Population genetics models with skewed fertilities: a forward and backward analysis

Thierry Huillet, Martin Moehle

POPULATION GENETICS MODELS WITH SKEWED FERTILITIES: A FORWARD AND BACKWARD ANALYSIS

THIERRY HUILLET, MARTIN MÖHLE

Abstract. Discrete population genetics models with unequal fertilities are considered, with an emphasis on skewed Cannings models, skewed conditional branching process models in the spirit of Karlin and McGregor, and skewed compound Poisson models. Three particular classes of models with skewed fertilities are investigated, the skewed Wright-Fisher model, the skewed Dirichlet model, and the skewed Kimura model. For each class the asymptotic behavior as the total population size N tends to infinity is investigated for power law fertilities and for geometric fertilities. This class of models can exhibit a rich variety of sub-linear or even constant effective population sizes. Therefore, the models are not even necessarily in the domain of attraction of the Kingman coalescent. For a substantial range of the parameters, discrete-time coalescent processes with simultaneous multiple collisions arise in the limit.

Running title: Population genetics models with skewed fertilities.

Keywords: Ancestral process; Cannings model; Compound Poisson model; Dirichlet model; Duality; Evolutionary processes; Exchangeable coalescent; Karlin and McGregor model; Kimura model; Kingman coalescent; Population dynamics; Simultaneous multiple collisions; Wright-Fisher model

1. Introduction

The well-known neutral discrete Wright-Fisher model describes the evolution of a population of constant size N from generation to generation forwards in time. One of the main features of this model (and similar models such as the discrete Moran model) is that each individual has the same propensity to produce offspring.

In this paper discrete population models are studied, in which individuals may have unequal propensities to reproduce. We shall speak of models with skewed fertilities. These models turn out to be of interest mainly because of the following two reasons. First of all, it turns out that the effective population size N_e of these models may differ significantly from the actual total population size N and may not even depend linearly on N. The fact that models with this behavior may have importance for biological applications, is for example indicated in papers of Eldon and Wakeley [7] and Wakeley and Sargsyan [30]. Models with effective population sizes smaller than N occur also in a different context of age-structure of populations in a paper of Sagitov and Jagers [26]. Secondly, in the limit as the total population size N tends to infinity, these models are not necessarily in the domain of attraction (see Definition 2.3) of the Kingman coalescent. The limiting coalescent processes may have simultaneous multiple collisions of ancestral lineages. Ancestral processes of this form are a major research area in coalescent theory and have been studied intensively over the last twenty years (see, for example, [21] or [27]).

The models we are mainly interested in are obtained by conditioning a sequence $\xi_1, \xi_2, \ldots$ of independent random variables on the event that $\xi_1 + \cdots + \xi_N = N$, with the interpretation that the outcome of ξ_n, under the constraint that $\xi_1 + \cdots + \xi_N = N$, is the number of children of the nth individual, $n \in \{1, \ldots, N\}$. Since the random variables $\xi_1, \xi_2, \ldots$ are not necessarily assumed to be identically distributed, this construction results in unequal propensities of the individuals to produce offspring. In the spirit of Karlin and McGregor [11] we call this model the skewed conditional branching process model.

Schweinsberg [28] studies models, in which N individuals are sampled from $S_N := X_1 + \cdots + X_N$ offspring, where $X_1, \ldots, X_N$ are given i.i.d. random variables. The effective population sizes of these models satisfy $N_e = O(N)$ (see [28, Eq. (18)]), and, in particular, N_e can exhibit a rich variety of sub-linear behavior, however, these models are different from our models, which are based on conditioning instead of sampling.

The paper is organized as follows. In Section 2 a skewed Cannings model is introduced and analyzed. The analysis of this model essentially boils down to a comparison with an associated standard exchangeable Cannings model [4, 5]. Section 2 therefore heavily gains from the theory on Cannings models. The results of Section 2 are applied in Section 3 to the skewed conditional branching process model. Since this huge class of conditional models has in general a quite complicated probabilistic structure, we further specialize in Section 4 to a subclass of skewed compound Poisson models. Exact formulae for the transition probabilities of the forward and the backward process are derived. It is furthermore shown (Theorem 4.3) that, in the unbiased (non-skewed) case, the model is in the domain of attraction (see Definition 2.3) of the Kingman coalescent. In all cases we obtain exact and asymptotic formulae for the effective population size N_e, which can deviate substantially from the total population size N. The following Sections 5 and 6 are devoted to two particular compound Poisson models, the skewed Wright-Fisher model and the skewed Dirichlet model. In both models the effective population size N_e is less than or equal to the total population size N, and, depending on the parameter choices, can indeed be substantially smaller than N, for example $N_e \sim \rho N$ with $\rho \in (0, 1)$, $N_e = O(N^\beta)$ with $\beta \in (0, 1)$, $N_e = O(\log N)$, or even $N_e = O(1)$. These two models are in particular analyzed for power law skewed fertilities and for geometrically skewed fertilities. It turns out that for a wide range of parameter choices, these models are not anymore in the domain of attraction of the Kingman coalescent. Coalescents allowing for simultaneous multiple collisions arise in the limit as the total population size tends to infinity. The paper finishes in Section 7 with an analog analysis of the skewed Kimura model, a model which does not belong to the compound Poisson class, but nevertheless exhibits similarities with the previously studied models.

Throughout the paper the notation $\mathbb{N} := \{1, 2, \ldots\}$, $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$, and $[N] := \{1, \ldots, N\}$, $N \in \mathbb{N}$, is used. We furthermore use, for fixed $N \in \mathbb{N}$, the symbol $S := \{0, \ldots, N\}$ for the state space of several discrete processes considered in this paper. Moreover, for $k \in \mathbb{N}_0$ and $x \in \mathbb{R}$, $(x)_k := x(x-1)\cdots(x-k+1)$ and $[x]_k := x(x+1)\cdots(x+k-1)$ denote the descending and ascending factorials respectively, with the convention that $(x)_0 = [x]_0 = 1$.

2. Skewed Cannings models

Consider a population with constant population size $N \in \mathbb{N}$ evolving in discrete non-overlapping generations $t \in \mathbb{N}_0$. The nth individual, $n \in [N]$, of the tth generation gives birth to a random number $\mu_n(t)$ of children. Since the population is assumed to have constant population size N, for each fixed generation $t \in \mathbb{N}$, the random vector $\mu(t) := (\mu_1(t), \dots, \mu_N(t))$ must take values in the discrete N-simplex $\Delta(N)$ consisting of all $k = (k_1, \dots, k_N) \in \mathbb{N}_0^N$ satisfying $k_1 + \cdots + k_N = N$. Note that the random variable $\mu_n(t) = \mu_{n,N}(t)$ is allowed to depend on the total population size N. However, for simplicity, this dependence on N is usually not indicated throughout the paper. It is assumed that the model is time-homogeneous in the sense that the random vectors $\mu(0), \mu(1), \dots$ are independent and identically distributed. We write $\mu_n := \mu_n(0)$, $n \in [N]$, and $\mu := \mu(0)$ for convenience. Note that the model is in principle defined in the same way as an exchangeable Cannings model [4, 5]. The only generalization is that, for each fixed generation $t \in \mathbb{N}_0$, the offspring variables $\mu_1(t), \dots, \mu_N(t)$ are not necessarily assumed to be exchangeable. To the best of our knowledge [17] is the only reference dealing with reproduction models of this form even in the time-inhomogeneous context in which the total population size is not necessarily constant equal to N. Particular examples, such as the skewed Wright-Fisher model, which is studied in Section 5 in more detail, have been the source of recurrent interest in the literature (see, for example, [3] and [16]). A main tool exploited in this section is the following shuffling procedure. For each fixed generation $t \in \mathbb{N}_0$ let $\nu(t) = (\nu_1(t), \dots, \nu_N(t))$ be a random permutation (shuffling) of $\mu(t) = (\mu_1(t), \dots, \mu_N(t))$. Then $\nu(t)$ is exchangeable with distribution

$$\mathbb{P}(\nu(t) = k) = \frac{1}{N!} \sum_{\pi} \mathbb{P}(\pi\mu(t) = k), \qquad k \in \Delta(N), \tag{1}$$

where $\pi\mu(t) := (\mu_{\pi 1}(t), \dots, \mu_{\pi N}(t))$ and the sum extends over all permutations π of $[N]$. We interpret $\nu_n(t)$ as the number of offspring of the nth individual of the tth generation in an exchangeable Cannings model [4, 5]. As we will see soon, this exchangeable Cannings population model will turn out to be very helpful to analyze the original skewed Cannings model. Again we write $\nu_n := \nu_n(0)$, $n \in [N]$, and $\nu := \nu(0)$ for convenience.

Take a sample of $n \in S$ individuals of generation 0 and, for $t \in \mathbb{N}_0$, let X_t denote the number of descendants of these n individuals in generation t. The process $X := (X_t)_{t \in \mathbb{N}_0}$ is called the forward process.

Lemma 2.1. (forward structure) *Fix $N \in \mathbb{N}$. The forward process $X = (X_t)_{t \in \mathbb{N}_0}$ is a time-homogeneous Markov chain with state space S and initial state $X_0 = n$. The transition probabilities $P_{i,j} := \mathbb{P}(X_{t+1} = j \,|\, X_t = i)$, $i, j \in S$, do not depend on the initial state n and are given by*

$$P_{i,j} \;=\; \frac{1}{\binom{N}{i}} \sum_{1 \le n_1 < \cdots < n_i \le N} \mathbb{P}\Big(\sum_{k=1}^{i} \mu_{n_k} = j \Big), \qquad i, j \in S. \tag{2}$$

The states 0 and N are absorbing. Moreover, X is a bounded martingale and, hence, X_t converges almost surely as $t \to \infty$ to a limiting random variable X_∞. If $\mathbb{P}(\mu_k = 1) < 1$ for some $k \in \{1, \dots, N\}$, then X_∞ takes the two values 0 and N with probability $\mathbb{P}(X_\infty = 0) = 1 - n/N$ and $\mathbb{P}(X_\infty = N) = n/N$ respectively.

Proof. Fix $i,j \in S$. The transition probability $P_{i,j}$ must coincide with the corresponding transition probability of the associated exchangeable Cannings model with offspring distributions (1), since in the considered skewed model all assignments of offspring to parents are assumed to be equally likely. From the literature on exchangeable models (Cannings [4, p. 267]), it follows that $P_{i,j} = \mathbb{P}(\nu_1+\cdots+\nu_i = j) = \sum_k \mathbb{P}(\nu = k)$, where the sum extends over all $k \in \Delta(N)$ satisfying $k_1+\cdots+k_i = j$. Plugging in (1) and interchanging the two sums involved yields

$$P_{i,j} \;=\; \frac{1}{N!}\sum_{\pi}\sum_{k}\mathbb{P}(\pi\mu = k) \;=\; \frac{1}{N!}\sum_{\pi}\mathbb{P}(\mu_{\pi 1}+\cdots+\mu_{\pi i} = j).$$

Introducing $n_1 := \pi 1, \ldots, n_i := \pi i$, and noting that there exist exactly $(N-i)!$ permutations π leaving $\pi 1, \ldots, \pi i$ fixed, it follows that

$$P_{i,j} = \frac{(N-i)!}{N!}\sum_{\substack{n_1,\ldots,n_i=1\\ \text{all distinct}}}^{N}\mathbb{P}\Big(\sum_{k=1}^{i}\mu_{n_k} = j\Big) = \frac{1}{\binom{N}{i}}\sum_{1\le n_1<\cdots<n_i\le N}\mathbb{P}\Big(\sum_{k=1}^{i}\mu_{n_k} = j\Big).$$

The chain X is a martingale, since $\mathbb{E}(X_{t+1}\,|\,X_t = i) = \sum_{j\in S} jP_{i,j} = \sum_{j\in S} j\mathbb{P}(\nu_1 + \cdots + \nu_i = j) = \mathbb{E}(\nu_1+\cdots+\nu_i) = i$, $i \in S$. Since X is bounded ($0 \le X_t \le N$ for all $t \in \mathbb{N}_0$), X_t converges almost surely to some random variable X_∞ as $t \to \infty$, and $(X_t)_{t\in\mathbb{N}_0\cup\{\infty\}}$ is still a martingale. For the rest of the proof see, for example, Section 2.1 in [18]. □

Remarks. 1. One may rewrite (2) in condensed form as

$$P_{i,j} \;=\; \frac{1}{\binom{N}{i}}\sum_{M}\mathbb{P}\Big(\sum_{m\in M}\mu_m = j\Big), \qquad i,j \in S,$$

where the sum $\sum_M$ extends over all subsets M of $[N]$ satisfying $|M| = i$.

2. The associated shuffled Cannings model with exchangeable offspring variables $\nu_1, \ldots, \nu_N$ introduced in the proof of Lemma 2.1 is useful in many respects. For instance, in terms of the so-called coalescence probability (see, for example, [19])

$$c_N \;:=\; \frac{\mathbb{E}((\nu_1)_2)}{N-1} \;=\; \frac{\mathrm{Var}(\nu_1)}{N-1} \;=\; 1-\mathbb{E}(\nu_1\nu_2) \;=\; -\mathrm{Cov}(\nu_1,\nu_2), \tag{3}$$

the variance of X_{t+1}, given $X_t = i$, can be expressed as

$$\begin{aligned}\mathrm{Var}(X_{t+1}^2\,|\,X_t = i) &= \mathrm{Var}((\nu_1+\cdots+\nu_i)^2)\\ &= i\mathrm{Var}(\nu_1) + i(i-1)\mathrm{Cov}(\nu_1,\nu_2)\\ &= i(N-1)c_N - i(i-1)c_N \;=\; i(N-i)c_N.\end{aligned}$$

Defining the heterozygosity of the population at generation t as $2(X_t/N)(1-X_t/N)$, we have

$$2\mathbb{E}\Big(\frac{X_t}{N}\Big(1-\frac{X_t}{N}\Big)\Big|X_0 = i\Big) \;=\; 2(1-c_N)^t\frac{i}{N}\Big(1-\frac{i}{N}\Big), \qquad t \in \mathbb{N}_0, i \in S,$$

showing that, if $c_N > 0$, the mean heterozygosity tends to 0 exponentially fast as $t \to \infty$. We will soon provide further information on c_N when the model is studied backwards in time.

3. Suppose that $\mathbb{P}(\mu_k = 1) < 1$ for some $k \in \{1, \ldots, N\}$. For $j \in S$ let $\tau_{n,j}$ denote the first time the process X hits the state j given that it starts in $X_0 = n$. The first time X hits the boundaries 0 or N, which is $\tau_n := \min(\tau_{n,0}, \tau_{n,N})$, is finite

with probability 1. Let $q_{n,N} := \mathbb{P}(\tau_{n,0} < \tau_{n,N})$ denote the extinction probability. The previous Lemma shows that $q_{n,N} = \mathbb{P}(\tau_{n,0} < \tau_{n,N}) = \mathbb{P}(X_\infty = 0) = 1 - n/N$, $n \in \{0, \dots, N\}$.

Instead of looking forwards in time let us now look backwards in time and count, starting with all N individuals from some generation $t_0 \in \mathbb{N}_0$, the number of ancestors of these N individuals $t \in \{0, \dots, t_0\}$ generations backward in time. More precisely, let $\widehat{X}_t$ denote the number of ancestors of the N individuals of generation t_0 in generation $t_0 - t$. The process $\widehat{X} := (\widehat{X}_t)_{t\in\{0,\dots,t_0\}}$ is called the backward process.

Lemma 2.2. (backward structure) *The backward process $\widehat{X} = (\widehat{X}_t)_{t\in\{0,\dots,t_0\}}$ is a time-homogeneous Markov chain with state space S and initial state N. The transition probabilities $\widehat{P}_{i,j} := \mathbb{P}(\widehat{X}_{t+1} = j \,|\, \widehat{X}_t = i)$, $i, j \in S$, are given by*

$$\widehat{P}_{i,j} \;=\; \frac{1}{\binom{N}{i}} \sum_{1\le n_1<\dots<n_j\le N} \sum_{\substack{l_1,\dots,l_j\in\mathbb{N}\\ l_1+\dots+l_j=i}} \mathbb{E}\Big(\prod_{k=1}^{j}\binom{\mu_{n_k}}{l_k}\Big), \qquad i,j\in S, \tag{4}$$

with the convention that $\widehat{P}_{i,0} = \delta_{i0}$ (Kronecker symbol). The states 0 and 1 are absorbing.

Proof. As in the proof of Lemma 2.1 we make use of the associated exchangeable Cannings model. The backward transition probability $\widehat{P}_{i,j}$ coincides with the analog backward transition probability of the associated Cannings model with exchangeable offspring vector ν having distribution (1). Therefore, from the literature on exchangeable models (Cannings [4, Theorem 11], Gladstien [9, Examples]), it follows that

$$\widehat{P}_{i,j} \;=\; \frac{\binom{N}{j}}{\binom{N}{i}} \sum_{\substack{l_1,\dots,l_j\in\mathbb{N}\\ l_1+\dots+l_j=i}} \mathbb{E}\Big(\prod_{k=1}^{j}\binom{\nu_k}{l_k}\Big), \qquad i,j\in S,$$

with the convention that $\widehat{P}_{i,0} = \delta_{i0}$. Using (1), a similar argument as given in the proof of Lemma 2.1 for $P_{i,j}$ shows that $\widehat{P}_{i,j}$ can be expressed in terms of the original offspring vector μ via (4). □

Remark. (eigenvalues, duality) The stochastic matrix $\widehat{P} = (\widehat{P}_{i,j})_{i,j\in S}$ is lower left triangular and has hence eigenvalues $\lambda_i := \widehat{P}_{i,i}$, $i \in S$. From Lemma 2.2 it follows that $\lambda_0 = 1$ and that

$$\lambda_i \;=\; \widehat{P}_{i,i} \;=\; \frac{1}{\binom{N}{i}} \sum_{1\le n_1<\dots<n_i\le N} \mathbb{E}(\mu_{n_1}\cdots\mu_{n_i}), \qquad i\in\{1,\dots,N\}.$$

Note that $\lambda_1 = 1$ and that $\lambda_2 = ((N)_2)^{-1}\sum_{1\le n_1<n_2\le N}\mathbb{E}(\mu_{n_1}\mu_{n_2}) = \mathbb{E}(\nu_1\nu_2) = 1-c_N$. Since the forward and backward transition matrices $P = (P_{i,j})_{i,j\in S}$ and $\widehat{P} = (\widehat{P}_{i,j})_{i,j\in S}$ coincide with those of the associated Cannings model with exchangeable offspring vector ν, it is allowed to apply results on exchangeable Cannings models, in particular duality results such as the duality relation $PH = H\widehat{P}'$, where $H = (H_{ij})_{i,j\in S}$ is (see, for example, [18]) the matrix with entries $H_{ij} = \binom{i}{j}/\binom{N}{j}$, $i, j \in S$.

Since H is non-singular, P has the same eigenvalues as $\widehat{P}$. In particular, $c_N = 1-\lambda_2$ is the spectral gap of P and $\widehat{P}$.

The random variable $\widehat{X}_t$ counts the number of ancestors in generation $t_0 - t$, but it gives no information about whether two individuals i and j, randomly picked from generation t_0, share a common parent in generation $t_0 - t$. In order to encode this information, a more enriched ancestral process has to be considered, which is now described. Take a random sample of $n \in [N]$ individuals from generation $t_0 \in \mathbb{N}_0$ and, for $t \in \{0, \dots, t_0\}$, define a random relation $\mathcal{R}_t$ on $\{1, \dots, n\}$ by saying that $(i, j) \in \mathcal{R}_t$ if and only if the individuals i and j have a common parent in generation $t_0 - t$. Note that $\mathcal{R}_t = {}_{(N)}\mathcal{R}_t^{(n)}$ depends on the sample size n and on the total population size N. The process $(\mathcal{R}_t)_{t\in\{0,\dots,t_0\}}$ is called the ancestral process or a discrete coalescent process. It is well known (see, for example, [17]) that $(\mathcal{R}_t)_{t\in\{0,\dots,t_0\}}$ is a Markov chain with state space $\mathcal{E}_n$, the set of equivalence relations (partitions) on $\{1, \dots, n\}$, and transition probabilities

$$\mathbb{P}(\mathcal{R}_{t+1} = \eta \,|\, \mathcal{R}_t = \xi) \;=\; \Phi_j(l_1, \dots, l_j), \qquad \xi, \eta \in \mathcal{E}_n, \xi \subseteq \eta, \tag{5}$$

where

$$\Phi_j(l_1, \dots, l_j) \;:=\; \frac{1}{(N)_{l_1+\cdots+l_j}} \sum_{\substack{n_1,\dots,n_j=1 \\ \text{all distinct}}}^{N} \mathbb{E}((\mu_{n_1})_{l_1} \cdots (\mu_{n_j})_{l_j}). \tag{6}$$

Here $j := |\eta|$ denotes the number of equivalence classes (blocks) of η and $l_1, \dots, l_j \in \mathbb{N}$ are the group sizes of merging classes of ξ. Note that $l_1 + \cdots + l_j$ is the number of classes (blocks) of ξ. Comparing (6) with (4) shows that

$$\widehat{P}_{i,j} \;=\; \frac{i!}{j!} \sum_{\substack{l_1,\dots,l_j\in\mathbb{N} \\ l_1+\cdots+l_j=i}} \frac{\Phi_j(l_1, \dots, l_j)}{l_1! \cdots l_j!}, \qquad i, j \in \{1, \dots, N\}. \tag{7}$$

For $j = 1$, (4) and (7) reduce to

$$\widehat{P}_{i,1} \;=\; \Phi_1(i) \;=\; \frac{1}{(N)_i} \sum_{n=1}^{N} \mathbb{E}((\mu_n)_i), \qquad i \in \{1, \dots, N\}.$$

The so-called coalescence probability c_N, i.e. the probability that two individuals, randomly chosen from some generation, have a common parent, is hence

$$c_N \;:=\; \widehat{P}_{2,1} \;=\; \Phi_1(2) \;=\; \frac{1}{(N)_2} \sum_{n=1}^{N} \mathbb{E}((\mu_n)_2),$$

in agreement with (3), and the effective population size is $N_e := 1/c_N$. We will later also make use of the probability that three individuals, randomly chosen from some generation, share a common parent, which is given by

$$d_N \;:=\; \widehat{P}_{3,1} \;=\; \Phi_1(3) \;=\; \frac{1}{(N)_3} \sum_{n=1}^{N} \mathbb{E}((\mu_n)_3).$$

The transition probabilities (5) do not depend on t and t_0. It is hence allowed to choose t_0 arbitrary large. We can therefore think of a process $(\mathcal{R}_t)_{t\in\mathbb{N}_0}$ with transition probabilities (5) and time $t \in \mathbb{N}_0$. It is well known (see [21]) that, if $d_N/c_N \to 0$, then, for each sample size $n \in \mathbb{N}$, the time-scaled process $(\mathcal{R}^{(n)}_{[t/c_N]})_{t\in[0,\infty)}$ converges

weakly to Kingman's n-coalescent $(R_t^{(n)})_{t\in[0,\infty)}$ as the total population size N tends to infinity. In [21] there is also a criterion in terms of the quantities (6) provided ensuring that, for each $n \in \mathbb{N}$, the time-scaled process $(\mathcal{R}^{(n)}_{[t/c_N]})_{t\in[0,\infty)}$ converges weakly as $N \to \infty$ to a more general process $(\varrho_n R_t)_{t\in[0,\infty)}$, where ϱ_n denotes the restriction from $\mathcal{E}$, the set of all equivalence relations on $\mathbb{N}$, to $\mathcal{E}_n$, and $R = (R_t)_{t\geq 0}$ is a continuous-time coalescent process allowing for simultaneous multiple collisions of ancestral lineages. This asymptotic behavior can only occur if $c_N \to 0$. If, instead, c_N converges to a positive constant, then, a similar criterion for the quantities (6), also provided in [21], ensures that, for each sample size $n \in \mathbb{N}$, the process $(\mathcal{R}_t^{(n)})_{t\in\mathbb{N}_0}$ (without any time-scaling involved) converges weakly to a discrete-time process $(\varrho_n R_t)_{t\in\mathbb{N}_0}$, where $(R_t)_{t\in[0,\infty)}$ is a discrete-time coalescent allowing for simultaneous multiple collisions of ancestral lineages.

Definition 2.3. *a) Let $R = (R_t)_{t\in[0,\infty)}$ be a continuous-time coalescent with simultaneous multiple collisions. We say that the considered population model is in the domain of attraction of R, if, for each sample size $n \in \mathbb{N}$, the time-scaled ancestral process $(\mathcal{R}^{(n)}_{[t/c_N]})_{t\in[0,\infty)}$ converges weakly to $(\varrho_n R_t)_{t\in[0,\infty)}$ as $N \to \infty$.*

b) Let $R = (R_t)_{t\in\mathbb{N}_0}$ be a discrete-time coalescent with simultaneous multiple collisions. We say that the considered population model is in the domain of attraction of R, if, for each sample size $n \in \mathbb{N}$, the ancestral process $(\mathcal{R}_t^{(n)})_{t\in\mathbb{N}_0}$ converges weakly to $(\varrho_n R_t)_{t\in\mathbb{N}_0}$ as $N \to \infty$.

In both cases we call the coalescent R the attractor of the considered population model.

Example 2.4. (extended Moran model) Let U be a random variable taking values in $\{0,\dots,N\}$ and define the offspring vector $\mu = (\mu_1,\dots,\mu_N)$ via $\mu_n := 1$ for $n \in \{1,\dots,N-U\}$, $\mu_n := 0$ for $n \in \{N-U+1,\dots,N-1\}$, and $\mu_N := U$. The associated shuffled exchangeable Cannings model was considered by Eldon and Wakeley [7]. For $U \equiv 0$, we obtain the trivial model in which each individual has exactly one offspring ($\mu_n = 1$ for all $n \in [N]$). For $U \equiv N$ we obtain a simple model for which the Nth individual is the parent of all the N children of the next generation ($\mu_N = N$). From (2), it follows that the forward chain X has transition probabilities

$$
\begin{aligned}
&P_{i,j}\\
&= \frac{1}{\binom{N}{i}}\Bigg(\sum_{1\leq n_1<\dots<n_i<N}\mathbb{P}\Big(\sum_{k=1}^{i}\mu_{n_k}=j\Big)+\sum_{1\leq n_1<\dots<n_i=N}\mathbb{P}\Big(\sum_{k=1}^{i-1}\mu_{n_k}=j-U\Big)\Bigg)\\
&= \frac{1}{\binom{N}{i}}\mathbb{E}\Bigg(\binom{N-U}{j}\binom{U-1}{i-j}+\binom{N-U}{j-U}\binom{U-1}{(i-1)-(j-U)}\Bigg)\\
&= \begin{cases}
\frac{1}{\binom{N}{i}}\mathbb{E}\Big(\binom{N-U}{j}\binom{U-1}{i-j}\Big) & \text{if } j<i,\\
\frac{1}{\binom{N}{i}}\mathbb{E}\Big(\binom{N-U}{j}+\binom{N-U}{N-j}\Big) \;=\; \mathbb{E}\Big(\frac{(i)_U+(N-i)_U}{(N)_U}\Big) & \text{if } j=i,\\
\frac{1}{\binom{N}{i}}\mathbb{E}\Big(\binom{N-U}{N-j}\binom{U-1}{j-i}\Big) & \text{if } j>i.
\end{cases}
\end{aligned}
$$

For $U \equiv 2$ this model reduces to the standard skewed Moran model with forward transition probabilities $P_{i,i-1} = i(N-i)/(N(N-1))$, $i \in \{1,\dots,N\}$, $P_{i,i+1} = i(N-i)/(N(N-1))$, $i \in \{0,\dots,N-1\}$, $P_{i,i} = 1-2i(N-i)/(N(N-1))$, $i \in \{0,\dots,N\}$, and $P_{i,j} = 0$ otherwise. From (4), it follows similarly that, for $i,j \in \{1,\dots,N\}$,

$$\widehat{P}_{i,j} \;=\; \begin{cases} \dfrac{\mathbb{E}\left(\binom{N-U}{j-1}\binom{U}{i-j+1}\right)}{\binom{N}{i}} & \text{if } j<i, \\ \dfrac{\mathbb{E}\left(\binom{N-U}{i}+U\binom{N-U}{i-1}\right)}{\binom{N}{i}} & \text{if } j=i, \\ 0 & \text{if } j>i. \end{cases}$$

Note that $\widehat{P}_{i,1} = \mathbb{E}((U)_i)/(N)_i$, $i \in \{2,\dots,N\}$. In particular, the coalescence probability is $c_N = \widehat{P}_{2,1} = \mathbb{E}(U(U-1))/(N(N-1))$, in agreement with [7, Equation (2)]. For example, if U is binomially distributed with parameters N and $p \in [0,1]$, then $c_N = p^2$. Note that $p = p_N$ may depend on N, so this model can have a wide variety of effective population sizes $N_e = 1/p_N^2$. For instance, if $p_N = N^{-\alpha}$, $\alpha > 0$, then $N_e = N^{2\alpha}$ is sub-linear for $\alpha < 1/2$ and super-linear for $\alpha > 1/2$. If $p_N = \lambda^N$, $\lambda < 1$, then $N_e = \lambda^{-2N}$ grows exponentially. We will come back to this extended Moran model in the following section.

In the following section we will introduce a skewed conditional branching process model, which can be viewed as a particular skewed Cannings model. We will later identify the attractor R of several concrete such population models.

3. Skewed conditional branching process models

Let $\xi_1, \xi_2, \dots$ be independent non-negative integer valued random variables and let f_n denote the probability generating function (pgf) of ξ_n, $n \in \mathbb{N}$. For any pgf g and $l \in \mathbb{N}_0$ we use the standard notation g^l for the lth power of g ($g^0 = 1$) and the notation $g^{(l)}$ for the lth derivative of g. Moreover, $[x^l]g(x)$ denotes the coefficient in front of x^l in the Taylor expansion of g around zero. For $N \in \mathbb{N}$ assume that $\mathbb{P}(\xi_1 + \cdots + \xi_N = N) > 0$ and let $\mu = (\mu_1,\dots,\mu_N)$ be a random vector with distribution

$$\mathbb{P}(\mu = k) \;:=\; \frac{\mathbb{P}(\xi_1 = k_1)\cdots P(\xi_N = k_N)}{\mathbb{P}(\xi_1 + \cdots + \xi_N = N)} \;=\; \frac{[x^{k_1}]f_1(x)\cdots[x^{k_N}]f_N(x)}{[x^N](f_1\cdots f_N)(x)},$$

$k = (k_1,\dots,k_N) \in \Delta(N)$. The distribution of μ is hence that of $(\xi_1,\dots,\xi_N)$ conditioned on the event that $\xi_1 + \cdots + \xi_N = N$. Note that, for $n \in [N]$, the marginal variable μ_n has distribution

$$\begin{aligned} \mathbb{P}(\mu_n = k) &= \frac{\mathbb{P}(\xi_n = k)\mathbb{P}(\sum_{m\in[N]\setminus\{n\}}\xi_m = N-k)}{\mathbb{P}(\xi_1+\cdots+\xi_N = N)} \\ &= \frac{([x^k]f_n(x))([x^{N-k}](\prod_{m\in[N]\setminus\{n\}} f_m(x)))}{[x^N](f_1\cdots f_N)(x)}, \qquad k \in \{0,\dots,N\}. \end{aligned}$$

For each $n \in [N]$ one may interpret μ_n as the number of offspring of individual n in a population with non-overlapping generations of constant population size N. Note that μ has pgf

$$\mathbb{E}(z_1^{\mu_1}\cdots z_N^{\mu_N}) \;=\; \frac{[x^N](f_1(xz_1)\cdots f_N(xz_N))}{[x^N](f_1\cdots f_N)(x)}, \qquad |z_1|,\dots,|z_N| \le 1. \tag{8}$$

Let $l_1, \ldots, l_N \in \mathbb{N}_0$ and put $l := l_1 + \cdots + l_N$. Applying the 'derivative operator' $(\partial^{l_1}/\partial^{l_1} z_1) \cdots (\partial^{l_N}/\partial^{l_N} z_N)$ to the left-hand side and the right-hand side of (8), and noting that it is allowed to interchange this derivative operator with the 'coefficient operator' $[x^N]$ in the numerator on the right-hand side of (8), it follows that

$$\mathbb{E}((\mu_1)_{l_1} z_1^{\mu_1 - l_1} \cdots (\mu_N)_{l_N} z_N^{\mu_N - l_N}) \;=\; \frac{[x^{N-l}](\prod_{i=1}^N f_i^{(l_i)}(x z_i))}{[x^N](f_1 \cdots f_N)(x)}.$$

Taking the limit $z_1 \nearrow 1, \ldots, z_N \nearrow 1$ shows that μ has descending factorial moments

$$\mathbb{E}((\mu_1)_{l_1} \cdots (\mu_N)_{l_N}) \;=\; \frac{[x^{N-l}](\prod_{i=1}^N f_i^{(l_i)}(x))}{[x^N](f_1 \cdots f_N)(x)}, \quad l_1, \ldots, l_N \in \mathbb{N}_0. \tag{9}$$

If the random variables $\xi_1, \xi_2, \ldots$ are identically distributed, then the model reduces to the conditional branching process model first introduced by Moran and Watterson [23] and further investigated for example by Karlin and McGregor [11]. In this case, for each fixed $N \in \mathbb{N}$, the random variables $\mu_1, \ldots, \mu_N$ are exchangeable. Note however, that in general it is not assumed here that the random variables $\xi_1, \xi_2, \ldots$ are identically distributed. In this sense the model is a bit more general than the conditional branching process model of [23]. Particular classes and examples of the model, which to the best of our knowledge are new or only briefly mentioned in the literature, are introduced and analyzed in the following sections, among them the skewed Wright-Fisher model (Section 5), the skewed Dirichlet model (Section 6), and the skewed Kimura model (Section 7). One may think that, at least for fixed N, the model can be reduced to a simpler model by randomly permutating the random variables $\xi_1, \ldots, \xi_N$. However, this is not the case. If $\eta_1, \ldots, \eta_N$ denotes a random permutation of $\xi_1, \ldots, \xi_N$, then the random variables $\eta_1, \ldots, \eta_N$ are identically distributed but in general not independent (not even uncorrelated) anymore. In this sense the model is indeed more general than the conditional branching process model of Karlin and McGregor.

On the other hand, for arbitrary but fixed $N \in \mathbb{N}$, the model is well known from the literature in the following sense. Fix $N \in \mathbb{N}$ and let $\nu = (\nu_1, \ldots, \nu_N)$ be a random permutation of $\mu = (\mu_1, \ldots, \mu_N)$. As explained in Section 2, the model can be interpreted as an exchangeable Cannings model with population size N and offspring vector ν. Therefore, essentially all results known for exchangeable Cannings models apply to our model, which simplifies our further analysis significantly. For example, as explained in Section 2, the transition matrix of the forward process has entries (2). Noting that, for any subset M of $[N]$,

$$\begin{aligned} \mathbb{P}(\sum_{m \in M} \mu_m = j) \;&=\; \frac{\mathbb{P}(\sum_{m \in M} \xi_m = j)\, \mathbb{P}(\sum_{m \in [N] \setminus M} \xi_m = N - j)}{\mathbb{P}(\xi_1 + \cdots + \xi_N = N)} \\ &=\; \frac{([x^j](\prod_{m \in M} f_m)(x))([x^{N-j}](\prod_{m \in [N] \setminus M} f_m)(x))}{[x^N](f_1 \cdots f_N)(x)}, \end{aligned}$$

it follows that the forward transition matrix $P = (P_{i,j})_{i,j \in S}$ has entries

$$P_{i,j} \;=\; \frac{1}{\binom{N}{i}} \sum_M \frac{([x^j](\prod_{m \in M} f_m)(x))\,([x^{N-j}](\prod_{m \in [N] \setminus M} f_m)(x))}{[x^N](f_1 \cdots f_N)(x)}, \quad i, j \in S, \tag{10}$$

where the sum $\sum_M$ extends over all subsets M of $[N]$ satisfying $|M| = i$. We now turn to the backward chain. From (4) and (9), it follows that the backward

transition probabilities are of the form

$$\widehat{P}_{i,j} = \frac{1}{\binom{N}{i}} \sum_{1\le n_1<\cdots<n_j\le N} \sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}} \frac{[x^{N-i}](f_{n_1}^{(l_1)}(x)\cdots f_{n_j}^{(l_j)}(x)\prod_{m\in[N]\setminus\{n_1,\ldots,n_j\}} f_m(x))}{l_1!\cdots l_j![x^N](f_1\cdots f_N)(x)}, \quad i,j\in S, \tag{11}$$

with the convention that $\widehat{P}_{i,0}=\delta_{i0}$, $i\in S$. In particular,

$$\widehat{P}_{i,1} = \Phi_1(i) = \frac{1}{(N)_i}\sum_{n=1}^{N}\frac{[x^{N-i}](f_n^{(i)}(x)\prod_{m\in[N]\setminus\{n\}} f_m(x))}{[x^N](f_1\cdots f_N)(x)}, \qquad i\in S,$$

and the coalescence probability is given by

$$c_N = \widehat{P}_{2,1} = \frac{1}{(N)_2}\sum_{n=1}^{N}\frac{[x^{N-2}](f_n''(x)\prod_{m\in[N]\setminus\{n\}} f_m(x))}{[x^N](f_1\cdots f_N)(x)}.$$

Moreover, the eigenvalues $\lambda_i := \widehat{P}_{i,i} = \Phi_i(1,\ldots,1)$, $i\in S$, of the matrix $\widehat{P}$ are given by $\lambda_0=1$ and

$$\lambda_i = \frac{1}{\binom{N}{i}}\sum_{1\le n_1<\cdots<n_i\le N}\frac{[x^{N-i}](f_{n_1}'(x)\cdots f_{n_i}'(x)\prod_{m\in[N]\setminus\{n_1,\ldots,n_i\}} f_m(x))}{[x^N](f_1\cdots f_N)(x)}$$

for $i\in\{1,\ldots,N\}$.

Remark. One may write $\widehat{P}_{i,j}$ in the form

$$\widehat{P}_{i,j} = \frac{1}{\binom{N}{i}}\frac{1}{[x^N](f_1\cdots f_N)(x)} \sum_{1\le n_1<\cdots<n_j\le N}[x^{N-i}](S_{f_{n_1},\ldots,f_{n_j},x}(i,j)\prod_{m\in[N]\setminus\{n_1,\ldots,n_j\}} f_m(x)),$$

where (see, for example, [20, Eq. (18)])

$$S_{g_1,\ldots,g_j,x}(i,j) := \frac{i!}{j!}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\frac{g_1^{(l_1)}(x)}{l_1!}\cdots\frac{g_j^{(l_j)}(x)}{l_j!}$$

are some sort of generalized Stirling numbers of the second kind, with the convention that, for $j=0$, $S_{g_1,\ldots,g_j,x}(i,j)=\delta_{i0}$, $i\in\mathbb{N}_0$.

Remark. The class of the conditional branching process models covers a wide variety of Cannings models. However (see the following proposition), not all Cannings models are conditional branching process models.

Proposition 3.1. *The extended Moran model (see Example 2.4) is not a conditional branching process model, provided that* $\mathbb{P}(1\le U\le N-1)=1$.

Proof. Suppose that the extended Moran model with population size N is a conditional branching process model. Then, there exist independent and identically distributed random variables $\xi_1,\xi_2,\ldots$ such that $\mathbb{P}(\xi_1+\cdots+\xi_N=N)\neq 0$ and

$$\mathbb{P}(\nu_1=k_1,\ldots,\nu_N=k_N) = \frac{p_{k_1}\cdots p_{k_N}}{\mathbb{P}(\xi_1+\cdots+\xi_N=N)}, \qquad k\in\Delta(N),$$

where $p_k := \mathbb{P}(\xi_1 = k)$, $k \in \mathbb{N}_0$. In particular,

$$\begin{aligned} 0 &\neq \mathbb{E}\Big(\frac{1}{N\binom{N-1}{U-1}}\Big) \\ &= \mathbb{P}(\nu_1 = \cdots = \nu_{N-U} = 1, \nu_{N-U+1} = \cdots = \nu_{N-1} = 0, \nu_N = U) \\ &= \frac{\mathbb{E}(p_0^{U-1} p_1^{N-U} p_U)}{\mathbb{P}(\xi_1 + \cdots + \xi_N = N)}, \end{aligned}$$

and, hence, $p_1 \neq 0$, since $N - U \geq 1$ almost surely by assumption. On the other hand,

$$0 \;=\; \mathbb{P}(\nu_1 = \cdots = \nu_N = 1) \;=\; \frac{\mathbb{P}(\xi_1 = 1) \cdots \mathbb{P}(\xi_N = 1)}{\mathbb{P}(\xi_1 + \cdots + \xi_N = N)} \;=\; \frac{p_1^N}{\mathbb{P}(\xi_1 + \cdots + \xi_N = N)},$$

and, hence, $p_1 = 0$, an obvious contradiction. In particular, for $N \geq 3$, the standard exchangeable Moran model ($U \equiv 2$) is not a conditional branching process model. □

It seems to be hard to derive further exact results or asymptotic results as $N \to \infty$ for the general skewed conditional branching process model. We therefore focus in the following sections on important subclasses.

4. The skewed compound Poisson class

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 $\phi_m \geq 0$, $m \in \mathbb{N}$. We also assume that $\phi_1 > 0$. Let furthermore $\theta_1, \theta_2, \ldots \in (0, \infty)$ be given strictly positive real parameters. In this section it is assumed that, for each $n \in \mathbb{N}$, the random variable ξ_n of the skewed conditional branching process model (as described in the previous section) has pgf

$$f_n(x) \;=\; \mathbb{E}(x^{\xi_n}) \;=\; \exp\bigg(-\theta_n \phi(z) \Big(1 - \frac{\phi(zx)}{\phi(z)}\Big)\bigg), \qquad |x| \leq 1. \tag{12}$$

In (12), z is viewed as a fixed parameter. However, for the following approach we also see z as a variable satisfying $|z| < r$. In order to state the following lemma we need to introduce, for $\theta > 0$, the Taylor expansion

$$\exp(\theta\phi(z)) \;=\; \sum_{k=0}^{\infty} \frac{\sigma_k(\theta)}{k!} z^k, \qquad |z| < r, \tag{13}$$

of $\exp(\theta\phi(z))$, seen as a function of z. Note that the coefficients $\sigma_k(\theta)$ are strictly positive and they depend on the sequence $\phi_\cdot := (\phi_n)_{n\in\mathbb{N}}$. More precisely, the coefficients $\sigma_k(\theta)$ satisfy the recursion $\sigma_0(\theta) = 1$ and

$$\sigma_{k+1}(\theta) \;=\; \theta \sum_{l=0}^{k} \binom{k}{l} \phi_{k-l+1} \sigma_l(\theta), \qquad k \in \mathbb{N}_0, \tag{14}$$

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. Note that, for each fixed $k \in \mathbb{N}$, $\sigma_k(\theta) \sim (\theta\phi_1)^k$ as $\theta \to \infty$.

Proposition 4.1. *If, for each $n \in \mathbb{N}$, the random variable ξ_n has a pgf of the form (12), then the forward process X of the associated skewed conditional branching process model has transition probabilities*

$$P_{i,j} \;=\; \frac{\binom{N}{j}}{\binom{N}{i}} \sum_M \frac{\sigma_j(\sum_{m\in M}\theta_m)\,\sigma_{N-j}(\sum_{m\in[N]\setminus M}\theta_m)}{\sigma_N(\Theta_N)}, \qquad i,j\in S, \tag{15}$$

where $\Theta_N := \theta_1+\cdots+\theta_N$, the sum $\sum_M$ extends over all subsets $M \subseteq [N]$ satisfying $|M| = i$, and the coefficients $\sigma_k(\theta)$ are recursively defined via (14).

Proof. For $j \in [N]$, $\theta > 0$ and $|x| \le 1$, it follows from (13) that

$$[x^j]\exp(\theta\phi(zx)) \;=\; [x^j]\sum_{k=0}^{\infty}\frac{\sigma_k(\theta)}{k!}(zx)^k \;=\; z^j\frac{\sigma_j(\theta)}{j!}.$$

Using, for $M \subseteq [N]$, the shortage $\theta := \sum_{m\in M}\theta_m$, it follows that

$$\begin{aligned}[x^j](\sum_{m\in M} f_m)(x) &= [x^j]\exp\bigg(-\theta\phi(z)\Big(1-\frac{\phi(zx)}{\phi(z)}\Big)\bigg)\\ &= \exp(-\theta\phi(z))[x^j]\exp(\theta\phi(zx)) \;=\; \exp(-\theta\phi(z))z^j\frac{\sigma_j(\theta)}{j!}.\end{aligned}$$

Thus, (15) follows from (10). □

Remark. For the unbiased case, when the parameter $\theta_n = \theta$ does not depend on $n \in \mathbb{N}$, (15) reduces to

$$P_{i,j} \;=\; \binom{N}{j}\frac{\sigma_j(i\theta)\,\sigma_{N-j}((N-i)\theta)}{\sigma_N(N\theta)}, \qquad i,j\in S.$$

Let us now turn to the backward process.

Proposition 4.2. *If, for each $n \in \mathbb{N}$, the random variable ξ_n has a pgf of the form (12), then the backward process $\widehat{X}$ of the associated skewed conditional branching process model has transition probabilities*

$$\widehat{P}_{i,j} \;=\; \frac{i!}{\sigma_i(\Theta_N)}\sum_{1\le n_1<\cdots<n_j\le N}\;\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\frac{\sigma_{l_1}(\theta_{n_1})\cdots\sigma_{l_j}(\theta_{n_j})}{l_1!\cdots l_j!}, \qquad i,j\in S, \tag{16}$$

with the convention that $\widehat{P}_{i,0} = \delta_{i0}$, $i \in S$. Here $\Theta_N := \theta_1 + \cdots + \theta_N$ and the coefficients $\sigma_k(\theta)$ are recursively defined via (14). In particular,

$$\widehat{P}_{i,1} \;=\; \frac{1}{\sigma_i(\Theta_N)}\sum_{n=1}^{N}\sigma_i(\theta_n), \qquad i\in\{1,\ldots,N\}.$$

Proof. From Kolchin's representation formula [13] (see also [25, Chapter 1, Theorem 1.2]), it follows that $\mu = (\mu_1,\ldots,\mu_N)$ has distribution

$$\mathbb{P}(\mu = k) \;=\; \frac{N!}{\sigma_N(\Theta_N)}\prod_{n=1}^{N}\frac{\sigma_{k_n}(\theta_n)}{k_n!}, \qquad k=(k_1,\ldots,k_N)\in\Delta(N),$$

Therefore, μ has joint descending factorial moments

$$\mathbb{E}((\mu_1)_{l_1}\cdots(\mu_N)_{l_N}) \;=\; \frac{(N)_{l_1+\cdots+l_N}}{\sigma_{l_1+\cdots+l_N}(\Theta_N)}\prod_{n=1}^{N}\sigma_{l_n}(\theta_n), \qquad l_1,\ldots,l_N\in\mathbb{N}_0.$$

The probability (6) is therefore of the form

$$\begin{aligned}\Phi_j(l_1,\ldots,l_j) &= \frac{1}{(N)_{l_1+\cdots+l_j}}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\mathbb{E}((\mu_{n_1})_{l_1}\cdots(\mu_{n_j})_{l_j})\\ &= \frac{1}{\sigma_{l_1+\cdots+l_j}(\Theta_N)}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\sigma_{l_1}(\theta_{n_1})\cdots\sigma_{l_j}(\theta_{n_j}), \quad l_1,\ldots,l_j\in\mathbb{N}.\end{aligned}$$

Using (7), it follows that

$$\begin{aligned}\widehat{P}_{i,j} &= \frac{i!}{j!}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\frac{\Phi_j(l_1,\ldots,l_j)}{l_1!\cdots l_j!}\\ &= \frac{i!}{j!\sigma_i(\Theta_N)}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^{N}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\frac{\sigma_{l_1}(\theta_{n_1})\cdots\sigma_{l_j}(\theta_{n_j})}{l_1!\cdots l_j!}\\ &= \frac{i!}{\sigma_i(\Theta_N)}\sum_{1\le n_1<\cdots<n_j\le N}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\frac{\sigma_{l_1}(\theta_{n_1})\cdots\sigma_{l_j}(\theta_{n_j})}{l_1!\cdots l_j!}.\end{aligned}$$

□

Let us now focus on the unbiased case, when all the parameters $\theta_n=\theta$ are equal to some constant $\theta\in(0,\infty)$. It turns out to be convenient to introduce, for any formal series $a(z)=\sum_{k=1}^{\infty}a_kz^k/k!$ or, equivalently, for any sequence $a_\cdot:=(a_1,a_2,\ldots)$, the Bell polynomials

$$B_{i,j}(a_\cdot) \;:=\; \frac{i!}{j!}[z^i](a(z))^j \;=\; \frac{i!}{j!}\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\frac{a_{l_1}\cdots a_{l_j}}{l_1!\cdots l_j!}, \qquad i,j\in\mathbb{N}_0. \tag{17}$$

Note that $B_{0,0}(a_\cdot)=1$, $B_{i,0}(a_\cdot)=B_{0,i}(a_\cdot)=0$, $B_{i,1}(a_\cdot)=a_i$ and $B_{i,i}(a_\cdot)=a_1^i$ for $i\in\mathbb{N}$. For more information on these polynomials we refer the reader to [2] and Chapter 1 of Pitman [25].

The following theorem provides exact and asymptotic formulae for the transition probabilities $\widehat{P}_{i,j}$ and clarifies that the unbiased compound Poisson class is in the domain of attraction of the Kingman coalescent.

Theorem 4.3. *If $\theta_n=\theta\in(0,\infty)$ for all $n\in\mathbb{N}$, then*

$$\widehat{P}_{i,j} \;=\; \frac{(N)_j}{\sigma_i(\theta N)}B_{i,j}(\sigma_\cdot(\theta)) \qquad i,j\in S, \tag{18}$$

where the $B_{i,j}(\sigma_\cdot(\theta))$ are the Bell polynomials of $\sigma_\cdot(\theta) := (\sigma_1(\theta), \sigma_2(\theta), \ldots)$. In particular,

$$\widehat{P}_{i,1} \;=\; \frac{N\sigma_i(\theta)}{\sigma_i(\theta N)}, \qquad i \in \{1, \ldots, N\}. \tag{19}$$

Moreover, for $i, j \in \mathbb{N}$ with $i \geq j$, the asymptotics

$$\widehat{P}_{i,j} \;\sim\; \frac{B_{i,j}(\sigma_\cdot(\theta))}{(\theta\phi_1)^i}\frac{1}{N^{i-j}}, \qquad N \to \infty, \tag{20}$$

holds, and, in the sense of Definition 2.3 a), the model is in the domain of attraction of the Kingman coalescent in the sense of Definition 2.3 a).

Proof. Since $\theta_n = \theta$ does not depend on n, the formula (16) reduces to (18) thanks to the formula (17) for the Bell polynomials. For $j = 1$, (18) reduces to (19), since $B_{i,1}(\sigma_\cdot(\theta)) = \sigma_i(\theta)$. For $i, j \in \mathbb{N}$ with $i \geq j$ and for all $N \geq i$ it follows that

$$\widehat{P}_{i,j} \;=\; \frac{(N)_j}{\sigma_i(\theta N)}B_{i,j}(\sigma_\cdot(\theta)) \;\sim\; \frac{N^j}{(\theta N\phi_1)^i}B_{i,j}(\sigma_\cdot(\theta)) \;=\; \frac{B_{i,j}(\sigma_\cdot(\theta))}{(\theta\phi_1)^i}\frac{1}{N^{i-j}},$$

which is (20). For $j = 1$, Eq. (20) reduces to

$$\widehat{P}_{i,1} \;\sim\; \frac{\sigma_i(\theta)}{(\theta\phi_1)^i}\frac{1}{N^{i-1}}, \qquad N \to \infty.$$

Thus, we obtain

$$\frac{d_N}{c_N} \;=\; \frac{\widehat{P}_{3,1}}{\widehat{P}_{2,1}} \;\sim\; \frac{\sigma_3(\theta)}{\theta\phi_1\sigma_2(\theta)}\frac{1}{N} \;\to\; 0.$$

By [19, Theorem 4 (b)], the model is hence in the domain of attraction of the Kingman coalescent. □

Remark. Theorem 4.3 in particular provides explicit exact formulae for the coalescence probability $c_N = \widehat{P}_{2,1}$ and for $d_N = \widehat{P}_{3,1}$, namely

$$c_N \;=\; \frac{N\sigma_2(\theta)}{\sigma_2(\theta N)} \;=\; \frac{N(\theta\phi_2 + \theta^2\phi_1^2)}{\theta N\phi_2 + (\theta N)^2\phi_1^2} \tag{21}$$

and

$$d_N \;=\; \frac{N\sigma_3(\theta)}{\sigma_3(\theta N)} \;=\; \frac{N(\theta\phi_3 + 3\theta^2\phi_1\phi_2 + \theta^3\phi_1^3)}{\theta N\phi_3 + 3(\theta N)^2\phi_1\phi_2 + (\theta N)^3\phi_1^3}.$$

Note that only the first three coefficients ϕ_1, ϕ_2, and ϕ_3 of the function ϕ are involved here.

Example 4.4. (Wright-Fisher model) For the Wright-Fisher model, $\phi(z) = z$, $\sigma_k(\theta) = \theta^k$. From $B_{i,j}(\theta, \theta^2, \ldots) = \theta^i B_{i,j}(1, 1, \ldots) = \theta^i S(i, j)$, where the $S(i, j)$ are the Stirling numbers of the second kind, it follows that

$$\widehat{P}_{i,j} \;=\; \frac{(N)_j}{\sigma_i(\theta N)}B_{i,j}(\sigma_\cdot(\theta)) \;=\; \frac{(N)_j}{(\theta N)^i}\theta^i S(i, j) \;=\; (N)_j N^{-i} S(i, j),$$

which is well known. For results concerning the skewed Wright-Fisher model we refer the reader to Section 5.

Example 4.5. (Dirichlet model) Suppose that $\phi(z) = -\log(1-z)$, $|z| < 1$, i.e. $\phi_m = (m-1)!$, $m \in \mathbb{N}$. Then, $\sigma_k(\theta) = [\theta]_k$, $k \in \mathbb{N}$, and, hence,

$$\widehat{P}_{i,j} \;=\; \frac{(N)_j}{[\theta N]_i} B_{i,j}([\theta]_\cdot) \;=\; \frac{i!}{[\theta N]_i}\binom{N}{j} \sum_{\substack{l_1,\ldots,l_j \in \mathbb{N} \\ l_1+\cdots+l_j=i}} \frac{[\theta]_{l_1}\cdots[\theta]_{l_j}}{l_1!\cdots l_j!}.$$

Results on the skewed Dirichlet model are presented in Section 6.

We briefly mention two further examples, showing the wide variety of models we are concerned with.

Example 4.6. Let $\alpha \in (0,1]$ and assume that $\phi(z) = 1-(1-z)^\alpha$, $|z| < 1$. Then, $\phi_m = (-1)^{m-1}(\alpha)_m = \alpha[1-\alpha]_{m-1} \geq 0$, $m \in \mathbb{N}$. Note that $\phi(z) \to 1$ as $z \to 1$, a smoothness property of ϕ. For $z \to 1$ the random variable ξ_1 has a discrete stable distribution with pgf $\mathbb{E}(x^{\xi_1}) = \exp(-\theta(1-x)^\alpha)$ (see, for example, Steutel and van Harn [29, Eq. (3.7)]) and tail asymptotics $\mathbb{P}(\xi_1 > y) \sim \theta c_\alpha y^{-\alpha}$ as $y \to \infty$, where $c_\alpha := \sin(\pi\alpha/2)\cos(\pi\alpha/2)\Gamma(\alpha)/(\pi/2)$ (apply, for example, Pitman [24, Theorem 1]). Note that $\phi_1 = \alpha$ and that $\phi_2 = \alpha\beta$ with $\beta := 1-\alpha$. Thus, from (21), it follows that

$$c_N \;=\; \widehat{P}_{2,1} \;=\; \frac{N(\theta\alpha\beta + \theta^2\alpha^2)}{\theta N\alpha\beta + (\theta N)^2\alpha^2} \;\sim\; \left(1 + \frac{\beta}{\theta\alpha}\right)\frac{1}{N}.$$

Thus, $N_e = 1/c_N \sim \rho N$ with $\rho := (1+\beta/(\theta\alpha))^{-1} < 1$.

Example 4.7. Let $\alpha \in (0,\infty)$ and assume that $\phi(z) = (1-z)^{-\alpha} - 1$, $|z| < 1$. Then, $\phi_m = [\alpha]_m$, $m \in \mathbb{N}$. Note that $\phi(z) \to \infty$ as $z \to 1$. The random variable ξ_1 has a compound Poisson distribution with negative binomial jumps and pgf $\mathbb{E}(x^{\xi_1}) = \exp(-\theta\phi(z)(1-\phi(zx)/\phi(z)))$, $|x| \leq 1$. By (21),

$$c_N \;=\; \frac{N(\theta\alpha(\alpha+1) + \theta^2\alpha^2)}{\theta N\alpha(\alpha+1) + (\theta N)^2\alpha^2} \;\sim\; \left(1 + \frac{\alpha+1}{\theta\alpha}\right)\frac{1}{N}.$$

Thus, $N_e \sim \rho N$ with $\rho := (1+(1+\alpha)/(\theta\alpha))^{-1} < 1$.

Theorem 4.3 clarifies that a large class of unbiased conditional branching process models is in the domain of attraction of the Kingman coalescent. For the skewed situation the asymptotical behavior of (16) as $N \to \infty$ is much more involved. In the following sections we focus on particular skewed population models. It will turn out that these models are not necessarily in the domain of attraction of the Kingman coalescent.

5. Skewed Wright-Fisher model

Let $m_1, m_2, \ldots \in (0,\infty)$ be given parameters and assume that ξ_n is Poisson distributed with parameter m_n, $n \in \mathbb{N}$. Since ξ_n has pgf $f_n(x) = \mathbb{E}(x^{\xi_n}) = e^{-m_n(1-x)}$, $n \in \mathbb{N}$, it follows that this model belongs to the skewed compound Poisson class (12) with $\phi(z) := z$ and $\theta_n := m_n/z$. Moreover, from (8) and (9) it follows that μ has a multinomial distribution with parameters N and $s_{1,N}, \ldots, s_{N,N}$, where $s_{n,N} := m_n/(m_1 + \cdots + m_N)$, $n \in \{1,\ldots,N\}$. We shall often drop the index N and use s_n instead of $s_{n,N}$ for notational convenience. Fertile individuals (with large s_n) are viewed as fitter in a much debated propensity interpretation

of fitness, [1]. This fitness of individuals is closer to the Darwinian view of fitness than to the Fisherian view of fitness as a fitness of types. Note that μ has pgf $\mathbb{E}(z_1^{\mu_1}\cdots z_N^{\mu_N}) = (s_1z_1+\cdots+s_Nz_N)^N$, $|z_1|,\ldots,|z_N| \le 1$, and descending factorial moments $\mathbb{E}((\mu)_{l_1}\cdots(\mu_N)_{l_N}) = (N)_l s_1^{l_1}\cdots s_N^{l_N}$, $l_1,\ldots,l_N \in \mathbb{N}_0$, where $l := l_1+\cdots+l_N$. The pgf of the nth marginal μ_n is $\mathbb{E}(z_n^{\mu_n}) = (1-s_n+z_ns_n)^N$, $|z_n| \le 1$, showing that μ_n has a binomial distribution with parameters N and s_n. In particular, $\mathbb{E}(\mu_n) = Ns_n$ and $\sigma_n^2 := \mathrm{Var}(\mu_n) = Ns_n(1-s_n)$, $n \in \{1,\ldots,N\}$. From the above formula for the descending factorial moments of μ, it follows that $\mathbb{E}(\mu_{n_1}\mu_{n_2}) = N(N-1)s_{n_1}s_{n_2}$ for $n_1,n_2 \in \{1,\ldots,N\}$ with $n_1 \ne n_2$. In particular, $\mathrm{Cov}(\mu_{n_1},\mu_{n_2}) = -Ns_{n_1}s_{n_2}$ for $n_1 \ne n_2$.

The expression (10) for the forward transition probabilities $P_{i,j}$ simplifies to

$$P_{i,j} = \frac{1}{\binom{N}{i}} \sum_{1\le n_1<\cdots<n_i\le N} \binom{N}{j} \Big(\sum_{k=1}^i s_{n_k}\Big)^j \Big(1-\sum_{k=1}^i s_{n_k}\Big)^{N-j}, \qquad i,j \in S. \tag{22}$$

Similarly, (11) reduces to

$$\widehat{P}_{i,j} = i! \sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}} \frac{1}{l_1!\cdots l_j!} \sum_{1\le n_1<\cdots<n_j\le N} s_{n_1}^{l_1}\cdots s_{n_j}^{l_j}, \qquad i,j \in S, \tag{23}$$

which, for $j=1$, yields $\widehat{P}_{i,1} = \sum_{n=1}^N s_n^i$, $i \in \{1,\ldots,N\}$. For $i=2$ and $i=3$, we obtain

$$c_N = \widehat{P}_{2,1} = \sum_{n=1}^N s_n^2 \quad \text{and} \quad d_N = \widehat{P}_{3,1} = \sum_{n=1}^N s_n^3.$$

Note that $c_N = \sum_{n=1}^N s_n^2 \ge 1/N$. In particular, $N_e := 1/c_N \le N$, so the effective population size is smaller than or equal to the effective population size in the unbiased case. Applying the Hölder inequality $\sum_n |a_nb_n| \le (\sum_n a_n^2)^{1/2}(\sum_n b_n^2)^{1/2}$ with $a_n := s_n^{1/2}$ and $b_n := s_n^{3/2}$ shows that $c_N \le d_N^{1/2}$, or, equivalently, that $c_N \le d_N/c_N$. On the other hand, since the Euclidian 2-norm is larger than or equal to the Euclidian 3-norm, we obtain $c_N^{1/2} \ge d_N^{1/3}$, or, equivalently, $d_N/c_N \le c_N^{1/2}$. Thus, $c_N \le d_N/c_N \le c_N^{1/2}$ and, as a consequence, $c_N \to 0$ is equivalent to $d_N/c_N \to 0$. Thus, if $c_N \to 0$, then the skewed Wright-Fisher model is in the domain of attraction of the Kingman coalescent as N tends to infinity.

Choosing $i=j$ in (23) shows that the backward matrix $\widehat{P}$ has eigenvalues $\lambda_0 = 1$ and

$$\lambda_i := \widehat{P}_{i,i} = i! \sum_{1\le n_1<\cdots<n_i\le N} s_{n_1}\cdots s_{n_i}, \qquad i \in \{1,\ldots,N\}$$

By duality [18], these are also the eigenvalues of the forward transition matrix P. Note that $1 = \lambda_0 = \lambda_1 > \lambda_2 = 1-c_N > \lambda_2 > \cdots > \lambda_N$. In particular, the eigenvalues $\lambda_1,\ldots,\lambda_N$ are pairwise distinct. Coming back to the random variables $\mu_1,\ldots,\mu_N$ we see that the average covariances of the μ_n's is

$$\frac{1}{\binom{N}{2}} \sum_{1\le n_1<n_2\le N} \mathrm{Cov}(\mu_{n_1},\mu_{n_2}) = -\frac{2}{N-1} \sum_{1\le n_1<n_2\le N} s_{n_1}s_{n_2} = -\frac{1-c_N}{N-1}.$$

Let

$$S_N := \frac{1}{N}\sum_{n=1}^N \frac{\mathbb{E}(\mu_n)}{N}\mu_n = \frac{1}{N}\sum_{n=1}^N s_n\mu_n$$

be the size-biased relative mean offspring fertility. The random variable S_N has pgf

$$\mathbb{E}(z^{S_N}) \;=\; \mathbb{E}((z^{s_n/N})^{\mu_1}\cdots(z^{s_N/N})^{\mu_N}) \;=\; \Big(\sum_{n=1}^{N} s_n z^{s_n/N}\Big)^N,$$

showing that $\mathbb{E}(S_N) = c_N$ and that $\mathrm{Var}(S_N) = (d_N - c_N^2)/N$, which provides an alternative proof of the inequality $d_N \geq c_N^2$.

For the Wright-Fisher model, the moments of X_{t+1}, given $X_t = i$, can be related to the backward probabilities as follows. For $j \in \mathbb{N}$, we have

$$\begin{aligned}\mathbb{E}(X_{t+1}^j \mid X_t = i) &= \frac{1}{\binom{N}{i}} \sum_{1\leq n_1<\cdots<n_i\leq N} \mathbb{E}\Big(\Big(\sum_{k=1}^{i}\mu_{n_k}\Big)^j\Big)\\ &= \frac{1}{\binom{N}{i}} \sum_{l=1}^{j} S(l,j)(N)_l \sum_{1\leq n_1<\cdots<n_i\leq N} \Big(\sum_{k=1}^{i} s_{n_k}\Big)^l.\end{aligned}$$

Using

$$\sum_{1\leq n_1<\cdots<n_i\leq N} \Big(\sum_{k=1}^{i} s_{n_k}\Big)^l \;=\; \sum_{k=1}^{l}\binom{N-k}{i-k}\widehat{P}_{l,k},$$

where $\widehat{P}_{l,k}$ is defined in (23), we get

$$\begin{aligned}\mathbb{E}(X_{t+1}^j \mid X_t = i) &= \frac{1}{\binom{N}{i}} \sum_{l=1}^{j} S(l,j)(N)_l \sum_{k=1}^{l}\binom{N-k}{i-k}\widehat{P}_{l,k}\\ &= \sum_{l=1}^{j} S(l,j)(N)_l \sum_{k=1}^{l}\frac{(i)_k}{(N)_k}\widehat{P}_{l,k},\end{aligned}$$

which is a polynomial of degree j in i.

For the Wright-Fisher model, there are the following two alternative representations of the backward transition probabilities (23).

(i) It can easily be checked that

$$\widehat{P}_{i,j} \;=\; N^{-i}(N)_j S(i,j,Ns), \tag{24}$$

where $s := (s_1,\ldots,s_N)$ and

$$S(i,j,s) \;:=\; \frac{i!}{j!}[x^i]\frac{1}{\binom{N}{j}} \sum_{1\leq n_1<\cdots<n_j\leq N}\prod_{k=1}^{j}(e^{xs_{n_k}}-1) \tag{25}$$

are generalized second kind Stirling numbers. Note that, when $\mathbb{E}(\mu_n) = Ns_n = 1$ for all $n \in [N]$, (25) reduces to the usual second kind Stirling numbers $S(i,j) = (i!/j!)[x^i](e^x-1)^j$ (see [6, vol. I, p. 144]). The unbiased version of (24) is thus $\widehat{P}_{i,j} = N^{-i}(N)_j S(i,j)$, which is well-known. Note that (24) and (25) also yield

$$\widehat{P}_{i,j} \;=\; i![x^i z^j]\prod_{n=1}^{N}(1+z(e^{xs_n}-1))$$

showing that, as shown in [16],

$$\mathbb{E}(z^{\widehat{X}_{t+1}}|\widehat{X}_t = i) \;=\; i![x^i]\prod_{n=1}^{N}(1+z(e^{xs_n}-1)),$$

and for the double pgf

$$\sum_i \frac{x^i}{i!}\mathbb{E}(z^{\widehat{X}_{t+1}}|\widehat{X}_t=i) \;=\; \prod_{n=1}^N (1+z(e^{xs_n}-1)).$$

This leads in particular to $\mathbb{E}(\widehat{X}_{t+1}|\widehat{X}_t=i)=\sum_{n=1}^N(1-(1-s_n)^i)$ and

$$\begin{aligned}&\mathbb{E}\left(\binom{\widehat{X}_{t+1}}{2} \mid \widehat{X}_t=i\right)\\ &\quad= \sum_{n=1}^N (n-1)(1-(1-s_n)^i) - \sum_{1\le n_2<n_1\le N}\big((1-s_{n_1})^i-(1-(s_{n_1}+s_{n_2}))^i\big).\end{aligned}$$

(ii) There is a second obvious representation taking into account repetitions, in the spirit of the Ewens sampling formula [8]. Assume there are a_l individuals at generation t, numbered $1\le n_{1,l}<\cdots<n_{a_l,l}\le N$, producing exactly l offspring, $l\in\{0,\dots,N\}$. Clearly, there is no overlap of the above number sequences for different values of l. Then, using (23)

$$\widehat{P}_{i,j} \;=\; i! \sum_{\substack{\sum_{l=1}^i la_l=i;\\ \sum_{l=1}^i a_l=j}} \sum_{l=1}^i \sum_{n_{1,l}<\cdots<n_{a_l,l}} \prod_{l=1}^i\prod_{k=1}^{a_l}\left(\frac{s_{n_{k,l}}^l}{l!}\right). \tag{26}$$

In (26), there are

$$\prod_{l=1}^i \binom{N-\sum_{k=1}^{l-1}a_k}{a_l} \;=\; \frac{(N)_j}{\prod_{l=1}^i a_l!}$$

sums of the type $\sum_{l=1}^i\sum_{n_{1,l}<\cdots<n_{a_l,l}}1$ so that when $s_n=1/N$ for all $n\in[N]$ (the unbiased case),

$$\widehat{P}_{i,j} \;=\; N^{-i}(N)_j \sum_{\substack{\sum_{l=1}^i la_l=i;\\ \sum_{l=1}^i a_l=j}} \frac{i!}{\prod_{l=1}^i a_l! l!^{a_l}},$$

where the last sum is an alternative representation of the second kind Stirling numbers $S(i,j)$ (see [6, vol. I, p. 145]).

Example 5.1. (power law growth) Fix a constant $\alpha\in\mathbb{R}$ and assume that $m_n:=\mathbb{E}(\xi_n)=n^{-\alpha}$, $n\in\mathbb{N}$. The unbiased case ($m_n=1$ for all $n\in\mathbb{N}$) corresponds to $\alpha=0$. In the following seven ranges for the parameter α are distinguished.

(i) If $\alpha<1/3$, then $M_N:=\sum_{n=1}^N m_n\sim N\int_0^1(Nx)^{-\alpha}dx=N^{1-\alpha}/(1-\alpha)$,

$$c_N \;=\; \frac{1}{M_N^2}\sum_{n=1}^N n^{-2\alpha} \;\sim\; \left(\frac{1-\alpha}{N^{1-\alpha}}\right)^2\frac{N^{1-2\alpha}}{1-2\alpha} \;=\; \frac{(1-\alpha)^2}{(1-2\alpha)N} \tag{27}$$

and, similarly,

$$d_N \;=\; \frac{1}{M_N^3}\sum_{n=1}^N n^{-3\alpha} \;\sim\; \left(\frac{1-\alpha}{N^{1-\alpha}}\right)^3\frac{N^{1-3\alpha}}{1-3\alpha} \;=\; \frac{(1-\alpha)^3}{1-3\alpha}\frac{1}{N^2}.$$

(ii) If $\alpha = 1/3$, then (27) still holds, i.e. $c_N \sim 4/(3N)$, but

$$d_N \;=\; \frac{1}{M_N^3}\sum_{n=1}^{N}\frac{1}{n} \;\sim\; \left(\frac{1-\alpha}{N^{1-\alpha}}\right)^3 \log N \;=\; \frac{8}{27}\frac{\log N}{N^2}.$$

(iii) If $1/3 < \alpha < 1/2$, then (27) still holds, but

$$d_N \;=\; \frac{1}{M_N^3}\sum_{n=1}^{N} n^{-3\alpha} \;\sim\; \left(\frac{1-\alpha}{N^{1-\alpha}}\right)^3 \zeta(3\alpha) \;=\; (1-\alpha)^3\zeta(3\alpha)\frac{1}{N^{3(1-\alpha)}},$$

where ζ denotes the Riemann zeta function.

In all the three cases, i.e. for $\alpha < 1/2$, we have $N_e = 1/c_N \sim \rho N$ with $0 < \rho := (1-2\alpha)/(1-\alpha)^2 \leq 1$. The effective population size is hence asymptotically of a factor ρ smaller than the effective population size N in the unbiased case ($\alpha = 0$).

Moreover, from (23), it follows that the eigenvalues satisfy

$$\lambda_i \;=\; \widehat{P}_{i,i} \;=\; \frac{i!}{M_N^i}\sum_{1\leq n_1<\cdots<n_i\leq N}(n_1\cdots n_i)^{-\alpha} \;\sim\; \left(\frac{1-\alpha}{N^{1-\alpha}}\right)^i i!\sum_{n=i!}^{(N)_i} n^{-\alpha}M_N(n,i),$$

where $M_N(n,i)$ is the number of multiplicative partitions of n into i ordered distinct factors each belonging to $\{1,\ldots,N\}$.

We will now see, that N_e can increase, but of order slower than N.

(iv) When $\alpha = 1/2$, it is readily checked that $M_N \sim 2N^{1/2}$, $c_N \sim (\log N)/(4N)$, and $d_N \sim \zeta(3/2)/(8N^{3/2})$. Thus, $N_e = 1/c_N \sim (4N)/(\log N)$ is asymptotically of a factor $4/(\log N)$ smaller than the standard effective population size N in the unbiased case.

(v) Assume now that $1/2 < \alpha < 1$. Then, still $M_N = \sum_{n=1}^{N} n^{-\alpha} \sim N^{1-\alpha}/(1-\alpha)$,

$$c_N \;=\; \frac{1}{M_N^2}\sum_{n=1}^{N} n^{-2\alpha} \;\sim\; \frac{(1-\alpha)^2}{N^{2(1-\alpha)}}\zeta(2\alpha)$$

and, similarly,

$$d_N \;=\; \frac{1}{M_N^3}\sum_{n=1}^{N} n^{-3\alpha} \;\sim\; \frac{(1-\alpha)^3}{N^{3(1-\alpha)}}\zeta(3\alpha)$$

Note that, with $\rho := 1/((1-\alpha)^2\zeta(2\alpha)) < 1$, $N_e \sim \rho N^{2(1-\alpha)}$ grows algebraically and the order is slower than N.

(vi) For $\alpha = 1$ it is straightforward to check that $M_N \sim \log N$, $c_N \sim \zeta(2)/(\log N)^2$, and $d_N \sim \zeta(3)/(\log N)^3$. Note that $N_e \sim (\log N)^2/\zeta(2)$ grows quite slow (logarithmically).

For all the six cases (i) - (vi) considered so far, i.e. for $\alpha \leq 1$, we have $c_N \to 0$ and $d_N/c_N \to 0$. Thus, in the sense of Definition 2.3 a), the model is in the domain of attraction of the Kingman coalescent. Thus, for each sample size $n \in \mathbb{N}$, the time-scaled backward process $(\mathcal{R}^{(n)}_{[t/c_N]})_{t\in[0,\infty)}$ converges weakly to the Kingman n-coalescent as $N \to \infty$. All that is left about the details of the original discrete fertility model ($m_n = n^{-\alpha}$) is enclosed in the parameter $N_e = 1/c_N \leq N$, where N is the time-scale of the unbiased Wright-Fisher model ($m_n = 1$).

(vii) Assume now that $\alpha > 1$. Then, $M_N = \sum_{n=1}^N n^{-\alpha} \to \zeta(\alpha) =: M$, and,

$$\widehat{P}_{i,1} \;=\; \Phi_1(i) \;=\; \sum_{n=1}^N s_n^i \;=\; \frac{1}{M_N^i} \sum_{n=1}^N n^{-i\alpha} \;\to\; \frac{\zeta(i\alpha)}{(\zeta(\alpha))^i}, \qquad i \in \mathbb{N}.$$

In particular $c_N \to \zeta(2\alpha)/(\zeta(\alpha))^2 > 0$ and $d_N \to \zeta(3\alpha)/(\zeta(\alpha))^3 > 0$. The regime $\alpha > 1$ thus differs significantly from all the previous studied cases, since the coalescence probability converges to a positive constant as $N \to \infty$. More generally, for $l_1, \ldots, l_j \in \mathbb{N}$, we have to analyze the behavior of

$$\Phi_j^{(N)}(l_1, \ldots, l_j) \;=\; \sum_{\substack{n_1,\ldots,n_j=1 \\ \text{all distinct}}}^N s_{n_1,N}^{l_1} \cdots s_{n_j,N}^{l_j}$$

as $N \to \infty$. For each fixed $n \in \mathbb{N}$ we have $s_{n,N} = n^{-\alpha}/M_N \to n^{-\alpha}/M = n^{-\alpha}/\zeta(\alpha) =: p_n$ as $N \to \infty$. Note that $\sum_{n=1}^\infty p_n = 1$. Moreover, $s_{n,N} \le n^{-\alpha}$ uniformly for all N, since $M_N \ge 1$. Thus, by dominated convergence, for each $l_1, \ldots, l_j \in \mathbb{N}$, the limit $\phi_j(l_1, \ldots, l_j) := \lim_{N\to\infty} \Phi_j^{(N)}(l_1, \ldots, l_j)$ exists and is of the form

$$\phi_j(l_1, \ldots, l_j) \;=\; \sum_{\substack{n_1,\ldots,n_j \in \mathbb{N} \\ \text{all distinct}}} p_{n_1}^{l_1} \cdots p_{n_j}^{l_j}. \tag{28}$$

For arbitrary $i, j \in \mathbb{N}$ it therefore follows from (23) that

$$\lim_{N\to\infty} \widehat{P}_{i,j} \;=\; \frac{i!}{j!} \sum_{\substack{l_1,\ldots,l_j \in \mathbb{N} \\ l_1+\cdots+l_j=i}} \frac{\phi_j(l_1, \ldots, l_j)}{l_1! \cdots l_j!}. \tag{29}$$

It is convenient (see, Schweinsberg [27]) to rewrite (28) in integral form as

$$\phi_j(l_1, \ldots, l_j) \;=\; \int_\Delta \sum_{\substack{n_1,\ldots,n_j \in \mathbb{N} \\ \text{all distinct}}} x_{n_1}^{l_1} \cdots x_{n_j}^{l_j} \frac{\Xi(dx)}{(x,x)},$$

where $\Delta := \{(x_1, x_2, \ldots) : x_1 \ge x_2 \ge \cdots \ge 0, \sum_{n=1}^\infty x_n \le 1\}$, $(x,x) := \sum_{n=1}^\infty x_n^2$ for $x \in \Delta$ and the measure Ξ on Δ assigns its total mass $\Xi(\Delta) := (p,p) = \sum_{n=1}^\infty p_n^2 = \zeta(2\alpha)/(\zeta(\alpha))^2$ to the single point $p = (p_1, p_2, \ldots) \in \Delta$. Theorem 2.1 of [21] ensures that the model, without any time-scaling involved, is in the domain of attraction (in the sense of Definition 2.3 b)) of the discrete-time Ξ-coalescent with the measure Ξ as just defined.

Example 5.2. (geometric growth) Fix a constant $\lambda \in (0, \infty)$ and assume that $m_n = \lambda^n$, $n \in \mathbb{N}$. The unbiased case corresponds to $\lambda = 1$, so without loss of generality we assume that $\lambda \ne 1$. Two cases need to be distinguished.

(i) Suppose that $\lambda \in (0,1)$. Then, $M_N = \sum_{n=1}^N \lambda^n = \lambda(1-\lambda^N)/(1-\lambda) \to \lambda/(1-\lambda) =: M$ as $N \to \infty$ and

$$\begin{aligned} \widehat{P}_{i,1} \;&=\; \frac{1}{M_N^i} \sum_{n=1}^N \lambda^{in} \;=\; \left(\frac{1-\lambda}{\lambda(1-\lambda^N)}\right)^i \frac{\lambda^i(1-\lambda^{iN})}{1-\lambda^i} \\ &=\; \frac{(1-\lambda)^i}{1-\lambda^i} \frac{1-\lambda^{iN}}{(1-\lambda^N)^i} \;\sim\; \frac{(1-\lambda)^i}{1-\lambda^i}. \end{aligned} \tag{30}$$

In particular,

$$c_N \;=\; \widehat{P}_{2,1} \;=\; \frac{(1-\lambda)^2}{1-\lambda^2}\frac{1-\lambda^{2N}}{(1-\lambda^N)^2} \;\sim\; \frac{1-\lambda}{1+\lambda}.$$

For arbitrary $i,j\in\mathbb{N}$ it follows similarly as in the previous Example 5.1 (vii) that the limiting formula (29) for $\widehat{P}_{i,j}$ holds, but in the formula (28) for $\phi_j(l_1,\ldots,l_j)$, the parameter p_n has to be replace by $p_n := \lambda^n/M = (1-\lambda)\lambda^{n-1}$, $n\in\mathbb{N}$. For instance, the diagonal entries $\widehat{P}_{i,i}$ of the matrix $\widehat{P}$ satisfy

$$\begin{aligned}\lim_{N\to\infty}\widehat{P}_{i,i} \;&=\; \phi_i(1,\ldots,1) \;=\; \sum_{\substack{n_1,\ldots,n_i\in\mathbb{N}\\ \text{all distinct}}} p_{n_1}\cdots p_{n_i}\\ &=\; \left(\frac{1-\lambda}{\lambda}\right)^i \sum_{\substack{n_1,\ldots,n_i\in\mathbb{N}\\ \text{all distinct}}} \lambda^{n_1+\cdots+n_i} \;=\; \left(\frac{1-\lambda}{\lambda}\right)^i \sum_{n\geq i(i+1)/2} \lambda^n A(n,i),\end{aligned}$$

where $A(n,i)$ is the number of vectors $(n_1,\ldots,n_i)\in\mathbb{N}^i$ with pairwise distinct components satisfying $n_1+\cdots+n_i=n$. This example essentially coincides with Example 5.1 (vii). Again (see [21, Theorem 2.1]) the model is the domain of attraction (in the sense of Definition 2.3 b)) of a discrete-time Ξ-coalescent, where the measure Ξ assigns its total mass $\Xi(\Delta)=(p,p)=(1-\lambda)/(1+\lambda)$ to the single point $p:=(p_1,p_2,\ldots)\in\Delta$.

(ii) Suppose that $\lambda\in(1,\infty)$. Then, $M_N=\sum_{n=1}^N\lambda^n=\lambda(\lambda^N-1)/(\lambda-1)\sim\lambda^{N+1}/(\lambda-1)$. The situation is hence a bit more complicated than in the previous case (i), since M_N does not converge anymore. We have

$$\widehat{P}_{i,1} \;=\; \frac{1}{M_N^i}\sum_{n=1}^N\lambda^{in} \;=\; \frac{1}{M_N^i}\lambda^i\frac{\lambda^{iN}-1}{\lambda^i-1} \;\sim\; \left(\frac{\lambda-1}{\lambda^{N+1}}\right)^i\lambda^i\frac{\lambda^{iN}}{\lambda^i-1} \;\sim\; \frac{(\lambda-1)^i}{\lambda^i-1}.$$

In terms of the inverse parameter $b:=1/\lambda$, this formula takes the form $\widehat{P}_{i,1}\to(1-b)^i/(1-b^i)$, which has the same structure as (30), but with λ replaced by its inverse b. It is hence reasonable to proceed in the same way as in the previous case (i), but with λ replaced by b. Thus, for $n\in\mathbb{N}$, define $p_n:=(1-b)b^{n-1}=(\lambda-1)/\lambda^n$, $n\in\mathbb{N}$. For arbitrary $i,j\in\mathbb{N}$ it follows from (23) that (29) holds with

$$\phi_j(l_1,\ldots,l_j) \;:=\; \lim_{N\to\infty}\frac{1}{M_N^i}\sum_{\substack{n_1,\ldots,n_j=1\\ \text{all distinct}}}^N\lambda^{n_1l_1+\cdots+n_jl_j} \;=\; \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\ \text{all distinct}}} p_{n_1}^{l_1}\cdots p_{n_j}^{l_j}.$$

The last equality is slightly more involved than in Example 5.1 (vii) and Example 5.2 (i), since M_N does not converge. We suggest to check this equality first for $j=2$, which should provide sufficient insight for arbitrary $j\in\mathbb{N}$. We provide in the following the explicit expression for $\phi_2(l_1,l_2)$. Note that $\sum_{n=1}^\infty p_n^l=(1-b)^l/(1-b^l)=(\lambda-1)^l/(\lambda^l-1)$, $l\in\mathbb{N}$. Thus,

$$\begin{aligned}\phi_2(l_1,l_2) \;&=\; \sum_{n_1\neq n_2} p_{n_1}^{l_1}p_{n_2}^{l_2} \;=\; \sum_{n_1=1}^\infty p_{n_1}^{l_1}\sum_{n_2=1}^\infty p_{n_2}^{l_2} - \sum_{n=1}^\infty p_n^{l_1+l_2}\\ &=\; \frac{(\lambda-1)^{l_1}}{\lambda^{l_1}-1}\frac{(\lambda-1)^{l_2}}{\lambda^{l_2}-1} - \frac{(\lambda-1)^{l_1+l_2}}{\lambda^{l_1+l_2}-1}\\ &=\; (\lambda-1)^{l_1+l_2}\left(\frac{1}{(\lambda^{l_1}-1)(\lambda^{l_2}-1)} - \frac{1}{\lambda^{l_1+l_2}-1}\right).\end{aligned}$$

The conclusion is that (28) and (29) are satisfied, so the model is in the domain of attraction of a Ξ-coalescent, where the measure Ξ assigns its total mass $\Xi(\Delta) := (p,p) = \sum_{n=1}^{\infty} p_n^2 = (1-b)/(1+b) = (\lambda-1)/(\lambda+1)$ to the single point $p = (p_1, p_2, \ldots) \in \Delta$.

Remarks. 1. Other particular examples could be investigated, for instance a Wright-Fisher model with oscillating fertilities $m_n := \sin^2 n$, $n \in \mathbb{N}$. For this model, $M_N = \sum_{n=1}^{N} m_n \sim N/2$, $s_n \sim 2(\sin^2 n)/N$,

$$c_N \;=\; \widehat{P}_{2,1} \;=\; \sum_{n=1}^{N} s_n^2 \;\sim\; \frac{4}{N}\int_0^1 \sin^4(Nx)dx \;\sim\; \frac{3}{2N} \;\to\; 0,$$

and $d_N = O(N^{-2})$. Thus, $N_e = \rho N$ with $\rho = 2/3$. This model is in the domain of attraction of the Kingman coalescent. Note that $M_N/N \to 1/2$. This examples therefore belongs to the class of models for which $M_N/N \to m$ for some constant $m \in (0,\infty)$. For skewed Wright-Fisher models having this property, it follows that $(\mu_1, \ldots, \mu_N, 0, 0, \ldots)$ converges weakly to $(\eta_1, \eta_2, \ldots)$, where $\eta_1, \eta_2, \ldots$ are independent random variables and η_n is Poisson distributed with parameter m_n/m, $n \in \mathbb{N}$. The random variables $\mu_1, \ldots, \mu_N$ are therefore asymptotically independent.

2. Assume that $m_n = \lambda^{n^2}$ for some $\lambda > 1$ ((super-exponential growth of fertilities). Then, $M_N = \sum_{n=1}^{N} m_n \sim \lambda^{N^2}$ has the same asymptotic behavior as the single last summand m_N, and it follows that

$$\widehat{P}_{i,1} \;=\; \frac{1}{M_N^i}\sum_{n=1}^{N} m_n^i \;\to\; 1, \qquad i \in \mathbb{N}.$$

In particular, $c_N \to 1$. For $i \in \mathbb{N}$ and $j \in \mathbb{N}\setminus\{1\}$ we conclude that $\widehat{P}_{i,j} \le 1 - \widehat{P}_{i,1} \to 0$. This model is hence in the domain of attraction of the star-shaped coalescent, i.e. the Ξ-coalescent with Ξ being the Dirac-measure at the point $(1,0,0,\ldots) \in \Delta$.

3. One could, for example, also be interested in a more general Wright-Fisher model with multiplicative mixed power-geometric growing fertilities $m_n = n^{-\alpha}\lambda^n$, where $\alpha \in \mathbb{R}$ and $\lambda > 0$. We leave the analysis of such more complicated models for the future or the interested reader.

6. Skewed Dirichlet model

Consider the model where ξ_n has a negative binomial distribution with pgf $f_n(x) = (p/(1-qx))^{a_n}$, where $a_n > 0$, $p \in (0,1)$, $q := 1-p$. Note that $m_n := \mathbb{E}(\xi_n) = (p/q)a_n$, $n \in \mathbb{N}$, and that this model belongs to the skewed compound Poisson class (12) with $\theta_n := a_n$, $\phi(z) := -\log(1-z) = \sum_{m=1}^{\infty} z^m/m$, and $z := q$. In this case μ has the asymmetric Dirichlet multinomial distribution

$$\mathbb{P}(\mu = k) \;=\; \frac{\binom{a_1+k_1-1}{k_1}\cdots\binom{a_N+k_N-1}{k_N}}{\binom{A_N+N-1}{N}} \;=\; \frac{N!}{k_1!\cdots k_N!}\frac{[a_1]_{k_1}\cdots[a_N]_{k_N}}{[A_N]_N},$$

$k = (k_1, \ldots, k_N) \in \Delta(N)$, where $A_N := a_1 + \cdots + a_N$. From (10), it follows that the forward process X has transition probabilities

$$P_{i,j} \;=\; \frac{1}{\binom{N}{i}}\sum_{1\le n_1<\cdots<n_j\le N}\frac{\binom{a_{n_1}+\cdots+a_{n_j}}{j}\binom{A_N-(a_{n_1}+\cdots+a_{n_j})}{N-j}}{\binom{A_N+N-1}{N}}, \qquad i,j \in S.$$

Note that μ has joint descending factorial moments

$$\mathbb{E}((\mu_1)_{l_1}\cdots(\mu_N)_{l_N}) \;=\; \frac{(N)_l}{[A_N]_l}\prod_{n=1}^{N}[a_n]_{l_n}, \qquad l_1,\ldots,l_N\in\mathbb{N}_0,$$

where $l := l_1+\cdots+l_N$. In particular, $\mathbb{E}(\mu_n) = Na_n/A_N$, $n\in\{1,\ldots,N\}$. We conclude from (11) that the backward process $\widehat{X}$ has transition probabilities

$$\widehat{P}_{i,j} \;=\; \frac{i!}{[A_N]_i}\sum_{1\le n_1<\cdots<n_j\le N}\;\sum_{\substack{l_1,\ldots,l_j\in\mathbb{N}\\ l_1+\cdots+l_j=i}}\prod_{k=1}^{j}\frac{[a_{n_k}]_{l_k}}{l_k!}, \qquad i,j\in S,$$

with the convention that $\widehat{P}_{i,0}=\delta_{i0}$. In particular, $\widehat{P}_{i,1} = ([A_N]_i)^{-1}\sum_{n=1}^{N}[a_n]_i$, $i\in\{1,\ldots,N\}$, and hence,

$$c_N \;=\; \widehat{P}_{2,1} \;=\; \frac{1}{A_N(A_N+1)}\sum_{n=1}^{N}a_n(a_n+1) \;>\; 0.$$

From $a_n/A_N\le(a_n+1)/(A_N+1)$ it follows that $c_N\ge\sum_{n=1}^{N}(a_n/A_n)^2\ge 1/N$, or, equivalently, $N_e = 1/c_N\le N$. Moreover,

$$d_N \;=\; \widehat{P}_{3,1} \;=\; \frac{1}{A_N(A_N+1)(A_N+2)}\sum_{n=1}^{N}a_n(a_n+1)(a_n+2) \;>\; 0.$$

It is now verified that $c_N\to 0$ if and only if $d_N/c_N\to 0$. The basic idea of the proof is the same as for the skewed Kimura model, however, the technical details are a bit more involved. We have

$$\max_{1\le n\le N}\left(\frac{a_n}{A_N}\right)^2 \;\le\; \sum_{n=1}^{N}\left(\frac{a_n}{A_N}\right)^2 \;\le\; c_N.$$

Thus, $\max_{1\le n\le N}(a_n/A_N)\le c_N^{1/2}$. Moreover,

$$\begin{aligned} d_N \;&=\; \sum_{n=1}^{N}\frac{a_n(a_n+1)(a_n+2)}{A_N(A_N+1)(A_N+2)} \;\le\; \sum_{n=1}^{N}\frac{a_n(a_n+1)}{A_N(A_N+1)}\left(\frac{a_n}{A_N}+\frac{2}{A_N}\right)\\ &\le\; \left(\max_{1\le n\le N}\frac{a_n}{A_N}\right)\sum_{n=1}^{N}\frac{a_n(a_n+1)}{A_N(A_N+1)}+\frac{2}{A_N}\sum_{n=1}^{N}\frac{a_n(a_n+1)}{A_N(A_N+1)}\\ &=\; \left(\max_{1\le n\le N}\frac{a_n}{A_N}\right)c_N+\frac{2}{A_N}c_N \;\le\; c_N^{3/2}+\frac{2}{A_N}c_N. \end{aligned}$$

Thus, $d_N/c_N\le c_N^{1/2}+2/A_N$. Suppose now that $c_N\to 0$. From the formula for c_N it is readily seen that $c_N\ge a_1^2/[A_N]_2$. Thus, $[A_N]_2\ge a_1^2/c_N\to\infty$. It follows that $A_N\to\infty$ and, hence, $d_N/c_N\le c_N^{1/2}+2/A_N\to 0$. Conversely, if $d_N/c_N\to 0$, then $c_N\to 0$ (see [19, p. 989)] or [22, Lemma 5.5]). Thus, the two conditions $c_N\to 0$ and $d_N/c_N\to 0$ are equivalent.

Example 6.1. (power law growth) Suppose that $a_n := n^{-\alpha}$ for some constant $\alpha\in\mathbb{R}$. Five ranges for the parameter α are distinguished. For $\alpha\in(-\infty,0)$,

$A_N \sim N^{1-\alpha}/(1-\alpha) \to \infty$,

$$c_N \;\sim\; \frac{1}{A_N^2}\sum_{n=1}^{N} a_n^2 \;\sim\; \frac{(1-\alpha)^2}{1-2\alpha}\frac{1}{N} \quad \text{and} \quad d_N \;\sim\; \frac{1}{A_N^3}\sum_{n=1}^{N} a_n^3 \;\sim\; \frac{(1-\alpha)^3}{1-3\alpha}\frac{1}{N^2}.$$

For $\alpha = 0$ we have $c_N = 2/(N+1) \sim 2/N$ and $d_N = 6/((N+1)(N+2)) \sim 6/N^2$. If $\alpha \in (0,1)$, then $c_N \sim 1/A_N \sim (1-\alpha)/N^{1-\alpha}$ and $d_N \sim 2/A_N^2 \sim 2(1-\alpha)^2/N^{2-2\alpha}$. If $\alpha = 1$, then $c_N \sim 1/\log N \to 0$ and $d_N \sim 2/\log^2 N$. In all these four cases considered so far we have $c_N \to 0$ and $d_N/c_N \to 0$, so the model is in the domain of attraction of the Kingman coalescent. Suppose now that $\alpha \in (1,\infty)$. Then, $A_N = \sum_{n=1}^N n^{-\alpha} \to A := \zeta(\alpha) > 1$, $\sum_{n=1}^N a_n(a_n+1) = \sum_{n=1}^N (n^{-2\alpha} + n^{-\alpha}) \to \zeta(2\alpha) + \zeta(\alpha)$, and, hence, $c_N \to (\zeta(2\alpha)+\zeta(\alpha))/[\zeta(\alpha)]_2 > 0$. Similarly, it follows that

$$\lim_{N\to\infty} d_N \;=\; \frac{\zeta(3\alpha) + 3\zeta(2\alpha) + 2\zeta(\alpha)}{\zeta(\alpha)(\zeta(\alpha)+1)(\zeta(\alpha)+2)}.$$

For $l_1, \ldots, l_j \in \mathbb{N} \setminus \{1\}$ and $N \geq i := l_1 + \cdots + l_j$,

$$\begin{aligned}
\Phi_j(l_1,\ldots,l_j) &= \frac{1}{[A_N]_i} \sum_{\substack{n_1,\ldots,n_j=1 \\ \text{all distinct}}}^{N} [a_{n_1}]_{l_1} \cdots [a_{n_j}]_{l_j} \\
&\to \frac{1}{[A]_i} \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N} \\ \text{all distinct}}} [a_{n_1}]_{l_1} \cdots [a_{n_j}]_{l_j} \\
&= \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N} \\ \text{all distinct}}} \frac{[a_{n_1}]_2 \cdots [a_{n_j}]_2}{[A]_{2j}} \frac{[a_{n_1}+2]_{l_1-2} \cdots [a_{n_j}+2]_{l_j-2}}{[A+2j]_{i-2j}} \\
&= \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N} \\ \text{all distinct}}} w(n_1,\ldots,n_j)\, \mathbb{E}(D_{n_1}^{l_1-2} \cdots D_{n_j}^{l_j-2}),
\end{aligned}$$

where $w(n_1,\ldots,n_j) := [a_{n_1}]_2 \cdots [a_{n_j}]_2/[A]_{2j}$ and $(D_{n_1},\ldots,D_{n_j}, 1-\sum_{k=1}^j D_{n_k})$ is Dirichlet distributed with parameters $2+a_{n_1},\ldots,2+a_{n_j}$ and $A-\sum_{k=1}^j a_{n_k}$. Thus, the limit $\phi_j(l_1,\ldots,l_j) := \lim_{N\to\infty} \Phi_j(l_1,\ldots,l_j)$ exists and is of the form

$$\phi_j(l_1,\ldots,l_j) \;=\; \int_{\Delta_j} x_1^{l_1-2} \cdots x_j^{l_j-2}\, \Lambda_j(dx_1,\ldots,dx_j),$$

where $\Delta_j := \{(x_1,\ldots,x_j) \in [0,1]^j : x_1 + \cdots + x_j \leq 1\}$ and the measure

$$\Lambda_j \;:=\; \sum_{\substack{n_1,\ldots,n_j\in\mathbb{N} \\ \text{all distinct}}} w(n_1,\ldots,n_j) P(n_1,\ldots,n_j) \tag{31}$$

on the simplex Δ_j is an infinite mixture of the distributions $P(n_1,\ldots,n_j)$ of $(D_{n_1},\ldots,D_{n_j})$. Note that the weights $w(n_1,\ldots,n_j)$ sum up to a finite value, so Λ_j is a finite measure. The measures $\Lambda_1, \Lambda_2, \ldots$ completely characterize the limiting coalescent process with simultaneous multiple collisions. The model is in the domain of attraction of this coalescent process. So far we have not been able to identify the characterizing measure Ξ of this coalescent.

Example 6.2. (geometric growth) Suppose that $a_n = \lambda^n$ for some constant $\lambda \in (0,\infty)$. The unbiased case corresponds to $\lambda = 1$.

(i) If $\lambda < 1$, then $A_N \to A := \lambda/(1-\lambda)$, so A_N converges. We are hence essentially in the situation of Example 6.1 with $\alpha > 1$. All results there are valid, but now with $A = \lambda/(1-\lambda)$ (instead of $A = \zeta(\alpha)$) and with $a_n = \lambda^n$ (instead of $a_n = n^\alpha$). In particular, the characterizing measures $\Lambda_1, \Lambda_2, \ldots$ of the limiting coalescent process with simultaneous multiple collisions are again infinite mixtures of the form (31).

(ii) If $\lambda > 1$, then $A_N = \sum_{n=1}^N \lambda^n \sim \lambda^{N+1}/(\lambda-1)$, so A_N does not converge. We are hence essentially in the situation of Example 5.2 (ii), and it follows in the same way that the measure Ξ of the limiting Ξ-coalescent assigns its total mass $\Xi(\Delta) = (\lambda-1)/(\lambda+1)$ to the single point $p = (p_1, p_2, \ldots) \in \Delta$ defined via $p_n := (\lambda-1)/\lambda^n$, $n \in \mathbb{N}$.

7. Skewed Kimura model

In this section an example is presented which — in contrast to the skewed Wright-Fisher model studied in Section 5 and the skewed Dirichlet model considered in the previous Section 6 — does not belong to the skewed compound Poisson class. Suppose that for each $n \in \mathbb{N}$ the random variable ξ_n has a binomial distribution with pgf $f_n(x) = (px+q)^{a_n}$, where $a_n \in \mathbb{N}$, $p \in (0,1)$ and $q := 1-p$. Note that $m_n := \mathbb{E}(\xi_n) = pa_n$, $n \in \mathbb{N}$. Then, $\mu = (\mu_1, \ldots, \mu_N)$ has a multi-hypergeometric distribution of the form

$$\mathbb{P}(\mu = k) \;=\; \frac{\binom{a_1}{k_1} \cdots \binom{a_N}{k_N}}{\binom{A_N}{N}}, \qquad k = (k_1, \ldots, k_N) \in \Delta(N),$$

where $A_N := a_1 + \cdots + a_N$. Note that $\mathbb{P}(\mu_n \le a_n) = 1$, i.e. the number μ_n of offspring of individual n is almost surely bounded by a_n, $n \in \{1, \ldots, N\}$. From (10), it follows that the forward process X has transition probabilities

$$P_{i,j} \;=\; \frac{1}{\binom{N}{i}} \sum_{1 \le n_1 < \cdots < n_i \le N} \frac{\binom{a_{n_1} + \cdots + a_{n_i}}{j} \binom{A_N - (a_{n_1} + \cdots + a_{n_i})}{N-j}}{\binom{A_N}{N}}, \qquad i, j \in S.$$

Note that μ has joint factorial moments

$$\mathbb{E}\left(\binom{\mu_1}{l_1} \cdots \binom{\mu_N}{l_N}\right) \;=\; \frac{\binom{N}{l_1 + \cdots + l_N}}{\binom{A_N}{l_1 + \cdots + l_N}} \binom{a_1}{l_1} \cdots \binom{a_N}{l_N}, \qquad l_1, \ldots, l_N \in \mathbb{N}_0,$$

In particular, $\mathbb{E}(\mu_n) = Na_n/A_N$, $n \in \{1, \ldots, N\}$. From (11), it follows that the backward process $\widehat{X}$ has transition probabilities

$$\widehat{P}_{i,j} \;=\; \frac{1}{\binom{A_N}{i}} \sum_{1 \le n_1 < \cdots < n_j \le N} \sum_{\substack{l_1, \ldots, l_j \in \mathbb{N} \\ l_1 + \cdots + l_j = i}} \binom{a_{n_1}}{l_1} \cdots \binom{a_{n_j}}{l_j}, \qquad i, j \in S,$$

with the convention that $\widehat{P}_{i,0} = \delta_{i0}$, $i \in S$. Note that $\widehat{P}_{i,1} = ((A_N)_i)^{-1} \sum_{n=1}^N (a_n)_i$, $i \in \{1, \ldots, N\}$, and, hence,

$$c_N \;=\; \widehat{P}_{2,1} \;=\; \frac{1}{(A_N)_2} \sum_{n=1}^N (a_n)_2 \quad \text{and} \quad d_N \;=\; \widehat{P}_{3,1} \;=\; \frac{1}{(A_N)_3} \sum_{n=1}^N (a_n)_3.$$

The coalescence probability c_N can be smaller than $1/N$ (choose for example $a_n = 2$ for all $n \in \mathbb{N}$ leading to $c_N = 1/(2N-1)$), so the effective population size $N_e = 1/c_N$ can be larger than N. We have $c_N = 0$ if and only if $a_n = 1$ for all $n \in \{1, \ldots, N\}$.

For the rest of this section it is assumed that $a_n > 1$ for at least one index $n \in \mathbb{N}$ such that $c_N > 0$ for sufficiently large N. In the following it is verified that $c_N \to 0$ if and only if $d_N/c_N \to 0$. From $(a_n-2)/(A_N-2) \le (a_n-1)/(A_N-1) \le a_n/A_N$ it follows that

$$\max_{1\le n\le N}\left(\frac{a_n-2}{A_N-2}\right)^2 \;\le\; \max_{1\le n\le N}\frac{a_n(a_n-1)}{A_N(A_N-1)} \;\le\; \sum_{n=1}^{N}\frac{a_n(a_n-1)}{A_N(A_N-1)} \;=\; c_N.$$

Thus, $\max_{1\le n\le N}((a_n-2)/(A_N-2)) \le c_N^{1/2}$, and consequently

$$\begin{aligned} d_N &= \sum_{n=1}^{N}\frac{a_n(a_n-1)(a_n-2)}{A_N(A_N-1)(A_N-2)} \\ &\le \left(\max_{1\le n\le N}\frac{a_n-2}{A_N-2}\right)\sum_{n=1}^{N}\frac{a_n(a_n-1)}{A_N(A_N-1)} = \left(\max_{1\le n\le N}\frac{a_n-2}{A_N-2}\right)c_N \le c_N^{3/2}, \end{aligned}$$

or, equivalently, $d_N/c_N \le c_N^{1/2}$. Therefore, if $c_N \to 0$ then $d_N/c_N \to 0$. Conversely, if $d_N/c_N \to 0$, then a fundamental result from coalescent theory (see, for example, [19, p. 989)] or [22, Lemma 5.5]) ensures that the condition $c_N \to 0$ is satisfied (even for arbitrary Cannings models). Thus, whenever $c_N \to 0$ as $N \to \infty$, the model is in the domain of attraction of the Wright-Fisher diffusion (forwards in time) and of the Kingman coalescent (backwards in time).

Example 7.1. (power law growth) Fix $K \in \mathbb{N}_0$ and suppose that $a_n = n^K$, $n \in \mathbb{N}$. The unbiased case corresponds to $K = 0$. Then, $A_N = \sum_{n=1}^{N} n^K \sim N^{K+1}/(K+1)$ and, hence,

$$c_N \;=\; \frac{1}{(A_N)_2}\sum_{n=1}^{N}(n^K)_2 \;\sim\; \frac{1}{A_N^2}\sum_{n=1}^{N}n^{2K} \;\sim\; \left(\frac{K+1}{N^{K+1}}\right)^2\frac{N^{2K+1}}{2K+1} \;=\; \frac{(K+1)^2}{2K+1}\frac{1}{N}.$$

Thus $N_e = \rho N$ with $\rho := (2K+1)/((K+1)^2) < 1$. Similarly,

$$d_N \;\sim\; \frac{1}{A_N^3}\sum_{n=1}^{N}n^{3K} \;\sim\; \left(\frac{K+1}{N^{K+1}}\right)^3\frac{N^{3K+1}}{3K+1} \;=\; \frac{(K+1)^3}{3K+1}\frac{1}{N^2}.$$

In particular, $c_N \to 0$, so the model is in the domain of attraction of the Wright-Fisher diffusion (forwards in time) and of the Kingman coalescent (backwards in time). This example essentially coincides with Example 5.1 (i) with $\alpha := -K$.

Example 7.2. (geometric growth) Fix an integer parameter $\lambda \ge 2$ and suppose that $a_n = \lambda^n$, $n \in \mathbb{N}$. Then, $A_N = \sum_{n=1}^{N}\lambda^n = \lambda(\lambda^N-1)/(\lambda-1) \sim \lambda^{N+1}/(\lambda-1)$. It is therefore reasonable to proceed as in Example 5.2 (ii). Thus, define $b := 1/\lambda$ and $p_n := (1-b)b^{n-1} = (\lambda-1)/\lambda^n$, $n \in \mathbb{N}$. We have

$$\begin{aligned} \widehat{P}_{i,1} &= \Phi_1(i) = \frac{1}{(A_N)_i}\sum_{n=1}^{N}(\lambda^n)_i \;\sim\; \frac{1}{A_N^i}\sum_{n=1}^{N}\lambda^{ni} \;=\; \frac{1}{A_N^i}\lambda^i\frac{\lambda^{iN}-1}{\lambda^i-1} \\ &\sim \left(\frac{\lambda-1}{\lambda^{N+1}}\right)^i\lambda^i\frac{\lambda^{iN}}{\lambda^i-1} \;\sim\; \frac{(\lambda-1)^i}{\lambda^i-1} \;=\; \frac{(1-b)^i}{1-b^i} \;=\; \sum_{n=1}^{\infty}p_n^i. \end{aligned}$$

More generally, for $l_1, \dots, l_j \in \mathbb{N}$, it follows that

$$\begin{aligned}\Phi_j(l_1,\dots,l_j) &= \frac{1}{(A_N)_i} \sum_{\substack{n_1,\dots,n_j=1 \\ \text{all distinct}}}^{N} (a_{n_1})_{l_1} \cdots (a_{n_j})_{l_j} \\ &\sim \frac{1}{A_N^i} \sum_{\substack{n_1,\dots,n_j=1 \\ \text{all distinct}}}^{N} \lambda^{n_1 l_1 + \cdots + n_j l_j} \;\to \sum_{\substack{n_1,\dots,n_j \in \mathbb{N} \\ \text{all distinct}}} p_{n_1}^{l_1} \cdots p_{n_j}^{l_j}.\end{aligned}$$

where the proof of the last convergence works exactly as in Example 5.2 (ii). Thus, we are indeed in the same situation as in Example 5.2 (ii). The measure Ξ of the limiting Ξ-coalescent assigns its total mass $\Xi(\Delta) = (\lambda - 1)/(\lambda + 1)$ to the single point $p = (p_1, p_2, \dots) \in \Delta$.

References

[1] Ariew, A.; Ernst, Z. (2009). What fitness can't be. *Erkenntnis* **71**, 289–301.

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

[3] Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**, 657–679.

[4] 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.

[5] 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.

[6] Comtet, L. (1970). *Analyse combinatoire*. Tomes 1 et 2. Presses Universitaires de France, Paris.

[7] Eldon, B.; Wakeley, J. (2006). Coalescent processes when the distribution of offspring number of individuals is highly skewed. *Genetics* **172**, 2621–2633.

[8] Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112.

[9] Gladstien, K. (1978). The characteristic values and vectors for a class of stochastic matrices arising in genetics. *SIAM J. Appl. Math.* **34**, 630–642.

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

[11] Karlin, S.; 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.

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

[13] Kolchin, V. F. (1986). *Random Mappings*. Translation Series in Mathematics and Engineering. Optimization Software Inc. Publications Division, New York.

[14] Liggett, T. M. (1985). *Interacting particle systems*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **276**. Springer-Verlag, New York.

[15] Maruyama, T. (1977). *Stochastic problems in population genetics*. Lecture Notes in Biomathematics **17**. Springer-Verlag, Berlin-New York.

[16] McSweeney, J. K.; Pittel, B. G. (2008). Expected coalescence time for a nonuniform allocation process. *Adv. Appl. Probab.* **40**, 1002–1032.

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

[18] Möhle, M. (1999). The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli* **5**, 761–777.

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

[20] Möhle, M. (2010). Looking forwards and backwards in the multi-allelic neutral Cannings population model. *J. Appl. Probab.* **47**, to appear.

[21] Möhle, M.; Sagitov, S. (2001). A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562.
[22] Möhle, M.; Sagitov, S. (2003) Coalescent patterns in diploid exchangeable population models. *J. Math. Biol.* **47**, 337–352.
[23] Moran, P. A. P.; Watterson, G. A. (1959). The genetic effects of family structure in natural populations. *Austral. J. Biol. Sci.* **12**, 1–15.
[24] Pitman, E. J. G. (1968) On the behaviour of the characteristic function of a probability distribution in the neighbourhood of the origin. *J. Austral. Math. Soc.* **8**, 422–443.
[25] Pitman, J. (2006) *Combinatorial Stochastic Processes*. Lecture Notes in Mathematics 1875, 13–35.
[26] Sagitov, S.; Jagers, P. (2005). The coalescent effective size of age-structured populations. *Ann. Appl. Probab.* **15**, 1778–1797.
[27] Schweinsberg, J. (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, Paper no. 12, 50 pp.
[28] Schweinsberg, J. (2003). Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch. Process. Appl.* **106**, 107–139.
[29] Steutel, F. W. and van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**, 893–899.
[30] Wakeley, J.; Sargsyan, O. (2009). Extensions of the coalescent effective population size. *Genetics* **181**, 341–345.

Thierry Huillet
Laboratoire de Physique Théorique et Modélisation
CNRS-UMR 8089 et Université de Cergy-Pontoise
2 Avenue Adolphe Chauvin
95302 Cergy-Pontoise, FRANCE
E-mail: thierry.huillet@u-cergy.fr

Martin Möhle
Mathematisches Institut
Eberhard Karls Universität Tübingen
Auf der Morgenstelle 10
72076 Tübingen, Germany
E-mail: martin.moehle@uni-tuebingen.de