1}{4} & \frac{2M-1}{8M} \\
\frac{M-1}{4M(2M-1)} & \frac{M-1}{4M(2M-1)} & \frac{M-1}{4M} & \frac{M}{2(2M-1)} & \frac{M}{2(2M-1)} & \frac{M-1}{4M}
\end{array}\right).$$

and $P_N = A + c_N B + o(c_N)$ as $N_{\min} \to \infty$, where

$$A := \left(\begin{array}{cc|cccc} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array}\right) \quad \text{and} \quad B := \left(\begin{array}{cc|cccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & 0 & 0 & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & 0 & 0 & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \end{array}\right).$$

We refer the reader to Hössjer [13, Theorem 4.1] for a similar convergence result for the process counting the total number of ancestral lineages. Related convergence statements and proof methods based on Lemma 1 of [20] can be found in Kaj et al. [16], Nordborg and Krone [24] and Sagitov and Jagers [30]. We leave similar convergence results for the associated multi-type block counting process for the interested reader.

One may modify the multi-type Wright–Fisher model by assuming in addition that the $N_{k,\ell}$ are random. For example, one may assume that, for each $\ell \in E$, the vector $(N_{k,\ell})_{k\in E}$ has a multinomial distribution with parameters N_ℓ and $(u_{k,\ell})_{k\in E}$ for some given parameters $u_{k,\ell}$ and that these vectors are independent over $\ell \in E$. We leave the technical analysis of this modified Wright–Fisher model for the interested reader.

7.2 A multi-type mutation model without merging events

Suppose that in each time step a given number $N_{k,\ell} \in \mathbb{N}_0$ of individuals of type $k \in E$ mutate to type $\ell \in E$, where the numbers $N_{k,\ell}$, $k, \ell \in E$, satisfy the constrains

$$\sum_{\ell\in E} N_{k,\ell} \;=\; N_k \;=\; \sum_{\ell\in E} N_{\ell,k} \;\in\; \mathbb{N}, \qquad k \in E. \tag{34}$$

This mutation model can be formulated in terms of a model with offspring sizes $\nu_{k,\ell,i}$ as described in Section 1 using the interpretation that 'mutation to type ℓ' is equivalent to 'having one offspring of type ℓ'. Since, with this interpretation, each individual has exactly one offspring, the family sizes $\nu_{k,\ell,i}$ satisfy

$$\sum_{\ell\in E} \nu_{k,\ell,i} \;=\; 1, \qquad k \in E, i \in [N_k].$$

In particular, each $\nu_{k,\ell,i}$ takes values in $\{0, 1\}$. In contrast to the Wright–Fisher model, for arbitrary but fixed $k \in E$ and $i \in [N_k]$, the offspring sizes $\nu_{k,\ell,i}$, $\ell \in E$, are not independent. The joint distribution of the offspring sizes is obtained as follows. For $k \in E$ define $\nu_k := (\nu_{k,\ell,i})_{i\in[N_k],\ell\in E}$. It is readily seen that the ν_k, $k \in E$, are independent, where each ν_k is uniformly distributed on the set Ω_k consisting of all $(n_{k,\ell,i})_{\ell\in E,i\in[N_k]}$ satisfying

- $n_{k,\ell,i} \in \{0, 1\}$ for all $\ell \in E$ and $i \in [N_k]$,
- $\sum_{i=1}^{N_k} n_{k,\ell,i} = N_{k,\ell}$ for all $\ell \in E$, and
- $\sum_{\ell\in E} n_{k,\ell,i} = 1$ for all $i \in [N_k]$.

Note that $|\Omega_k| = N_k!/\prod_{\ell\in E} N_{k,\ell}!$. Thus, the joint offspring distribution is given by

$$\mathbb{P}\Big(\bigcap_{k,\ell\in E}\bigcap_{i=1}^{N_k}\{\nu_{k,\ell,i} = n_{k,\ell,i}\}\Big) \;=\; \prod_{k\in E}\frac{1}{|\Omega_k|} \;=\; \prod_{k\in E}\frac{\prod_{\ell\in E} N_{k,\ell}!}{N_k!} \tag{35}$$

for all $n_{k,\ell,i} \in \{0,1\}$, $k,\ell \in E$, $i \in [N_k]$, satisfying $\sum_{i=1}^{N_k} n_{k,\ell,i} = N_{k,\ell}$ for all $k,\ell \in E$ and $\sum_{\ell\in E} n_{k,\ell,i} = 1$ for all $k \in E$ and $i \in [N_k]$. In particular, Assumptions (A), (A1) and (A2) hold.

Looking backwards in time no merging events are possible. Thus, a transition of the discrete ancestral process from $\pi \in \mathcal{P}_{n,E}$ to $\pi' \in \mathcal{P}_{n,E}$ is only possible if π' has the same blocks as π (types of the blocks disregarded here). In this case, by (3),

$$\begin{aligned} p_{\pi,\pi'} \;&=\; \prod_{k\in E}\Big(\frac{(N_k)_{j_k}}{(N_k)_{i_k}}\mathbb{E}\Big(\prod_{\ell\in E}\prod_{s=1}^{j_k}(\nu_{k,\ell,s})_{i_{k,\ell,s}}\Big)\Big) \\ &=\; \prod_{k\in E}\Big(\frac{(N_k)_{j_k}}{(N_k)_{i_k}}\mathbb{P}\Big(\bigcap_{\ell\in E}\bigcap_{\substack{s=1\\ i_{k,\ell,s}=1}}^{j_k}\{\nu_{k,\ell,s}=1\}\Big)\Big) \;=\; \prod_{k\in E}\Big(\frac{(N_k)_{j_k}}{(N_k)_{i_k}}\frac{\frac{(N_k-j_k)!}{\prod_{\ell\in E}(N_{k,\ell}-i_{k,\ell})!}}{|\Omega_k|}\Big) \\ &=\; \prod_{k\in E}\frac{\prod_{\ell\in E}(N_{k,\ell})_{i_{k,\ell}}}{(N_k)_{i_k}} \;=\; \prod_{\ell\in E}\frac{\prod_{k\in E}(N_{k,\ell})_{i_{k,\ell}}}{(N_\ell)_{i_\ell}}, \end{aligned} \tag{36}$$

where, for $k,\ell \in E$, $i_{k,\ell} := \sum_{s=1}^{j_k} i_{k,\ell,s}$ denotes the number of blocks being a k-block of π' and an ℓ-block of π. Note that $i_\ell := \sum_{k\in E} i_{k,\ell}$ is the number of ℓ-blocks of π, $\ell \in E$, and that $j_k := \sum_{\ell\in E} i_{k,\ell}$ is the number of k-blocks of π'. In particular, the transition matrix of the ancestral process has diagonal entries

$$p_{\pi,\pi} \;=\; \sum_{k\in E}\frac{(N_{k,k})_{i_k}}{(N_k)_{i_k}}, \qquad \pi \in \mathcal{P}_{n,E}. \tag{37}$$

This model is a reformulation of the migration step considered around Figure 2 on p. 105 of [22] in terms of a multi-type Cannings model with offspring distribution (35). A migration from colony k to colony ℓ in the model in [22] is interpreted as a mutation from type k to type ℓ. Note that (36) is in agreement with [22, Eq. (12)]. Figure 2 of [22] shows a graphical representation of one generation step of this model for $E = \{1,2,3\}$, subpopulation sizes $N_1 := 4$, $N_2 := 5$ and $N_3 := 7$, and mutation numbers $N_{1,1} := 1$, $N_{1,2} := 2$, $N_{1,3} := 1$, $N_{2,1} := 1$, $N_{2,2} := 0$, $N_{2,3} := 4$, $N_{3,1} := 2$, $N_{3,2} := 3$ and $N_{3,3} := 2$.

Clearly, the transition probabilities (36) satisfy $\sum_{\pi'\in\mathcal{P}_{n,E}} p_{\pi,\pi'} = 1$ for all $\pi \in \mathcal{P}_{n,E}$, which can be also formally verified as follows. Fix $\pi \in \mathcal{P}_{n,E}$. For $\ell \in E$ let i_ℓ denote the number of ℓ-blocks of π. For any given matrix $A := (i_{k,\ell})_{k,\ell\in E}$ with nonnegative integer entries $i_{k,\ell} \in \mathbb{N}_0$ satisfying $\sum_{k\in E} i_{k,\ell} = i_\ell$ for all $\ell \in E$ let us denote with $\mathcal{P}(A)$ the subset of all $\pi' \in \mathcal{P}_{n,E}$ having the same (un-labeled) blocks as π and with the property that the number of blocks being a k-block of π' and an ℓ-block of π is equal to $i_{k,\ell}$. Note that $\mathcal{P}(A)$ depends on n, E and π, but this dependence is suppressed in our notation, since it is not important in the following. By (36),

$$\sum_{\pi'\in\mathcal{P}_{n,E}} p_{\pi,\pi'} \;=\; \sum_A\sum_{\pi'\in\mathcal{P}(A)} p_{\pi,\pi'} \;=\; \sum_A |\mathcal{P}(A)|\prod_{\ell\in E}\frac{\prod_{k\in E}(N_{k,\ell})_{i_{k,\ell}}}{(N_\ell)_{i_\ell}},$$

where the sum $\sum_A$ extents over all matrices $A = (i_{k,\ell})_{k,\ell\in E}$ as described above. Since $|\mathcal{P}(A)| = \prod_{\ell\in E} \frac{i_\ell!}{\prod_{k\in E} i_{k,\ell}!}$, it follows that

$$\sum_{\pi'\in\mathcal{P}_{n,E}} p_{\pi,\pi'} \;=\; \sum_A \prod_{\ell\in E} \frac{\prod_{k\in E}\binom{N_{k,\ell}}{i_{k,\ell}}}{\binom{N_\ell}{i_\ell}} \;=\; \prod_{\ell\in E}\Bigg(\sum_{\substack{(i_{k,\ell})_{k\in E}\in\mathbb{N}_0^E\\ \sum_{k\in E} i_{k,\ell}=i_\ell}} \frac{\prod_{k\in E}\binom{N_{k,\ell}}{i_{k,\ell}}}{\binom{N_\ell}{i_\ell}}\Bigg) \;=\; 1,$$

since the sum inside the last brackets is equal to 1 (total mass of a multi-hypergeometric distribution).

For every $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$, the support of the function Φ_j is concentrated on tensors T satisfying the constrains (32). Due to these constrains, the general consistency equations in Proposition 1 simplify to

$$\Phi_j(T) \;=\; \Phi_{j+e_\ell}(T(\ell,\ell)) + \sum_{s=1}^{j_\ell} \Phi_j(T(\ell,\ell,s)), \qquad \ell\in E.$$

These consistency equations are, except that we are in the discrete model before taking any limit, of the reduced form (22), although the support of the functions Φ_j is not concentrated on diagonal tensors.

If, for all $k,\ell\in E$, $N_{k,\ell}/N_\ell \to \rho_{k,\ell}$ as $N_{\min}\to\infty$ for some constants $\rho_{k,\ell}$, then Theorem 1 is applicable, since the transition probability in (36) converges to $\prod_{k,\ell\in E}\rho_{k,\ell} =: a_{\pi,\pi'}$ as $N_{\min}\to\infty$. The limiting process $(\Pi_r^{(n)})_{r\in\mathbb{N}_0}$ in Theorem 1 has transition probabilities $a_{\pi,\pi'}$ whenever π and π' have the same blocks (types of the blocks disregarded here).

Since all coalescence probabilities (6) and (7) are equal to zero in this model, Theorem 2 is not applicable with the standard choice (16) for the time-scaling c_N.

8 Proofs

We start this section with the proof of the formula (3) for the transition probability $p_{\pi,\pi'}$.

Proof of Eq. (3). Let $\mathcal{G}$ denote the σ-algebra generated by all the offspring sizes $\nu_{k,\ell,i}$, $k,\ell\in E$, $i\in[N_k]$. Since, independently for each type $\ell\in E$, individuals of type ℓ in the child generation are randomly assigned to parental offspring lineages of type ℓ, it follows conditional on the offspring sizes that

$$\mathbb{P}(\mathcal{A}_r = \pi' \,|\, \mathcal{A}_{r-1} = \pi, \mathcal{G}) \;=\; \sum_n \prod_{\ell\in E} \frac{\prod_{k\in E}\prod_{s=1}^{j_k}(\nu_{k,\ell,n(k,s)})_{i_{k,\ell,s}}}{(N_\ell)_{i_\ell}},$$

where the sum $\sum_n$ extents over all $n := (n(k,s))_{k\in E, s\in[j_k]}$ with $n(k,s)\in[N_k]$ for all $k\in E$ and $s\in[j_k]$ and $n(k,s)\neq n(k,r)$ for all $k\in E$ and $s,r\in[j_k]$ with $s\neq r$. Taking expectation yields

$$p_{\pi,\pi'} \;=\; \frac{1}{\prod_{\ell\in E}(N_\ell)_{i_\ell}} \sum_n \mathbb{E}\Bigg(\prod_{k,\ell\in E}\prod_{s=1}^{j_k}(\nu_{k,\ell,n(k,s)})_{i_{k,\ell,s}}\Bigg).$$

Exploiting the joint exchangeability (A) it follows that

$$p_{\pi,\pi'} \;=\; \frac{1}{\prod_{\ell\in E}(N_\ell)_{i_\ell}} \sum_n \mathbb{E}\Big(\prod_{k,\ell\in E} \prod_{s=1}^{j_k} (\nu_{k,\ell,s})_{i_{k,\ell,s}} \Big).$$

The last expectation does not depend on n. Thus, (3) follows from $\sum_n 1 = \prod_{k\in E}(N_k)_{j_k}$. $\square$

Next we provide a short and elegant proof of the crucial consistency relation (9).

Proof of Proposition 1. Define the random variable $X := \prod_{k,\ell\in E}\prod_{s\in[j_k]}(\nu_{k,\ell,s})_{i_{k,\ell,s}}$ and the two constants $A := \prod_{k\in E}(N_k)_{j_k}$ and $B := \prod_{\ell\in E}(N_\ell)_{i_\ell}$. Now fix $\ell\in E$ with $i_\ell < N_\ell$. For all $k\in E$ it follows from $N_{k,\ell} = \sum_{s=1}^{N_k}\nu_{k,\ell,s}$ and with the notation $i_{k,\ell} := \sum_{s=1}^{j_k} i_{k,\ell,s}$ that

$$\begin{aligned}
\mathbb{E}\big(X(N_{k,\ell} - i_{k,\ell})\big) \;&=\; \mathbb{E}\Big(X \sum_{s=j_k+1}^{N_k} \nu_{k,\ell,s}\Big) + \sum_{s=1}^{j_k} \mathbb{E}\big(X(\nu_{k,\ell,s} - i_{k,\ell,s})\big) \\
&=\; (N_k - j_k)\mathbb{E}(X\nu_{k,\ell,j_k+1}) + \sum_{s=1}^{j_k} \mathbb{E}\big(X(\nu_{k,\ell,s} - i_{k,\ell,s})\big)
\end{aligned}$$

by the joint exchangeability (A). Summation over all $k\in E$ yields

$$(N_\ell - i_\ell)\mathbb{E}(X) \;=\; \sum_{k\in E}(N_k - j_k)\mathbb{E}(X\nu_{k,\ell,j_k+1}) + \sum_{k\in E}\sum_{s=1}^{j_k}\mathbb{E}\big(X(\nu_{k,\ell,s} - i_{k,\ell,s})\big).$$

Multiplication of both sides of this equation with $A/((N_\ell - i_\ell)B)$ (>0) shows that

$$\frac{A}{B}\mathbb{E}(X) \;=\; \sum_{k\in E}\frac{(N_k - j_k)A}{(N_\ell - i_\ell)B}\mathbb{E}(X\nu_{k,\ell,j_k+1}) + \sum_{k\in E}\sum_{s=1}^{j_k}\frac{A}{(N_\ell - i_\ell)B}\mathbb{E}\big(X(\nu_{k,\ell,s} - i_{k,\ell,s})\big),$$

which is the desired Eq. (9), since $\frac{A}{B}\mathbb{E}(X) = \Phi_j(T)$ by the definition (8) of Φ_j, and, similarly, $\frac{(N_k-j_k)A}{(N_\ell-i_\ell)B}\mathbb{E}(X\nu_{k,\ell,j_k+1}) = \Phi_{j+e_k}\big(T(k,\ell)\big)$ and $\frac{A}{(N_\ell-i_\ell)B}\mathbb{E}\big(X(\nu_{k,\ell,s} - i_{k,\ell,s})\big) = \Phi_j\big(T(k,\ell,s)\big)$. $\square$

Proof of Corollary 1. For $j = j'$, (12) follows from the consistency equation (9). The result thus follows by induction on $\sum_{k\in E}(j'_k - j_k)$ and exploiting in each induction step the consistency equation (9) again. $\square$

Proof of Corollary 2. The proof exploits ideas going back at least to Burke and Rosenblatt [5]. Since $\varrho_{n,m} = \varrho_{m+1,m}\circ\cdots\circ\varrho_{n,n-1}$, we can and do assume without loss of generality that $m = n-1$. Define $f := \varrho_{n,m}$. Fix $\pi\in\mathcal{P}_{n,E}$ and $\tau'\in\mathcal{P}_{m,E}$. In the following an expression for

$$\mathbb{P}(f\circ\mathcal{A}_r^{(n)} = \tau' \,|\, \mathcal{A}_{r-1}^{(n)} = \pi) \;=\; \sum_{\pi'\in f^{-1}(\tau')} p_{\pi,\pi'} \tag{38}$$

is derived, which in particular shows that (38) depends on π only via $\tau := f(\pi)$.

If $\tau \not\subseteq \tau'$, then $\pi \not\subseteq \pi'$, i.e., $p_{\pi,\pi'} = 0$ for all $\pi' \in f^{-1}(\tau')$ and (38) is equal to zero. Assume now that $\tau \subseteq \tau'$.

Clearly, the partition τ' is of the form $\tau' = \{(B_1, k_1), \ldots, (B_a, k_a)\}$ with $a \in [m]$ blocks having types $k_1, \ldots, k_a \in E$ respectively. For $k \in E$ let $j_k := |\{\alpha \in [a] : k_\alpha = k\}|$ denote the number of k-blocks of τ'.

Since $\tau \subseteq \tau'$, the partition τ has blocks $B_{\alpha\beta}$, $\alpha \in [a]$, $\beta \in [b_\alpha]$, satisfying $B_\alpha = \bigcup_{\beta=1}^{b_\alpha} B_{\alpha\beta}$ for all $\alpha \in [a]$. Moreover, each block $B_{\alpha\beta}$ of τ is labeled with some type $\ell_{\alpha\beta} \in E$. For $\ell \in E$ let $i_\ell := |\{(\alpha, \beta) : \ell_{\alpha\beta} = \ell\}|$ denote the number of ℓ-blocks of τ. Furthermore, for $k, \ell \in E$ and $s \in [j_k]$, let $i_{k,\ell,s}$ denote the groups size of ℓ-blocks of τ merging to the s-th k-block of τ' and define the tensor $T := (t_{k,\ell})_{k,\ell \in E}$ via $t_{k,\ell} := (i_{k,\ell,s})_{s \in [j_k]}$ for all $k, \ell \in E$.

For $k \in E$ define $\pi'_0(k) := \{(B_1, k_1), \ldots, (B_a, k_a), (\{n\}, k)\} \in \mathcal{P}_{n,E}$ and for $\alpha \in [a]$ define $\pi'_\alpha := \{(B_1, k_1), \ldots, (B_\alpha \cup \{n\}, k_\alpha), \ldots, (B_a, k_a)\} \in \mathcal{P}_{n,E}$.

If $\{n\}$ is not a block of π, then there exist $\alpha \in [a]$ and $\beta \in [b_\alpha]$ such that $B_{\alpha\beta} \cup \{n\}$ is a block of π. Then, $\pi' := \pi'_\alpha$ is the only partition satisfying $\pi \subseteq \pi'$ and $f(\pi') = \tau'$. Hence (38) is in this case equal to $p_{\pi,\pi'_\alpha} = \Phi_j(T)$.

Assume now that $\{n\}$ is a block of π, i.e., there exists $\ell \in E$ such that $(\{n\}, \ell)$ is a labeled block of π. In this case exactly the partitions $\pi' \in \{\pi'_0(k) : k \in E\} \cup \{\pi'_1, \ldots, \pi'_a\}$ satisfy $\pi \subseteq \pi'$ and $f(\pi') = \tau'$. Therefore,

$$\begin{aligned}
\mathbb{P}(f \circ \mathcal{A}_r^{(n)} = \tau' \,|\, \mathcal{A}_{r-1}^{(n)} = \pi) \;&=\; \sum_{\pi' \in f^{-1}(\tau')} p_{\pi,\pi'} \;=\; \sum_{k \in E} p_{\pi,\pi'_0(k)} + \sum_{\alpha=1}^{a} p_{\pi,\pi'_\alpha} \\
&=\; \sum_{k \in E} \Phi_{j+e_k}(T(k,\ell)) + \sum_{k \in E} \sum_{s=1}^{j_k} \Phi_j(T(k,\ell,s)) \;=\; \Phi_j(T),
\end{aligned}$$

where the last equality holds by Lemma 9, the tensor $T(k,\ell)$ is obtained from T by replacing the vector $t_{k,\ell} = (i_{k,\ell,1}, \ldots, i_{k,\ell,j_k})$ by the vector $(i_{k,\ell,1}, \ldots, i_{k,\ell,j_k}, 1)$ and the vector $t_{k,\ell'} = (i_{k,\ell',1}, \ldots, i_{k,\ell',j_k})$ by the vector $(i_{k,\ell',1}, \ldots, i_{k,\ell',j_k}, 0)$ for all $\ell' \neq \ell$ and the tensor $T(k,\ell,s)$ is obtained from T by replacing the single entry $i_{k,\ell,s}$ by $i_{k,\ell,s} + 1$. Thus,

$$\mathbb{P}(f \circ \mathcal{A}_r^{(n)} = \tau' \,|\, \mathcal{A}_{r-1}^{(n)} = \pi) \;=\; \begin{cases} \Phi_j(T) & \text{if } f(\pi) \subseteq \tau', \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

In particular, (39) depends on π only via $\tau := f(\pi)$. Now, since $\{f \circ \mathcal{A}_{r-1}^{(n)} = \tau\} = \bigcup_{\pi \in f^{-1}(\tau)} \{\mathcal{A}_{r-1}^{(n)} = \pi\}$, an elementary calculation shows that (39) implies

$$\mathbb{P}(f \circ \mathcal{A}_r^{(n)} = \tau' \,|\, f \circ \mathcal{A}_{r-1}^{(n)} = \tau) \;=\; \begin{cases} \Phi_j(T) & \text{if } \tau \subseteq \tau', \\ 0 & \text{otherwise.} \end{cases}$$

Since $(\mathcal{A}_r^{(n)})_{r \in \mathbb{N}_0}$ is Markovian, all these calculations remain valid if in the condition the event $\{f \circ \mathcal{A}_{r-1}^{(n)} = \tau\}$ is replaced by $\{f \circ \mathcal{A}_{r-1}^{(n)} = \tau, f \circ \mathcal{A}_{r-2}^{(n)} = \tau_{r-2}, \ldots, f \circ \mathcal{A}_0^{(n)} = \tau_0\}$. Thus, $(f \circ \mathcal{A}_r^{(n)})_{r \in \mathbb{N}_0}$ is a Markov chain with the same transition probabilities (and the same initial state) as $(\mathcal{A}_r^{(m)})_{r \in \mathbb{N}_0}$. □

Let us now turn to the proofs concerning symmetry and exchangeability issues. For the proof of Lemma 1 the following result is needed.

Lemma 3. *For all $\pi, \pi' \in \mathcal{P}_{n,E}$ and all $\sigma \in S_n$ we have $p_{\pi,\pi'} = p_{\sigma(\pi),\sigma(\pi')}$.*

Proof. Two cases are distinguished. Assume first that $\pi \not\subseteq \pi'$. Then there exists some block $B \in \pi'$ that is not a union of blocks of π. Thus, for all blocks $A_1, \ldots, A_m$ of π, where $m \leq |\pi|$ ($|\pi|$ denotes the number of blocks of π) we have $\bigcup_{i=1}^m A_i \neq B$. Thus, also $\bigcup_{i=1}^m \sigma(A_i) \neq \sigma(B)$. Since every block of $\sigma(\pi)$ is of the form $\sigma(A)$ for some block A of π, it follows that $\sigma(\pi) \not\subseteq \sigma(\pi')$. Thus, $p_{\pi,\pi'} = 0 = p_{\sigma(\pi),\sigma(\pi')}$.

Assume now that $\pi \subseteq \pi'$. Let $T = ((i_{k,\ell,s})_{s\in[j_k]})_{k,\ell\in E}$, where j_k is the number of k-blocks of π' and $i_{k,\ell,s}$ denotes the number of ℓ-blocks of π merging to the s-th k-block of π'. Note that the number i_ℓ of ℓ-blocks of π can be recovered from T via $i_\ell = \sum_{k\in E}\sum_{s=1}^{j_k} i_{k,\ell,s}$. Thus, $p_{\pi,\pi'} = \Phi_j(T)$ depends only via T on π and π'.

Fix $k, \ell \in E$. For $s \in [j_k]$ let $(A_1, \ell), \ldots, (A_{i_{k,\ell,s}}, \ell)$ denote the ℓ-blocks merging to the s-th k-block (B, k) of π'. Then $(\sigma(B), k)$ is a k-block of $\sigma(\pi')$ and the ℓ-blocks of $\sigma(\pi)$ merging to $(\sigma(B), k)$ are exactly $(\sigma(A_1), \ell), \ldots, (\sigma(A_{i_{k,\ell,s}}), \ell)$, since it is straightforward to check that $\sigma(A_m) \subseteq \sigma(B)$ and $A \cap \sigma(B) = \emptyset$ for all ℓ-blocks (A, ℓ) of $\sigma(\pi)$, where $A \neq A_m$ for all $m \in [i_{k,\ell,s}]$. Thus, we have exactly $i_{k,\ell,s}$ ℓ-blocks of $\sigma(\pi)$ merging to this k-block of $\sigma(\pi')$. Firstly, this shows that $\sigma(\pi) \subseteq \sigma(\pi')$, since every k-block of $\sigma(\pi')$ is of the form $(\sigma(B), k)$ for some k-block (B, k) of π'.

In the same way as above, let $T' = ((a_{k,\ell,s})_{s\in[j_k]})_{k,\ell\in E}$, where $a_{k,\ell,s}$ denotes the number of ℓ-blocks of $\sigma(\pi)$ merging to the s-th k-block of $\sigma(\pi')$. The above secondly shows that $(a_{k,\ell,s})_{s\in[j_k]}$ is a permutation of $(i_{k,\ell,s})_{s\in[j_k]}$. By assumption (A) and writing $j = (j_k)_{k\in E}$, we thus obtain $p_{\pi,\pi'} = \Phi_j(T) = \Phi_j(T') = p_{\sigma(\pi),\sigma(\pi')}$. □

Thanks to Lemma 3 the proof of Lemma 1 is now straightforward and works as follows.

Proof of Lemma 1. We use induction on $r \in \mathbb{N}_0$. By assumption $\mathcal{A}_0$ is exchangeable. The induction step from $r-1$ to $r \in \mathbb{N}$ works as follows. For all $\pi' \in \mathcal{P}_{n,E}$ and $\sigma \in S_n$,

$$\mathbb{P}(\mathcal{A}_r = \pi') \;=\; \sum_{\pi\in\mathcal{P}_{n,E}} p_{\pi,\pi'}\mathbb{P}(\mathcal{A}_{r-1} = \pi) \;=\; \sum_{\pi\in\mathcal{P}_{n,E}} p_{\pi,\pi'}\mathbb{P}\big(\mathcal{A}_{r-1} = \sigma(\pi)\big)$$

by the induction hypothesis. Lemma 3 therefore yields

$$\mathbb{P}(\mathcal{A}_r = \pi') \;=\; \sum_{\pi\in\mathcal{P}_{n,E}} p_{\sigma(\pi),\sigma(\pi')}\mathbb{P}\big(\mathcal{A}_{r-1} = \sigma(\pi)\big).$$

With π also $\tau := \sigma(\pi)$ extents over all labeled partitions in $\mathcal{P}_{n,E}$. Therefore,

$$\mathbb{P}(\mathcal{A}_r = \pi') \;=\; \sum_{\tau\in\mathcal{P}_{n,E}} p_{\tau,\sigma(\pi')}\mathbb{P}(\mathcal{A}_{r-1} = \tau) \;=\; \mathbb{P}\big(\mathcal{A}_r = \sigma(\pi')\big).$$

Thus, $\mathcal{A}_r$ is exchangeable. □

We now turn to the proofs concerning asymptotic considerations as $N_{\min} := \min_{k\in E} N_k$ tends to infinity.

Proof of Lemma 2. Since $n \geq 2$, the transition matrix P_N has as entries in particular all the mean backward mutation probabilities $\mathbb{E}(N_{k,\ell})/N_\ell$, $k, \ell \in E$, and as well all coalescence

probabilities (6) and (7). Exploiting the monotonicity (Corollary 1) it follows that $P_N \to I$ as $N_{\min} \to \infty$ if and only if $\frac{N_k}{N_\ell}\mathbb{E}(\nu_{k,\ell,1}) = \mathbb{E}(N_{k,\ell})/N_\ell \to \delta_{k,\ell}$ (Kronecker symbol) for all $k, \ell \in E$ and if all coalescence probabilities tend to zero as $N_{\min} \to \infty$. From

$$c_{k,k}(N_k, N_k) \;=\; \frac{\mathbb{E}(\nu_{k,k,1}^2) - \mathbb{E}(\nu_{k,k,1})}{N_k - 1}$$

and the bounds in Lemma 6 in the appendix for the coalescence probabilities it therefore follows that $P_N \to I$ is equivalent to

$$\frac{\mathbb{E}(N_{k,\ell})}{N_\ell} \;\to\; \delta_{k,\ell} \quad \text{and} \quad \frac{\mathbb{E}(\nu_{k,k,1}^2)}{N_k - 1} \;\to\; 0 \qquad \text{for all } k, \ell \in E. \tag{40}$$

Thus, it suffices to verify that (40) is equivalent to

$$\mathbb{E}(\nu_{k,k,1}) \;\to\; 1 \quad \text{and} \quad \mathbb{E}(\nu_{k,k,1}\nu_{k,k,2}) \;\to\; 1 \qquad \text{for all } k \in E. \tag{41}$$

Assume first that (40) holds. Then, in particular $\mathbb{E}(\nu_{k,k,1}) = \mathbb{E}(N_{k,k})/N_k \to 1$. Moreover, since $N_{k,k}/N_k \leq 1$, we obtain for the second moment of $N_{k,k}/N_k$ the upper bound

$$\frac{\mathbb{E}(N_{k,k}^2)}{N_k^2} \;\leq\; \frac{\mathbb{E}(N_{k,k})}{N_k} \;\to\; 1 \tag{42}$$

and, by Jensen's inequality, the lower bound

$$\frac{\mathbb{E}(N_{k,k}^2)}{N_k^2} \;\geq\; \left(\frac{\mathbb{E}(N_{k,k})}{N_k}\right)^2 \;\to\; 1. \tag{43}$$

Thus, $\mathbb{E}(N_{k,k}^2)/(N_k)_2 \to 1$ and, hence,

$$\mathbb{E}(\nu_{k,k,1}\nu_{k,k,2}) \;=\; \frac{\mathbb{E}(N_{k,k}^2)}{(N_k)_2} - \frac{\mathbb{E}(\nu_{k,k,1}^2)}{N_k - 1} \;\to\; 1 - 1 \;=\; 0.$$

Thus, (41) holds.

Conversely, assume that (41) holds. Then $\mathbb{E}(N_{k,k})/N_k = \mathbb{E}(\nu_{k,k,1}) \to 1$ for all $k \in E$. Thus, for all $\ell \in E$,

$$\mathbb{E}\left(\sum_{k \neq \ell} \frac{N_{k,\ell}}{N_\ell}\right) \;=\; \mathbb{E}\left(1 - \frac{N_{\ell,\ell}}{N_\ell}\right) \;=\; 1 - \frac{\mathbb{E}(N_{\ell,\ell})}{N_\ell} \;\to\; 1 - 1 \;=\; 0.$$

In particular, $\mathbb{E}(N_{k,\ell})/N_\ell \to 0$ for all $k, \ell \in E$ with $k \neq \ell$. For the second moment of $N_{k,k}/N_k$ the bounds (42) and (43) are still valid. Thus, again $\mathbb{E}(N_{k,k}^2)/(N_k)_2 \to 1$ and, hence,

$$\frac{\mathbb{E}(\nu_{k,k,1}^2)}{N_k - 1} \;=\; \frac{\mathbb{E}(N_{k,k}^2)}{(N_k)_2} - \mathbb{E}(\nu_{k,k,1}\nu_{k,k,2}) \;\to\; 1 - 1 \;=\; 0.$$

Therefore, (40) holds.

To verify the last statement let $k, \ell \in E$ with $k \neq \ell$ and assume that $P_N \to I$. Then, as seen before, $\mathbb{E}(N_{k,\ell})/N_\ell \to 0$ and, hence,

$$\frac{(N_k)_2}{(N_\ell)_2}\mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2}) \;=\; \frac{N_k}{(N_\ell)_2}\mathbb{E}\bigg(\nu_{k,\ell,1}\sum_{i=2}^{N_k}\nu_{k,\ell,i}\bigg) \;\leq\; \frac{N_k}{N_\ell - 1}\mathbb{E}(\nu_{k,\ell,1}) \;=\; \frac{\mathbb{E}(N_{k,\ell})}{N_\ell - 1} \;\to\; 0.$$

Thus, $(N_k/N_\ell)^2\mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2}) \to 0$ for all $k, \ell \in E$ with $k \neq \ell$. □

Let us now turn to the proofs of the two convergence results, Theorem 1 and Theorem 2, respectively. Although the two proofs have much in common, we provide these proofs separately, since the proof of Theorem 1 is less technical and does not rely on any time-scaling.

Proof of Theorem 1. Let $\pi, \pi' \in \mathcal{P}_{n,E}$. If $\pi \not\subseteq \pi'$, then $p^{(N)}_{\pi,\pi'} = 0 =: a_{\pi,\pi'}$. Assume now that $\pi \subseteq \pi'$, i.e., each block of π' is a union of some blocks of π (types of the blocks disregarded here). Let i_ℓ and j_k denote the number of ℓ-blocks of π and k-blocks of π' respectively and let $i_{k,\ell,s}$, $k, \ell \in E$, $s \in [j_k]$, denote the group sizes of ℓ-blocks of π merging to the s-th k-block of π'. Furthermore, let $T = (t_{k,\ell})_{k,\ell\in E}$ denote the tensor with entries $t_{k,\ell} := (i_{k,\ell,s})_{s\in[j_k]}$, $k, \ell \in E$. Then, $p^{(N)}_{\pi,\pi'} = \Phi^{(N)}_j(T)$. If $T \neq \mathbf{1}_j$, then, by assumption, $p^{(N)}_{\pi,\pi'} = \Phi^{(N)}_j(T) \to \phi_j(T) =: a_{\pi,\pi'}$ as $N_{\min} \to \infty$. If $T = \mathbf{1}_j$, then $\pi = \pi'$ and

$$p^{(N)}_{\pi,\pi} \;=\; 1 - \sum_{\pi'\neq\pi} p^{(N)}_{\pi,\pi'} \;\to\; 1 - \sum_{\pi'\neq\pi} a_{\pi,\pi'} \;=:\; a_{\pi,\pi}, \qquad N_{\min} \to \infty.$$

Thus, the transition matrix $P_N = (p^{(N)}_{\pi,\pi'})_{\pi,\pi'\in\mathcal{P}_{n,E}}$ of the ancestral process converges to A as $N_{\min} \to \infty$, with the matrix A as defined in the statement of the theorem. The convergence of the finite-dimensional distributions of the ancestral process $(\mathcal{A}^{(n,N)}_r)_{r\in\mathbb{N}_0}$ and the convergence in $D_{\mathcal{P}_{n,E}}(\mathbb{N}_0)$ to a Markov chain $\Pi^{(n)} = (\Pi^{(n)}_r)_{r\in\mathbb{N}_0}$ with transition matrix A now follows immediately. □

Proof of Theorem 2. Let $\pi, \pi' \in \mathcal{P}_{n,E}$. If $\pi \not\subseteq \pi'$, then $p^{(N)}_{\pi,\pi'} = 0$ and, hence, $p^{(N)}_{\pi,\pi'}/c_N = 0 =: q_{\pi,\pi'}$. Assume now that $\pi \subseteq \pi'$, i.e., each block of π' is a union of some blocks of π (types of the blocks disregarded here). Let i_ℓ and j_k denote the number of ℓ-blocks of π and k-blocks of π' respectively and let $i_{k,\ell,s}$, $k, \ell \in E$, $s \in [j_k]$, denote the group sizes of ℓ-blocks of π merging to the s-th k-block of π'. Furthermore, let $T = (t_{k,\ell})_{k,\ell\in E}$ denote the tensor with entries $t_{k,\ell} := (i_{k,\ell,s})_{s\in[j_k]}$, $k, \ell \in E$. Then, $p^{(N)}_{\pi,\pi'} = \Phi^{(N)}_j(T)$. If $T \neq \mathbf{1}_j$, then, by assumption, $p^{(N)}_{\pi,\pi'}/c_N = \Phi^{(N)}_j(T)/c_N \to \phi_j(T) =: q_{\pi,\pi'}$ as $N_{\min} \to \infty$. If $T = \mathbf{1}_j$, then $\pi = \pi'$ and

$$\frac{1 - p^{(N)}_{\pi,\pi}}{c_N} \;=\; \sum_{\pi'\neq\pi} \frac{p^{(N)}_{\pi,\pi'}}{c_N} \;\to\; \sum_{\pi'\neq\pi} q_{\pi,\pi'} \;=:\; -q_{\pi,\pi}, \qquad N_{\min} \to \infty.$$

Thus, the transition matrix $P_N = (p^{(N)}_{\pi,\pi'})_{\pi,\pi'\in\mathcal{P}_{n,E}}$ of the ancestral process $(\mathcal{A}^{(n,N)}_r)_{r\in\mathbb{N}_0}$ satisfies $P_N = I + c_N Q + o(c_N)$ as $N_{\min} \to \infty$, with the matrix Q as defined in the statement of the theorem. The convergence of the finite-dimensional distributions of the time-scaled

ancestral process $(\mathcal{A}^{(n,N)}_{\lfloor t/c_N \rfloor})_{t\geq 0}$ as $N_{\min} \to \infty$ and the convergence in $D_{\mathcal{P}_{n,E}}([0,\infty))$ to a Markov process $\Pi^{(n)} = (\Pi^{(n)}_t)_{t\geq 0}$ with generator Q now follows as in the proof of Theorem 1 of [22] by applying Ethier and Kurtz [9, p. 168, Theorem 2.6]. □

We now turn to the proofs concerning integral representations.

Proof of Proposition 2. We proceed as in the proof of [23, Lemma 3.1]. Fix $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$. The particular diagonal tensor $(t_{k,\ell})_{k,\ell\in E} \in \mathcal{T}_j$ with diagonal entries $t_{k,k} := (2,\ldots,2) \in \mathbb{R}^{j_k}$ for all $k \in E$ is denoted by $\mathbf{2}_j$. If $\phi_j(\mathbf{2}_j) = 0$, then the statement holds with Λ_j being the zero measure on Δ_j. Assume now that $\phi_j(\mathbf{2}_j) > 0$. Then, $\mathbb{E}\big(\prod_{k\in E}\prod_{s=1}^{j_k}(\nu_{k,k,s})_2\big) > 0$ for all sufficiently large values of $N_{\min} := \min_{k\in E} N_k$. Let $Y_{k,s}$, $k \in E$, $s \in [j_k]$, be random variables with distribution

$$\mathbb{P}\Big(\bigcap_{k\in E}\bigcap_{s=1}^{j_k}\{Y_{k,s} = i_{k,s}\}\Big) \;:=\; \frac{\prod_{k\in E}\prod_{s=1}^{j_k}(i_{k,s})_2}{\mathbb{E}\big(\prod_{k\in E}\prod_{s=1}^{j_k}(\nu_{k,k,s})_2\big)}\mathbb{P}\Big(\bigcap_{k\in E}\bigcap_{s=1}^{j_k}\{\nu_{k,k,s} = i_{k,s}\}\Big).$$

Note that each $Y_{k,s}$ in general depends on $j = (j_k)_{k\in E}$, but this dependence is suppressed for convenience in our notation. The random variables $Y_{k,s}$ are a size biased modification of the offspring variables $\nu_{k,k,s}$. Note that $Y_{k,k,s}$ takes values in $\{2,\ldots,N_k\}$ almost surely. For all $i_{k,s} \in \mathbb{N}_0$, $k \in E$, $s \in [j_k]$, we have

$$\mathbb{E}\Big(\prod_{k\in E}\prod_{s=1}^{j_k} Y_{k,s}^{i_{k,s}}\Big) \;=\; \frac{\mathbb{E}\big(\prod_{k\in E}\prod_{s=1}^{j_k}(\nu_{k,k,s}^{i_{k,s}+2} - \nu_{k,k,s}^{i_{k,s}+1})\big)}{\mathbb{E}\big(\prod_{k\in E}\prod_{s=1}^{j_k}(\nu_{k,k,s})_2\big)}.$$

Using that $t^i = \sum_{m=0}^{i}(t)_m S(i,m)$ for all $t \in \mathbb{R}$ and $i \in \mathbb{N}_0$, where the $S(.,.)$ denote the Stirling numbers of the second kind, it follows from the assumptions of Theorem 2 that

$$\mathbb{E}\Big(\prod_{k\in E}\prod_{s=1}^{j_k}\Big(\frac{Y_{k,s}}{N_k}\Big)^{i_{k,s}}\Big) \;\to\; \frac{\phi_j(T+\mathbf{2}_j)}{\phi_j(\mathbf{2}_j)} \tag{44}$$

as $N_{\min} \to \infty$, where $T = (t_{k,\ell})_{k,\ell\in E}$ denotes the diagonal tensor with diagonal entries $t_{k,k} := (i_{k,s})_{s\in[j_k]}$, $k \in E$. Note that $(Y_{k,s}/N_k)_{k\in E, s\in[j_k]}$ is concentrated on Δ_j. Since Δ_j is compact, the convergence of all the moments (44) implies the existence of a probability measure P_j on Δ_j such that $(Y_{k,s}/N_k)_{k\in E,s\in[j_k]}$ converges in distribution to P_j as $N_{\min} \to \infty$. Thus, (23) holds with $Q_j := \phi_j(\mathbf{2}_j)P_j$. The measure Q_j is uniquely determined, since the limiting moments (44) fully determine P_j. For all $j, j' \in \mathbb{N}_0^E$ with $j \leq j'$ it follows from the monotonicity property (19) that $Q_j(\Delta_j) = \phi_j(\mathbf{2}_j) \geq \phi_{j'}(\mathbf{2}_{j'}) = Q_{j'}(\Delta_{j'})$. □

9 Proof of Theorem 3

We extend the proof of the 'only if' part of Schweinsberg [31, Theorem 2] to the multi-type setting. Let $\Pi = (\Pi_t)_{t\geq 0}$ denote the multi-type exchangeable coalescent with rates $\phi_j(T)$, $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$, $T \in \mathcal{T}_j$. We furthermore denote with $\Pi_t^{(n)} := \varrho_n \circ \Pi_t$ the restriction of Π_t to $[n]$ and by $\Pi^{(n)} := (\Pi_t^{(n)})_{t\geq 0}$ the corresponding n-coalescent. Define the stopping time

$$S \;:=\; \inf\{t > 0 : 1 \text{ and } 2 \text{ are in the same block of } \Pi_t\}$$

and the events

$$E_n \;:=\; \{1,\ldots,n \text{ are in distinct blocks of } \Pi_{S-}\}, \quad n\in\mathbb{N}.$$

As in [31] it follows that $\mathbb{P}(E_n)>0$. For $n\in\mathbb{N}$ let Θ_n be a random labeled partition of $[n]$ whose distribution is the conditional distribution of Π_S given E_n. We claim that there exists a random labeled partition Θ of $\mathbb{N}$ such that $\varrho_n\circ\Theta$ has the same distribution as Θ_n for all $n\in\mathbb{N}$. To prove this, let us first verify that $\varrho_{n,m}\circ\Theta_n$ has the same distribution as Θ_m for all $m,n\in\mathbb{N}$ with $m<n$. Fix $m<n$. Let $\theta\in\mathcal{P}_{m,E}$ such that 1 and 2 are in the same block of θ. For $k\in E$ let j_k denote the number of k-blocks of θ and $i_{k,s}\in\mathbb{N}$, $s\in[j_k]$, the corresponding block sizes. Define $\phi_1(\mathbf{2}):=\sum_{k\in E}\phi_{e_k}(\mathbf{2}_k)$, where $\mathbf{2}_k\in\mathcal{T}_{e_k}$ is the tensor with $i_{k,1}=2$. Note that $\phi_1(\mathbf{2})<\infty$. Furthermore, let T denote the diagonal tensor with diagonal entries $t_{k,k}:=(i_{k,s})_{s\in[j_k]}$, $k\in E$. If $\Pi_t^{(m)}$ has m singleton blocks, then these blocks merge to the blocks of θ at the rate $\phi_j(T)$. The total rate of all coalescence events involving 1 and 2 is $\phi_1(\mathbf{2})$. Thus,

$$\mathbb{P}(\Theta_m=\theta) \;=\; \frac{\phi_j(T)}{\phi_1(\mathbf{2})}. \tag{45}$$

In the same way, if $\Pi_t^{(n)}$ consists of n singletons, then the total rate of all coalescence events of $[n]$, such that their restriction to $[m]$ merges to the blocks of θ, is also $\phi_j(T)$. Therefore, $\mathbb{P}(\varrho_{n,m}\circ\Theta_n=\theta)=\phi_j(T)/\phi_1(\mathbf{2})$. Thus, $\varrho_{n,m}\circ\Theta_n$ has the same distribution as Θ_m. A standard application of the Daniell–Kolmogorov theorem yields the existence of Θ as the projective limit of the sequence $(\Theta_n)_{n\in\mathbb{N}}$.

Now, denote by Θ' the restriction of Θ to $\{3,4,\ldots\}$. Since Π is exchangeable, so is Θ'. Since exchangeability in the single-type and multi-type case coincide, copying the proof of Aldous [2, Section 11, p. 84 ff.], we obtain the existence of the limiting frequencies of the blocks of Θ', equipped with a (random) label, the label of the respective block. Write $(P_1,K_1),(P_2,K_2),\ldots$ for the pairs of limiting frequencies and their labels, where the P_i are ordered decreasingly and K_i is the label of P_i, $i\in\mathbb{N}$. We have $P_n=0$ if Θ' has fewer than n blocks with non-zero limiting frequencies. Let $P_0:=1-\sum_{j=1}^\infty P_j$. Note that the blocks of Θ also have limiting relative frequencies that coincide with those of Θ'. Write $B_1,B_2,\ldots$ for the blocks of Θ, such that block B_i has limiting frequency P_i on $\{P_i>0\}$ and blocks with the same limiting relative frequencies are ordered at random, independently of Θ. The block B_i is undefined on $\{P_i=0\}$.

Let $(x_\ell)_{\ell\in E}$ be some sequence of distinct real numbers in $[0,1)$. Define a sequence of random variables $Z_1,Z_2,\ldots$ via $Z_m:=i+x_\ell$ on the event $\{m\in B_i,K_i=\ell\}$ and $Z_m:=0$ on the complement of $\bigcup_{i\in\mathbb{N}}\bigcup_{\ell\in E}\{m\in B_i,K_i=\ell\}$, i.e., if the block containing m has limiting frequency 0. Furthermore, let $\mathcal{G}$ denote the σ-algebra generated by (P,K), where $P:=(P_i)_{i\in\mathbb{N}}$ and $K:=(K_i)_{i\in\mathbb{N}}$, and let $(x,x):=\sum_{i=1}^\infty x_i^2$ for $x=(x_i)_{i\in\mathbb{N}}\in\Delta$.

Lemma 4. *For all $m\in\{3,4,\ldots\}$, $i\in\mathbb{N}$ and $\ell\in E$ we have $\mathbb{P}(Z_m=i+x_\ell\,|\,\mathcal{G})=1_{\{K_i=\ell\}}P_i$ almost surely and $\mathbb{P}(Z_m=0\,|\,\mathcal{G})=P_0$ almost surely. Furthermore, $\mathbb{P}(Z_1=i+x_\ell\,|\,\mathcal{G})=1_{\{K_i=\ell\}}P_i^2/(P,P)$ almost surely on $\{P_1>0\}$. Moreover, $Z_1,Z_3,Z_4,\ldots$ are conditionally independent given $\mathcal{G}$.*

Proof. Since Θ' is exchangeable, also $Z_3,Z_4,\ldots$ are exchangeable. The σ-field generated by its limiting empirical distribution is $\mathcal{G}$. Therefore, by [31, Lemma 38], for all $m\in\{3,4,\ldots\}$,

$i \in \mathbb{N}$ and $\ell \in E$,

$$\begin{aligned}\mathbb{P}(Z_m = i + x_\ell \,|\, \mathcal{G}) &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{\{Z_m = i + x_\ell\}} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{\{m \in B_i, K_i = \ell\}} \\ &= 1_{\{K_i=\ell\}} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} 1_{\{m\in B_i\}} = 1_{\{K_i=\ell\}} P_i \quad \text{almost surely.}\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{P}(Z_m = 0 \,|\, \mathcal{G}) &= 1 - \sum_{i=1}^{\infty} \sum_{\ell\in E} \mathbb{P}(Z_m = i + x_\ell \,|\, \mathcal{G}) \\ &= 1 - \sum_{i=1}^{\infty} \sum_{\ell\in E} P_i 1_{\{K_i=\ell\}} = 1 - \sum_{i=1}^{\infty} P_i = P_0 \quad \text{almost surely.}\end{aligned}$$

Since Π is exchangeable, the sequences $(Z_1, Z_3, Z_4, \ldots)$ and $(Z_1, Z_{\sigma(3)}, Z_{\sigma(4)}, \ldots)$ have the same distribution for all finite permutations σ of $\{3, 4, \ldots\}$. Be Lemma 39 in Appendix A of [31], $Z_1, Z_3, Z_4, \ldots$ are conditionally independent given $\mathcal{G}$. It remains to show that $\mathbb{P}(Z_1 = i + x_\ell \,|\, \mathcal{G}) = 1_{\{K_i=\ell\}} P_i^2/(P,P)$ almost surely on $\{P_1 > 0\}$. Fix $n, k \in \mathbb{N}$ with $4 \le k < n$ and define the stopping time

$$S_k := \inf\{t > 0 : k-1 \text{ and } k \text{ are in the same block of } \Pi_t\}$$

and the event

$$E_{n,k} := \{1, \ldots, n \text{ are in distinct blocks of } \Pi_{S_k-}\}.$$

Furthermore, for all $\ell, m \in E$ let

$$\begin{aligned}\theta_{k,1,m,\ell} &:= \{(\{1,2\}, m), (\{3,\ldots,k\}, \ell)\} \in \mathcal{P}_{k,E}, \\ \theta_{k,2,\ell,m} &:= \{(\{1,\ldots,k-2\}, \ell), (\{k-1,k\}, m)\} \in \mathcal{P}_{k,E}, \\ M_{1,\ell} &:= \{\theta_{k,1,m,\ell} : m \in E\}, \\ M_{2,\ell} &:= \{\theta_{k,2,\ell,m} : m \in E\}.\end{aligned}$$

Now, let $\Pi^{(k+1,n)}$ denote the restriction of Π to $\{k+1, \ldots, n\}$ and let $\pi \in \mathcal{P}_{\{k+1,\ldots,n\},E}$. The exchangeability of Π implies

$$\mathbb{P}(\Pi_S^{(k)} = \theta_{k,1,m,\ell}, \Pi_S^{(k+1,n)} = \pi) = \mathbb{P}(\Pi_{S_k}^{(k)} = \theta_{k,2,\ell,m}, \Pi_{S_k}^{(k+1,n)} = \pi)$$

and therefore, by summation over all $m \in E$,

$$\mathbb{P}(\Pi_S^{(k)} \in M_{1,\ell}, \Pi_S^{(k+1,n)} = \pi) = \mathbb{P}(\Pi_{S_k}^{(k)} \in M_{2,\ell}, \Pi_{S_k}^{(k+1,n)} = \pi).$$

Since $S = S_k$ on $\{\Pi_{S_k}^{(k)} \in M_{2,\ell}\} \cap E_{n,k}$ and $\{\Pi_S^{(k)} \in M_{2,\ell}\} \cap E_n$, we obtain

$$\mathbb{P}(\{\Pi_S^{(k)} \in M_{1,\ell}\} \cap E_n \cap \{\Pi_S^{(k+1,n)} = \pi\}) = \mathbb{P}(\{\Pi_S^{(k)} \in M_{2,\ell}\} \cap E_n \cap \{\Pi_{S_k}^{(k+1,n)} = \pi\}).$$

Therefore,

$$\mathbb{P}(\Pi_S^{(k)} \in M_{1,\ell}, \Pi_S^{(k+1,n)} = \pi \,|\, E_n) \;=\; \mathbb{P}(\Pi_S^{(k)} \in M_{2,\ell}, \Pi_{S_k}^{(k+1,n)} = \pi \,|\, E_n).$$

Let $\Theta_{k+1,n}$ denote the restriction of Θ to $\{k+1,\ldots,n\}$ and $\mathcal{G}_{k+1,n}$ its generated σ-field. Since Θ_n is by definition the conditional distribution of Π_S given E_n, it follows that

$$\mathbb{P}(\Theta_k \in M_{1,\ell}, \Theta_{k+1,n} = \pi) \;=\; \mathbb{P}(\Theta_k \in M_{2,\ell}, \Theta_{k+1,n} = \pi).$$

Because this holds for arbitrary $\pi \in \mathcal{P}_{\{k+1,\ldots,n\},E}$ we obtain

$$\mathbb{P}(\Theta_k \in M_{1,\ell} \,|\, \mathcal{G}_{k+1,n}) \;=\; \mathbb{P}(\Theta_k \in M_{2,\ell} \,|\, \mathcal{G}_{k+1,n}) \quad \text{almost surely.}$$

Now, let $\mathcal{G}_{k+1}$ denote the σ-field generated by the restriction of Θ' to $\{k+1,k+2,\ldots\}$. Because $\bigcup_{n\in\mathbb{N}} \mathcal{G}_{k+1,n} = \mathcal{G}_{k+1}$, a standard martingale argument yields

$$\mathbb{P}(\Theta_k \in M_{1,\ell} \,|\, \mathcal{G}_{k+1}) \;=\; \mathbb{P}(\Theta_k \in M_{2,\ell} \,|\, \mathcal{G}_{k+1}) \quad \text{almost surely.}$$

The limiting relative frequencies and (random) colors of blocks can be recovered from the restriction of Θ' to $\{k+1,k+2,\ldots\}$, so $(P_i,K_i)_{i\in\mathbb{N}}$ is $\mathcal{G}_{k+1}$-measurable. Therefore, $\mathcal{G} \subseteq \mathcal{G}_{k+1}$ and by the tower property of conditional expectations

$$\mathbb{P}(\Theta_k \in M_{1,\ell} \,|\, \mathcal{G}) \;=\; \mathbb{P}(\Theta_k \in M_{2,\ell} \,|\, \mathcal{G}) \quad \text{almost surely.} \tag{46}$$

In the following it is shown that (46) leads to the statement of the lemma: By the definition of $Z_1, Z_2, \ldots$ it follows by conditional independence that

$$\begin{aligned}
\mathbb{P}(\Theta_k \in M_{1,\ell} \,|\, \mathcal{G}) &= \sum_{i=1}^{\infty} \mathbb{P}(Z_3 = \cdots = Z_k = i + x_\ell, Z_1 = Z_2 \neq i + x_\ell \,|\, \mathcal{G}) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(Z_3 = \cdots = Z_k = i + x_\ell \,|\, \mathcal{G}) \mathbb{P}(Z_1 \neq i + x_\ell \,|\, \mathcal{G}) \\
&= \sum_{i=1}^{\infty} P_i^{k-2} 1_{\{K_i=\ell\}} (1 - Q_{i,\ell})
\end{aligned}$$

almost surely, where $Q_{i,\ell} := \mathbb{P}(Z_1 = i + x_\ell \,|\, \mathcal{G})$ for all $i \in \mathbb{N}$ and $\ell \in E$. Similarly,

$$\begin{aligned}
&\mathbb{P}(\Theta_k \in M_{2,\ell} \,|\, \mathcal{G}) \\
&= \sum_{i\neq j} \sum_{m\in E} \mathbb{P}(Z_1 = Z_2 = i + x_\ell, Z_3 = \cdots = Z_{k-2} = i + x_\ell, Z_{k-1} = Z_k = j + x_m | \mathcal{G}) \\
&= \sum_{i\neq j} \sum_{m\in E} \mathbb{P}(Z_1 = i + x_\ell \,|\, \mathcal{G}) \mathbb{P}(Z_3 = \cdots = Z_{k-2} = i + x_\ell \,|\, \mathcal{G}) \mathbb{P}(Z_{k-1} = Z_k = j + x_m \,|\, \mathcal{G}) \\
&= \sum_{i\neq j} \sum_{m\in E} Q_{i,\ell} P_i^{k-4} 1_{\{K_i=\ell\}} P_j^2 1_{\{K_j=m\}} \\
&= \sum_{i=1}^{\infty} Q_{i,\ell} P_i^{k-4} 1_{\{K_i=\ell\}} \sum_{j\neq i} P_j^2 \sum_{m\in E} 1_{\{K_j=m\}} \\
&= \sum_{i=1}^{\infty} Q_{i,\ell} P_i^{k-4} 1_{\{K_i=\ell\}} \sum_{j\neq i} P_j^2 \quad \text{almost surely.}
\end{aligned}$$

It follows from (46) that

$$\begin{aligned}
0 &= \mathbb{P}(\Theta_k \in M_{1,\ell}\,|\,\mathcal{G}) - \mathbb{P}(\Theta_k \in M_{2,\ell}\,|\,\mathcal{G}) \\
&= \sum_{i=1}^{\infty}\left(P_i^{k-2}1_{\{K_i=\ell\}}(1-Q_{i,\ell}) - Q_{i,\ell}P_i^{k-4}1_{\{K_i=\ell\}\sum_{j\neq i}P_j^2}\right) \\
&= \sum_{i=1}^{\infty}P_i^{k-4}1_{\{K_i=\ell\}}\left(P_i^2 - P_i^2Q_{i,\ell} - Q_{i,\ell}\sum_{j\neq i}P_j^2\right) \\
&= \sum_{i=1}^{\infty}P_i^{k-4}1_{\{K_i=\ell\}}(P_i^2 - Q_{i,\ell}(P,P)) \quad \text{almost surely.}
\end{aligned}$$

From [31, Lemma 20] we obtain $Q_{i,\ell} = P_i^2/(P,P)$ on $\{P_i > 0, K_i = \ell\}$. Thus,

$$Q_{i,\ell} \;=\; 1_{\{K_i=\ell\}}\frac{P_i^2}{(P,P)} \tag{47}$$

on $\{P_i > 0\}$. On $\{P_1 > 0\}$ we have $\sum_{j\in\mathbb{N}}\sum_{\ell\in E}Q_{j,\ell} = 1$ almost surely by (47). Thus, $Q_{i,\ell} = 0$ almost surely on $\{P_1 > 0, P_i = 0\}$. It follows from the paragraph after (39) of [31] that (47) holds also on $\{P_1 > 0\}$ for all $i \in \mathbb{N}$. □

Thanks to Lemma 4, the proof of Theorem 3 is now completed as follows. Let $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ and T be a diagonal tensor with diagonal entries $t_{k,k} := (i_{k,s})_{s\in[j_k]} \in \{2,3,\ldots\}^{j_k}$. Define $|T| := \sum_{k\in E}\sum_{s=1}^{j_k} i_{k,s}$. Let $\theta \in \mathcal{P}_{|T|,E}$ with j_k k-blocks $B_{k,1},\ldots,B_{k,j_k}$ of sizes $i_{k,1},\ldots,i_{k,j_k}$ respectively, and such that 1 and 2 are in the same block of θ. It is already shown in (45) that

$$\phi_j(T) \;=\; \phi_1(\mathbf{2})\mathbb{P}(\Theta_{|T|} = \theta).$$

Lemma 40 of [31] carries over to the multi-type setting and implies that almost surely every block of Θ' having limiting relative frequency zero is a singleton. So if $i,j \geq 3$, i and j are in the same block if and only if $Z_i = Z_j \neq 0$ almost surely. On $\{P_1 > 0\}$, Lemma 4 implies $Z_1 = Z_2 > 0$ almost surely, so $i,j \in \mathbb{N}$ are in the same block if and only if $Z_i = Z_j \neq 0$. This shows in particular that, on $\{P_1 > 0\}$, K_i is the label of P_i also of Θ. Therefore, on $\{P_1 > 0\}$, the event $\{\Theta_{|T|} = \theta\}$ coincides, up to a null set, with the event that there exist pairwise distinct $m_{k,s}$, $k \in E$, $s \in [j_k]$, satisfying

$$Z_m \;=\; m_{k,s} + x_k \quad \text{for all } k \in E,\ s \in [j_k] \text{ and } m \in B_{k,s}.$$

Let $\ell_0 \in E$ with $1 \in D_{1,\ell_0}$. Then, by Lemma 4, for all $k \neq \ell_0$ and all $s \in [j_k]$,

$$\mathbb{P}(Z_m = m_{k,s} + x_k \text{ for all } m \in B_{k,s}\,|\,\mathcal{G}) \;=\; P_{m_{k,s}}^{i_{k,s}}1_{\{K_{m_{k,s}}=k\}}$$

and, furthermore, for all $s \in \{2,\ldots,j_{\ell_0}\}$,

$$\mathbb{P}(Z_m = m_{\ell_0,s} + x_{\ell_0} \text{ for all } m \in B_{\ell_0,s}\,|\,\mathcal{G}) \;=\; P_{m_{\ell_0},s}^{i_{\ell_0,s}}1_{\{K_{m_{\ell_0,s}}=k\}}$$

and

$$\begin{aligned}
&\mathbb{P}(Z_m = m_{\ell_0,1} + x_{\ell_0} \text{ for all } m \in B_{\ell_0,1} \,|\, \mathcal{G}) \\
&\quad = \; P_{m_{\ell_0},1}^{i_{\ell_0,1}} 1_{\{K_{m_{\ell_0,s}}=\ell_0\}}^{i_{\ell_0,1}-2} \mathbb{P}(Z_1 = m_{\ell_0,1} + x_{\ell_0} \,|\, \mathcal{G}) \\
&\quad = \; P_{m_{\ell_0},1}^{i_{\ell_0,1}} 1_{\{K_{m_{\ell_0,s}}=\ell_0\}}^{i_{\ell_0,1}-1} / (P,P) \;=\; P_{m_{\ell_0},1}^{i_{\ell_0,1}} 1_{\{K_{m_{\ell_0,s}}=\ell_0\}} / (P,P)
\end{aligned}$$

almost surely on $\{P_1 > 0\}$, where the last equality holds since $i_{\ell_0,1} \geq 2$. Summation over all possible values of $m_{k,s}$, $k \in E$, $s \in [j_k]$, yields

$$\mathbb{P}(\Theta_{|T|} = \theta | \mathcal{G}) \;=\; \frac{1}{(P,P)} \sum_{m_{k,s}} \prod_{k \in E} \prod_{s=1}^{j_k} P_{m_{k,s}}^{i_{k,s}} 1_{\{k\}}(K_{m_{k,s}})$$

almost surely on $\{P_1 > 0\}$. On $\{P_1 = 0\}$, Lemma 4 implies that $\mathbb{P}(Z_m = 0 \,|\, \mathcal{G}) = P_0 = 1$ almost surely and, therefore, $Z_m = 0$ almost surely for all $m \in \{3, 4, \ldots\}$, so Θ' has only singletons almost surely. The exchangeability of Π implies that, conditional on the event that Θ' has only singletons, the probability that $1, 2$ and q are in the same block of Θ is the same for all $q \in \{3, 4, \ldots\}$ and must hence be zero. It follows that, on $\{P_1 = 0\}$, $\Theta_{|T|}$ takes values in the set of partitions of the form $\pi_{k_1,\ldots,k_{|T|-1}} := (\{1,2\}, \ell_1), (\{3\}, \ell_2) \ldots, (\{|T|\}, \ell_{|T|-1}))$ for some $\ell_1, \ldots, \ell_{|T|-1} \in E$ almost surely. In other words, $\Theta_{|T|} = ((\{1,2\}, L_1), (\{3\}, L_2), \ldots, (\{|T|\}, L_{|T|-1}))$ almost surely on $\{P_1 = 0\}$ for some E-valued random variables $L_1, \ldots, L_{|T|-1}$. Since all the blocks of $\theta \in \mathcal{P}_{|T|,E}$ have size at least 2, it follows that $\mathbb{P}(\Theta_{|T|} = \theta | \mathcal{G}) = 1_{\{L_1 = k\}}$ almost surely on $\{P_1 = 0\}$ if $j = e_k$ and $i_{k,1} = 2$ for some $k \in E$ and $\mathbb{P}(\Theta_{|T|} = \theta | \mathcal{G}) = 0$ almost surely on $\{P_1 = 0\}$ otherwise. Combining this, we obtain

$$\begin{aligned}
\mathbb{P}(\Theta_{|T|} = \theta | \mathcal{G}) \;&=\; 1_{\{P_1=0\}} \sum_{k \in E} 1_{\{L_1=k\}} 1_{\{j=\mathbf{e}_k, i_{k,1}=2\}} \\
&\quad + 1_{\{P_1>0\}} \frac{1}{(P,P)} \sum_{m_{k,s}} \prod_{k \in E} \prod_{s=1}^{j_k} P_{m_{k,s}}^{i_{k,s}} 1_{\{k\}}(K_{m_{k,s}}) \qquad (48)
\end{aligned}$$

almost surely. Analogous to Aldous [2] it follows that (P, K) takes values in $\Delta \times E^{\mathbb{N}}$ almost surely. Let Q denote the distribution of (P, K) restricted to $(\Delta \setminus \{0\}) \times E^{\mathbb{N}}$. Furthermore, define $\alpha_k := \mathbb{P}(P_1 = 0, L_1 = k)$ for all $k \in E$. Taking the expectation in (48) yields

$$\mathbb{P}(\Theta_{|T|} = \theta) \;=\; \sum_{k \in E} \alpha_k 1_{\{j=\mathbf{e}_k, i_{k,1}=2\}} + \int_{(\Delta \setminus \{\mathbf{0}\}) \times E^{\mathbb{N}}} \sum_{m_{k,s}} \prod_{k \in E} \prod_{s=1}^{j_k} x_{m_{k,s}}^{i_{k,s}} 1_{\{k\}}(y_{m_{k,s}}) \, \frac{Q(\mathrm{d}(x,y))}{(x,x)}.$$

Since $\phi_j(T) = c\mathbb{P}(\Theta_{|T|} = \theta)$ with $c := \phi_j(\mathbf{2})$, the statement of Theorem 3 follows with $\Xi := cQ$ and $a_k := c\alpha_k$ for all $k \in E$.

10 Appendix

This appendix collects some basic definitions, illustrations and results used in the article.

10.1 Some formulas concerning the number of offspring

We provide some straightforward formulas and bounds for certain (joint) moments of numbers of offspring.

Lemma 5. *Under (A),*

$$\mathbb{E}(\nu_{k,\ell,i}) \;=\; \frac{\mathbb{E}(N_{k,\ell})}{N_k}, \qquad k,\ell \in E, i \in [N_k], \tag{49}$$

and

$$\mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2}) \;=\; \frac{\mathrm{Var}(N_{k,\ell})}{(N_k)_2} + \left(\frac{\mathbb{E}(N_{k,\ell})}{N_k}\right)^2 - \frac{\mathrm{Var}(\nu_{k,\ell,1})}{N_k - 1}, \qquad k,\ell \in E, \tag{50}$$

provided that $N_k > 1$.

Proof. Fix $k,\ell \in E$. Eq. (49) follows from $\mathbb{E}(N_{k,\ell}) = \mathbb{E}(\sum_{i=1}^{N_k} \nu_{k,\ell,i}) = N_k\mathbb{E}(\nu_{k,\ell,1})$. Moreover, $\mathbb{E}(N_{k,\ell}^2) = \mathbb{E}(\sum_{i=1}^{N_k} \nu_{k,\ell,i} \sum_{j=1}^{N_k} \nu_{k,\ell,j}) = N_k\mathbb{E}((\nu_{k,\ell,1})^2) + (N_k)_2\mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2})$. Assume now that $N_k > 1$. Solving for $\mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2})$ yields

$$\begin{aligned}\mathbb{E}(\nu_{k,\ell,1}\nu_{k,\ell,2}) \;&=\; \frac{\mathbb{E}(N_{k,\ell}^2) - N_k\mathbb{E}(\nu_{k,\ell,1}^2)}{(N_k)_2} \;=\; \frac{\mathbb{E}(N_{k,\ell}^2) - N_k(\mathbb{E}(\nu_{k,\ell,1}))^2 - N_k\mathrm{Var}(\nu_{k,\ell,1})}{(N_k)_2} \\ &=\; \frac{\mathbb{E}(N_{k,\ell}^2) - N_k(\frac{\mathbb{E}(N_{k,\ell})}{N_k})^2}{(N_k)_2} - \frac{\mathrm{Var}(\nu_{k,\ell,1})}{N_k - 1} \;=\; \frac{\mathrm{Var}(N_{k,\ell})}{(N_k)_2} + \left(\frac{\mathbb{E}(N_{k,\ell})}{N_k}\right)^2 - \frac{\mathrm{Var}(\nu_{k,\ell,1})}{N_k - 1},\end{aligned}$$

which is (50). □

The following lemma provides upper bounds for the coalescence probabilities defined in (6) and (7), respectively.

Lemma 6. *For all* $k,\ell \in E$,

$$c_{k,\ell}(N_k, N_\ell) \;\le\; \frac{\mathbb{E}((N_{k,\ell})_2)}{(N_\ell)_2} \;\le\; \frac{\mathbb{E}(N_{k,\ell})}{N_\ell}$$

and, for all $k,\ell_1,\ell_2 \in E$ *with* $\ell_1 \neq \ell_2$,

$$c_{k,\ell_1,\ell_2}(N_k, N_{\ell_1}, N_{\ell_2}) \;\le\; \frac{\mathbb{E}(N_{k,\ell_1}N_{k,\ell_2})}{N_{\ell_1}N_{\ell_2}} \;\le\; \min\left(\frac{\mathbb{E}(N_{k,\ell_1})}{N_{\ell_1}}, \frac{\mathbb{E}(N_{k,\ell_2})}{N_{\ell_2}}\right).$$

Proof. For all $k,\ell \in E$, by (6),

$$\begin{aligned}c_{k,\ell}(N_k, N_\ell) \;&=\; \frac{1}{(N_\ell)_2}\mathbb{E}\left(\sum_{i=1}^{N_k} \nu_{k,\ell,i}(\nu_{k,\ell,i} - 1)\right) \;\le\; \frac{1}{(N_\ell)_2}\mathbb{E}\left(\sum_{i=1}^{N_k} \nu_{k,\ell,i}(N_{k,\ell} - 1)\right) \\ &=\; \frac{\mathbb{E}((N_{k,\ell})_2)}{(N_\ell)_2} \;\le\; \frac{\mathbb{E}(N_{k,\ell}(N_\ell - 1))}{(N_\ell)_2} \;=\; \frac{\mathbb{E}(N_{k,\ell})}{N_\ell}.\end{aligned}$$

Similarly, for all $k, \ell_1, \ell_2 \in E$ with $\ell_1 \neq \ell_2$, by (7),

$$\begin{aligned} c_{k,\ell_1,\ell_2}(N_k, N_{\ell_1}, N_{\ell_2}) &= \frac{1}{N_{\ell_1} N_{\ell_2}} \mathbb{E}\Big(\sum_{i=1}^{N_k} \nu_{k,\ell_1,i} \nu_{k,\ell_2,i} \Big) \\ &\leq \frac{1}{N_{\ell_1} N_{\ell_2}} \mathbb{E}\Big(\sum_{i=1}^{N_k} \nu_{k,\ell_1,i} N_{k,\ell_2} \Big) = \frac{\mathbb{E}(N_{k,\ell_1} N_{k,\ell_2})}{N_{\ell_1} N_{\ell_2}}. \end{aligned}$$

The last inequality follows from $N_{k,\ell_1}/N_{\ell_1} \leq 1$ and $N_{k,\ell_2}/N_{\ell_2} \leq 1$. □

10.2 Block structure of the backward transition matrix

If the states π of $\mathcal{P}_{n,E}$ are ordered with respect to the number $|\pi|$ of blocks of π, then the transition matrix P of the ancestral process $(\mathcal{A}_r^{(n,N)})_{r \in \mathbb{N}_0}$ is a left lower triangular block matrix of the form $P = (P_{i,j})_{1 \leq i,j \leq n}$, where each $P_{i,j}$ is a $(d^i S(n,i) \times d^j S(n,j))$-matrix with $P_{i,j} = 0$ for $i < j$, where $S(.,.)$ denote the Stirling numbers of the second kind. For all $i \geq j$ the block matrix $P_{i,j}$ contains all the transition probabilities of transitions from states π having i blocks to states π' having j blocks. In the following this block structure is provided in detail for sample sizes $n = 1$ and $n = 2$. It is also assumed that $d := |E| < \infty$ for simplicity. For $n = 1$ the state space $\mathcal{P}_{1,E}$ of the ancestral process has size $|\mathcal{P}_{1,E}| = d$ and consists of all labeled partitions of the form $\pi_k := \{(\{1\}, k)\}$, $k \in E$. By (3), $p_{\pi_k,\pi_\ell} = \mathbb{E}(N_{\ell,k})/N_k$, $k, \ell \in E$. Thus, for $n = 1$ the transition matrix $P = (p_{\pi,\pi'})_{\pi,\pi' \in \mathcal{P}_{1,E}}$ of the ancestral process coincides with the mean backward mutation matrix $M := (m_{k,\ell})_{k,\ell \in E}$, where $m_{k,\ell} := \mathbb{E}(N_{\ell,k})/N_k$ is the mean backward mutation probability, i.e. the mean proportion of the individuals in subpopulation k after the reproduction step, who were born in subpopulation ℓ.

For $n = 2$, $P = \begin{pmatrix} M & 0 \\ C & D \end{pmatrix}$ where $M = (m_{k,\ell})_{k,\ell \in E}$ is again the mean backward mutation matrix, C contains the coalescence probabilities and D the remaining transition probabilities $p_{\pi,\pi'}$ for all states π and π' having two blocks, which can be calculated using (3). For illustration, for $E := \{1, 2\}$, and the state space ordered by

$$\pi_1 := \{(\{1,2\}, 1)\}, \qquad \pi_2 := \{(\{1,2\}, 2)\},$$

and

$$\begin{aligned} \pi_3 &:= \{(\{1\}, 1), (\{2\}, 1)\}, & \pi_4 &:= \{(\{1\}, 1), (\{2\}, 2)\}, \\ \pi_5 &:= \{(\{1\}, 2), (\{2\}, 1)\}, & \pi_6 &:= \{(\{1\}, 2), (\{2\}, 2)\}, \end{aligned}$$

the transition matrix $P = (p_{\pi_i,\pi_j})_{1 \leq i,j \leq 6}$ of the ancestral process is provided in Figure 3.

Figure 3: Transition matrix P of the ancestral process for sample size $n = 2$ and $E = \{1, 2\}$

$$P = \left(\begin{array}{cc|cccc}
\mathbb{E}(\nu_{1,1,1}) & \frac{N_2\mathbb{E}(\nu_{2,1,1})}{N_1} & 0 & 0 & 0 & 0 \\
\frac{N_1\mathbb{E}(\nu_{1,2,1})}{N_2} & \mathbb{E}(\nu_{2,2,1}) & 0 & 0 & 0 & 0 \\
\hline
\frac{\mathbb{E}((\nu_{1,1,1})_2)}{N_1-1} & \frac{N_2\mathbb{E}((\nu_{2,1,1})_2)}{(N_1)_2} & \mathbb{E}(\nu_{1,1,1}\nu_{1,1,2}) & \frac{N_2\mathbb{E}(\nu_{1,1,1}\nu_{2,1,1})}{N_1-1} & \frac{N_2\mathbb{E}(\nu_{1,1,1}\nu_{2,1,1})}{N_1-1} & \frac{(N_2)_2\mathbb{E}(\nu_{2,1,1}\nu_{2,1,2})}{(N_1)_2} \\
\frac{\mathbb{E}(\nu_{1,1,1}\nu_{1,2,1})}{N_2} & \frac{\mathbb{E}(\nu_{2,1,1}\nu_{2,2,1})}{N_1} & \frac{N_1-1}{N_2}\mathbb{E}(\nu_{1,1,1}\nu_{1,2,1}) & \mathbb{E}(\nu_{1,1,1}\nu_{2,2,1}) & \mathbb{E}(\nu_{1,2,1}\nu_{2,1,1}) & \frac{N_2-1}{N_1}\mathbb{E}(\nu_{2,1,1}\nu_{2,2,2}) \\
\frac{\mathbb{E}(\nu_{1,1,1}\nu_{1,2,1})}{N_2} & \frac{\mathbb{E}(\nu_{2,1,1}\nu_{2,2,1})}{N_1} & \frac{N_1-1}{N_2}\mathbb{E}(\nu_{1,1,2}\nu_{1,2,1}) & \mathbb{E}(\nu_{1,2,1}\nu_{2,1,1}) & \mathbb{E}(\nu_{1,1,1}\nu_{2,2,1}) & \frac{N_2-1}{N_1}\mathbb{E}(\nu_{2,1,1}\nu_{2,2,1}) \\
\frac{N_1\mathbb{E}((\nu_{1,2,1})_2)}{(N_2)_2} & \frac{\mathbb{E}((\nu_{2,2,1})_2)}{N_2-1} & \frac{(N_1)_2}{(N_2)_2}\mathbb{E}(\nu_{1,2,1}\nu_{1,2,2}) & \frac{N_1}{N_2-1}\mathbb{E}(\nu_{1,2,1}\nu_{2,2,1}) & \frac{N_1}{N_2-1}\mathbb{E}(\nu_{1,2,1}\nu_{2,2,1}) & \mathbb{E}(\nu_{2,2,1}\nu_{2,2,2})
\end{array}\right)$$

10.3 Multi-type exchangeable partition probability function

Let E be some at most countable set. We call E the type space. For $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ let $\mathcal{T}_j$ denote the set of all tensors $T = (t_{k,\ell})_{k,\ell\in E}$ with vector entries $t_{k,\ell} \in \mathbb{N}_0^{j_k}$ for all $k, \ell \in E$. Note that if $j_k = 0$ then $t_{k,\ell} = () =: \mathbf{0} \in \mathbb{R}^0$ is the empty vector for all $\ell \in E$. Furthermore, define $\mathcal{T} := \bigcup_{j\in\mathbb{N}_0^E} \mathcal{T}_j$. The particular (and only) tensor in $\mathcal{T}_0$ with all its entries $t_{k,\ell} = ()$ being the empty vector is denoted by T_0. The following definitions are inspired from the properties of the transition functions Φ_j, $j \in \mathbb{N}_0^E$, of multi-type Cannings models studied in Section 3.

Definition 2 (normalization)**.** *A function $p : \mathcal{T} \to [0, 1]$ is called normalized, if $p(T_0) = 1$.*

Definition 3 (consistency)**.** *A function $p : \mathcal{T} \to [0, 1]$ is called consistent, if for all $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$ and all tensors $T \in \mathcal{T}_j$ the equality*

$$p(T) \;=\; \sum_{k\in E} p(T(k,\ell)) + \sum_{k\in E}\sum_{s=1}^{j_k} p(T(k,\ell,s)) \tag{51}$$

holds for each $\ell \in E$, where the tensor $T(k, \ell)$ is obtained from $T = (t_{k,\ell})_{k,\ell\in E}$ by replacing the (possibly empty) vector $t_{k,\ell} = (i_{k,\ell,1}, \ldots, i_{k,\ell,j_k})$ by $(i_{k,\ell,1}, \ldots, i_{k,\ell,j_k}, 1)$ and the (possibly empty) vector $t_{k,\ell'} = (i_{k,\ell',1}, \ldots, i_{k,\ell',j_k})$ by $(i_{k,\ell',1}, \ldots, i_{k,\ell',j_k}, 0)$ for all $\ell' \neq \ell$, and the vector $T(k, \ell, s)$ is obtained from T by replacing the single entry $i_{k,\ell,s}$ by $i_{k,\ell,s} + 1$.

Definition 4 (symmetry)**.** *A function $p : \mathcal{T} \to [0, 1]$ is called symmetric, if for all $j = (j_k)_{k\in E} \in \mathbb{N}_0^E$, all tensors $T = (t_{k,\ell})_{k,\ell\in E} \in \mathcal{T}_j$ and all permutations $\sigma_{k,\ell} \in S_{j_k}$, $k, \ell \in E$, the symmetry relation*

$$p(T) \;=\; p(\sigma(T)) \tag{52}$$

holds, where $\sigma(T) := (\sigma_{k,\ell} t_{k,\ell})_{k,\ell\in E}$.

A function $p : \mathcal{T} \to [0, 1]$ is called a *multi-type partition probability function* (M-PPF), if p is normalized and consistent. If p is in addition symmetric, then p is called a *multi-type exchangeable partition probability function* (M-EPPF). For the single-type case $|E| = 1$, this terminology is in agreement with the notion of the exchangeable partition probability function (EPPF) of Pitman [25].

References

[1] Alberti, F. (2024) Labeled partitions in action: recombination, selection, mutation and more. *Stoch. Dyn.* **24**(2), Paper No. 2450014, 28pp. MR4767757

[2] Aldous, D. (1985) *Exchangeability and Related Topics.* Lecture Notes in Mathematics **1117**, Springer, Berlin. MR0883646

[3] Applebaum, D. (2009) *Lévy Processes and Stochastic Calculus.* Cambridge Stud. Adv. Math. **116**, Second Edition, Cambridge University Press, Cambridge. MR2512800

[4] Blath, J., Kurt, N., Slowik, M. and Wilke-Berenguer, M. (2024) Dormancy in stochastic population models. *Jahresberichte der DMV* **126**(4), 249–281. MR4832518

[5] Burke, C. J. and Rosenblatt, M. (1958) A Markovian function of a Markov chain. *Ann. Math. Statist.* **29**, 1112–1122. MR0101557

[6] Cannings, C. (1974) The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models. *Adv. in Appl. Probab.* **6**(2), 260–290. MR0343949

[7] Cannings, C. (1975) The latent roots of certain Markov chains arising in genetics: A new approach, II. Further haploid models. *Adv. in Appl. Probab.* **7**(2), 260–290. MR0371430

[8] Cannings, C. (1976) Multitype drift models. *Adv. in Appl. Probab.* **8**(1), 1–2.

[9] Ethier, S. N. and Kurtz, T. G. (1986) *Markov Processes, Characterization and Convergence.* Wiley. MR0838085

[10] González Casanova, A., Kurt, N., Nuñez Morales, I. and Pérez, J. L. (2024) Multitype Λ-coalescents and continuous state branching processes. arXiv preprint 2501.00329v1

[11] Goyt, A. M. and Pudwell, L. K. (2011) Avoiding colored partitions of lengths two and three. *Pure Math. Appl.* **22**(2), 107–128. MR2924742

[12] Goyt, A. M. and Pudwell, L. K. (2012) Avoiding colored partitions of two elements in the pattern sense. *J. Integer Seq.* **15**(6), Article 12.6.2, 17 pp. MR2954657

[13] Hössjer, O. (2011) Coalescence theory for a general class of structured populations with fast migration. *Adv. in Appl. Probab.* **43**(4), 1027–1047. MR2867944

[14] Hsu, L. C. and Shiue, P. J.-S. (1998) A unified approach to generalized Stirling numbers. *Adv. in Appl. Math.* **20**(3), 366–384. MR1618435

[15] Johnston, S. G. G., Kyprianou, A. and Rogers, T. (2023) Multitype Λ-coalescents. *Ann. Appl. Probab.* **33**(6A), 4210–4237. MR4674049

[16] Kaj, I., Krone, S. M. and Lascoux, M. (2001) Coalescent theory for seed bank models. *J. Appl. Probab.* **38**(2), 285–300. MR1834743

[17] Kallenberg, O. (2005) *Probabilistic Symmetries and Invariance Principles.* Probability and Its Applications. Springer. MR2161313

[18] Kingman, J. F. C. (1982a) On the genealogy of large populations. *J. Appl. Probab.* **19**, 27–43. MR0633178

[19] Mansour, T. (2013) *Combinatorics of Set Partitions.* Discrete Mathematics and its Applications, CRC Press, Chapman & Hall. MR2953184

[20] Möhle, M. (1998) A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. in Appl. Probab.* **30**(2), 493–512. MR1642850

[21] Möhle, M. (2002) The coalescent in population models with time-inhomogeneous environment. *Stochastic Process. Appl.* **97**(2), 199–227. MR1875333

[22] Möhle, M. (2024) On multi-type Cannings models and multi-type exchangeable coalescents. *Theoret. Population Biol.* **156**, 103–116.

[23] Möhle, M. and Sagitov, S. (2001) A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**(4), 1547–1562. MR1880231

[24] Nordborg, M. and Krone, S. M. (2001) Separation of time scales and convergence to the coalescent in structured populations. In: Modern Developments in Theoretical Population Genetics. eds. M. Slatkin and M. Veuille, Oxford University Press, pp. 194–232.

[25] Pitman, J. (1995) Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102**(2), 145–158. MR1337249

[26] Pitman, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* **27**(4), 1870–1902. MR1742892

[27] Pitman, J. (2006) Combinatorial Stochastic Processes, Lecture Notes in Mathematics **1875**, Springer, Berlin. MR2245368

[28] Sagitov, S. (1999) The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**(4), 1116–1125. MR1742154

[29] Sagitov, S. (2003) Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40**(4), 839–854. MR2012671

[30] Sagitov, S. and Jagers, P. (2005) The coalescent effective size of age-structured populations. *Ann. Appl. Probab.* **15**(3), 1778–1797. MR2152245

[31] Schweinsberg, J. (2000) Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, 1–50. MR1781024