DOI: 10.1214/ECP.v16-1654

Elect. Comm. in Probab. **16** (2011), 567–582

ELECTRONIC COMMUNICATIONS in PROBABILITY

COALESCENT PROCESSES DERIVED FROM SOME COMPOUND POISSON POPULATION MODELS

MARTIN MÖHLE
Mathematisches Institut, Eberhard Karls Universität Tübingen, Auf der Morgentelle 10, 72076 Tübingen, Germany
email: martin.moehle@uni-tuebingen.de

Submitted 25 February 2011, *accepted in final form* 17 August 2011

AMS 2000 Subject classification: Primary 92D25, 60J70 Secondary 92D15, 60F17
Keywords: Ancestral process; Cannings model; coalescent; compound Poisson model; conditional branching process model; Dirichlet model; exchangeability; neutrality; simultaneous multiple collisions; weak convergence; Wright–Fisher model

Abstract

A particular subclass of compound Poisson population models is analyzed. The models in the domain of attraction of the Kingman coalescent are characterized and it is shown that these models are never in the domain of attraction of any other continuous-time coalescent process. Results are obtained characterizing which of these models are in the domain of attraction of a discrete-time coalescent with simultaneous multiple mergers of ancestral lineages. The results extend those obtained by Huillet and the author in 'Population genetics models with skewed fertilities: a forward and backward analysis', *Stochastic Models* **27** (2011), 521–554.

1 Introduction and model description

We study a certain class of haploid population models with non-overlapping generations and fixed population size $N \in \mathbb{N} := \{1,2,\ldots\}$. Each model in this class is a particular Cannings model [2, 3]. Cannings models are characterized by exchangeable random variables $\nu_1,\ldots,\nu_N$, where ν_i denotes the number of offspring of the ith individual. The class of models we are interested here, is defined as follows. We start with independent but not necessarily identically distributed random variables $\xi_1,\xi_2,\ldots$ taking values in $\mathbb{N}_0 := \{0,1,2,\ldots\}$. Assuming that $\mathbb{P}(\xi_1+\cdots+\xi_N=N)>0$ for all $N \in \mathbb{N}$, let $\mu_1,\ldots,\mu_N$ be random variables such that the joint distribution of $\mu_1,\ldots,\mu_N$ coincides with that of $\xi_1,\ldots,\xi_N$ conditioned on the event that $\xi_1+\cdots+\xi_N=N$. Finally, we randomly permutate these N random variables $\mu_1,\ldots,\mu_N$, which leads to the desired exchangeable random variables $\nu_1,\ldots,\nu_N$. Cannings models of this form are called conditional branching process models, since they are obtained from an independent sequence $\xi_1,\xi_2,\ldots$ by conditioning on the event that $\xi_1+\cdots+\xi_N=N$ (and random shuffling). Models of this form for the situation when the random variables $\xi_1,\xi_2,\ldots$ are additionally assumed to be i.i.d. are at least known since the works of Karlin and McGregor [7, 8]. The most prominent example is the symmetric Wright–Fisher model, which is obtained by choosing all the ξ_n to have a Poisson distribution with some parameter $\alpha > 0$.

The class of models differs from those considered in [14], which are based on sampling instead of conditioning.

In this paper we restrict our attention to a particular class of random variables ξ_n, $n \in \mathbb{N}$. Let ϕ be a given power series of the form $\phi(z) = \sum_{m=1}^{\infty} \phi_m z^m/m!$, $|z| < r$, with positive radius $r \in (0, \infty]$ of convergence and with non-negative coefficients $\phi_m \geq 0$, $m \in \mathbb{N}$. It is also assumed that $\phi_1 > 0$. Let furthermore $\theta_1, \theta_2, \ldots \in (0, \infty)$ be given real parameters. We assume that the random variables ξ_n, $n \in \mathbb{N}$, have probability generating functions (pgf)

$$f_n(x) \;:=\; \mathbb{E}(x^{\xi_n}) \;=\; \exp\Big(-\theta_n \phi(z)\Big(1 - \frac{\phi(zx)}{\phi(z)}\Big)\Big), \qquad |x| \leq 1, n \in \mathbb{N}. \tag{1}$$

In (1), z is viewed as a fixed parameter, however, it is also useful to see z as a variable satisfying $|z| < r$. If M_n is a random variable having a Poisson distribution with parameter $\theta_n \phi(z)$ and if $X_1, X_2, \ldots$ are independent random variables and independent of M_n each with pgf $x \mapsto \phi(zx)/\phi(z)$, $|x| \leq 1$, then $\sum_{j=1}^{M_n} X_j$ has pgf (1). This subclass of Cannings models is therefore called the compound Poisson class. Note that

$$\mathbb{E}(x^{\sum_{n=1}^{N} \xi_n}) = \exp(-(\sum_{n=1}^{N} \theta_n)\phi(z)(1 - \phi(zx)/\phi(z))), \quad N \in \mathbb{N}$$

and that $\mathbb{E}(\xi_n) = \theta_n z \phi'(z)$, $n \in \mathbb{N}$. For basic properties of these models we refer the reader to [6].

We are interested in the behavior of the ancestral structure of these models when the total population size N tends to infinity. The analysis of the full class of compound Poisson models seems to be quite involved. We therefore focus on a particular subclass of compound Poisson models satisfying an additional constrain, which is described precisely in Eq. (6) below. It turns out (see Lemma 2.1 and the remarks thereafter) that these models are generalized Wright–Fisher models and generalized Dirichlet models.

For the symmetric case ($\theta_n = \theta$ for all $n \in \mathbb{N}$) the asymptotical behavior is clarified by Theorem 4.3 of [6]. For the asymmetric case, particular examples have been studied in Sections 5 and 6 of [6], but the authors did not provide more general asymptotic results for the asymmetric case. The asymptotic results presented in the following Section 2 extend those obtained in [6]. The proofs, provided in Section 3, rely on well known general convergence-to-the-coalescent results for Cannings models. For more information on these convergence results and on the arising limiting coalescent processes allowing for simultaneous multiple collisions of ancestral lineages we refer the reader to [11] and [13].

2 Results

In order to state the results we need to introduce, for $\theta > 0$, the Taylor expansion

$$\exp(\theta\phi(z)) \;=\; \sum_{k=0}^{\infty} \frac{\sigma_k(\theta)}{k!} z^k, \qquad |z| < r. \tag{2}$$

The coefficients $\sigma_k(\theta)$ are strictly positive and they satisfy the recursion

$$\sigma_0(\theta) = 1 \quad \text{and} \quad \sigma_{k+1}(\theta) \;=\; \theta \sum_{l=0}^{k} \binom{k}{l} \phi_{k-l+1} \sigma_l(\theta), \qquad k \in \mathbb{N}_0, \tag{3}$$

i.e. $\sigma_1(\theta)=\theta\phi_1$, $\sigma_2(\theta)=\theta\phi_2+\theta^2\phi_1^2$, $\sigma_3(\theta)=\theta\phi_3+3\theta^2\phi_1\phi_2+\theta^3\phi_1^3$, and so on. In particular, σ_k is a polynomial in θ of degree k, since $\phi_1>0$ by assumption. For more information on the coefficients $\sigma_k(\theta)$, in particular their relation to Bell polynomials, we refer the reader to [1]. The coefficients $\sigma_k(\theta)$ are in general not simple to compute. They are mainly introduced, since, by (1), the distribution of ξ_n, $n\in\mathbb{N}$, satisfies

$$\mathbb{P}(\xi_n=k) \;=\; \sigma_k(\theta_n)\frac{z^k}{k!}\exp(-\theta_n\phi(z)), \qquad k\in\mathbb{N}_0.$$

The distribution of $\mu:=(\mu_1,\ldots,\mu_N)$ is therefore of the form

$$\mathbb{P}(\mu_1=j_1,\ldots,\mu_N=j_N) \;=\; \frac{N!}{j_1!\cdots j_N!}\frac{\sigma_{j_1}(\theta_1)\cdots\sigma_{j_N}(\theta_N)}{\sigma_N(\theta_1+\cdots+\theta_N)}, \tag{4}$$

$j_1,\ldots,j_N\in\mathbb{N}_0$ with $j_1+\cdots+j_N=N$, and μ has joint factorial moments

$$\mathbb{E}((\mu_1)_{k_1}\cdots(\mu_N)_{k_N}) \;=\; \frac{N!}{\sigma_N(\theta_1+\cdots+\theta_N)}\sum_{\substack{j_1\ge k_1,\ldots,j_N\ge k_N\\ j_1+\cdots+j_N=N}}\frac{\sigma_{j_1}(\theta_1)\cdots\sigma_{j_N}(\theta_N)}{(j_1-k_1)!\cdots(j_N-k_N)!}, \tag{5}$$

$k_1,\ldots,k_N\in\mathbb{N}_0$, where, for $x\in\mathbb{R}$ and $k\in\mathbb{N}_0$, the notation $(x)_k:=x(x-1)\cdots(x-k+1)$ is used with the convention that $(x)_0:=1$. In particular, the distribution of μ does not depend on the auxiliary parameter z. The expression (5) for the joint factorial moments of μ is quite involved and not very simple to analyze. We therefore focus on a particular subclass of compound Poisson models satisfying the relation

$$\frac{\sigma_{k+1}(\theta)}{\sigma_k(\theta)}+\frac{\sigma_{k'+1}(\theta')}{\sigma_{k'}(\theta')} \;=\; \frac{\sigma_{k+k'+1}(\theta+\theta')}{\sigma_{k+k'}(\theta+\theta')}, \qquad k,k'\in\mathbb{N}_0,\theta,\theta'\in(0,\infty). \tag{6}$$

The following lemma clarifies which compound Poisson models satisfy (6). Its proof is provided in Section 3.

Lemma 2.1. *A compound Poisson model (with given fixed power series ϕ) satisfies (6) if and only if $\phi_m=(m-1)!\phi_1(\phi_2/\phi_1)^{m-1}$ for all $m\in\mathbb{N}$. If (6) holds, then μ has joint factorial moments*

$$\mathbb{E}((\mu_1)_{k_1}\cdots(\mu_N)_{k_N}) \;=\; (N)_k\frac{\sigma_{k_1}(\theta_1)\cdots\sigma_{k_N}(\theta_N)}{\sigma_k(\theta_1+\cdots+\theta_N)}, \qquad k_1,\ldots,k_N\in\mathbb{N}_0. \tag{7}$$

Remarks. Relation (6) thus determines the coefficients ϕ_m, $m\ge 3$, of the power series ϕ completely. For $\phi_2=0$ the power series ϕ is of the form $\phi(z)=\phi_1 z$ corresponding to a generalized Wright–Fisher model, whereas for $\phi_2>0$ we obtain the power series

$$\phi(z) \;=\; \sum_{m=1}^\infty\frac{\phi_m}{m!}z^m \;=\; \frac{\phi_1^2}{\phi_2}\sum_{m=1}^\infty\frac{1}{m}\left(\frac{\phi_2}{\phi_1}z\right)^m \;=\; -\frac{\phi_1^2}{\phi_2}\log\left(1-\frac{\phi_2}{\phi_1}z\right), \quad |z|<\frac{\phi_1}{\phi_2}, \tag{8}$$

corresponding to a generalized Dirichlet model. Formula (7) for the joint factorial moments of μ is considerably simpler than the general formula (5), which is the main reason why we restrict our considerations to the special subclass of compound Poisson models satisfying (6). In the following it is always assumed that (6) holds.

For $N,k\in\mathbb{N}$ define the partial sums $\Theta_k(N):=\sum_{n=1}^N\theta_n^k$ for convenience. Furthermore, for $n\in\mathbb{N}$ let $\mathscr{E}_n$ denote the set of all equivalence relations on $[n]:=\{1,\ldots,n\}$. In the following we are interested in the so-called ancestral process $(\mathscr{R}_t^{(n)})_{t\in\mathbb{N}_0}$ of a sample of $n\in\mathbb{N}$ individuals taken from some generation. This is a Markovian process taking values in $\mathscr{E}_n$ with the interpretation that $(i,j)\in\mathscr{R}_t^{(n)}$ if and only if the ith and the jth individual of the sample have a common parent t generations backwards in time. It is well known (see, for example, the proof of [6, Proposition 4.2]) that the ancestral process $(\mathscr{R}_t^{(n)})_{t\in\mathbb{N}_0}$ has transition probabilities

$$\mathbb{P}(\mathscr{R}_{t+1}^{(n)}=\eta\,|\,\mathscr{R}_t^{(n)}=\xi)=\Phi_j^{(N)}(k_1,\ldots,k_j),\quad \xi,\eta\in\mathscr{E}_n$$

with $\xi\subseteq\eta$, where

$$\begin{aligned}\Phi_j^{(N)}(k_1,\ldots,k_j) &:= \frac{1}{(N)_{k_1+\cdots+k_j}}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\mathbb{E}((\mu_{n_1})_{k_1}\cdots(\mu_{n_j})_{k_j})\\ &= \frac{1}{\sigma_{k_1+\cdots+k_j}(\Theta_1(N))}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\sigma_{k_1}(\theta_{n_1})\cdots\sigma_{k_j}(\theta_{n_j}),\end{aligned}$$

by (7). Here $j:=|\eta|$ denotes the number of equivalence classes (blocks) of η and $k_1,\ldots,k_j\in\mathbb{N}$ are the group sizes of merging blocks of ξ. Note that $k:=k_1+\cdots+k_j$ is the number of blocks of ξ. In particular, the coalescence probability c_N, i.e. the probability that two individuals, randomly chosen from some generation, have a common parent, is

$$\begin{aligned}c_N &:= \Phi_1^{(N)}(2) \;=\; \frac{1}{\sigma_2(\Theta_1(N))}\sum_{n=1}^N\sigma_2(\theta_n)\\ &= \frac{\phi_2\Theta_1(N)+\phi_1^2\Theta_2(N)}{\phi_2\Theta_1(N)+\phi_1^2(\Theta_1(N))^2} \;=\; \frac{\phi_2+\phi_1^2\frac{\Theta_2(N)}{\Theta_1(N)}}{\phi_2+\phi_1^2\Theta_1(N)}.\end{aligned}\tag{9}$$

Since $N\Theta_2(N)\geq(\Theta_1(N))^2$, it follows that $c_N\geq 1/N$. For all compound Poisson models, the effective population size $N_e:=1/c_N$ is therefore smaller than or equal to N. The equality $N_e=N$ holds if and only if $(N-1)\phi_2/\phi_1^2=\Theta_1(N)-N\Theta_2(N)/\Theta_1(N)$. In particular, $N_e=N$ if $\phi_2=0$ and $\theta_1=\cdots=\theta_N$. We will also need the probability that three individuals, randomly sampled from some generation, have a common parent, which is given by

$$\begin{aligned}d_N &:= \Phi_1^{(N)}(3) \;=\; \frac{1}{\sigma_3(\Theta_1(N))}\sum_{n=1}^N\sigma_3(\theta_n)\\ &= \frac{\phi_3\Theta_1(N)+3\phi_1\phi_2\Theta_2(N)+\phi_1^3\Theta_3(N)}{\phi_3\Theta_1(N)+3\phi_1\phi_2(\Theta_1(N))^2+\phi_1^3(\Theta_1(N))^3}\\ &= \frac{\phi_3+3\phi_1\phi_2\frac{\Theta_2(N)}{\Theta_1(N)}+\phi_1^3\frac{\Theta_3(N)}{\Theta_1(N)}}{\phi_3+3\phi_1\phi_2\Theta_1(N)+\phi_1^3(\Theta_1(N))^2}.\end{aligned}\tag{10}$$

For $n\in\mathbb{N}$ let ϱ_n denote the restriction from $\mathscr{E}$, the set of all equivalence relations on $\mathbb{N}$, to $\mathscr{E}_n$. As in Definition 2.1 of [6], we say that the considered population model is in the domain of attraction of a continuous-time coalescent process $R=(R_t)_{t\in[0,\infty)}$, if, for each sample size $n\in\mathbb{N}$,

the time-scaled ancestral process $(\mathscr{R}^{(n)}_{[t/c_N]})_{t\in[0,\infty)}$ weakly converges to $(\varrho_n R_t)_{t\in[0,\infty)}$ as $N\to\infty$. Analogously, we say that the considered population model is in the domain of attraction of a discrete-time coalescent process $(R_t)_{t\in\mathbb{N}_0}$, if for each sample size $n\in\mathbb{N}$, the ancestral process $(\mathscr{R}^{(n)}_t)_{t\in\mathbb{N}_0}$ weakly converges to $(\varrho_n R_t)_{t\in\mathbb{N}_0}$ as $N\to\infty$.
In order to state the first theorem it is helpful to introduce, for each $j\in\mathbb{N}$, the simplex

$$\Delta_j := \{(x_1,\ldots,x_j)\,|\,x_1,\ldots,x_j\ge 0, x_1+\cdots+x_j\le 1\}$$

and as well the infinite simplex $\Delta := \{(x_1,x_2,\ldots) : x_1\ge x_2\ge\cdots\ge 0, \sum_{n=1}^\infty x_n\le 1\}$. Moreover, for $x=(x_1,x_2,\ldots)\in\Delta$, the notation $|x| := \sum_{n=1}^\infty x_n$ and $(x,x) := \sum_{n=1}^\infty x_n^2$ will be used. Before we state the theorem, let us briefly recall the definition of a discrete-time Ξ-coalescent. A discrete-time Ξ-coalescent $R=(R_t)_{t\in\mathbb{N}_0}$ is a time-homogeneous Markovian process with state space $\mathscr{E}$ characterized by a finite measure Ξ on Δ having no atom at zero and satisfying

$$\int_\Delta \Xi(dx)/(x,x)\le 1$$

as follows. If the process R is in a state $\xi\in\mathscr{E}$ with k equivalence classes (blocks), then transitions to $\eta\in\mathscr{E}$ with $\xi\subseteq\eta$ occur with probability (see [13, Eq. (81)])

$$\mathbb{P}(R_{t+1}=\eta\,|\,R_t=\xi) \;=\; \int_\Delta \sum_{l=0}^{s}\binom{s}{l}(1-|x|)^{s-l}\sum_{\substack{i_1,\ldots,i_{r+l}\in\mathbb{N}\\ \text{all distinct}}} x_{i_1}^{k_1}\cdots x_{i_{r+l}}^{k_{r+l}}\,\frac{\Xi(dx)}{(x,x)},$$

where $j:=|\eta|$ denotes the number of blocks of η, $k_1,\ldots,k_j$ are the group sizes of merging classes of ξ, $s:=|\{1\le i\le j : k_i=1\}|$, and $r:=j-s$. We are now able to state the first main result.

Theorem 2.2. *Suppose that (6) holds. If $\sum_{n=1}^\infty\theta_n<\infty$, then the compound Poisson model is in the domain of attraction of a discrete-time coalescent process $R=(R_t)_{t\in\mathbb{N}_0}$ with simultaneous multiple collisions (Ξ-coalescent). The characterizing measure Ξ on Δ of R is obtained from the parameters of the compound Poisson model as follows. There exists a consistent sequence $(Q_j)_{j\in\mathbb{N}}$ of probability distribution Q_j on Δ_j uniquely determined via their moments*

$$\int_{\Delta_j} x_1^{k_1}\cdots x_j^{k_j}\,Q_j(dx_1,\ldots,dx_j) \;=\; \frac{\sigma_{k_1}(\theta_1)\cdots\sigma_{k_j}(\theta_j)}{\sigma_{k_1+\cdots+k_j}(\sum_{n=1}^\infty\theta_n)}, \tag{11}$$

$j\in\mathbb{N}$, $k_1,\ldots,k_j\in\mathbb{N}_0$. Let Q denote the projective limit of the sequence $(Q_j)_{j\in\mathbb{N}}$, let $X_1,X_2,\ldots$ be random variables with joint distribution Q, let $X_{(1)},X_{(2)},\ldots$ denote the $X_1,X_2,\ldots$ in descending order, and let ν be the joint distribution of the random variables $X_{(n)}$, $n\in\mathbb{N}$. Then, the characterizing measure Ξ on Δ of R has density $x\mapsto(x,x)$ with respect to ν. The probability measure ν (and hence also Ξ) is concentrated on the subset Δ^* of points $x=(x_1,x_2,\ldots)\in\Delta$ satisfying $|x|=1$.

Remarks. Note that $\sum_{n=1}^\infty\theta_n<\infty$ automatically implies that $\sum_{n=1}^\infty\theta_n^2<\infty$. The sequence $(X_n)_{n\in\mathbb{N}}$ in Theorem 2.2 satisfies $\sum_{n=1}^\infty X_n=1$ almost surely and can hence be viewed as a random partition of the unit interval. The proof of Theorem 2.2 is provided in Section 3.

Examples. Suppose that $\theta:=\sum_{n=1}^\infty\theta_n<\infty$.
1. (asymmetric Wright–Fisher models) If $\phi(z)=\phi_1 z$, then $\sigma_k(\theta)=\theta^k\phi_1^k$, $k\in\mathbb{N}_0$. Note that $\mu=(\mu_1,\ldots,\mu_N)$ has a multinomial distribution with parameters N and $\theta_n/(\theta_1+\cdots+\theta_N)$,

$n \in \{1, \ldots, N\}$, corresponding to an asymmetric Wright–Fisher model because of the constrain $\sum_{n=1}^{\infty} \theta_n < \infty$. It is readily seen that in this case ν is the Dirac measure at

$$p = (\theta_1/\theta, \theta_2/\theta, \ldots) \in \Delta^*.$$

The measure Ξ assigns its total mass $\Xi(\Delta) = (p, p) = (\sum_{n=1}^{\infty} \theta_n^2)/\theta^2$ to the single point p. We refer to Example 5.1 (vii) and Example 5.2 of [6] for the particular examples where $\theta_n = n^{-\alpha}$ with $\alpha > 1$ or $\theta_n = \lambda^n$ with $0 < \lambda < 1$.
2. (Dirichlet models) If $\phi(z) = -\log(1 - z)$, then $\phi_m = (m-1)!$, $m \in \mathbb{N}$, and

$$\sigma_k(\theta) = [\theta]_k := \theta(\theta + 1) \cdots (\theta + k - 1), \quad k \in \mathbb{N}_0.$$

In this case the limiting coalescent is the discrete-time Dirichlet–Kingman coalescent with parameter $(\theta_n)_{n \in \mathbb{N}}$. We refer the reader to the remark after Example 6.2 of [6] for more details on this particular coalescent process.

We now come to the second theorem, which covers the situation when the series $\sum_{n=1}^{\infty} \theta_n$ diverges, but so slowly that the series $\sum_{n=1}^{\infty} \theta_n^2$ still converges.

Theorem 2.3. *Suppose that (6) holds. If $\sum_{n=1}^{\infty} \theta_n = \infty$ and if $\sum_{n=1}^{\infty} \theta_n^2 < \infty$, then the compound Poisson model is in the domain of attraction of the Kingman coalescent [9]. The time-scaling c_N satisfies $c_N = \Theta_2(N)/(\Theta_1(N))^2$ if $\phi_2 = 0$ and $c_N \sim \phi_2/(\phi_1^2 \Theta_1(N))$ if $\phi_2 > 0$.*

In contrast to the situation in Theorem 2.2, the limiting coalescent in Theorem 2.3 (Kingman coalescent) does not depend on the particular function ϕ of the compound Poisson model. Theorem 2.3 is for example applicable if $\theta_n = n^{-\alpha}$, $n \in \mathbb{N}$, with $\alpha \in (1/2, 1]$.
It remains to focus on the situation when the series $\sum_{n=1}^{\infty} \theta_n$ and $\sum_{n=1}^{\infty} \theta_n^2$ both diverge. This situation turns out to be more involved than it seems at the first glance. We provide at least partial solutions for this case.

Theorem 2.4. *Suppose that (6) holds, that $\sum_{n=1}^{\infty} \theta_n = \infty$ and that $\sum_{n=1}^{\infty} \theta_n^2 = \infty$. Then the compound Poisson model is in the domain of attraction of the Kingman coalescent if and only if $\Theta_2(N)/(\Theta_1(N))^2 \to 0$ as $N \to \infty$. In this case the time-scaling c_N satisfies*

$$c_N \sim \phi_2/(\phi_1^2 \Theta_1(N)) + \Theta_2(N)/(\Theta_1(N))^2.$$

The following corollary, which is known from the literature (see Theorem 4.3 of [6]), is a direct consequence of Theorem 2.4. It covers the symmetric case, when all the parameters θ_n are equal to a given constant $\theta \in (0, \infty)$.

Corollary 2.5. *If (6) holds and if $\theta_n = \theta \in (0, \infty)$ for all $n \in \mathbb{N}$, then the compound Poisson model is in the domain of attraction of the Kingman coalescent. The time-scaling c_N satisfies*

$$c_N \sim (1 + \phi_2/(\phi_1^2 \theta))/N.$$

The proofs of Theorem 2.3 and Theorem 2.4 are provided in Section 3. These proofs give a bit more information than stated so far. For example, they show that compound Poisson models satisfying (6) are never in the domain of attraction of any continuous-time coalescent process different from the Kingman coalescent. Therefore, only the Kingman coalescent or discrete-time coalescent processes can arise in the limit as the total population size N tends to infinity. The proofs also show that the Kingman coalescent pops up in the limit if and only if $\lim_{N\to\infty} c_N = 0$ or, equivalently (see Lemma 3.1), if and only if

$$\lim_{N\to\infty} \frac{\Theta_2(N)}{(\Theta_1(N))^2} \;=\; 0. \tag{12}$$

There exist obviously compound Poisson models satisfying (6) which are not covered by the three theorems presented so far. Take for example $\theta_n := \lambda^n$ for some constant $\lambda \in (0,\infty)$. For $\lambda < 1$ we are in the situation of Theorem 2.2 and the case $\lambda = 1$ is covered by Corollary 2.5. Suppose now that $\lambda > 1$. Then, the series $\sum_{n=1}^\infty \theta_n$ and $\sum_{n=1}^\infty \theta_n^2$ both diverge, but

$$\Theta_2(N)/(\Theta_1(N))^2 \to (\lambda-1)/(\lambda+1) > 0.$$

By Theorem 2.4, this model cannot be in the domain of attraction of the Kingman coalescent and, due to the remarks made above, not in the domain of attraction of any continuous-time Ξ-coalescent. The following theorem covers this example.

Theorem 2.6. *Suppose that (6) holds and that all the limits*

$$p_1(k) \;:=\; \lim_{N\to\infty} \frac{\Theta_k(N)}{(\Theta_1(N))^k}, \qquad k \in \mathbb{N} \tag{13}$$

exist. Then, all the limits

$$p_j(k_1,\ldots,k_j) \;:=\; \lim_{N\to\infty} \frac{1}{(\Theta_1(N))^{k_1+\cdots+k_j}} \sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N} \theta_{n_1}^{k_1}\cdots\theta_{n_j}^{k_j}, \tag{14}$$

$k_1,\ldots,k_j \in \mathbb{N}$, *exist. Suppose now in addition that* $\sum_{n=1}^\infty \theta_n = \infty$ *and that* $p_1(2) > 0$. *Then the compound Poisson model is in the domain of attraction of a discrete-time* Ξ*-coalescent* $R = (R_t)_{t\in\mathbb{N}_0}$ *whose distribution is uniquely determined via the transition probabilities*

$$\mathbb{P}(\varrho_n R_{t+1} = \eta \,|\, \varrho_n R_t = \xi) = p_j(k_1,\ldots,k_j), \quad n \in \mathbb{N}$$

and $\xi,\eta \in \mathscr{E}_n$ *with* $\xi \subseteq \eta$, *where* $j := |\eta|$ *and* $k_1,\ldots,k_j$ *are the group sizes of merging classes of* ξ. *The characterizing measure* $\nu(dx) := \Xi(dx)/(x,x)$ *of* R *is the Dirac measure* $\nu = \delta_x$, *where* $x = (x_1,x_2,\ldots) \in \Delta$ *is given recursively via* $x_1 := \lim_{k\to\infty}(p_1(k))^{1/k}$ *and*

$$x_{n+1} := \lim_{k\to\infty}(p_1(k) - (x_1^k + \cdots + x_n^k))^{1/k}, \quad n \in \mathbb{N}.$$

Remarks. 1. Since the Euclidian 1-norm is greater than or equal to the Euclidian k-norm, it follows that $p_1(k) \le 1$ for all $k \in \mathbb{N}$.
2. For $N \in \mathbb{N}$ let Z_N be a random variable taking the value $\theta_n/\Theta_1(N)$ with probability $\theta_n/\Theta_1(N)$, $n \in \{1,\ldots,N\}$. It is readily verified that the existence of all the limits (13) is equivalent to the convergence $Z_N \to Z$ in distribution as $N \to \infty$, where Z is a random variable taking values in $[0,1]$ with characteristic function $\varphi(t) := \mathbb{E}(e^{itZ}) = \sum_{k=0}^\infty (t^k/k!)p_1(k+1)$, $t \in \mathbb{R}$.
3. In contrast to the situation in Theorem 2.2, the limiting discrete-time Ξ-coalescent R in Theorem 2.6 does not depend on the function ϕ of the compound Poisson model.

Examples. Fix $\lambda > 1$.
1. Suppose that $\theta_n = \lambda^n$ for all $n \in \mathbb{N}$. Then, for all $k \in \mathbb{N}$,

$$\Theta_k(N) = \lambda^k(\lambda^{kN} - 1)/(\lambda^k - 1) \sim \lambda^{k(N+1)}/(\lambda^k - 1)$$

as $N \to \infty$, and hence $p_1(k) = (\lambda - 1)^k/(\lambda^k - 1) > 0$, $k \in \mathbb{N}$. In this case Theorem 2.6 yields that the measure Ξ of the limiting Ξ-coalescent assigns its total mass $\Xi(\Delta) = p_1(2) = (\lambda - 1)/(\lambda + 1)$ to the single point $x = (x_1, x_2, \ldots) \in \Delta^*$, defined via $x_n := (\lambda-1)/\lambda^n = (1-1/\lambda)(1/\lambda)^{n-1}$, $n \in \mathbb{N}$.
2. If $\theta_n = \lambda^{n^2}$ for all $n \in \mathbb{N}$, then $\Theta_k(N) = \sum_{n=1}^N \theta_n^k \sim \theta_N^k = \lambda^{kN^2}$. It follows that $p_1(k) = 1$ for all $k \in \mathbb{N}$. The limiting Ξ-coalescent R is the discrete-time star-shaped coalescent, where Ξ is the Dirac measure at $(1, 0, 0, \ldots) \in \Delta$. In other words, if R is in a state with k blocks, then after one time step simply all k blocks have already merged together.
Conclusion. The results can be roughly summarized as follows. Compound Poisson models satisfying (6) are in the domain of attraction of the Kingman coalescent if and only if the sequence $(\theta_n)_{n\in\mathbb{N}}$ is balanced in the sense that (12) holds. If the sequence $(\theta_n)_{n\in\mathbb{N}}$ is unbalanced in the sense that it converges too fast to zero or too fast to infinity, then compound Poisson models tend to be in the domain of attraction of a discrete-time Ξ-coalescent (Theorem 2.2 and Theorem 2.6). It remains open to provide similar results for the full class of compound Poisson models which do not necessarily satisfy the constraint (6).

3 Proofs

Proof of Lemma 2.1. By induction on $m \in \mathbb{N}$ it follows from (6) that

$$\sum_{j=1}^m \frac{\sigma_{k_j+1}(\theta_j)}{\sigma_{k_j}(\theta_j)} = \frac{\sigma_{k_1+\cdots+k_m+1}(\theta_1 + \cdots + \theta_m)}{\sigma_{k_1+\cdots+k_m}(\theta_1 + \cdots + \theta_m)}, \quad k_1, \ldots, k_m \in \mathbb{N}_0, \theta_1, \ldots, \theta_m \in (0, \infty). \tag{15}$$

Choosing $k_j := 1$ for all $j \in \{1, \ldots, m\}$ in (15) leads to

$$\frac{\sigma_{m+1}(\theta_1 + \cdots + \theta_m)}{\sigma_m(\theta_1 + \cdots + \theta_m)} = \sum_{j=1}^m \frac{\sigma_2(\theta_j)}{\sigma_1(\theta_j)} = \sum_{j=1}^m \frac{\phi_2\theta_j + \phi_1^2\theta_j^2}{\phi_1\theta_j} = m\frac{\phi_2}{\phi_1} + (\theta_1 + \cdots + \theta_m)\phi_1$$

for all $m \in \mathbb{N}$ and all $\theta_1, \ldots, \theta_m \in (0, \infty)$. Choosing $\theta_j := \theta/m$ for all $j \in \{1, \ldots, m\}$ with $\theta \in (0, \infty)$ it follows that $\sigma_{m+1}(\theta) = \sigma_m(\theta)(m\phi_2/\phi_1 + \theta\phi_1)$ for all $m \in \mathbb{N}$ and all $\theta \in (0, \infty)$. The solution of this recursion with initial conditions $\sigma_0(\theta) = 1$ and $\sigma_1(\theta) = \phi_1\theta$ is

$$\sigma_m(\theta) = \prod_{i=0}^{m-1} \left(i\frac{\phi_2}{\phi_1} + \theta\phi_1 \right), \qquad m \in \mathbb{N}_0, \theta \in (0, \infty). \tag{16}$$

In particular, $\sigma_m(\theta)$ is a polynomial in θ of degree N and the coefficient in front of θ is

$$\phi_1 \prod_{i=1}^{m-1} (i\phi_2/\phi_1) = (m-1)!\phi_1(\phi_2/\phi_1)^{m-1}.$$

In general, the coefficient in front of θ of the polynomial $\sigma_m(\theta)$ is ϕ_m, which shows that the coefficients are of the form $\phi_m = (m-1)!\phi_1(\phi_2/\phi_1)^{m-1}$, $m \in \mathbb{N}$. Conversely, it is readily checked that if the coefficients ϕ_m are of this form, then $\sigma_m(\theta)$ is given by (16) and, hence, (6) holds.

Let us now verify (7) by backward induction on $k := k_1 + \cdots + k_N$. For $k > N$ both sides of (7) are equal to zero. For $k = N$ we have

$$\mathbb{E}((\mu_1)_{k_1} \cdots (\mu_N)_{k_N}) \;=\; k_1! \cdots k_N! \mathbb{P}(\mu_1 = k_1, \ldots, \mu_N = k_N) \;=\; N! \frac{\sigma_{k_1}(\theta) \cdots \sigma_{k_N}(\theta)}{\sigma_N(\theta_1 + \cdots + \theta_N)},$$

which is (7) for $k = N$. The induction step from $k+1$ to k works as follows. From $\mu_1 + \cdots + \mu_N = N$ and by induction it follows that

$$\begin{aligned}
(N-k)\mathbb{E}((\mu_1)_{k_1} \cdots (\mu_N)_{k_N}) \;&=\; \mathbb{E}((\mu_1)_{k_1} \cdots (\mu_N)_{k_N} \sum_{j=1}^N (\mu_j - k_j)) \\
&=\; \sum_{j=1}^N \mathbb{E}((\mu_1)_{k_1} \cdots (\mu_j)_{k_j+1} \cdots (\mu_N)_{k_N}) \;=\; \sum_{j=1}^N (N)_{k+1} \frac{\sigma_{k_1}(\theta_1) \cdots \sigma_{k_j+1}(\theta_j) \cdots \sigma_{k_N}(\theta_N)}{\sigma_{k+1}(\theta_1 + \cdots + \theta_N)} \\
&=\; (N)_{k+1} \frac{\sigma_{k_1}(\theta_1) \cdots \sigma_{k_N}(\theta_N)}{\sigma_{k+1}(\theta_1 + \cdots + \theta_N)} \sum_{j=1}^N \frac{\sigma_{k_j+1}(\theta_j)}{\sigma_{k_j}(\theta_j)} \;=\; (N)_{k+1} \frac{\sigma_{k_1}(\theta_1) \cdots \sigma_{k_N}(\theta_N)}{\sigma_k(\theta_1 + \cdots + \theta_N)},
\end{aligned} \tag{17}$$

where the last equality holds by (15). Division of (17) by $N - k$ shows that (7) holds for k which completes the induction. □

Proof of Theorem 2.2. For all $j, k_1, \ldots, k_j \in \mathbb{N}$ it follows by dominated convergence that

$$\begin{aligned}
\Phi_j^{(N)}(k_1, \ldots, k_j) \;&:=\; \frac{1}{\sigma_{k_1+\cdots+k_j}(\theta_1 + \cdots + \theta_N)} \sum_{\substack{n_1,\ldots,n_j=1 \\ \text{all distinct}}}^N \sigma_{k_1}(\theta_{n_1}) \cdots \sigma_{k_j}(\theta_{n_j}) \\
&\to\; \frac{1}{\sigma_{k_1+\cdots+k_j}(\sum_{n=1}^\infty \theta_n)} \sum_{\substack{n_1,\ldots,n_j \in \mathbb{N} \\ \text{all distinct}}} \sigma_{k_1}(\theta_{n_1}) \cdots \sigma_{k_j}(\theta_{n_j}) \\
&=:\; p_j(k_1, \ldots, k_j).
\end{aligned}$$

Note that $p_j(k_1, \ldots, k_j) \in [0,1]$, since $\Phi_j^{(N)}(k_1, \ldots, k_j) \in [0,1]$ for all $N \in \mathbb{N}$. Moreover,

$$\lim_{N\to\infty} c_N \;=\; \lim_{N\to\infty} \Phi_1^{(N)}(2) \;=\; p_1(2) \;=\; \frac{\sum_{n=1}^\infty \sigma_2(\theta_n)}{\sigma_2(\sum_{n=1}^\infty \theta_n)} \;>\; 0,$$

since $\sigma_2(\theta) = \theta\phi_2 + \theta^2\phi_1^2 \geq \theta^2\phi_1^2 > 0$ for all $\theta > 0$ by the general assumption $\phi_1 > 0$. The convergence $\Phi_j^{(N)}(k_1, \ldots, k_j) \to p_j(k_1, \ldots, k_j)$ as $N \to \infty$ for all $j, k_1, \ldots, k_j \in \mathbb{N}$ ensures (see, for example, [11, Theorem 2.1]) that for each sample size $n \in \mathbb{N}$ the ancestral process $(\mathscr{R}_t^{(n)})_{t\in\mathbb{N}_0}$ weakly converges to $(\varrho_n R_t)_{t\in\mathbb{N}_0}$ as $N \to \infty$, where $R = (R_t)_{t\in\mathbb{N}_0}$ is a discrete-time coalescent such that, if R is in a state with k blocks, any transition involving a $(k_1, \ldots, k_j)$-collision occurs with probability $p_j(k_1, \ldots, k_j)$. Thus it is shown that the model is in the domain of attraction of R. Let us now determine the characterizing measure Ξ of R. For $j \in \mathbb{N}$ and $k_1, \ldots, k_j \in \mathbb{N}_0$ define

$$m_j(k_1, \ldots, k_j) \;:=\; \frac{\sigma_{k_1}(\theta_1) \cdots \sigma_{k_j}(\theta_j)}{\sigma_{k_1+\cdots+k_j}(\sum_{n=1}^\infty \theta_n)} \;=\; \lim_{N\to\infty} \frac{\mathbb{E}((\mu_1)_{k_1} \cdots (\mu_j)_{k_j})}{(N)_{k_1+\cdots+k_j}},$$

where the last equality follows from (7). For every fixed $j \in \mathbb{N}$ the multi sequence $m_j(k_1, \ldots, k_j)$, $k_1, \ldots, k_j \in \mathbb{N}_0$ is completely monotone in the sense of Gupta [4, p. 287], since for arbitrary but fixed $i_1,\ldots,i_j \in \mathbb{N}_0$ with $i_1+\cdots+i_j \le N$, the multi-sequence $(i_1)_{k_1}\cdots(i_j)_{k_j}/(N)_{k_1+\cdots+k_j}$, $k_1,\ldots,k_j \in \mathbb{N}_0$ with $k_1+\cdots+k_j \le N$, is completely monotone and this property carries over to the convex combined multi-sequence

$$\frac{\mathbb{E}((\mu_1)_{k_1}\cdots(\mu_j)_{k_j})}{(N)_{k_1+\cdots+k_j}} \;=\; \sum_{\substack{i_1,\ldots,i_j\in\mathbb{N}_0\\ i_1+\cdots+i_j\le N}} \mathbb{P}(\mu_1=i_1,\ldots,\mu_j=i_j)\frac{(i_1)_{k_1}\cdots(i_j)_{k_j}}{(N)_{k_1+\cdots+k_j}},$$

$k_1,\ldots,k_j \in \mathbb{N}_0$ with $k_1+\cdots+k_j \le N$ and, hence, to the limiting sequence $m_j(k_1,\ldots,k_j)$, $k_1,\ldots,k_j \in \mathbb{N}_0$. Thus (see, for example, [4]), for each $j\in\mathbb{N}$, there exists a measure Q_j on the j-dimensional simplex $\Delta_j := \{(x_1,\ldots,x_j) : x_1,\ldots,x_j \ge 0, x_1+\cdots+x_j \le 1\}$ uniquely determined via its moments

$$\int_{\Delta_j} x_1^{k_1}\cdots x_j^{k_j}\, Q_j(dx_1,\ldots,dx_j) \;=\; m_j(k_1,\ldots,k_j).$$

Since $m_{j+1}(k_1,\ldots,k_j,0) = m_j(k_1,\ldots,k_j)$ for all $k_1,\ldots,k_j \in \mathbb{N}_0$, it follows that the sequence of measures $(Q_j)_{j\in\mathbb{N}}$ is consistent. Thus, by Kolmogorov's extension theorem there exists a probability measure Q on $\mathbb{R}^{\mathbb{N}}$, the projective limit of the sequence $(Q_j)_{j\in\mathbb{N}}$. Let $X_1, X_2,\ldots$ be random variables with joint distribution Q, let $X_{(1)} \ge X_{(2)} \ge \cdots$ denote the $X_1, X_2,\ldots$ in decreasing order, and let ν be the joint distribution of the ordered random variables $X_{(n)}$, $n\in\mathbb{N}$. Then,

$$\begin{aligned}
&\int_\Delta \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\ \text{all distinct}}} x_{n_1}^{k_1}\cdots x_{n_j}^{k_j}\,\nu(dx)\\
&\qquad= \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\ \text{all distinct}}} \mathbb{E}(X_{(n_1)}^{k_1}\cdots X_{(n_j)}^{k_j}) \;=\; \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\ \text{all distinct}}} \mathbb{E}(X_{n_1}^{k_1}\cdots X_{n_j}^{k_j})\\
&\qquad= \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\ \text{all distinct}}} \frac{\sigma_{k_1}(\theta_{n_1})\cdots\sigma_{k_j}(\theta_{n_j})}{\sigma_{k_1+\cdots+k_j}(\sum_{n=1}^\infty \theta_n)} \;=\; p_j(k_1,\ldots,k_j),
\end{aligned}$$

showing that $\Xi(dx) := (x,x)\nu(dx)$ is the characterizing measure of the coalescent R in the spirit of Schweinsberg [13]. Note that $\int_\Delta |x|\,\nu(dx) = p_1(1) = \sum_{n_1=1}^\infty \sigma_1(\theta_{n_1})/\sigma_1(\sum_{n=1}^\infty \theta_n) = 1$, since $\sigma_1(\theta) = \phi_1\theta$ for all $\theta>0$. Thus, $\int_\Delta (1-|x|)\nu(dx) = 0$, showing that ν is concentrated on the subset Δ^* of points $x\in\Delta$ satisfying $|x|=1$. □

Remark. For pairwise distinct $n_1,\ldots,n_j\in\mathbb{N}$ let $P_{n_1,\ldots,n_j}$ denote the distribution of $X_{n_1},\ldots,X_{n_j}$, and define the symmetric measure M_j on Δ_j via $M_j := \sum_{n_1,\ldots,n_j} P_{n_1,\ldots,n_j}$, where the sum extends over all pairwise distinct $n_1,\ldots,n_j\in\mathbb{N}$. Then, for all M_j-integrable functions g,

$$\mathbb{E}(\sum_{n_1,\ldots,n_j} g(X_{n_1},\ldots,X_{n_j})) = \int_{\Delta_j} g\,dM_j,$$

showing that M_j is the correlation measure (see, for example, [5, Eq. (2.1)]) of the point process $\sum_{n=1}^\infty \delta_{X_n}$.

We now come to the proofs of Theorem 2.3 and Theorem 2.4. They are based on the following technical but fundamental lemma. Once this lemma is established the proofs of both theorems follow with only little effort. Recall that $\Theta_k(N) := \sum_{n=1}^N \theta_n^k$ for $N,k\in\mathbb{N}$.

Lemma 3.1. *If (6) holds, then the following four conditions are equivalent.*

(i) $\lim_{N\to\infty} \frac{\Theta_2(N)}{(\Theta_1(N))^2} = 0.$ *(ii)* $\lim_{N\to\infty} \frac{\Theta_3(N)}{\Theta_1(N)\Theta_2(N)} = 0.$

(iii) $\lim_{N\to\infty} c_N = 0.$ *(iv)* $\lim_{N\to\infty} \frac{d_N}{c_N} = 0.$

Remark. For the equivalence of (i) and (ii) the constraint (6) is not needed.

Proof. '(i) ⇒ (ii)': Since the Euclidian 2-norm is larger than or equal to the Euclidian 3-norm, we have $(\Theta_2(N))^{1/2} \ge (\Theta_3(N))^{1/3}$. Thus, $\Theta_3(N) \le (\Theta_2(N))^{3/2}$, and, consequently,

$$\frac{\Theta_3(N)}{\Theta_1(N)\Theta_2(N)} \le \Big(\frac{\Theta_2(N)}{(\Theta_1(N))^2}\Big)^{1/2}.$$

'(ii) ⇒ (i)': An application of the Hölder inequality

$$\left(\sum_{n=1}^N |a_n b_n|\right)^2 \le \left(\sum_{n=1}^N a_n^2\right)\left(\sum_{n=1}^N b_n^2\right)$$

to $a_n := \theta_n^{1/2}$ and $b_n := \theta_n^{3/2}$ leads to $(\Theta_2(N))^2 \le \Theta_1(N)\Theta_3(N)$, or, equivalently,

$$\frac{\Theta_2(N)}{(\Theta_1(N))^2} \le \frac{\Theta_3(N)}{\Theta_1(N)\Theta_2(N)}.$$

'(iii) ⇒ (i)': Since the Euclidian 1-norm is larger than or equal to the Euclidian 2-norm, we have $(\Theta_1(N))^2 \ge \Theta_2(N)$. Using (9) it is easily checked that this inequality implies that

$$c_N = \frac{\phi_2 + \phi_1^2\frac{\Theta_2(N)}{\Theta_1(N)}}{\phi_2 + \phi_1^2\Theta_1(N)} \ge \frac{\Theta_2(N)}{(\Theta_1(N))^2}.$$

'(i) ⇒ (iii)': Let us first verify that $\Theta_1(N) \to \infty$ as $N \to \infty$. Suppose this is not the case. Then, $\sum_{n=1}^\infty \theta_n < \infty$. In particular, $\theta_n \to 0$ as $n \to \infty$, and, as a consequence, the series $\sum_{n=1}^\infty \theta_n^2 < \infty$ converges as well. It follows that $\lim_{N\to\infty} \Theta_2(N)/(\Theta_1(N))^2 = (\sum_{n=1}^\infty \theta_n^2)/(\sum_{n=1}^\infty \theta_n)^2 > 0$, an obvious contradiction to (i). Thus, we have $\Theta_1(N) \to \infty$ as $N \to \infty$. Therefore,

$$c_N \le \frac{\phi_2 + \phi_1^2\frac{\Theta_2(N)}{\Theta_1(N)}}{\phi_1^2\Theta_1(N)} = \frac{\phi_2}{\phi_1^2}\frac{1}{\Theta_1(N)} + \frac{\Theta_2(N)}{(\Theta_1(N))^2} \to 0 \text{ by (i).}$$

'(iv) ⇒ (iii)': A fundamental theorem from coalescent theory (see, for ex., [10, p. 989] or [12, Lemma 5.5]) states that, for arbitrary exchangeable Cannings models, if $d_N/c_N \to 0$, then $c_N \to 0$.
'(i) ⇒ (iv)': Note first (as already shown in the part '(i) ⇒ (iii)') that (i) implies that $\Theta_1(N) \to \infty$ as $N \to \infty$. Furthermore, we have already shown that (i) is equivalent to (ii), so we are allowed to use (ii). Two cases are now distinguished.
Case 1: Suppose that $\sum_{n=1}^\infty \theta_n^2 = \infty$. Then, using the inequality $c_N \ge \Theta_2(N)/(\Theta_1(N))^2$ (which is always valid as shown in the part '(iii) ⇒ (i)'), and using (10), it follows that

$$\begin{aligned}\frac{d_N}{c_N} &\le d_N\frac{(\Theta_1(N))^2}{\Theta_2(N)} \le \frac{\phi_3 + 3\phi_1\phi_2\frac{\Theta_2(N)}{\Theta_1(N)} + \phi_1^3\frac{\Theta_3(N)}{\Theta_1(N)}}{\phi_1^3(\Theta_1(N))^2}\frac{(\Theta_1(N))^2}{\Theta_2(N)}\\ &= \frac{\phi_3}{\phi_1^3}\frac{1}{\Theta_2(N)} + \frac{3\phi_2}{\phi_1^2}\frac{1}{\Theta_1(N)} + \frac{\Theta_3(N)}{\Theta_1(N)\Theta_2(N)} \to 0,\end{aligned}$$

by (ii) and since $\Theta_1(N) \to \infty$ and $\Theta_2(N) \to \sum_{n=1}^{\infty} \theta_n^2 = \infty$.
Case 2: Suppose that $\sum_{n=1}^{\infty} \theta_n^2 < \infty$. Then, we also have $\sum_{n=1}^{\infty} \theta_n^3 < \infty$. If $\phi_2 > 0$ and $\phi_3 > 0$, then

$$c_N = \frac{\phi_2 + \phi_1^2 \frac{\Theta_2(N)}{\Theta_1(N)}}{\phi_2 + \phi_1^2 \Theta_1(N)} \sim \frac{\phi_2}{\phi_1^2} \frac{1}{\Theta_1(N)} \tag{18}$$

and analogously $d_N \sim \phi_3/(\phi_1^3(\Theta_1(N))^2)$, and it follows that $d_N/c_N \sim \phi_3/(\phi_1\phi_2\Theta_1(N)) \to 0$. If $\phi_2 > 0$ and $\phi_3 = 0$, then (18) still holds, but

$$d_N \sim \frac{3\phi_1\phi_2\Theta_2(N) + \phi_1^3\Theta_3(N)}{\phi_1^3(\Theta_1(N))^3}.$$

Since the sequences $(\Theta_2(N))_{N\in\mathbb{N}}$ and $(\Theta_3(N))_{N\in\mathbb{N}}$ are bounded, it follows that

$$d_N/c_N \leq C/(\Theta_1(N))^2$$

for some constant $C > 0$, and, in particular, $d_N/c_N \to 0$.
Finally, if $\phi_2 = \phi_3 = 0$, then $d_N/c_N = \Theta_3(N)/(\Theta_1(N)\Theta_2(N)) \to 0$ by (ii), which completes the proof of the lemma. □

Thanks to Lemma 3.1, the following proofs of Theorem 2.3 and 2.4 are short and straightforward.

Proof of Theorem 2.3. By assumption, $\Theta_1(N) \to \sum_{n=1}^{\infty} \theta_n = \infty$ as $N \to \infty$, whereas the sequence $(\Theta_2(N))_{N\in\mathbb{N}}$ is bounded. The condition (i) of Lemma 3.1 and hence all four conditions of Lemma 3.1 are therefore satisfied. In particular, $d_N/c_N \to 0$ as $N \to \infty$. By [11] or [10, Theorem 4 (b)], for each sample size $n \in \mathbb{N}$, the time-scaled ancestral process $(\mathscr{R}^{(n)}_{[t/c_N]})_{t\in[0,\infty)}$ weakly converges to the Kingman n-coalescent as $N \to \infty$. Thus, the model is in the domain of attraction of the Kingman coalescent. From (9) it follows that $c_N = \Theta_2(N)/(\Theta_1(N))^2$ if $\phi_2 = 0$ and that $c_N \sim \phi_2/(\phi_1^2\Theta_1(N))$ if $\phi_2 > 0$. □

Proof of Theorem 2.4. By Lemma 3.1, the condition (i) is equivalent to $d_N/c_N \to 0$ as $N \to \infty$, which in turn (see [11] or [10, Theorem 4 (b)]) is equivalent to the fact that the model is in the domain of attraction of the Kingman coalescent. Moreover, since $\Theta_1(N) \to \infty$, it follows that

$$c_N = \frac{\phi_2 + \phi_1^2 \frac{\Theta_2(N)}{\Theta_1(N)}}{\phi_2 + \phi_1^2\Theta_1(N)} \sim \frac{\phi_2 + \phi_1^2 \frac{\Theta_2(N)}{\Theta_1(N)}}{\phi_1^2\Theta_1(N)} = \frac{\phi_2}{\phi_1^2}\frac{1}{\Theta_1(N)} + \frac{\Theta_2(N)}{(\Theta_1(N))^2}.$$

□

Proof of Theorem 2.6. Let us first verify by induction on j that all the limits (14) exist. For $j = 1$, these limits exist by assumption. Now fix $j \in \mathbb{N}$ and assume that the limits

$$p_1(k_1), p_2(k_1, k_2), \ldots, p_j(k_1, \ldots, k_j)$$

exist for all $k_1,\ldots,k_j\in\mathbb{N}$. Then, for all $k_1,\ldots,k_{j+1}\in\mathbb{N}$,

$$\begin{aligned}
&\frac{1}{(\Theta_1(N))^{k_1+\cdots+k_{j+1}}}\sum_{\substack{n_1,\ldots,n_{j+1}=1\\ \text{all distinct}}}^{N}\theta_{n_1}^{k_1}\cdots\theta_{n_{j+1}}^{k_{j+1}}\\
&= \frac{1}{(\Theta_1(N))^{k_1+\cdots+k_{j+1}}}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\theta_{n_1}^{k_1}\cdots\theta_{n_j}^{k_j}\Big(\Theta_{k_{j+1}}(N)-\sum_{i=1}^{j}\theta_{n_i}^{k_{j+1}}\Big)\\
&= \frac{\Theta_{k_{j+1}}(N)}{(\Theta_1(N))^{k_{j+1}}}\frac{1}{(\Theta_1(N))^{k_1+\cdots+k_j}}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\theta_{n_1}^{k_1}\cdots\theta_{n_j}^{k_j}\\
&\qquad -\sum_{i=1}^{j}\frac{1}{(\Theta_1(N))^{k_1+\cdots+k_{j+1}}}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\theta_{n_1}^{k_1}\cdots\theta_{n_i}^{k_i+k_{j+1}}\cdots\theta_{n_j}^{k_j}\\
&\to p_1(k_{j+1})p_j(k_1,\ldots,k_j)-\sum_{i=1}^{j}p_j(k_1,\ldots,k_{i-1},k_i+k_{j+1},k_{i+1},\ldots,k_j)
\end{aligned}$$

as $N\to\infty$, which shows that the limits $p_{j+1}(k_1,\ldots,k_{j+1})$ exist for all $k_1,\ldots,k_{j+1}\in\mathbb{N}$ and which completes the induction. It is in particular shown that the limits (14) satisfy the recursion

$$\begin{aligned}
p_{j+1}(k_1,\ldots,k_{j+1}) &= p_j(k_1,\ldots,k_j)p_1(k_{j+1})\\
&\quad -\sum_{i=1}^{j}p_j(k_1,\ldots,k_{i-1},k_i+k_{j+1},k_{i+1},\ldots,k_j), \qquad (19)
\end{aligned}$$

$j,k_1,\ldots,k_{j+1}\in\mathbb{N}$. Now fix $k_1,\ldots,k_j\in\mathbb{N}$ and define $k:=k_1+\cdots+k_j$ for convenience. Since $\Theta_1(N)\to\sum_{n=1}^{\infty}\theta_n=\infty$ by assumption, we have

$$\begin{aligned}
\Phi_j^{(N)}(k_1,\ldots,k_j) &= \frac{1}{\sigma_k(\Theta_1(N))}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\sigma_{k_1}(\theta_{n_1})\cdots\sigma_{k_j}(\theta_{n_j})\\
&\sim \frac{1}{\phi_1^k(\Theta_1(N))^k}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\phi_1^{k_1}\theta_{n_1}^{k_1}\cdots\phi_1^{k_j}\theta_{n_j}^{k_j}\\
&= \frac{1}{(\Theta_1(N))^k}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\theta_{n_1}^{k_1}\cdots\theta_{n_j}^{k_j}\to p_j(k_1,\ldots,k_j)
\end{aligned}$$

as $N\to\infty$. Suppose now that $p_1(2)>0$. Note that in the situation of Theorem 2.2 the condition $p_1(2)>0$ is automatically satisfied whereas in the present proof it has to be assumed that $p_1(2)>0$ in order to be able to apply [11, Theorem 2.1] with $c:=\lim_{N\to\infty}c_N>0$. As in the proof of Theorem 2.2, the convergence $\Phi_j^{(N)}(k_1,\ldots,k_j)\to p_j(k_1,\ldots,k_j)$ for all $j,k_1,\ldots,k_j\in\mathbb{N}$ ensures that the model is in the domain of attraction of a discrete-time Ξ-coalescent $R=(R_t)_{t\in\mathbb{N}_0}$, whose distribution is uniquely determined via the transition probabilities

$$\mathbb{P}(\varrho_n R_{t+1}=\eta\,|\,\varrho_n R_t=\xi)=p_j(k_1,\ldots,k_j),\quad n\in\mathbb{N},$$

$\xi,\eta\in\mathscr{E}_n$ with $\xi\subseteq\eta$, where $j=|\eta|$ and $k_1,\ldots,k_j$ are the merging classes of ξ. It remains to characterize the measure Ξ of R. By Proposition 35 of [13], the measure $\nu(dx):=\Xi(dx)/(x,x)$ satisfies $\nu(\Delta)\le 1$ and (choose $s=0$ in Eq. (81) of [13])

$$p_j(k_1,\ldots,k_j) \;=\; \int_\Delta \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\ \text{all distinct}}} x_{n_1}^{k_1}\cdots x_{n_j}^{k_j}\nu(dx) \tag{20}$$

for $j\in\mathbb{N}$ and $k_1,\ldots,k_j\ge 2$. In particular, for all $k\ge 2$,

$$\begin{aligned} p_2(k,k) &= \int_\Delta \sum_{\substack{n_1,n_2\in\mathbb{N}\\ n_1\ne n_2}} x_{n_1}^k x_{n_2}^k\,\nu(dx) \;=\; \int_\Delta \Big(\Big(\sum_{n=1}^\infty x_n^k\Big)^2-\sum_{n=1}^\infty x_n^{2k}\Big)\,\nu(dx)\\ &= \int_\Delta g_k^2(x)\,\nu(dx)-p_1(2k), \end{aligned} \tag{21}$$

with $g_k:\Delta\to[0,1]$ defined via $g_k(x):=\sum_{n=1}^\infty x_n^k$, $k\ge 2$, $x\in\Delta$. On the other hand, by (19), $p_2(k,k)=(p_1(k))^2-p_1(2k)$. Comparison with (21) shows that

$$\int_\Delta g_k^2(x)\,\nu(dx) \;=\; (p_1(k))^2 \;=\; \Big(\int_\Delta g_k(x)\,\nu(dx)\Big)^2. \tag{22}$$

Let Q_k denote the probability measure on $([0,1],\mathscr{B}([0,1]))$ defined via

$$Q_k(B) \;:=\; \frac{\nu_{g_k}(B)}{\nu(\Delta)} \;:=\; \frac{\nu(g_k^{-1}(B))}{\nu(\Delta)}, \qquad B\in\mathscr{B}([0,1]),$$

and let Y_k be a random variable with distribution Q_k. Since Y_k has moments

$$\mathbb{E}(Y_k^j) \;=\; \frac{1}{\nu(\Delta)}\int_{[0,1]} y^j\,\nu_{g_k}(dy) \;=\; \frac{1}{\nu(\Delta)}\int_\Delta (g_k(x))^j\,\nu(dx), \qquad j\in\mathbb{N}_0,$$

it follows from (22) that $\mathbb{E}(Y_k^2)=\nu(\Delta)(\mathbb{E}(Y_k))^2$. Since $\mathbb{E}(Y_k^2)\ge(\mathbb{E}(Y_k))^2$ it follows that $\nu(\Delta)\ge 1$. Thus $\nu(\Delta)=1$ and, hence, $\mathbb{E}(Y_k^2)=(\mathbb{E}(Y_k))^2$, showing that each measure Q_k assigns its total mass 1 to a single point, say $q_k\in[0,1]$. In other words, the measure ν is concentrated on points $x=(x_1,x_2,\ldots)\in\Delta$ satisfying $g_k(x)=q_k$ for all $k\in\{2,3,\ldots\}$. Since the map $g:\Delta\to\mathbb{R}^\mathbb{N}$, $g(x):=(g_2(x),g_3(x),\ldots)$, is injective (see Corollary 4.2 in the appendix), the measure ν assigns its total mass 1 to a single point, say $x=(x_1,x_2,\ldots)\in\Delta$. By (20), $p_1(k)=\sum_{n=1}^\infty x_n^k=g_k(x)$ for $k\in\{2,3,\ldots\}$. For $j\in\mathbb{N}$ define $f_j:\mathbb{N}\to\mathbb{R}$ via $f_j(n):=x_{n+j-1}$ for $n\in\mathbb{N}$. By Lemma 4.1 (applied with $\mu:=\delta_\mathbb{N}$, the counting measure on $\mathbb{N}$) it follows that $(p_1(k))^{1/k}=\|f_1\|_k\to\|f_1\|_\infty=x_1$ as $k\to\infty$, $(p_1(k)-x_1^k)^{1/k}=\|f_2\|_k\to\|f_2\|_\infty=x_2$, and so on, which shows that the point x is obtained recursively from the $p_1(k)$, $k\in\mathbb{N}$, via $x_1:=\lim_{k\to\infty}(p_1(k))^{1/k}$ and

$$x_{n+1}=(p_1(k)-(x_1^k+\cdots+x_n^k))^{1/k}, \quad n\in\mathbb{N}.$$

□

4 Appendix

Versions of the following Lemma 4.1 are well known from the literature. Standard proofs (see, for example, [15, p. 34, Theorem 1]) usually work under the assumption that the underlying measure μ is finite. We state a slightly different version, which is valid for arbitrary (not necessarily finite) measures μ.

Lemma 4.1. *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and let $f : \Omega \to \mathbb{R}$ be a measurable function satisfying $\|f\|_r := (\int |f|^r \, d\mu)^{1/r} < \infty$ for all $r \geq r_0 \geq 1$. Then,*

$$\lim_{r\to\infty} \|f\|_r = \|f\|_\infty := \inf\{a \in [0,\infty) : |f| \leq a \ \mu\text{-almost everywhere}\} \in [0,\infty].$$

Proof. We have

$$\|f\|_r^r = \int |f|^r \, d\mu = \int |f|^{r_0} |f|^{r-r_0} \, d\mu \leq \|f\|_\infty^{r-r_0} \int |f|^{r_0} \, d\mu.$$

Thus, $\|f\|_r \leq \|f\|_\infty^{1-r_0/r} \|f\|_{r_0}^{r_0/r} \to \|f\|_\infty$ as $r \to \infty$, and, consequently, $\limsup_{r\to\infty} \|f\|_r \leq \|f\|_\infty$. In order to verify that $\liminf_{r\to\infty} \|f\|_r \geq \|f\|_\infty$ fix $0 < a < \|f\|_\infty$ and define $A := \{|f| > a\}$. Note that $\mu(A) > 0$. Consider

$$\|f\|_r \ = \ a\left(\int \left(\frac{|f|}{a}\right)^r d\mu\right)^{\frac{1}{r}} \ \geq \ a\left(\int_A \left(\frac{|f|}{a}\right)^r d\mu\right)^{\frac{1}{r}}.$$

On A we have $|f|/a > 1$. Thus, for $r \to \infty$ the term below the last integral converges to infinity. From $\mu(A) > 0$ it follows by monotone convergence that $\int_A (|f|/a)^r \, d\mu \to \infty$ as $r \to \infty$. For all sufficiently large r this integral is in particular greater than 1, which implies that

$$(\int_A (|f|/a)^r \, d\mu)^{1/r} > 1$$

for all sufficiently large r. It follows that $\|f\|_r > a$ for all sufficiently large r. Since a can be chosen arbitrarily close to $\|f\|_\infty$, it follows that $\liminf_{r\to\infty} \|f\|_r \geq \|f\|_\infty$. □

The following corollary is needed in the proof of Theorem 2.6.

Corollary 4.2. *The map $g : \Delta \to \mathbb{R}^{\mathbb{N}}$, $g(x) := (\sum_{n=1}^\infty x_n^2, \sum_{n=1}^\infty x_n^3, \ldots)$, is injective.*

Proof. For $x \in \Delta$ define $\|x\|_k := (\sum_{n=1}^\infty x_n^k)^{1/k}$ and $\|x\|_\infty := \sup_{n\in\mathbb{N}} x_n = x_1$. Suppose that $x, y \in \Delta$ with $g(x) = g(y)$. Then, $\|x\|_k = \|y\|_k$ for all $k \in \{2, 3, \ldots\}$. By Lemma 4.1 (applied with $\mu := \delta_{\mathbb{N}}$, the counting measure on $\mathbb{N}$, and $f(n) := x_n$ for $n \in \mathbb{N}$) it follows that

$$x_1 = \|x\|_\infty = \lim_{k\to\infty} \|x\|_k = \lim_{k\to\infty} \|y\|_k = \|y\|_\infty = y_1.$$

Define $\tilde{x} := (x_2, x_3, \ldots) \in \Delta$ and $\tilde{y} := (y_2, y_3, \ldots) \in \Delta$. Since $g(x) = g(y)$ and $x_1 = y_1$, it follows that $g(\tilde{x}) = g(\tilde{y})$. Proceeding in the same way as before, but with $\tilde{x}$ and $\tilde{y}$ instead of x and y, yields $x_2 = y_2$. Inductively it follows that $x_n = y_n$ for all $n \in \mathbb{N}$. Therefore, $x = y$. □

References

[1] Berestycki, N. and Pitman, J. (2007) Gibbs distributions for random partitions generated by a fragmentation process. *J. Stat. Phys.* **127**, 381–418. MR2314353

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

[3] Cannings, C. (1975) The latent roots of certain Markov chains arising in genetics: a new approach, II. Further haploid models. *Adv. Appl. Probab.* **7**, 264–282. MR0371430

[4] Gupta, J. C. (1999) The moment problem for the standard k-dimensional simplex. *Sankhyā A* **61**, 286–291. MR1714879

[5] Handa, K. (2009) The two-parameter Poisson–Dirichlet point process. *Bernoulli* **15**, 1082–1116. MR2597584

[6] Huillet, T. and Möhle, M. (2011) Population genetics models with skewed fertilities: a forward and backward analysis. *Stochastic Models* **27**, 521–554. Math. Review number not yet available.

[7] Karlin, S. and McGregor, J. (1964) Direct product branching processes and related Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* **51**, 598–602. MR0163362

[8] Karlin, S. and McGregor, J. (1965) Direct product branching processes and related Markov chains. I. Calculations of rates of approach to homozygosity. *Proc. Internat. Res. Sem.*, Springer, Berlin, pp. 111–145. MR0217892

[9] Kingman, J. F. C. (1982) The coalescent. *Stoch. Process. Appl.* **13**, 235–248. MR0671034

[10] Möhle, M. (2000) Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.* **32**, 983–993. MR1808909

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

[12] Möhle, M. and Sagitov, S. (2003) Coalescent patterns in diploid exchangeable population models. *J. Math. Biol.* **47**, 337–352. MR2024501

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

[14] Schweinsberg, J. (2003) Coalescent processes obtained from supercritical Galton–Watson processes. *Stoch. Process. Appl.* **106**, 107–139. MR1983046

[15] Yosida, K. (1980) *Functional Analysis*, Sixth Edition, Springer, Berlin. MR0617913