

Coalescent results for diploid exchangeable population models^{*†‡}

Matthias Birkner[§] Huili Liu[¶] Anja Sturm^{||}

Abstract

We consider diploid bi-parental analogues of Cannings models: in a population of fixed size N the next generation is composed of $V_{i,j}$ offspring from parents i and j , where $V = (V_{i,j})_{1 \leq i \neq j \leq N}$ is a (jointly) exchangeable (symmetric) array. Every individual carries two chromosome copies, each of which is inherited from one of its parents. We obtain general conditions, formulated in terms of the vector of the total number of offspring to each individual, for the convergence of the properly scaled ancestral process for an n -sample of genes towards a (Ξ) -coalescent. This complements Möhle and Sagitov’s (2001) result for the haploid case and sharpens the profile of Möhle and Sagitov’s (2003) study of the diploid case, which focused on fixed couples, where each row of V has at most one non-zero entry.

We apply the convergence result to several examples, in particular to two diploid variations of Schweinsberg’s (2003) model, leading to Beta-coalescents with two-fold and with four-fold mergers, respectively.

Keywords: diploid population model; diploid ancestral process; coalescent with simultaneous multiple collisions.

AMS MSC 2010: Primary 60J17; 92D10, Secondary 60J70; 92D25.

Submitted to EJP on September 9, 2017, final version accepted on April 30, 2018.

Supersedes arXiv:1709.02563v2.

1 Introduction and main results

For haploid population models, in which every individual (gene) has one parent (gene), coalescent processes have been used widely in order to describe the ancestral

*This project received financial support from the German Research Foundation (DFG), within the Priority Programme 1590 “Probabilistic Structures in Evolution” through Grants BI 1058/2-1, BI 1058/2-2 and STU 527/1-1, STU 527/1-2.

†H. L. acknowledges research support by National Natural Science Foundation of China (No. 11501164).

‡M. B. and A. S. thank the Institute for Mathematical Sciences, National University of Singapore for hospitality and support during the program Genealogies of interacting particle systems in 2017, where part of this work was completed.

§Johannes-Gutenberg-Universität Mainz, Germany. E-mail: birkner@mathematik.uni-mainz.de

¶Hebei Normal University, China. E-mail: l.huili@live.concordia.ca

||Georg-August-Universität Göttingen, Germany. E-mail: asturm@math.uni-goettingen.de

structure of a sample of n genes when the total population size N is sufficiently large. The purpose of this work is to extend the coalescent theory to general diploid population models, in which individuals carry two copies of each gene which they inherit from two distinct parental individuals. In this context, we derive the diploid analogue of Möhle and Sagitov’s classification of the ancestral processes in exchangeable haploid population models [22]. This gives a unified picture of studying genealogies in an exchangeable diploid setting, which has up to now only been available in special cases.

We consider a general diploid exchangeable population model with fixed constant population size $N \in \mathbb{N} := \{1, 2, \dots\}$ and non-overlapping generations $m \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$ without explicit sexes (see however our remarks in Section 2.3.3). The generations are labeled backwards in time. That is, $m = 0$ is the current generation; $m = 1$ is one generation backwards in time and so on. Each individual possesses two chromosome copies, each inherited from one of its two parents. Which parental chromosome is inherited is a uniform random pick, independently for each child. For $m \in \mathbb{N}$, let $V_{i,j}^{(m)}$ be the number of children by individuals i and j (for $i < j$) in the m -th generation. We call these quantities *pairwise offspring numbers* and implicitly define $V_{k,j}^{(m)} = V_{j,k}^{(m)}$ for $k > j$ when notationally necessary. We exclude the possibility of self-fertilisation, i.e., $V_{i,i}^{(m)} = 0$. We assume that the reproduction law is independent and identically distributed from generation to generation, i.e., the matrices $(V_{i,j}^{(m)})_{1 \leq i < j \leq N}$, $m \in \mathbb{N}$ are i.i.d. We will often write $V_{i,j} = V_{i,j}^{(1)}$ for simplicity. We have $\sum_{1 \leq i < j \leq N} V_{i,j} = N$ because the population size is fixed. Note that despite this dependence on N of the law of $(V_{i,j})_{1 \leq i < j \leq N}$ we will suppress the N -dependence in the notation. Our fundamental assumption is the following exchangeability condition:

$$(V_{i,j})_{1 \leq i < j \leq N} \stackrel{d}{=} (V_{\sigma(i),\sigma(j)})_{1 \leq i < j \leq N} \quad \text{for any permutation } \sigma \text{ of } \{1, \dots, N\}, \quad (1.1)$$

i.e., $(V_{i,j})_{1 \leq i < j \leq N}$ is a finite jointly exchangeable array (see also Remark 4 in Section 1.1).

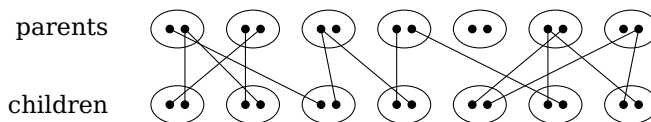


Figure 1: An example for the assignment of parental genes in a population of size $N = 7$, each individual has two gene copies (the filled circles).

Generally, we are interested in tracking the genealogy of a sample of $n \in \mathbb{V}_N := \{1, 2, \dots, N\}$ genes from the present population of size N . We will follow the customary approach of describing ancestral relations among n sampled genes by partitions characterising which genes are descended from the same parental gene. Unless specified otherwise, asymptotic relations refer to letting $N \rightarrow \infty$ throughout the paper.

Let \mathcal{E}_n be the collection of partitions of \mathbb{V}_n and \mathcal{E}_∞ be the collection of partitions of \mathbb{N} . Any element in \mathcal{E}_n can be expressed by $\xi = \{C_1, C_2, \dots, C_b\}$ where $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^b C_i = \mathbb{V}_n$ with $b = |\xi|$ the number of partition elements in ξ . When it is necessary to make the representation unique we order the $C_i, i = 1, \dots, b$ by their smallest element in ascending order. In the following we will also refer to the partition elements as blocks. For any $\xi, \eta \in \mathcal{E}_n$, write $\xi \subseteq \eta$ if and only if every block of η is a union of (one or more) blocks of ξ .

In order to also specify which ancestral genes belong to the same ancestral individuals we use notation introduced by Möhle and Sagitov in [23] and consider the state space

$$\mathcal{S}_n = \left\{ \{ \{C_1, C_2\}, \dots, \{C_{2x-1}, C_{2x}\}, C_{2x+1}, \dots, C_b \} : b \in \mathbb{V}_n, x \in \mathbb{V}_{\lfloor \frac{b}{2} \rfloor}, \{C_1, \dots, C_b\} \in \mathcal{E}_n \right\},$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x . We equip the space \mathcal{E}_n as well as \mathcal{S}_n with the discrete topology. For later use we define a map $\text{cd} : \mathcal{S}_n \rightarrow \mathcal{E}_n$ such that for any

$$\begin{aligned} \xi &= \{\{C_1, C_2\}, \dots, \{C_{2x-1}, C_{2x}\}, C_{2x+1}, \dots, C_b\} \in \mathcal{S}_n, \\ \text{cd}(\xi) &:= \{C_1, C_2, \dots, C_{2x-1}, C_{2x}, C_{2x+1}, \dots, C_b\} \in \mathcal{E}_n. \end{aligned}$$

Following [4], we call $\text{cd}(\xi)$ the *complete dispersion* of ξ .

Now sample n genes randomly from the current generation. We can think of sampling $n/2$ individuals and looking at both of their genes or sampling n individuals and inspecting only one randomly chosen gene in each or something in-between, this will not matter in the limit we are interested in. For $m \in \mathbb{N}_0$, let $\xi^{n,N}(m)$ be the configuration of the genealogical structure for the sampled genes when looking m generations backwards in time: i and j are in the same block of $\xi^{n,N}(m)$ if and only if the i -th and the j -th sampled genes have the same ancestral gene m generations ago. We also keep track of the grouping of these ancestral genes into ancestral (diploid) individuals (this is necessary so that the dynamics of $\xi^{n,N}$ is Markovian). For example, in Figure 1 if we sample all genes of the children then the leftmost parent carries two ancestral genes while the third parent from the left only carries one. We are interested in the convergence of the (suitably time-scaled) *ancestral process* $(\xi^{n,N}(m))_{m \in \mathbb{N}_0}$, which is a Markov chain with state space \mathcal{S}_n .

For the description of the possible limit processes as $N \rightarrow \infty$ the *total offspring numbers*

$$V_i := \sum_{1 \leq j \leq N} V_{i,j}, \tag{1.2}$$

giving the total number of offspring of individual i for $1 \leq i \leq N$ will play a crucial role. Note that these V_i children may be full or half siblings. We have $\sum_{i=1}^N V_i = 2N$ and the vector $(V_i)_{1 \leq i \leq N}$ inherits exchangeability from the array $(V_{i,j})$. Indeed, for any permutation σ on \mathbb{V}_N , we have

$$\begin{aligned} (V_{\sigma(1)}, \dots, V_{\sigma(N)}) &= \left(\sum_{j=1}^N V_{\sigma(i),j} \right)_{i \in \mathbb{V}_N} = \left(\sum_{j=1}^N V_{\sigma(i),\sigma(j)} \right)_{i \in \mathbb{V}_N} \\ &\stackrel{d}{=} \left(\sum_{j=1}^N V_{i,j} \right)_{i \in \mathbb{V}_N} = (V_1, \dots, V_N). \end{aligned} \tag{1.3}$$

Thus, $(V_i)_{1 \leq i \leq N}$ can essentially be viewed as an offspring distribution for a Cannings model with population size $2N$ (in which always only N individuals are parents to offspring in the following generation).

In order to consider a suitable scaling for the large population limit a key quantity is the probability that two genes (picked at random) from two distinct individuals, which are chosen randomly without replacement from the same generation, have a common ancestor (gene) in the previous generation. In our model this quantity is given by

$$c_N = \frac{1}{8} \mathbb{E}[V_{1,2}^2 - \frac{2}{N-1}] + \frac{N-2}{8} \mathbb{E}[V_{1,2}V_{1,3}] = \frac{1}{8} \frac{1}{N-1} \mathbb{E}[(V_1)_2] \tag{1.4}$$

where $(v)_k := v(v-1) \cdots (v-k+1)$ denotes the k -th falling factorial (see Lemma 3.1 below, where also alternative expressions for c_N are given). If $c_N \rightarrow 0$ as $N \rightarrow \infty$, the correct time scaling is $1/c_N$ and any limiting genealogical process will be a continuous-time Markov chain. We will assume that $c_N \rightarrow 0$ as $N \rightarrow \infty$ throughout this paper.

We write $V_{(1)} \geq V_{(2)} \geq \dots \geq V_{(N)}$ for the ranked version of (V_1, \dots, V_N) and

$$\Phi_N := \mathcal{L}\left(\frac{V_{(1)}}{2N}, \frac{V_{(2)}}{2N}, \dots, \frac{V_{(N)}}{2N}, 0, 0, \dots\right)$$

for the law of their ranked (total) offspring frequencies, viewed as a probability measure on the infinite dimensional simplex $\Delta := \{(x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, \sum_{i=1}^{\infty} x_i \leq 1\}$. For all $x = (x_1, x_2, \dots) \in \Delta$, denote by $|x| := \sum_{i=1}^{\infty} x_i$ and $(x, x) := \sum_{i=1}^{\infty} x_i^2$, put $\mathbf{0} := (0, 0, \dots) \in \Delta$. We equip Δ with the topology of coordinate-wise convergence, metrised e.g. via $d_{\Delta}(x, y) = \sum_{i=1}^{\infty} 2^{-i} |x_i - y_i|$ for any $x, y \in \Delta$.

We assume that

$$\frac{1}{2c_N} \Phi_N(dx) \xrightarrow{N \rightarrow \infty} \frac{1}{(x, x)} \Xi'(dx) \quad \text{vaguely on } \Delta \setminus \{\mathbf{0}\} \tag{1.5}$$

where Ξ' is a probability measure on Δ . It is possible that the vague limit of the l.h.s. in (1.5) is a strict sub-probability measure, we then add the remaining mass to Ξ' as the weight of an atom at $\mathbf{0}$, i.e., we put $\Xi'(\{\mathbf{0}\}) = 1 - \Xi'(\Delta \setminus \{\mathbf{0}\})$. It is known that (1.5) is equivalent to the following condition (see Lemma A.6 and Remark 3 in Section 1.1):

$$\phi_j(k_1, \dots, k_j) := \lim_{N \rightarrow \infty} \frac{1}{c_N} \frac{\mathbb{E}[(V_1)_{k_1} \dots (V_j)_{k_j}]}{N^{k_1 + \dots + k_j - j} 2^{k_1 + \dots + k_j}} \quad \text{exists for all } j \in \mathbb{N} \text{ and } k_1, \dots, k_j \geq 2. \tag{1.6}$$

Other characterisations of the equivalent conditions (1.5) and (1.6) are recalled in the Appendix (see Conditions I, II and III). If either of the conditions (1.5) and (1.6) holds, the two limiting objects are connected via (see Theorem A.5)

$$\phi_j(k_1, \dots, k_j) = \mathbb{1}_{\{j=1, k_1=2\}} \cdot 2\Xi'(\{\mathbf{0}\}) + \int_{\Delta \setminus \{\mathbf{0}\}} \sum_{\substack{i_1, \dots, i_j=1 \\ \text{distinct}}}^{\infty} x_{i_1}^{k_1} x_{i_2}^{k_2} \dots x_{i_j}^{k_j} \frac{2\Xi'(dx)}{(x, x)}. \tag{1.7}$$

Furthermore, in this case also the limits

$$\psi_{j,s}(k_1, \dots, k_j) = \lim_{N \rightarrow \infty} \frac{1}{c_N} \frac{\mathbb{E}[(V_1)_{k_1} \dots (V_j)_{k_j} V_{j+1} \dots V_{j+s}]}{N^{k_1 + \dots + k_j - j} 2^{k_1 + \dots + k_j + s}} \tag{1.8}$$

exist, see [22, Lemma 3.5].

The objects and conditions appearing in (1.5), (1.6) and (1.7) are familiar from the theory of coalescents with simultaneous multiple mergers, so called Ξ -coalescents (see [22, 28]). Let Ξ be a finite measure on Δ . An n - Ξ -coalescent is a continuous time (jump-hold) Markov chain $(\xi^n(t))_{t \geq 0}$ on \mathcal{E}_n where in each move, possibly several groups of blocks are merged. If $\eta \in \mathcal{E}_n$ has b blocks and η' with $a < b$ blocks arises from η by merging j groups of sizes $k_1, \dots, k_j \geq 2$ (in particular, there are $s = b - k_1 - \dots - k_j$ "singleton" blocks in η which do not participate in any merger), the transition from η to η' occurs at rate

$$r_{\eta, \eta'} = \lambda_{b; k_1, \dots, k_j; s} = \mathbb{1}_{\{j=1, k_1=2\}} \Xi(\{\mathbf{0}\}) + \int_{\Delta \setminus \{\mathbf{0}\}} \sum_{\ell=0}^s \sum_{\substack{i_1, \dots, i_{j+\ell}=1 \\ \text{distinct}}}^{\infty} \binom{s}{\ell} x_{i_1}^{k_1} \dots x_{i_j}^{k_j} x_{i_{j+1}} \dots x_{i_{j+\ell}} (1 - |x|)^{s-\ell} \frac{\Xi(dx)}{(x, x)}. \tag{1.9}$$

With this notation we can now state our main result.

Theorem 1.1. Assume that $c_N \rightarrow 0$ and that the laws of (V_1, \dots, V_N) , derived from $(V_{i,j})$ via (1.2), satisfy one of the conditions (1.5) and (1.6). Assume also that $\xi^{n,N}(0) = \xi_0 \in \mathcal{S}_n$ for all N . Then

$$(\xi^{n,N}(\lfloor t/c_N \rfloor))_{t \geq 0} \longrightarrow (\xi^n(t))_{t \geq 0}$$

as $N \rightarrow \infty$ in the sense of finite-dimensional distributions. The limit process ξ^n is an n - Ξ -coalescent starting from $\xi^n(0) = \text{cd}(\xi_0)$ with $\Xi = \Xi' \circ \varphi^{-1}$ where Ξ' is the probability measure on Δ appearing on the r.h.s. of (1.5) and (1.7) and $\varphi : \Delta \rightarrow \Delta$ is given by $\varphi(x_1, x_2, x_3, \dots) = (x_1/2, x_1/2, x_2/2, x_2/2, \dots)$.

Note that (1.9) shows that $\Xi(\{\mathbf{0}\})$ corresponds to the rate for binary mergers of two blocks, which is the dynamics of Kingman’s coalescent. We remark that since $\varphi(\mathbf{0}) = \mathbf{0}$ the measures Ξ' and Ξ give the same mass to $\mathbf{0}$ and so have the same Kingman coalescent component.

We also point out that in Theorem 1.1 we only state f.d.d. convergence since one cannot expect weak convergence on $D([0, \infty), \mathcal{S}_n)$, the set of \mathcal{S}_n -valued càdlàg paths equipped with Skorohod’s J_1 -topology (see, e.g., [10, Ch. 3.4]). This is because whenever two ancestral genes descend from the same parental individual the probability that they descend from different ancestral genes (carried by the parental individual) is $1/2$, as is the probability that they descend from the same ancestral gene (resulting in a coalescent event). We have chosen our scaling such that the latter event happens at a finite rate in the limit. Thus, also the former event, which creates some partition $\xi \in \mathcal{S}_n \setminus \mathcal{E}_n$ happens at a positive rate in the limit. But the reason we have f.d.d. convergence in \mathcal{E}_n in Theorem 1.1 is that in the limit any $\xi \in \mathcal{S}_n \setminus \mathcal{E}_n$ transitions instantaneously to $\text{cd}(\xi) \in \mathcal{E}_n$. Thus, due to the discrete topology on \mathcal{S}_n we always have a non-vanishing probability of an accumulation of jumps of finite size which precludes weak convergence in Skorohod’s J_1 -topology. However, if we instead consider the process which tracks the succession of complete dispersion states then weak convergence on $D([0, \infty), \mathcal{E}_n)$ holds:

Corollary 1.2. Let $\tilde{\xi}^{n,N}(m) := \text{cd}(\xi^{n,N}(m)) \in \mathcal{E}_n$ be the ancestral partition of the n sampled genes m generations in the past, irrespective of the grouping into diploid individuals. Under the assumptions of Theorem 1.1, we have

$$\left(\tilde{\xi}^{n,N}(\lfloor t/c_N \rfloor)\right)_{t \geq 0} \xrightarrow{N \rightarrow \infty} (\xi^n(t))_{t \geq 0} \quad \text{weakly on } D([0, \infty), \mathcal{E}_n)$$

and the limit process is the n - Ξ -coalescent from Theorem 1.1.

Before continuing we briefly outline the structure of the remaining paper. After discussing our main result and its relation to the literature in Section 1.1 we consider various examples and discuss their biological motivation in Section 2. In particular, we study two diploid variations of a model by Schweinsberg [29] in Section 2.1 and 2.2. We also discuss the relation to previous results on coalescents for diploid population models and possible extensions in more detail in Section 2.3.

The final Section 3 is dedicated to the proofs of our main results. In Section 3.1 we prove Theorem 1.1 and Corollary 1.2. In Sections 3.2 and 3.3 we prove the more technical convergence results for the models considered in Section 2: Proposition 2.3 of Section 2.1 is proven in Section 3.2 and Proposition 2.5 of Section 2.2 in Section 3.3.

1.1 Discussion

In this section we first give a brief overview over existing coalescent theory in the haploid and diploid setting. Subsequently, we make several remarks regarding our main results.

Classical large population approximation results in the haploid setting can be found in Kingman [15, 16], where a convergence theorem to the classical coalescent (nowadays

known as Kingman's coalescent) is established for a class of exchangeable populations. In recent years, there has been a tremendous development in coalescent theory. We refer to Pitman [25], Sagitov [26] and Donnelly & Kurtz [8] for coalescents with multiple mergers and to Schweinsberg [28] and Möhle & Sagitov [22] for coalescents with simultaneous multiple mergers. At the same time, coalescent theory has been applied to more complex population models. Sagitov [26] deduced a necessary condition for the convergence of the haploid ancestral process to coalescents with multiple mergers. Möhle and Sagitov [22] then fully classified haploid exchangeable population models, so called Cannings models, in terms of the convergence of their ancestral lines to coalescents with simultaneous multiple mergers. They characterised the coalescent generators in terms of the joint moments of offspring sizes as well as in terms of a sequence of measures defined on the infinite dimensional simplex. Subsequently, Sagitov [27] presented a criterion of weak convergence to the coalescent with simultaneous multiple mergers by a scaled vector of the ranked offspring sizes which constitute a given generation.

For diploid population models, the available theory has been more limited. Möhle [17] introduced a diploid population model with selfing and studied the ancestral process in the Wright-Fisher case. He proved that in this case the limit is Kingman's coalescent. We recover Möhle's result without selfing as a special case of our general result, see Section 2. In Möhle [18] it was proved that the scaled ancestral process of n sampled genes in the two-sex Wright-Fisher model behaves like Kingman's coalescent. In this context, Möhle also derived coalescence estimates for general offspring mechanisms if only two genes are sampled. Subsequently, Möhle and Sagitov [23] completely classified the coalescent patterns in two-sex diploid exchangeable population models and established conditions for the limiting scaled ancestral process to either be Kingman's coalescent or the coalescent with (simultaneous) multiple mergers. In contrast to our set-up, individuals are either male or female (N individuals each) and in each generation N couples are formed that have children according to a general exchangeable offspring distribution. Sexes are again assigned randomly conditioned on there being again N males and N females. This is a special case of our result, see Section 2.3.1. In fact, Theorem 1.1 is in a sense an explicitly worked-out version of the remarks in [23, Section 7].

Birkner et al. [4] studied a diploid Moran type population model in which two individuals drawn uniformly at random contribute a (potentially) large number of offspring relative to the total population size. They proved that due to this property and the diploid inheritance the scaled ancestral process admits in the limit simultaneous multiple mergers in up to four groups. In Section 2.3.2 we give more details on the relationship to our main results. In particular, the single-locus analogues of Theorems 1.2 and 1.3 in [4] can be recovered as a special case of Theorem 1.1.

We would like to emphasize a number of points regarding our main results:

1. Broadly speaking, Theorem 1.1 says that we can (for N large) use "equivalent" sampling on the gene level and ignore the grouping of genes into diploid individuals. This phenomenon has been observed many times before (e.g. in [23] and [4]), it is explained by an asymptotic separation of time-scales: the "breaking up" of grouping into diploids is much faster than non-trivial coalescence on the gene level (see the proof of Theorem 1.1).

For finite N , the process $\tilde{\xi}^{n,N}$ is in general not a Markov chain, this is one of the reasons why we consider $\xi^{n,N}$ in Theorem 1.1. However, the limit process is Markovian.

2. $(\xi^n(t))_{t \geq 0}$ can in a natural way be interpreted as a tree describing the genealogy of n sampled genes. In population genetics applications, functionals of this tree, in particular the total length

$$L_{\text{tot}}(\xi^n) := \int_0^\tau \#\xi^n(t) dt \quad \text{with } \tau := \inf\{t \geq 0 : \#\xi^n(t) = 1\}$$

and the length of all branches subtending i leaves for $i \in \{1, 2, \dots, n-1\}$

$$L_i(\xi^n) := \int_0^\tau \#\{C \in \xi^n(t) : \#C = i\} dt$$

are of interest. By Corollary 1.2, the distribution of such functionals of $(\tilde{\xi}^{n,N}(\lfloor t/c_N \rfloor))_{t \geq 0}$ converges as well.

3. Note the normalisation with $2c_N$ in (1.5). The expression in (1.7) is the limit object related to sampling according to the (“haploid”) offspring vectors (V_1, \dots, V_N) and the asymptotically correct scaling of Φ_N (so that the corresponding limit object Ξ' is a probability measure on Δ) is given by $1/c'_N$ with

$$c'_N = \mathbb{E} \left[\frac{1}{2N(2N-1)} \sum_{i=1}^N V_i(V_i-1) \right] = \frac{\mathbb{E}[V_1(V_1-1)]}{2(2N-1)},$$

see e.g. [27, Eq. (1.5)]. We can interpret c'_N as referring to sampling directly on the level of chromosomes where c_N as defined in (1.4) refers to sampling on the level of diploid individuals. We have $c'_N \sim 2c_N$ for $N \rightarrow \infty$ (see also Lemma 3.1) and our normalisation in (1.7) entails $\phi_1(2) = 2$.

4. (1.1) says that $(V_{i,j})$ is a finite jointly exchangeable array. A related notion is that of “separately exchangeable arrays” where rows and columns may use different permutations. See also the discussion in Section 2.3 and see e.g. [14], [1] for general background on exchangeable arrays.

2 Examples

We will now apply Theorem 1.1 to various examples. Some of these have been considered in the literature before, and we recover the known limiting results in an efficient way, some of the examples are new or analysed here in a more general setting.

Arguably, the simplest diploid population model is the diploid Wright-Fisher model where each individual in the children’s generation is independently assigned two distinct parents by drawing twice without replacement from the parent’s generation; the joint distribution of $(V_{i,j})_{1 \leq i < j \leq N}$ is then a $\binom{N}{2}$ -dimensional multinomial distribution with uniform weights. This model was considered e.g. by Möhle [17] and to set the stage we briefly discuss how we recover his result for the case with no selfing ($s = 0$ in Möhle’s [17] notation) from Theorem 1.1.

Proposition 2.1. *In the diploid Wright-Fisher model without selfing, using time-scaling with $c_N = 1/(2N)$, the limiting coalescent (in the f.d.d. sense) is Kingman’s coalescent (which corresponds to the case $\Xi' = \Xi = \delta_0$ in Thm. 1.1).*

Proof. By choosing $W_i \equiv 1$ in Section 2.1 this model is a special case of the class considered there and the result follows from Proposition 2.3, case 1. Alternatively, one can easily check that the (sharp) criterion on the third factorial moment of V_1 from Möhle [20, Eq. (14) in Sect. 4] for convergence to Kingman’s coalescent is satisfied. \square

It is well known that the class of possible coalescent processes arising as limiting genealogies in population models is much richer than just Kingman’s coalescent. An important family of examples for the haploid case is given by the $\text{Beta}(2 - \alpha, \alpha)$ -coalescents, where $0 < \alpha < 2$. For the sub-case $1 \leq \alpha < 2$ these are well motivated by a class of models that were proposed and analysed by Schweinsberg [29], which work as follows: Let each generation consist of N haploid (adult) individuals. Individual i produces X_i juveniles, where X_1, \dots, X_N are independent copies of X with $\mathbb{E}[X] > 1$ and X has a (strictly) regularly varying tail,

$$\mathbb{P}(X > x) \sim cx^{-\alpha} \quad \text{as } x \rightarrow \infty \tag{2.1}$$

with $c \in (0, \infty)$. Then, N of the $S_N = X_1 + \dots + X_N$ ($> N$ typically) juveniles are drawn at random without replacement to form the next (adult) generation. It turns out (see [29, Thm. 4]) that in the limit $N \rightarrow \infty$, the suitably scaled genealogies of samples from such a population model converge to a $\text{Beta}(2 - \alpha, \alpha)$ -coalescent. This is a particular so-called Λ -coalescent. In the notation of (1.9) it is given by

$$\Xi(dx) = \int_{[0,1]} \delta_{(x,0,0,\dots)} \text{Beta}(2 - \alpha, \alpha)(dx)$$

where $\text{Beta}(2 - \alpha, \alpha)$ is the probability law on $[0, 1]$ with density

$$\frac{1}{B(2 - \alpha, \alpha)} x^{1-\alpha} (1 - x)^{\alpha-1}, \quad 0 < x < 1. \tag{2.2}$$

Here, for $a, b > 0$, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ denotes the Beta-function. There is also a rich mathematical structure linking these particular Λ -coalescents to stable branching processes, see e.g. [2, 3].

This model captures situations where occasionally some individuals can, for example due to environmental fluctuations, possibly produce many more offspring than others (note that (2.1) with $\alpha < 2$ implies $\text{Var}[X_i] = \infty$). It is thus a possible mathematical formalisation of the concept of “sweepstakes reproduction” that appears in the biological literature, see e.g. Eldon and Wakeley [9] and the discussion and references therein.

There are various possibilities how one can extend this model – literally or in spirit – to a diploid scenario. We explore two such possibilities below in more detail: In Section 2.1 we assign each individual i in a given generation independently a random “fitness value” $W_i \geq 0$ and decree that each child has a chance $\propto W_i W_j$ to descend from couple (i, j) , i.e. the joint law of offspring numbers is given by (2.4). In Section 2.2, each couple (i, j) independently produces $X_{i,j}$ juveniles and then N out of the $\sum_{i < j \leq N} X_{i,j}$ juveniles are drawn at random to form the next generation, analogous to [29].

It turns out that two distinct forms of “diploid $\text{Beta}(2 - \alpha, \alpha)$ -coalescents” arise from these two set-ups: The limiting Ξ in Section 2.1 arises from a $\text{Beta}(2 - \alpha, \alpha)$ -distributed x by replacing it with *two* equal weights $x/4$ (see Equation (2.10) in Proposition 2.3) whereas in Section 2.2 it is split into *four* equal weights (see (see Equation (2.20) in Proposition 2.5).

Intuitively, this can be understood as follows: In Section 2.1, the dominant contribution comes from situations when one W_i is exceptionally large ($\approx O(N)$) whereas all others are much smaller; then there is a large family of half-siblings and the two weights correspond to the two chromosome copies of the exceptional individual i ; all the other parents will typically have a total number of offspring which is negligible in comparison to N and none of their genes will be involved in a multiple merging event. On the other hand, in Section 2.2 the dominant contribution comes from cases when $X_{i,j} \approx O(N)$ for

exactly one couple (i, j) and all other $X_{k,\ell}$ ($\{k, \ell\} \neq \{i, j\}$) are much smaller; then there is a large family of full siblings and the four weights correspond to the four chromosome copies of the two individuals in the successful couple.

In particular, we see that the answer to the question which diploid coalescent is appropriate for a given biological population with potentially highly skewed individual reproductive success can depend on the typical mating behaviour.

We describe and analyse these diploid variations of Schweinsberg’s [29] model in more detail in Sections 2.1 and 2.2 below. In Section 2.3, we briefly discuss how results of previous studies of diploid population models, especially from [4] and [23], fit into our framework. We also mention possible extensions and additional examples there.

The perspicacious reader will observe that in Propositions 2.3 and 2.5 below (compare Assumptions (2.5) and (2.17), respectively), we have excluded the boundary cases $\alpha = 2$ and $\alpha = 1$. In view of Schweinsberg’s [29] results for the haploid case, we expect the following: For $\alpha = 2$, $c_N \sim c(\log N)/N$ in Lemmas 2.2 and 2.4 and convergence to Kingman’s coalescent; for $\alpha = 1$, $c_N \sim c/\log N$ and in both Proposition 2.3, 2. and Proposition 2.5, 2., the uniform distribution on $[0, 1]$ will appear, i.e. the limiting coalescent will then be a variation on the Bolthausen-Sznitman-coalescent where jumps are broken into two groups and into four groups, respectively.

We leave the details to future work.

2.1 Diploid population model with random individual fitness

Let $W_1, W_2, \dots \geq 0$ be independent copies of nonnegative random variables W with $\mu_W := \mathbb{E}[W] > 0$, put

$$Z_N := \sum_{1 \leq i < j \leq N} W_i W_j = \frac{1}{2} \left(\sum_{i=1}^N W_i \right)^2 - \frac{1}{2} \sum_{i=1}^N W_i^2. \tag{2.3}$$

Given the W_i ’s let

$$(V_{i,j})_{1 \leq i < j \leq N} \stackrel{d}{=} \text{Multinomial} \left(N, \frac{W_1 W_2}{Z_N}, \frac{W_1 W_3}{Z_N}, \dots, \frac{W_{N-1} W_N}{Z_N} \right). \tag{2.4}$$

(we will see in Section 3.2 that the event $Z_N = 0$ has negligible probability in the limit $N \rightarrow \infty$). Note that when the W_i ’s are identical, (2.4) coincides literally with our version of the diploid Wright-Fisher model, see Proposition 2.1.

The “fitness” in this section’s title is not based on an explicitly modelled genetic type and is not passed on to offspring as the values are drawn afresh in each generation. The offspring distribution in (2.4) may be appropriate for a population with high individual reproductive potential (thinking e.g. of plants or marine species that can in principle produce large numbers of seeds or eggs) in an environment that fluctuates rapidly both in space and time. In reality, there may be a very complex and highly variable interplay between ecological and genetic factors that determine the reproductive success of a given individual at a given time. All this would be subsumed in this model into a random “effective fitness parameter” W .

For the fitness parameter we will consider separately the finite variance case as well as the case of (strictly) regularly varying tails such that

$$\mathbb{P}(W \geq x) \sim c_W x^{-\alpha} \quad \text{as } x \rightarrow \infty \quad \text{for some } c_W \in (0, \infty) \text{ and } 1 < \alpha < 2. \tag{2.5}$$

Before stating the convergence result we specify the asymptotic behavior of the

scaling parameter c_N as specified in (1.4). A key quantity is

$$Q_N := \sum_{j=2}^N \frac{W_1 W_j}{Z_N}, \tag{2.6}$$

the probability that a randomly chosen child is an offspring of parent 1.

Lemma 2.2. *The pair coalescence probability over one generation for $V_{i,j}$'s as in (2.4) is given by*

$$c_N = \frac{N}{8} \mathbb{E} [Q_N^2]. \tag{2.7}$$

1. If $\mu_W^{(2)} := \mathbb{E}[W^2] < \infty$ we have

$$c_N \sim C_{\text{pair}}^{(\text{Kingm})} N^{-1} \quad \text{as } N \rightarrow \infty \quad \text{with } C_{\text{pair}}^{(\text{Kingm})} = \frac{\mu_W^{(2)}}{2\mu_W}. \tag{2.8}$$

2. If (2.5) holds we have

$$c_N \sim C_{\text{pair}}^{(\text{Beta})} N^{1-\alpha} \quad \text{as } N \rightarrow \infty \quad \text{with } C_{\text{pair}}^{(\text{Beta})} = c_W (2/\mu_W)^\alpha \alpha \Gamma(2-\alpha) \Gamma(\alpha) / 8. \tag{2.9}$$

By scaling with the appropriate c_N we obtain the following convergence results.

Proposition 2.3. 1. *If $\mu_W^{(2)} < \infty$ then $(\xi^{n,N}(cNt))_{t \geq 0}$ with $c = 1/C_{\text{pair}}^{(\text{Kingm})}$ (cf. (2.8)) converges in the f.d.d. sense to Kingman's coalescent.*

2. *If the tails of W vary (strictly) regularly as specified in (2.5) then $(\xi^{n,N}(cN^{\alpha-1}t))_{t \geq 0}$ with $c = 1/C_{\text{pair}}^{(\text{Beta})}$ (cf. (2.9)) converges in the f.d.d. sense to a Ξ -coalescent with $\Xi = \text{Beta}(2-\alpha, \alpha) \circ \varphi^{-1}$ where $\varphi : [0, 1] \rightarrow \Delta$ is given by $\varphi(x) = (x/4, x/4, 0, 0, \dots)$, that is*

$$\Xi(dx) = \int_{[0,1]} \delta_{(\frac{x}{4}, \frac{x}{4}, 0, 0, \dots)} \text{Beta}(2-\alpha, \alpha)(dx) \tag{2.10}$$

with the density of the $\text{Beta}(2-\alpha, \alpha)$ distribution given in (2.2).

The proof of Lemma 2.2 and Proposition 2.3 can be found in Section 3.2.

2.2 Diploid population model related to supercritical Galton-Watson processes

In this section, we consider another diploid version of the model introduced and studied by Schweinsberg in [29] in which an abundance of offspring is produced in each generation of which only a limited number survives. The model is similar to that of Section 2.1 in that large families may be produced. However, in contrast to the random individual fitness model of the last section, in which *individual parents* may have many offspring due to an unusual fitness, we here have *parent couples* that may produce a large family.

More concretely, let $X_{i,j} = X_{i,j}^{(N)}, 1 \leq i < j \leq N$, be the ‘‘potential offspring’’ of parent i and j in any given generation with a distribution that may depend on N . For notational convenience, we set $X_{j,i}^{(N)} = X_{i,j}^{(N)}$ if $j > i$ and $X_{i,i}^{(N)} = 0$. We also denote the number of potential offspring to parent i by

$$X_i = X_i^{(N)} = \sum_{j=1, j \neq i}^N X_{i,j}^{(N)}$$

for any $1 \leq i \leq N$. Let

$$S_N = \sum_{1 \leq i < j \leq N} X_{i,j}^{(N)} \tag{2.11}$$

be the total number of potential offspring. The actual total population size is always fixed at N . If $S_N \geq N$, we obtain the next generation by sampling N of these offspring at random without replacement. We use $V_{i,j} = V_{i,j}^{(N)}$ to denote the number of offspring sampled from $X_{i,j}$ for any $1 \leq i < j \leq N$. Our scaling will be such that there are enough offspring for resampling with sufficiently high probability so that the details of (any exchangeable) offspring assignment will not be relevant otherwise. We assume in the following that

$$\left(X_{i,j}^{(N)} \right)_{1 \leq i < j \leq N} \text{ are i.i.d.} \tag{2.12}$$

so that the potential offspring of parent pairs are generated as in a Galton-Watson process. In addition, we assume that

$$\mathcal{L} \left(X_{i,j}^{(N)} \mid X_{i,j}^{(N)} > 0 \right) = \mathcal{L}(X) \text{ and } p_N := \mathbb{P} \left(X_{i,j}^{(N)} > 0 \right) \sim c_{X,1}/N \text{ with } c_{X,1} \in (0, \infty) \tag{2.13}$$

where the law of X does not depend on N and satisfies

$$\mu_X := \mathbb{E}[X] \in (2/c_{X,1}, \infty). \tag{2.14}$$

Note that (2.14) implies that

$$\mathbb{E}[S_N] = \binom{N}{2} p_N \mu_X \underset{N \rightarrow \infty}{\sim} \mu N \text{ with } \mu := \frac{c_{X,1}}{2} \mu_X > 1. \tag{2.15}$$

(We will see in Lemma 3.13 below that this implies that the event $\{S_N < N\}$ has asymptotically negligible probability in the scaling regimes we consider.) Finally, we require one of the following assumptions:

$$\mathbb{E}[X^2] < \infty \tag{2.16}$$

or

$$\mathbb{P}(X > k) \underset{k \rightarrow \infty}{\sim} c_{X,2} k^{-\alpha} \text{ for some } \alpha \in (1, 2) \text{ and } c_{X,2} \in (0, \infty). \tag{2.17}$$

These assumptions might appear at first sight somewhat artificial, see however the discussion in Remark 2.6 below.

Before stating the convergence result we again specify the asymptotic behavior of the scaling parameter c_N given in (1.4).

Lemma 2.4. *1. If (2.16) holds then*

$$c_N \sim \tilde{C}_{\text{pair}}^{(\text{Kingm})} N^{-1} \text{ as } N \rightarrow \infty \text{ with } \tilde{C}_{\text{pair}}^{(\text{Kingm})} = \frac{1}{2} \left(\frac{\mathbb{E}[X(X-1)]}{c_{X,1} \mu_X^2} + 1 \right). \tag{2.18}$$

2. If (2.17) holds then

$$c_N \sim \tilde{C}_{\text{pair}}^{(\text{Beta})} N^{1-\alpha} \text{ as } N \rightarrow \infty \text{ with } \tilde{C}_{\text{pair}}^{(\text{Beta})} = \frac{1}{8} \frac{c_{X,1} c_{X,2} \alpha}{\mu^\alpha} B(2-\alpha, \alpha). \tag{2.19}$$

By scaling with c_N we obtain the following convergence result.

Proposition 2.5. *1. If (2.16) holds then $(\xi^{n,N}(cNt))_{t \geq 0}$ with $c = 1/\tilde{C}_{\text{pair}}^{(\text{Kingm})}$ (cf. (2.18)) converges in the f.d.d. sense to Kingman’s coalescent.*

2. If (2.17) holds then $(\xi^{n,N}(cN^{\alpha-1}t))_{t \geq 0}$ with $c = 1/\tilde{C}_{\text{pair}}^{(\text{Beta})}$ (cf. (2.19)) converges in the f.d.d. sense to a Beta-coalescent with simultaneous mergers of four groups. More precisely, the limiting coalescent is a Ξ -coalescent with

$$\Xi(dx) = \int_{(0,1]} \delta_{(\frac{x}{4}, \frac{x}{4}, \frac{x}{4}, \frac{x}{4}, 0, 0, \dots)} \text{Beta}(2-\alpha, \alpha)(dx), \tag{2.20}$$

where the density of the $\text{Beta}(2-\alpha, \alpha)$ distribution is given in (2.2).

The proof of Lemma 2.4 and Proposition 2.5 can be found in Section 3.3.

Remark 2.6. 1. The model considered in this section is appropriate for a large, unstructured population of N diploid individuals with promiscuous reproductive behaviour. It is intended to capture situations where there is potentially great variability between the number of juveniles produced by different mating couples and this is achieved in analogy to Schweinberg’s model [29] in the mathematically simplest way by assuming that the $X_{i,j}^{(N)}$ are independent with the same distribution.

At first sight, it may seem then that the structural assumptions (2.13) and (2.16) or (2.17) are artificial choices just to “make the mathematical theory work”. However, if we stipulate, as seems biologically reasonable, that the law $\mathcal{L}(X_1^{(N)})$ of the number of potential offspring of a typical individual (think e.g. of the number of gametes produced) should be roughly independent of the population size N – in particular, it should not diverge as N grows large – then together with the i.i.d.-assumption for the $X_{i,j}^{(N)}$ we see that there is essentially no other choice than to assume the first half of (2.13): $\mathbb{P}(X_{i,j}^{(N)} > 0)$ must be of order $1/N$. (2.13) also means that the average number of partners of a typical individual stabilises in distribution as $N \rightarrow \infty$, in fact, it is approximately Poisson distributed with mean $c_{X,1}$.

From the point of view of a biological model, we suggest to read (2.13) as follows: Given a large population of size $N \gg 1$ let

$$c_{X,1} = NP(\text{two randomly drawn individuals produce potential offspring together})$$

and let $\mathcal{L}(X)$ be the law of the number of potential offspring produced by two randomly drawn individuals, given that they do produce some. Then, if X satisfies (2.16) or (2.17), the genealogy of an n -sample is over time-scales $\propto 1/c_N$ approximately described by Proposition 2.5. See Section 2.3.3 for possible extensions.

2. We do not strive here to answer in full generality the mathematical question “if one only assumes that $X_{i,j}^{(N)}, 1 \leq i < j \leq N$ are i.i.d. with law ν_N such that $\sup_N \mathbb{E}[X_1^{(N)}] = \sup_N (N - 1)\mathbb{E}[X_{1,2}^{(N)}] < \infty$, what are sharp conditions on the family ν_N of probability measures on \mathbb{Z}_+ so that the genealogical processes of finite samples in such population models converge?”

Obviously, then necessarily $\sup_N NP(X_{i,j}^{(N)} > 0) < \infty$ and we see from the proofs of Lemma 2.4 and Proposition 2.5 that in the case of infinite variance $\mathbb{E}[(X_{i,j}^{(N)})^2] = \infty$ suitable control of the tail behaviour $\mathbb{P}(X_{i,j}^{(N)} > x | X_{i,j}^{(N)} > 0)$ uniformly in N is required for the limit in (3.50) to exist. In fact, one can cook up examples where $NP(X_{i,j}^{(N)} > 0)$ and $N^{\alpha-1}c_N$ oscillates as a function of N or where even though $\lim_{x \rightarrow \infty} x^\alpha \mathbb{P}(X_{i,j}^{(N)} > x | X_{i,j}^{(N)} > 0) =: c_{X,2}$ exists for all N one has convergence to different coalescents along different subsequences.

3. Our parametrisation in (2.13) enforces $\mathbb{P}(X = 0) = 0$. If one prefers to allow $0 < \mathbb{P}(X = 0) < 1$, one can replace p_N by $p_N \mathbb{P}(X > 0)$ and X by X' where $\mathbb{P}(X' \in \cdot) = \mathbb{P}(X \in \cdot | X > 0)$.

2.3 Relation to previous work and possible further extensions

2.3.1 Diploid population model with randomly chosen pairs as couples

We here recover the convergence result of [23, Theorem 4.2 and Corollary 4.3] concerning a diploid two-sex population model. In order to make comparisons to [23] easier

we assume here that the population size is given by $2N$. Now, let $\{J_1, \dots, J_N\}$ be a random and randomly ordered partition of $\{1, \dots, 2N\}$ into N subsets of size 2. This partition describes the grouping of individuals into N distinct couples which give birth to the individuals of the next generation. Let $\tilde{V}_1, \dots, \tilde{V}_N$ be a sequence of exchangeable non-negative random variables representing the number of children from each couple respectively with $\tilde{V}_1 + \dots + \tilde{V}_N = 2N$. Write $\tilde{V}_{(1)} \geq \tilde{V}_{(2)} \geq \dots \geq \tilde{V}_{(N)}$ for its ranked version. The offspring distribution $(V_{i,j})_{1 \leq i < j \leq 2N}$ is then given by

$$V_{i,j} = \begin{cases} \tilde{V}_\ell, & \text{if } \{i, j\} = J_\ell \text{ for some } \ell \in \{1, 2, \dots, N\}, \\ 0, & \text{else.} \end{cases}$$

It is clear that the $(V_{i,j})_{1 \leq i < j \leq 2N}$ are exchangeable as in (1.1). Note that the corresponding total offspring size vector (V_1, \dots, V_{2N}) is a random permutation of $(\tilde{V}_1, \tilde{V}_1, \dots, \tilde{V}_N, \tilde{V}_N)$. In particular, V_1 and \tilde{V}_1 have the same distribution. Thus, we obtain from (1.4) that (remember that we use population size $2N$ here)

$$c_{2N} = \frac{\mathbb{E}[(V_1)_2]}{8(2N-1)} = \frac{\mathbb{E}[(\tilde{V}_1)_2]}{8(2N-1)}.$$

Let us remark that this model can also be interpreted as a two-sex model, which is the formulation used in [23]. Here, in each generation, we randomly assign sexes to the offspring such that there are N male and N female offspring. Subsequently, the couples are formed at random between the males and the females. We have the following proposition:

Proposition 2.7 (Möhle & Sagitov [23]). *Assume that $c_{2N} \rightarrow 0$ as $N \rightarrow \infty$ and that*

$$\frac{1}{4c_{2N}} \mathcal{L} \left(\frac{\tilde{V}_{(1)}}{2N}, \frac{\tilde{V}_{(2)}}{2N}, \dots, \frac{\tilde{V}_{(N)}}{2N} \right) \xrightarrow{N \rightarrow \infty} \frac{1}{(x, x)} \Xi''(dx) \text{ vaguely on } \Delta \setminus \{\mathbf{0}\} \quad (2.21)$$

where Ξ'' is a probability measure on Δ . Then there is convergence in the f.d.d. sense to a Ξ -coalescent with $\Xi = \Xi'' \circ \varphi^{-1} \circ \varphi^{-1} = \Xi'' \circ \tilde{\varphi}^{-1}$ (recall the function φ from Thm. 1.1) with

$$\tilde{\varphi} : \Delta \rightarrow \Delta \text{ given by } \tilde{\varphi}(x_1, x_2, \dots) = (x_1/4, x_1/4, x_1/4, x_1/4, x_2/4, x_2/4, x_2/4, x_2/4, \dots).$$

Remark 2.8. We note that (2.21) can be equivalently formulated in terms of moment conditions as in (1.6) or as in Appendix A.

Proof. We use Theorem 1.1 with condition (1.5), which is in the present context of the form

$$\frac{1}{2c_{2N}} \mathcal{L} \left(\frac{\tilde{V}_{(1)}}{4N}, \frac{\tilde{V}_{(1)}}{4N}, \frac{\tilde{V}_{(2)}}{4N}, \frac{\tilde{V}_{(2)}}{4N}, \dots, \frac{\tilde{V}_{(N)}}{4N}, \frac{\tilde{V}_{(N)}}{4N}, 0, 0, \dots \right) \xrightarrow{N \rightarrow \infty} \frac{1}{(x, x)} \Xi'(dx) \quad (2.22)$$

vaguely on $\Delta \setminus \{(0, 0, \dots)\}$ for a certain probability measure Ξ' on Δ . Note that the measure on the left-hand side of (2.22) equals

$$\frac{1}{2} \frac{1}{c_{2N}} \mathcal{L} \left(\varphi \left(\frac{\tilde{V}_{(1)}}{2N}, \frac{\tilde{V}_{(2)}}{2N}, \dots, \frac{\tilde{V}_{(N)}}{2N} \right) \right)$$

and that (2.21) means that for every continuous function f with compact support in $\Delta \setminus \{\mathbf{0}\}$

$$\begin{aligned} & \frac{1}{c_{2N}} \mathbb{E} \left[f \left(\varphi \left(\frac{\tilde{V}_{(1)}}{2N}, \frac{\tilde{V}_{(2)}}{2N}, \dots, \frac{\tilde{V}_{(N)}}{2N} \right) \right) \right] \\ & \xrightarrow{N \rightarrow \infty} 4 \int_{\Delta} f(\varphi(y)) \frac{1}{(y, y)} \Xi''(dy) = 2 \int_{\Delta} f(x) \frac{1}{(x, x)} (\Xi'' \circ \varphi^{-1})(dx) \end{aligned}$$

(observe $(\varphi(y), \varphi(y)) = \frac{1}{2}(y, y)$). Thus (2.22) is equivalent to (2.21) with $\Xi' = \Xi'' \circ \varphi^{-1}$ and the result follows. \square

Note that we see clearly from the form of $\Xi = \Xi'' \circ \tilde{\varphi}^{-1}$ that the mass of each large family is split into four equal parts, representing the four chromosomes from a particular couple.

2.3.2 A diploid population model with occasional large families

Here, we briefly discuss how the class of continuous-time diploid population models from [4], which involve suitably rare but “large” reproduction events with a single large family, can be formulated in our present discrete-time context and thus the “single-locus” analogues of Theorems 1.2 and 1.3 there can be obtained from our main result.

The children’s generation arises as follows: Randomly choose two distinct individuals $\{I_1, I_2\}$ from $\mathbb{V}_N = \{1, 2, \dots, N\}$. Individuals I_1 and I_2 form a couple (as in [23]) and have a random number Ψ_N of children together but with no-one else, i.e. $V_{I_1, I_2} = \Psi_N$, $V_{I_i, j} = 0$ for $j \neq I_{3-i}$, $i = 1, 2$. The other $N - 2$ individuals in $\mathbb{V}_N \setminus \{I_1, I_2\}$ give birth to $N - \Psi_N$ children according to the diploid Wright-Fisher model. We have (see proof of Lemma 3.1)

$$\begin{aligned}
 c_N &= \mathbb{E} \left[\frac{\Psi_N (\Psi_N - 1)}{N(N-1)} \frac{1}{4} + \sum_{i < j \text{ in } \mathbb{V}_N \setminus \{I_1, I_2\}} \frac{V_{i,j} (V_{i,j} - 1)}{N(N-1)} \frac{1}{4} + \sum_{\substack{i,j,k \text{ in } \mathbb{V}_N \setminus \{I_1, I_2\} \\ \text{pairwise distinct}}} \frac{V_{i,j} V_{i,k}}{N(N-1)} \frac{1}{8} \right] \\
 &= \frac{\mathbb{E}[\Psi_N (\Psi_N - 1)]}{4N(N-1)} + O(N^{-1}).
 \end{aligned}
 \tag{2.23}$$

The limiting behaviour depends on the sequence of laws $\mathcal{L}(\Psi_N)$, $N \in \mathbb{N}$:

1. A simple choice, inspired by [9], is to assume that the “large” family constitutes always a fixed fraction of the total population: Given $\psi \in (0, 1)$, we assume

$$\mathbb{P}(\Psi_N = \lfloor \psi N \rfloor) = 1 - \mathbb{P}(\Psi_N = 1) = N^{-\gamma},$$

where $\gamma > 0$. Then

- I. If $\gamma \in (0, 1)$, we have $c_N \sim \frac{\psi^2}{4} N^{-\gamma}$ and (1.5) holds true with $\Xi' = \delta_{(\frac{\psi}{2}, \frac{\psi}{2}, 0, 0, \dots)}$. Thus Theorem 1.1 yields that the scaled ancestral process converges to a Ξ -coalescent process with $\Xi = \delta_{(\frac{\psi}{4}, \frac{\psi}{4}, \frac{\psi}{4}, \frac{\psi}{4}, 0, 0, \dots)}$.
- II. If $\gamma = 1$, we have $c_N \sim \frac{\psi^2}{4N} + \frac{1}{2N}$ and (1.5) holds true with $\Xi' = \frac{\psi^2}{\psi^2+2} \delta_{(\frac{\psi}{2}, \frac{\psi}{2}, 0, 0, \dots)} + \frac{2}{\psi^2+2} \delta_0$. Thus, Theorem 1.1 shows that the scaled ancestral process converges to a Ξ -coalescent process with $\Xi = \frac{\psi^2}{\psi^2+2} \delta_{(\frac{\psi}{4}, \frac{\psi}{4}, \frac{\psi}{4}, \frac{\psi}{4}, 0, 0, \dots)} + \frac{2}{\psi^2+2} \delta_0$.
- III. If $\gamma > 1$ the limit process is Kingman’s coalescent.

We note that alternatively, we could assume that with probability $1 - N^{-\gamma}$ there is just a Wright-Fisher reproduction step and with probability $N^{-\gamma}$ a reproduction step with one exceptionally fertile couple as above occurs. This yields the same limit process.

2. More generally, we can assume that the sequence of laws $\mathcal{L}(\Psi_N)$, $N \in \mathbb{N}$ satisfies $c_N \rightarrow 0$ as $N \rightarrow \infty$ with c_N from (2.23) and there exists a probability measure F on $[0, 1]$ such that

$$\frac{1}{2c_N} \mathbb{P}(\Psi_N > Nx) \longrightarrow \int_x^1 \frac{1}{y^2} F(dy) \text{ as } N \rightarrow \infty \text{ for any } x \in (0, 1) \text{ where } F \text{ is continuous.}$$

In this case, (1.5) holds true with $\Xi' = \int_{[0,1]} \delta_{(\frac{y}{2}, \frac{y}{2}, 0, 0, \dots)} F(dy)$ and Theorem 1.1 yields that the scaled ancestral process converges to a Ξ -coalescent process with $\Xi = \int_{[0,1]} \delta_{(\frac{y}{4}, \frac{y}{4}, \frac{y}{4}, \frac{y}{4}, 0, 0, \dots)} F(dy)$.

3. One can easily generalise this model to any number k of “large” families. Initially, we randomly choose k couples from \mathbb{V}_N . The number of children of those k couples are $\Psi_{N,1}, \Psi_{N,2}, \dots, \Psi_{N,k}$ with $\Psi_{N,i} = \lfloor \psi_i N \rfloor$ and suitable $\psi_i \in (0, 1)$ for any $i = 1, 2, \dots, k$. Assume c_N converges to 0, then the scaled ancestral process converges to a Ξ -coalescent process with

$$\Xi = \delta_{\left(\frac{\psi_1}{4}, \frac{\psi_1}{4}, \frac{\psi_1}{4}, \frac{\psi_1}{4}, \dots, \frac{\psi_k}{4}, \frac{\psi_k}{4}, \frac{\psi_k}{4}, \frac{\psi_k}{4}, 0, 0, \dots\right)}.$$

2.3.3 Further remarks and possible extensions

The matrix of offspring numbers $(V_{i,j})_{1 \leq i < j \leq N}$ can equivalently be viewed as a(n exchangeable) random multigraph on N nodes, by drawing $V_{i,j}$ undirected edges (one for each child) between nodes i and j . For example, we can interpret the model from Section 2.2 as a variation on Erdős-Rényi graphs (remembering (2.13)): First draw an Erdős-Rényi graph on N nodes with edge probability $c_{X,1}/N$, then replace the i -th of the resulting $M \approx c_{X,1}N/2$ edges by $\nu_{i,M}$ edges, where $(\nu_{1,M}, \dots, \nu_{M,M})$ are drawn as in Schweinsberg [29, Section 1.3] (except that we enforce $\nu_{1,M} + \dots + \nu_{M,M} = N$, not $= M$).

From the point of view of biological modelling, one might feel that the set-up in Section 2.2, which in particular enforces that the number of reproductive partners of a typical individual is essentially Poisson distributed (and hence the number of potential offspring has a compound Poisson distribution), is somewhat restrictive. A natural generalisation of (2.13) would be the following: Let D_i be independent copies of an \mathbb{N}_0 -valued random variable D with $\mathbb{E}[e^{\lambda D}] < \infty$ for λ in a neighbourhood of 0; we think of D_i as the number of reproductive partners of individual i . Use the configuration model to assign reproductive partners, i.e. attach D_i “half-edges” to node i and then randomly match all half-edges conditional on producing no self-loops (if $D_1 + \dots + D_N$ is odd, throw away the last half-edge, say). Finally replace each resulting edge e by a random number X_e of edges where X_e are independent copies of X . If $\mathbb{E}[X]\mathbb{E}[D] > 2$ and X satisfies (2.16) or (2.17), then a suitable analogue of Proposition 2.5 will hold. We do not go into detail here but note that by Theorem 1.1, asymptotically for our study of genealogies, only the joint law of $(V_i)_{1 \leq i \leq N}$, which in the language of random graphs corresponds to the empirical degree distribution, is important. In the extension of the model from Section 2.2 just sketched, this will again on the relevant time-scales be dominated by one exceptionally large value of X_e if (2.17) holds and negligible compared to N if (2.16) holds. See e.g. [30] for background on random graphs, which is currently a very active research topic.

Obviously, these models allow various generalisations where the “degree of promiscuity” can be chosen as a parameter: One could for example assign $a\%$ of the children to fixed couples as in the model from Section 2.3.1 and the remaining $(100 - a)\%$ of the children by using a “configuration model” as just discussed.

In most of the models discussed so far we did not include individuals of different sexes. However, as described in Section 2.3.1 two-sex models, possibly with unequal sex ratio $r : 1 - r$, can be in principle easily embedded into our set-up: take a random bi-partite “exchangeable” multigraph on $\lfloor rN \rfloor$ and $N - \lfloor rN \rfloor$ nodes with N edges (equivalently: a separately exchangeable $\lfloor rN \rfloor \times (N - \lfloor rN \rfloor)$ -matrix, with values in \mathbb{N}_0 , summing to N), assign individuals $i = 1, \dots, N$ randomly to the two “sex groups”. One can combine this with small variations of all the models discussed in Section 2 and one can for example also incorporate differences in the variance of reproductive success between the two sexes in this class of models.

One can also allow the possibility of selfing, i.e. $\mathbb{P}(V_{i,i} > 0) > 0$. Then complete dispersion will not happen (asymptotically) immediately but only after a certain random number of sampled genes have merged due to selfing, analogous to [17]. We leave the details to future work.

3 Proofs

3.1 Proof of Theorem 1.1 and Corollary 1.2

In this section we prove our main convergence result Theorem 1.1 and Corollary 1.2.

3.1.1 The pair coalescence probability

We start by analyzing the pair coalescence probability c_N . Recall that this is the probability that two genes (picked at random) from two distinct individuals, which are chosen randomly without replacement from the same generation (in the population of size N), have a common ancestor (gene) in the previous generation.

Lemma 3.1. *We have*

$$c_N = \frac{1}{8} \mathbb{E} \left[V_{1,2}^2 - \frac{2}{N-1} \right] + \frac{N-2}{8} \mathbb{E} [V_{1,2} V_{1,3}] = \frac{1}{8} \mathbb{E} [V_{1,2} (V_1 - 1)] = \frac{1}{8} \frac{1}{N-1} \mathbb{E} [(V_1)_2]. \quad (3.1)$$

Proof. Pick two distinct individuals at random from the current population and pick from each of them independently one of the two gene copies at random by a fair coin flip. These two genes may be descended from the same ancestral gene in the previous generation if the two individuals have both parents or just one parent in common (full siblings, half siblings). The probabilities that the two genes are descended from the same ancestral gene of one of the parent individuals is then $\frac{1}{4}$ and $\frac{1}{8}$ respectively. Thus

$$\begin{aligned} c_N &= \mathbb{E} \left[\sum_{1 \leq i < j \leq N} \left[\frac{V_{i,j}(V_{i,j} - 1)}{N(N-1)} \frac{1}{4} + \frac{V_{i,j}(V_i - V_{i,j})}{N(N-1)} \frac{1}{8} + \frac{V_{i,j}(V_j - V_{i,j})}{N(N-1)} \frac{1}{8} \right] \right] \\ &= \frac{1}{8} \frac{1}{N(N-1)} \sum_{1 \leq i < j \leq N} \mathbb{E} [V_{i,j} (V_i + V_j - 2)] = \frac{1}{8} \frac{\binom{N}{2}}{N(N-1)} \mathbb{E} [V_{1,2} (V_1 + V_2 - 2)] \\ &= \frac{1}{8} \mathbb{E} [V_{1,2} (V_1 - 1)] = \frac{1}{8(N-1)} \mathbb{E} \left[\sum_{j=2}^N V_{1,j} (V_1 - 1) \right] = \frac{1}{8} \frac{1}{N-1} \mathbb{E} [(V_1)_2] \end{aligned}$$

by the exchangeability assumptions (1.1). Alternatively write

$$\frac{1}{8} \mathbb{E} [V_{1,2} (V_1 - 1)] = \frac{1}{8} \sum_{j=2}^N \mathbb{E} [V_{1,2} V_{1,j}] - \frac{1}{8} \mathbb{E} [V_{1,2}] = \frac{1}{8} \mathbb{E} [V_{1,2}^2] - \frac{1}{8} \frac{2}{N-1} + \frac{N-2}{8} \mathbb{E} [V_{1,2} V_{1,3}]$$

to obtain the first equality in (3.1). □

Note that one can also express c_N in terms of variance and covariances of $V_{i,j}$, as follows:

$$c_N = \frac{1}{8} \text{Var}[V_{1,2}] + \frac{1}{8} (N - 2) \text{Cov}[V_{1,2}, V_{1,3}] + \frac{1}{4(N - 1)}.$$

3.1.2 Transition probabilities

Next, we analyze the transition probabilities of the ancestral process. Let $\Pi_{n,N} = (\pi_{n,N}(\xi, \eta))_{\xi, \eta \in \mathcal{S}_n}$ be the transition matrix of the Markov chain $(\xi^{n,N}(m))_{m \in \mathbb{N}_0}$. ($\Pi_{n,N}$ can be viewed as an $|\mathcal{S}_n| \times |\mathcal{S}_n|$ -matrix if we fix an order for \mathcal{S}_n , which we will do later). In particular, we see from the argument in Lemma 3.1 that $c_N = \pi_{2,N}(\{\{1\}, \{2\}\}, \{\{1, 2\}\})$.

It turns out that for our purposes it is sufficient to describe the $\pi_{n,N}(\xi, \eta)$ in case that $\xi \in \mathcal{E}_n$. For some $b \leq n$ we thus consider the transition probability from states $\xi \in \mathcal{E}_n$ to $\eta \in \mathcal{S}_n$ of the form

$$\xi = \{C_1, \dots, C_b\} \quad \text{and} \tag{3.2}$$

$$\eta = \{\{D_1, D_2\}, \dots, \{D_{2d-1}, D_{2d}\}, D_{2d+1}, \dots, D_a\}, \tag{3.3}$$

for some $a \leq b$ and $2d \leq a$ such that $\xi \subseteq \text{cd}(\eta)$. Assume that D_i is a union of $k_i \geq 1$ classes from ξ with $k_1 + \dots + k_a = b$.

Denote by $\mathcal{E}_{a,d}$ the collection of elements in \mathcal{E}_a with $a - d$ blocks, d of which have cardinality 2 while the other $a - 2d$ have cardinality 1. Then we can describe the grouping into diploid individuals in η via $\zeta = \{\zeta_1, \dots, \zeta_{a-d}\} \in \mathcal{E}_{a,d}$: Let D_i and D_j belong to the same diploid ancestral individual in η if and only if $\{i, j\} \in \zeta$. We put

$$\ell_j := \sum_{i \in \zeta_j} k_i, \quad j = 1, \dots, a - d, \tag{3.4}$$

which is the number of offspring classes in ξ that belong to the j -th ancestral individual described by η , and we have therefore that $\sum_{j=1}^{a-d} \ell_j = b$.

In order to calculate and describe the transition probabilities we will introduce another useful concept and corresponding notation. Recall that $V_{i,j}$ ($= V_{j,i}$) represents the random number of offspring of the parental individuals i and j . By definition, each offspring inherits one chromosome copy from each parent. Let us assume that we randomly and uniformly “mark” one of these chromosome copies as “relevant” (in the sense that it will be this copy that we possibly later examine in the child), let $\widehat{V}_{i,j}$ be the number of offspring with parents from i and j who inherited their relevant chromosome copy from parent i . Mathematically, this means that conditional on $(V_{u,w})_{1 \leq u, w \leq N}$, $\widehat{V}_{i,j}$ is $\text{Bin}(V_{i,j}, \frac{1}{2})$ -distributed, $\widehat{V}_{j,i} = V_{i,j} - \widehat{V}_{i,j}$ and $\widehat{V}_{i,j}$ and $\widehat{V}_{k,\ell}$ are independent when $\{i, j\} \neq \{k, \ell\}$. We will write

$$\widehat{V}_i := \sum_{j \neq i}^N \widehat{V}_{i,j}$$

for the total number of “relevant” offspring of individual i . Note that we have $\sum_{i=1}^N \widehat{V}_i = N$ by definition.

Lemma 3.2. *Both the array $(\widehat{V}_{i,j})_{1 \leq i \neq j \leq N}$ of relevant pairwise offspring numbers and the vector $(\widehat{V}_i)_{1 \leq i \leq N}$ of relevant total offspring numbers are exchangeable.*

Proof. Note that

$$\begin{aligned} & \mathbb{P}\left(\widehat{V}_{i,j} = \widehat{v}_{i,j}, 1 \leq i \neq j \leq N\right) \\ &= \mathbb{P}\left(V_{i,j} = \widehat{v}_{i,j} + \widehat{v}_{j,i}, 1 \leq i < j \leq N\right) \prod_{1 \leq i < j \leq n} \binom{\widehat{v}_{i,j} + \widehat{v}_{j,i}}{\widehat{v}_{i,j}} 2^{-(\widehat{v}_{i,j} + \widehat{v}_{j,i})}. \end{aligned} \tag{3.5}$$

For any permutation σ on \mathbb{V}_N , we have

$$\begin{aligned} & \mathbb{P}\left(\widehat{V}_{i,j} = \widehat{v}_{\sigma(i),\sigma(j)}, 1 \leq i \neq j \leq N\right) \\ &= \mathbb{P}\left(V_{i,j} = \widehat{v}_{\sigma(i),\sigma(j)} + \widehat{v}_{\sigma(j),\sigma(i)}, 1 \leq i < j \leq N\right) \\ & \quad \cdot \prod_{1 \leq i < j \leq n} \binom{\widehat{v}_{\sigma(i),\sigma(j)} + \widehat{v}_{\sigma(j),\sigma(i)}}{\widehat{v}_{\sigma(i),\sigma(j)}} 2^{-(\widehat{v}_{\sigma(i),\sigma(j)} + \widehat{v}_{\sigma(j),\sigma(i)})}. \end{aligned} \tag{3.6}$$

It follows from (1.1) that (3.5) equals (3.6), i.e. $\left(\widehat{V}_{\sigma(i),\sigma(j)}\right)_{1 \leq i \neq j \leq N} \stackrel{d}{=} \left(\widehat{V}_{i,j}\right)_{1 \leq i \neq j \leq N}$.

Exchangeability of $(\widehat{V}_i)_{1 \leq i \leq N}$ follows from this as in (1.3). □

The following lemma states that if the limits in (1.6) exist then they can also be expressed in terms of the quantities $(\widehat{V}_i)_{i=1,\dots,N}$ instead of the $(V_i)_{i=1,\dots,N}$.

Lemma 3.3. *Let $b, c \in \mathbb{N}$, $\ell_1, \dots, \ell_c \in \mathbb{N}$ with $\ell_1 + \dots + \ell_c = b$. Under Condition (1.6), we have*

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[(\widehat{V}_1)_{\ell_1} \cdots (\widehat{V}_c)_{\ell_c}]}{c_N N^{b-c}} = \lim_{N \rightarrow \infty} \frac{\mathbb{E}[(V_1)_{\ell_1} \cdots (V_c)_{\ell_c}]}{c_N N^{b-c} 2^b}.$$

Proof. Recall the combinatorial identity for choosing (without replacement) ℓ objects out of $\sum_{i=1}^n a_i$ objects

$$\binom{\sum_{i=1}^n a_i}{\ell} = \frac{1}{\ell!} (\sum_{i=1}^n a_i)_{\ell} = \sum_{\substack{(k_1, \dots, k_n) \in \mathbb{N}_0^n \\ k_1 + \dots + k_n = \ell}} \prod_{i=1}^n \frac{(a_i)_{k_i}}{k_i!}, \quad \ell, n \in \mathbb{N}, a_1, \dots, a_n \in \mathbb{N}_0, \tag{3.7}$$

which results from choosing exactly k_i out of a_i objects and from considering all the possible choices of $k_1 + \dots + k_n = \ell$. Now, set

$$\begin{aligned} & \mathcal{K}_N(\ell_1, \dots, \ell_c) \\ & := \left\{ (k_{i,j})_{i=1,\dots,c, j=1,\dots,N} \in \mathbb{N}_0^{c \times N} : k_{i,1} + \dots + k_{i,N} = \ell_i \text{ with } k_{i,i} = 0 \text{ for } i = 1, \dots, c \right\}. \end{aligned}$$

Then, expanding the definition of \widehat{V}_i and using (3.7) yield

$$\begin{aligned} \prod_{i=1}^c \frac{(\widehat{V}_i)_{\ell_i}}{\ell_i!} &= \prod_{i=1}^c \left\{ \sum_{\substack{(k_{i,1}, \dots, k_{i,N}) \in \mathbb{N}_0^N \\ k_{i,1} + \dots + k_{i,N} = \ell_i, k_{i,i} = 0}} \prod_{j \neq i} \frac{(\widehat{V}_{i,j})_{k_{i,j}}}{k_{i,j}!} \right\} = \sum_{(k_{i,j}) \in \mathcal{K}_N(\ell_1, \dots, \ell_c)} \prod_{i=1}^c \left\{ \prod_{j \neq i} \frac{(\widehat{V}_{i,j})_{k_{i,j}}}{k_{i,j}!} \right\} \\ &= \sum_{(k_{i,j}) \in \mathcal{K}_N(\ell_1, \dots, \ell_c)} \left\{ \prod_{1 \leq i < j \leq c} \frac{(\widehat{V}_{i,j})_{k_{i,j}}}{k_{i,j}!} \frac{(V_{i,j} - \widehat{V}_{i,j})_{k_{j,i}}}{k_{j,i}!} \right\} \times \left\{ \prod_{i=1}^c \prod_{j=c+1}^N \frac{(\widehat{V}_{i,j})_{k_{i,j}}}{k_{i,j}!} \right\} \end{aligned} \tag{3.8}$$

Thus,

$$\begin{aligned} & \mathbb{E} \left[\prod_{i=1}^c \frac{(\widehat{V}_i)^{\ell_i}}{\ell_i!} \mid (V_{u,w})_{1 \leq u,w \leq N} \right] \\ &= \sum_{(k_{i,j}) \in \mathcal{K}_N(\ell_1, \dots, \ell_c)} \left\{ \prod_{1 \leq i < j \leq c} \frac{1}{2^{k_{i,j} + k_{j,i}}} \frac{(V_{i,j})_{k_{i,j} + k_{j,i}}}{k_{i,j}! k_{j,i}!} \right\} \times \left\{ \prod_{i=1}^c \prod_{j=c+1}^N \frac{1}{2^{k_{i,j}}} \frac{(V_{i,j})_{k_{i,j}}}{k_{i,j}!} \right\} \\ &= \frac{1}{2^b} \sum_{(k_{i,j}) \in \mathcal{K}_N(\ell_1, \dots, \ell_c)} \left\{ \prod_{1 \leq i < j \leq c} \frac{(V_{i,j})_{k_{i,j} + k_{j,i}}}{k_{i,j}! k_{j,i}!} \right\} \times \left\{ \prod_{i=1}^c \prod_{j=c+1}^N \frac{(V_{i,j})_{k_{i,j}}}{k_{i,j}!} \right\} \end{aligned} \quad (3.9)$$

where we have used the fact that

$$\mathbb{E} \left[(\widehat{V}_{i,j})_{k_{i,j}} (V_{i,j} - \widehat{V}_{i,j})_{k_{j,i}} \mid (V_{u,w})_{1 \leq u,w \leq N} \right] = \frac{1}{2^{k_{i,j} + k_{j,i}}} (V_{i,j})_{k_{i,j} + k_{j,i}}$$

(a special case of an identity for mixed factorial moments of a multinomial vector, see also Lemma 3.11) and the conditional independence properties of $\widehat{V}_{i,j}$'s in the first equation. Note that if we replace in (3.9) the term

$$\prod_{1 \leq i < j \leq c} \frac{(V_{i,j})_{k_{i,j} + k_{j,i}}}{k_{i,j}! k_{j,i}!} \quad \text{by} \quad \prod_{1 \leq i < j \leq c} \frac{(V_{i,j})_{k_{i,j}} (V_{i,j})_{k_{j,i}}}{k_{i,j}! k_{j,i}!} \quad (3.10)$$

then we obtain (as in (3.8)),

$$\frac{1}{2^b} \sum_{(k_{i,j}) \in \mathcal{K}_N(\ell_1, \dots, \ell_c)} \prod_{i=1}^c \prod_{\substack{j=1 \\ j \neq i}}^N \frac{(V_{i,j})_{k_{i,j}}}{k_{i,j}!} = \frac{1}{2^b} \prod_{i=1}^c \frac{(V_i)^{\ell_i}}{\ell_i!}. \quad (3.11)$$

The difference of the two terms $(V_{i,j})_{k_{i,j} + k_{j,i}}$ and $(V_{i,j})_{k_{i,j}} (V_{i,j})_{k_{j,i}}$ which get replaced inside the product in (3.10) vanishes whenever $k_{i,j} + k_{j,i} \leq 1$ and is

$$O((V_{i,j})^{k_{i,j} + k_{j,i} - 1})$$

otherwise (with a combinatorial constant that depends on c and ℓ_1, \dots, ℓ_c but not on N). Thus,

$$\frac{1}{c_N N^{b-c}} \left(\mathbb{E}[\text{term in (3.9)}] - \mathbb{E}[\text{term on r.h.s. of (3.11)}] \right) \rightarrow 0 \quad \text{as } N \rightarrow \infty$$

because of Condition (1.6), which is the claim. \square

With the help of this lemma we can now prove the following:

Lemma 3.4. For ξ, η as in (3.2), (3.3) with ℓ_j from (3.4), we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{\pi_{n,N}(\xi, \eta)}{c_N} &= \lim_{N \rightarrow \infty} \sum_{\substack{i_1, \dots, i_{a-d}=1 \\ \text{distinct}}}^N \frac{\mathbb{E}[(V_{i_1})_{\ell_1} (V_{i_2})_{\ell_2} \cdots (V_{i_{a-d}})_{\ell_{a-d}}] 2^{a-d}}{c_N N^b 2^{2b}} \\ &= \lim_{N \rightarrow \infty} \frac{\mathbb{E}[(V_1)_{\ell_1} (V_2)_{\ell_2} \cdots (V_{a-d})_{\ell_{a-d}}]}{c_N 2^b (2N)^{b-a+d}}. \end{aligned} \quad (3.12)$$

If $\ell_1, \dots, \ell_{a-d} \geq 2$ we see from (1.6) that the limit in (3.12) equals

$$\phi_{a-d}(\ell_1, \dots, \ell_{a-d}) \cdot \frac{1}{2^{b-a+d}}.$$

When $s \geq 1$ of the ℓ_i are equal to 1, say $\ell_1, \dots, \ell_{a-d-s} \geq 2, \ell_{a-d-s+1} = \dots = \ell_{a-d} = 1$ we see from (1.8) that the limit in (3.12) equals

$$\psi_{a-d-s,s}(\ell_1, \dots, \ell_{a-d-s}) \cdot \frac{1}{2^{b-a+d}}.$$

Proof of Lemma 3.4. Since $\xi \in \mathcal{E}_n \subset \mathcal{S}_n$ is a completely dispersed state, its b classes belong to b distinct individuals in the offspring generation and we can think of the chromosome copies which belong to ξ as the relevant ones (in the corresponding individuals), i.e., the transition from ξ to η corresponds to drawing b times without replacement from an urn which contains \widehat{V}_i balls of colour i for $i = 1, \dots, N$. Thus

$$\pi_{n,N}(\xi, \eta) = \sum_{\substack{i_1, \dots, i_{a-d}=1 \\ \text{distinct}}}^N \mathbb{E} \left[\frac{\prod_{r=1}^{a-d} (\widehat{V}_{i_r})_{\ell_r}}{(N)_b} \times \prod_{r=1}^{a-d} 2^{-\ell_r+1} \right] = 2^{-b+a-d} \frac{(N)_{a-d}}{(N)_b} \mathbb{E} \left[\prod_{r=1}^{a-d} (\widehat{V}_r)_{\ell_r} \right]. \tag{3.13}$$

Note that the factor $\prod_{r=1}^{a-d} 2^{-\ell_r+1}$ accounts for the fact that we still have to assign for each $r = 1, \dots, a-d$ which of the ℓ_r classes in $\cup_{i \in \zeta_r} D_i$ descends from which of the two chromosome copies in the i_r -th ancestral individual (this makes ℓ_r assigned picks if we decree who descends from the “first” and who from the “second” chromosome in individual i_r but we gain a factor of 2 because the roles of the “first” and the “second” chromosome are arbitrary and can be swapped). The second equation is a consequence of exchangeability of the \widehat{V}_i . Finally, (3.12) follows from (3.13) and Lemma 3.3. \square

For $n \in \mathbb{N}$, $\xi, \eta \in \mathcal{E}_n$, put

$$\widetilde{\pi}_{n,N}(\xi, \eta) := \sum_{\{\eta' \in \mathcal{S}_n : \text{cd}(\eta') = \eta\}} \pi_{n,N}(\xi, \eta'). \tag{3.14}$$

Note that $\widetilde{\pi}_{n,N}$ is a Markov transition matrix on \mathcal{E}_n , a step according to $\widetilde{\pi}_{n,N}$ means first taking a step according to $\pi_{n,N}$ and then applying the complete dispersion operator, i.e., ignoring the grouping into diploid individuals.

In particular, there is sampling consistency: For $\xi \in \mathcal{E}_n$, $\eta = \{D_1, \dots, D_a\} \in \mathcal{E}_n$ we have

$$\begin{aligned} \widetilde{\pi}_{n,N}(\xi, \eta) &= \widetilde{\pi}_{n+1,N}(\xi \cup \{\{n+1\}\}, \eta \cup \{\{n+1\}\}) \\ &+ \sum_{i=1}^a \widetilde{\pi}_{n+1,N}(\xi \cup \{\{n+1\}\}, \{D_1, \dots, D_{i-1}, D_i \cup \{n+1\}, D_{i+1}, \dots, D_a\}). \end{aligned} \tag{3.15}$$

Furthermore, the transition probabilities depend on n only implicitly through the merger structure that the transition from ξ to η induces.

Lemma 3.5. *Let $n \in \mathbb{N}$, $\xi \in \mathcal{E}_n$, $\eta \in \mathcal{S}_n$ where ξ has b classes and $\text{cd}(\eta)$ arises from ξ by merging $j \geq 1$ groups of classes with sizes $k_1, k_2, \dots, k_j \geq 2$ from ξ and leaving $s \geq 0$ singleton classes (in particular, η has $a = j + s$ classes and $b = k_1 + \dots + k_j + s$). Then*

$$\lim_{N \rightarrow \infty} \frac{1}{c_N} \widetilde{\pi}_{n,N}(\xi, \text{cd}(\eta)) = \lambda_{b; k_1, \dots, k_j; s} \tag{3.16}$$

where $\lambda_{b; k_1, \dots, k_j; s}$ are the transition rates of the Ξ -coalescent defined in Theorem 1.1 (recalled in (1.9)).

Proof. Assume ξ and η are given by (3.2) and (3.3) with classes denoted by C_i and D_i , respectively. Recall that D_i is a union of $k_i \geq 1$ classes from ξ with $k_1 + \dots + k_a = b$ and that ζ describes the grouping into diploid individuals.

Now note that $\text{cd}^{-1}(\text{cd}(\eta)) := \{\eta' \in \mathcal{S}_n : \text{cd}(\eta') = \text{cd}(\eta)\}$ can be parametrised by choosing any $d \in \{0, 1, \dots, \lfloor a/2 \rfloor\}$ and $\zeta \in \mathcal{E}_{a,d}$ (d describes the number of ancestral individuals in η carrying two ancestral genes and ζ describes the grouping of the a

ancestral chromosomes in η into diploid individuals). For $\zeta \in \mathcal{E}_{a,d}$ and $i = 1, 2, \dots, a$, we define $\widehat{\zeta}(i) = k \in \{1, \dots, a-d\}$ if $i \in \zeta_k$.

First consider the case $s = 0$ and hence $a = j > 1$: From (3.14), Lemma 3.4, as well as (1.6) and (1.7),

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{c_N} \widetilde{\pi}_{n,N}(\xi, \text{cd}(\eta)) \\ &= \sum_{d=0}^{\lfloor a/2 \rfloor} \sum_{\zeta \in \mathcal{E}_{a,d}} 2^{-b+a-d} \int_{\Delta} \sum_{\substack{i_1, \dots, i_{a-d}=1 \\ \text{distinct}}}^{\infty} \prod_{\ell=1}^{a-d} x_{i_{\ell}}^{\left(\sum_{r \in \zeta_{\ell}} k_r\right)} \frac{2\Xi'(dx)}{(x, x)} \\ &= \int_{\Delta} \sum_{d=0}^{\lfloor a/2 \rfloor} 2^{a-d} \sum_{\substack{i_1, \dots, i_{a-d}=1 \\ \text{distinct}}}^{\infty} \sum_{\zeta \in \mathcal{E}_{a,d}} \left(\frac{1}{2}x_{i_{\widehat{\zeta}(1)}}\right)^{k_1} \left(\frac{1}{2}x_{i_{\widehat{\zeta}(2)}}\right)^{k_2} \cdots \left(\frac{1}{2}x_{i_{\widehat{\zeta}(a)}}\right)^{k_a} \frac{2\Xi'(dx)}{(x, x)} \\ &= \int_{\Delta} \sum_{\substack{i'_1, \dots, i'_a=1 \\ \text{distinct}}}^{\infty} (\varphi(x)_{i'_1})^{k_1} (\varphi(x)_{i'_2})^{k_2} \cdots (\varphi(x)_{i'_a})^{k_a} \frac{\Xi'(dx)}{(\varphi(x), \varphi(x))} \\ &= \int_{\Delta} \sum_{\substack{i'_1, \dots, i'_a=1 \\ \text{distinct}}}^{\infty} x_{i'_1}^{k_1} \cdots x_{i'_a}^{k_a} \frac{\Xi(dx)}{(x, x)} = \lambda_{b; k_1, \dots, k_a; 0} \end{aligned} \tag{3.17}$$

where we used that by definition of φ , for any function $F : [0, 1]^a \rightarrow [0, \infty)$ and $x = (x_1, x_2, \dots) \in \Delta$

$$\begin{aligned} & \sum_{\substack{i'_1, \dots, i'_a=1 \\ \text{distinct}}}^{\infty} F(\varphi(x)_{i'_1}, \varphi(x)_{i'_2}, \dots, \varphi(x)_{i'_a}) \\ &= \sum_{d=0}^{\lfloor a/2 \rfloor} 2^{a-d} \sum_{\substack{i_1, \dots, i_{a-d}=1 \\ \text{distinct}}}^{\infty} \sum_{\zeta \in \mathcal{E}_{a,d}} F\left(\frac{1}{2}x_{i_{\widehat{\zeta}(1)}}, \frac{1}{2}x_{i_{\widehat{\zeta}(2)}}, \dots, \frac{1}{2}x_{i_{\widehat{\zeta}(a)}}\right) \end{aligned}$$

and that $(\varphi(x), \varphi(x)) = \frac{1}{2}(x, x)$. In the case $j = 1$, we additionally have the Kingman term in (3.17) of the form

$$\mathbb{I}_{\{j=1, k_1=2\}} 2 \Xi'(\{\mathbf{0}\}) 2^{-b+a-d} = \mathbb{I}_{\{j=1, k_1=2\}} \Xi(\{\mathbf{0}\})$$

since $-b + a - d = -1$ in this case and $\Xi'(\{\mathbf{0}\}) = \Xi(\{\mathbf{0}\})$. Thus, (3.16) holds when $s = 0$.

For the general case $s > 0$, we can employ the consistency relations (3.15) in order to use induction on s : Assume that (3.16) holds whenever the number of “singleton classes” involved is at most s , and $b = k_1 + \dots + k_j + s$ with $k_1, \dots, k_j \geq 2$. Let $\text{cd}(\eta) \in \mathcal{E}_{n+1}$ arise from $\xi \in \mathcal{E}_{n+1}$ by a merger in j groups of sizes $k_1, \dots, k_j \geq 2$, leaving $s + 1$ singleton classes. By the symmetries of the model, we may (without changing the transition probability) assume that one of the relevant singleton classes in η is $\{n + 1\}$. Then, rearranging (3.15) and using the induction hypothesis we see that

$$\lim_{N \rightarrow \infty} \frac{\widetilde{\pi}_{n+1,N}(\xi, \text{cd}(\eta))}{c_N} = \lambda_{b; k_1, \dots, k_j; s} - \sum_{i=1}^j \lambda_{b+1; k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_j; s} - s \lambda_{b+1; k_1, \dots, k_j, 2; s-1}.$$

The term on the right-hand side equals $\lambda_{b+1; k_1, \dots, k_j; s+1}$ by the consistency relation for transition probabilities of Ξ -coalescents (implicit in [22, Eq. (11) and Lemma 3.4], explicitly spelled out for example in [27, Eq. (2.5)]). \square

A central ingredient in the proof of Theorem 1.1 is the following result from [17]. It makes the separation of time-scales behind Theorem 1.1 explicit: On the fast time scale $O(1)$, any diploid sample configuration is transformed into its complete dispersion whereas on the much slower time scale $O(1/c_N)$, non-trivial merging occurs.

Lemma 3.6 (Möhle [17]). *Let $X_N = (X_N(m))_{m \in \mathbb{N}_0}$ be a sequence of time homogeneous Markov chains on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the same finite state space S and let Π_N denote the transition matrix of X_N . Assume that the following conditions are satisfied:*

1. $A := \lim_{N \rightarrow \infty} \Pi_N$ exists and $\Pi_N \neq A$ for all sufficiently large N .
2. $P := \lim_{m \rightarrow \infty} A^m$ exists.
3. $G := \lim_{N \rightarrow \infty} PB_N P$ exists, where $B_N := (\Pi_N - A)/c_N$ and $c_N := \|\Pi_N - A\|$ for all $N \in \mathbb{N}$.

If the sequence of initial probability measures $\mathbb{P}_{X_N(0)}$ converge weakly to some probability measure μ , then the finite dimensional distributions of the process $(X_N(\lfloor t/c_N \rfloor))_{t \geq 0}$ converge to those of a time continuous Markov process $(X_t)_{t \geq 0}$ with initial distribution

$$X_0 \stackrel{d}{=} \mu,$$

transition matrix $\Pi(t) := P - I + e^{tG} = Pe^{tG}$, $t > 0$, and infinitesimal generator G .

Proof of Theorem 1.1: Applying Lemma 3.6, the strategy to prove our result is based on the decomposition of the transition matrix $\Pi_{n,N} = (\pi_{n,N}(\xi', \eta'))_{\xi', \eta' \in \mathcal{S}_n}$. In order to have these transitions be well defined as matrices we choose a specific order of \mathcal{S}_n : Namely, consider the standard order of $\mathcal{E}_n \subset \mathcal{S}_n$. We will then insert all remaining elements of $cd^{-1}(\xi) \subset \mathcal{S}_n$ directly following any $\xi \in \mathcal{E}_n$ (the order here is fixed in an arbitrary way). This way, the matrix $\Pi_{n,N}$ decomposes into sub-matrices $(\tilde{\Pi}_{n,N}(\xi, \eta))_{\xi, \eta \in \mathcal{E}_n}$ with $\tilde{\Pi}_{n,N}(\xi, \eta)$ a $|cd^{-1}(\xi)| \times |cd^{-1}(\eta)|$ matrix.

What we need to do is to find the decomposition such that

$$\Pi_{n,N} = A + c_N B_{n,N},$$

where $A = \lim_{N \rightarrow \infty} \Pi_{n,N}$ does not depend on N , $c_N \rightarrow 0$ and $B_{n,N}$ is bounded. In the sense of sub-matrix structure, it is necessary to find the decomposition such that

$$\tilde{\Pi}_{n,N}(\xi, \eta) = \tilde{A}(\xi, \eta) + c_N \tilde{B}_{n,N}(\xi, \eta),$$

for $\xi, \eta \in \mathcal{E}_n$, where both $\tilde{A}(\xi, \eta)$ and $\tilde{B}_{n,N}(\xi, \eta)$ are $|cd^{-1}(\xi)| \times |cd^{-1}(\eta)|$ sub-matrices of A and $B_{n,N}$. Set $\tilde{A}(\xi, \eta) = 0$ for $\xi \neq \eta$ and

$$\tilde{A}(\xi, \xi) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} := P_\xi.$$

Note that this matrix maps any $\xi' \in cd^{-1}(\xi)$ to $\xi \in \mathcal{E}_n$. It follows that

$$\lim_{m \rightarrow \infty} \tilde{A}^m(\xi, \xi) = \tilde{A}(\xi, \xi) := P_\xi.$$

Thus, we have $P := \lim_{m \rightarrow \infty} A^m$ with sub matrix structure $\tilde{P}(\xi, \eta) = \mathbf{0}$ for $\xi \neq \eta$ and $\tilde{P}(\xi, \xi) = P_\xi$. It is easy to show that

$$G := \lim_{N \rightarrow \infty} P B_{n,N} P = \lim_{N \rightarrow \infty} (P_\xi B_{n,N}(\xi, \eta) P_\eta)_{\xi, \eta \in \mathcal{E}_n}$$

$$= \lim_{N \rightarrow \infty} \left(\begin{pmatrix} g_{\xi, \eta}^{(N)} & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ g_{\xi, \eta}^{(N)} & 0 & \cdots & 0 \end{pmatrix} \right)_{\xi, \eta \in \mathcal{E}_n}$$

where $g_{\xi, \eta}^{(N)}$ is the sum of all the entries in the first row of matrix $B_{n,N}(\xi, \eta)$. Consequently,

$$g_{\xi, \eta}^{(N)} = \sum_{\eta' \in \text{cd}^{-1}(\eta)} \frac{\pi_{n,N}(\xi, \eta')}{c_N} - \frac{\delta_{\xi, \eta}}{c_N}.$$

Assume η arises from ξ by merging $j \geq 1$ groups of classes with sizes $k_1, k_2, \dots, k_j \geq 2$ and leaving $s \geq 0$ singleton classes ($b = k_1 + \dots + k_j + s$). Applying Lemma 3.5, we have

$$\lim_{N \rightarrow \infty} g_{\xi, \eta}^{(N)} = \lambda_{b; k_1, \dots, k_j; s}.$$

Note that a transition from any given state $\eta' \in \mathcal{S}_n$ with $a \leq n$ classes (as in (3.3)) to its complete dispersion state $\text{cd}(\eta')$ happens whenever none of the ancestral genes of distinct ancestral individuals in configurations η' have a common parental ancestral gene in the previous generation. (Ancestral genes of the same individual naturally have distinct parental ancestral genes as we have excluded selfing.) For any pair of such ancestral genes the probability to have a common parental ancestral gene is c_N and there are at most $\binom{a}{2}$ such pairs to consider. Thus, the transition probability satisfies

$$\pi_{n,N}(\eta', \text{cd}(\eta')) \geq 1 - \binom{a}{2} c_N.$$

It is clear that $\pi_{n,N}(\eta', \text{cd}(\eta')) \rightarrow 1$ as $N \rightarrow \infty$ or in other words that $A = \lim_{N \rightarrow \infty} \Pi_{n,N}$. Hence, complete dispersion happens instantaneously in the limit of the genealogical process for the diploid population model. By eliminating all those instantaneous states, we can get an \mathcal{E}_n -valued marginal process $(R_n(t))_{t \geq 0}$ whose generator is given by $(\lim_{N \rightarrow \infty} g_{\xi, \eta}^{(N)})_{\xi, \eta \in \mathcal{E}_n}$. The process $(R_n(t))_{t \geq 0}$ is exactly the n - Ξ -coalescent process. \square

Proof of Corollary 1.2: Our argument is essentially borrowed from the proof of Theorem 3.1 in Möhle [19]. Theorem 1.1 yields that the finite-dimensional distributions of

$$(\tilde{\xi}^{n,N}(\lfloor t/c_N \rfloor))_{t \geq 0} = (\text{cd}(\xi^{n,N}(\lfloor t/c_N \rfloor)))_{t \geq 0}$$

converge to those of ξ^n . Thus, in order to strengthen this to weak convergence on the path space $D([0, \infty), \mathcal{E}_n)$ we only have to verify tightness there. Since \mathcal{E}_n is finite (and so in particular compact) and $\tilde{\xi}^{n,N}$ can by construction only move by mergers, it suffices to check that in the limit $N \rightarrow \infty$, the jump times of $(\text{cd}(\xi^{n,N}(\lfloor \cdot/c_N \rfloor)))$ do not accumulate (see e.g. [10, Thm. 6.2 in Ch. 3] or [5]). Noting that for any $\xi \in \mathcal{S}_n$

$$\mathbb{P}(\text{cd}(\xi^{n,N}(m+1)) \neq \text{cd}(\xi) \mid \xi^{n,N}(m) = \xi)$$

$$= \mathbb{P}(\text{at least one pair of genes (necessarily from distinct individuals) merges in the next step} \mid \xi^{n,N}(m) = \xi) \leq \binom{n}{2} c_N$$

we see that the times between jumps of $(\tilde{\xi}^{n,N}(m))_{m \in \mathbb{N}_0}$ are stochastically larger than independent geometric random variables with success parameter $c_N n(n-1)/2$ which after time scaling converge in distribution to independent exponentials with rate $n(n-1)/2$. \square

3.2 Proof of Proposition 2.3

In this section we prove Lemma 2.2 and Proposition 2.3, the main convergence result for the diploid population model with random individual fitness of Section 2.1. Apart from Lemma 2.2 we need two auxiliary lemmas for the proof of Proposition 2.3. Their proofs are postponed to Section 3.2.1.

Note that while the set-up in Proposition 2.3 is quite similar in spirit to that from [29], the adaptation poses some additional technical difficulties. In particular, observe that Z_N , the normalising constant in the representation (2.4) of the law of $(V_{i,j})$ as a mixture of multinomials is not literally an i.i.d. sum as in [29].

Recall from (2.3) and (2.4) that in the random individual fitness model $V_{i,j}, 1 \leq i < j \leq N$ are multinomial with N trials and with success probabilities of individuals i and j proportional to $W_i W_j$ where the fitness parameters W_i are independent copies of W with mean μ_W .

Lemma 3.7. 1. If $\mu_W^{(2)} = \mathbb{E}[W^2] < \infty$ we have

$$\phi_1(3) = \lim_{N \rightarrow \infty} \frac{\mathbb{E}[(V_1)_3]}{c_N N^2} = 0. \tag{3.18}$$

2. If (2.5) holds we have

$$\phi_2(2, 2) = \lim_{N \rightarrow \infty} \frac{\mathbb{E}[(V_1)_2(V_2)_2]}{c_N N^2} = 0. \tag{3.19}$$

Lemma 3.8. If (2.5) holds we have

$$\frac{N}{c_N} \mathbb{P}(V_1 > Nx) \xrightarrow{N \rightarrow \infty} 8 \int_x^1 \frac{1}{y^2} \text{Beta}(2 - \alpha, \alpha)(dy) \quad \text{for } x \in (0, 1). \tag{3.20}$$

Proof of Proposition 2.3. 1. The scaling of the pair coalescence probability c_N is given in (2.8) in Lemma 2.2. Using Theorem 1.1 we should verify that $\mu_W^{(2)} < \infty$ implies that (cf. Condition (1.6)) for all $j \in \mathbb{N}$ and $k_1, \dots, k_j \geq 2$

$$\lim_{N \rightarrow \infty} \frac{1}{c_N} \frac{\mathbb{E}[(V_1)_{k_1} \cdots (V_j)_{k_j}]}{N^{k_1 + \dots + k_j - j} 2^{k_1 + \dots + k_j}} = 2 \mathbb{1}_{\{j=1, k_1=2\}}. \tag{3.21}$$

For $j = 1, k_1 = 2$ this follows from the fact that $c_N = \mathbb{E}[V_1(V_1 - 1)] / (8(N - 1))$ (see Lemma 3.1); for $j = 1, k_1 = 3$ it follows from Lemma 3.7; it is well known that the latter implies that (3.21) also holds for $j = 1, k_1 > 3$ and for $j \geq 2$ (see [22] and [27]).

2. In this case, the scaling of c_N is given by (2.9) in Lemma 2.2. To verify the claimed form of the limiting coalescent we should check that the probability measure Ξ' on Δ appearing in (1.5) and (1.7) is given by

$$\Xi'(A) = \int_0^1 \mathbb{1}_A(y/2, 0, 0, \dots) \frac{y^{1-\alpha}(1-y)^{\alpha-1}}{\Gamma(2-\alpha)\Gamma(\alpha)} dy,$$

i.e. Ξ' is the image measure of $\text{Beta}(2-\alpha, \alpha)$ under the mapping $[0, 1] \ni x \mapsto (x/2, 0, 0, \dots) \in \Delta$.

It is known that (3.19) from Lemma 3.7 implies that the vague limit measure of Φ_N from (1.5) is concentrated on $\tilde{\Delta} := \{(x_1, x_2, \dots) \in \Delta : x_2 = 0\}$, see [27, Cor. 2.1] (one can view $\tilde{\Delta}$ as the canonical embedding of $[0, 1]$ into Δ) so it suffices to observe that for any $x \in (0, 1/2)$ by Lemma 3.8

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{c_N} \mathbb{P}(V_{(1)} > 2Nx) &= \lim_{N \rightarrow \infty} \frac{N}{c_N} \mathbb{P}(V_1 > 2Nx) = 8 \int_{2x}^1 \frac{1}{y^2} \text{Beta}(2 - \alpha, \alpha)(dy) \\ &= 2 \int_{2x}^1 \frac{1}{(y/2)^2} \text{Beta}(2 - \alpha, \alpha)(dy) = 2 \int_0^1 \mathbb{1}_{A(x)}(y/2, 0, 0, \dots) \frac{1}{y^2/4} \frac{y^{1-\alpha}(1-y)^{\alpha-1}}{\Gamma(2-\alpha)\Gamma(\alpha)} dy \end{aligned}$$

with $A(x) = \{(y_1, y_2, \dots) \in \Delta : y_1 > x, y_2 = 0\}$. □

3.2.1 Proofs of auxiliary results

Here, we provide some details of the proofs of Lemmas 2.2, 3.7 and 3.8.

Lemma 3.9. Assume that $W \geq 0$ satisfies the tail assumption (2.5). For $k \in \{2, 3, 4, \dots\}$, we have

$$\lim_{M \rightarrow \infty} M^\alpha \mathbb{E} \left[\frac{W^k}{(W + M)^k} \right] = c_W \alpha B(k - \alpha, \alpha). \tag{3.22}$$

This is a small variation on Lemma 12 from [29], addressing the case $k = 2$ and W integer-valued. For a rough idea why the asymptotic decay rate of $\mathbb{E}[W^k/(M + W)^k]$ is $M^{-\alpha}$ note that on the event $\{W \geq M\}$, which has probability $\sim c_W M^{-\alpha}$, the integrand is almost constant.

Proof. For any bounded monotone $g \in C^1([0, \infty))$ with $g(0) = 0$ we have

$$\mathbb{E}[g(W)] = \mathbb{E} \left[\int_0^\infty \mathbb{1}(x \leq W) g'(x) dx \right] = \int_0^\infty g'(x) \mathbb{P}(W \geq x) dx.$$

Applying this with $g(x) = x^k/(x + M)^k$ hence $g'(x) = kMx^{k-1}/(x + M)^{k+1}$ we obtain

$$\mathbb{E} \left[\frac{W^k}{(W + M)^k} \right] = \int_0^\infty g'(x) \mathbb{P}(W \geq x) dx = \int_0^\infty Mk \frac{x^{k-1}}{(x + M)^{k+1}} \mathbb{P}(W \geq x) dx. \tag{3.23}$$

For every $L > 0$ we have

$$\begin{aligned} \limsup_{M \rightarrow \infty} M^\alpha \int_0^L g'(x) \mathbb{P}(W \geq x) dx &\leq \limsup_{M \rightarrow \infty} M^\alpha \int_0^L g'(x) dx \\ &= \lim_{M \rightarrow \infty} M^\alpha \frac{L^k}{(L + M)^k} = 0, \end{aligned} \tag{3.24}$$

using that $\alpha < 2$ and that $k \geq 2$. Furthermore,

$$\begin{aligned} \int_L^\infty \frac{x^{k-1}}{(x + M)^{k+1}} x^{-\alpha} dx &= \int_0^{M/(M+L)} \left(\frac{M(1-y)}{y} \right)^{k-1-\alpha} \left(\frac{y}{M} \right)^{k+1} M y^{-2} dy \\ &= M^{-\alpha-1} \int_0^{M/(M+L)} (1-y)^{k-1-\alpha} y^\alpha dy \end{aligned}$$

(we substituted $y = M/(M + x)$, hence $x = M(1 - y)/y$, $dx/dy = -My^{-2}$ for the first equation). Thus,

$$\begin{aligned} \lim_{M \rightarrow \infty} M^\alpha \int_0^\infty Mk \frac{x^{k-1}}{(x + M)^{k+1}} x^{-\alpha} dx &= k \int_0^1 (1-y)^{k-1-\alpha} y^\alpha dy \\ &= \frac{k\Gamma(\alpha + 1)\Gamma(k - \alpha)}{\Gamma(k + 1)} \\ &= \alpha \frac{\Gamma(\alpha)\Gamma(k - \alpha)}{\Gamma(k)} = \alpha B(k - \alpha, \alpha) \end{aligned} \tag{3.25}$$

For $\varepsilon > 0$ we can choose L so large that $(1 - \varepsilon)c_W x^{-\alpha} \leq \mathbb{P}(W \geq x) \leq (1 + \varepsilon)c_W x^{-\alpha}$ holds for all $x \geq L$. Combining (3.23)–(3.25) we see that

$$\limsup_{M \rightarrow \infty} M^\alpha \mathbb{E} \left[\frac{W^k}{(W + M)^k} \right] \leq (1 + \varepsilon)c_W \alpha B(k - \alpha, \alpha)$$

and similarly for the lim inf. (3.22) follows by taking $\varepsilon \downarrow 0$. □

Defining

$$S_{k,N} := \sum_{j=k}^N W_j, \quad S_{k,N}^{(2)} := \sum_{j=k}^N W_j^2 \tag{3.26}$$

(for $2 \leq k < N$) we can re-express Z_N from (2.3) as

$$Z_N = W_1 S_{2,N} + \frac{1}{2}((S_{2,N})^2 - S_{2,N}^{(2)}) \tag{3.27}$$

and also as

$$\begin{aligned} Z_N &= \frac{1}{2}(W_1 + W_2 + S_{3,N})^2 - \frac{1}{2}W_1^2 - \frac{1}{2}W_2^2 - \frac{1}{2}S_{3,N}^{(2)} \\ &= W_1 W_2 + (W_1 + W_2)S_{3,N} + \frac{1}{2}((S_{3,N})^2 - S_{3,N}^{(2)}). \end{aligned} \tag{3.28}$$

Lemma 3.10. Assume that W satisfies (2.5) or that $\mu_W^{(2)} < \infty$. For $0 < \delta < 1$ and $k \in \{2, 3\}$ let

$$A_\delta := \{S_{k,N} < (1 - \delta)\mu_W N\} \cup \{(S_{k,N})^2 - S_{k,N}^{(2)} < (1 - \delta)\mu_W^2 N^2\}$$

then there exists $r = r(\delta) > 0$ such that

$$\mathbb{P}(A_\delta) \leq e^{-rN} \quad \text{for all } N \text{ large enough.} \tag{3.29}$$

Furthermore, for all $r = r(\delta) > 0$ with $\delta \in (0, 1)$ we have

$$\mathbb{P}((S_{k,N})^2 - S_{k,N}^{(2)} < \mu_W N S_{k,N} / 2) \leq e^{-rN} \quad \text{for all } N \text{ large enough.} \tag{3.30}$$

For (3.30) (under Assumption (2.5)) note that while the typical size of $S_{k,N}$ is $\approx \mu_W N$ by the law of large numbers, conditioned on $S_{k,N} \gg \mu_W N$ there will typically be just one exceptionally large summand (by the tail assumption (2.5), this is much more likely than having many moderately large summands, cf. [24]). Then, the order of magnitude of $(S_{k,N})^2 - S_{k,N}^{(2)}$ will in fact be $\approx N S_{k,N}$ up to constants.

Proof. Write $W_i = U_i + O_i$ with $U_i := W_i \mathbb{I}(W_i \leq K)$, $O_i := W_i \mathbb{I}(W_i > K)$ where K is chosen so large that $\mathbb{E}[U_i] > (1 - \delta/5)\mu_W$. Since $U_i \geq 0$ are i.i.d. bounded random variables, we get from Cramér’s large deviations theorem (see, for example, Theorem 2.2.3 in [7]) that

$$\mathbb{P}\left(\sum_{i=k}^N U_i < (1 - \delta/4)\mu_W N\right) + \mathbb{P}\left(\sum_{i=k}^N U_i > (1 + \delta/4)\mu_W N\right) \leq e^{-rN} \tag{3.31}$$

for all N large enough with $r = r(\delta) > 0$. We now argue that $A_\delta \subset \{\sum_{i=k}^N U_i < (1 - \delta/4)\mu_W N\}$ for N large enough: Obviously, $\{S_{k,N} < (1 - \delta)\mu_W N\} \subset \{\sum_{i=k}^N U_i < (1 - \delta/4)\mu_W N\}$. On the event $\{\sum_{i=k}^N U_i \geq (1 - \delta/4)\mu_W N\}$ we have

$$\begin{aligned} (S_{k,N})^2 - S_{k,N}^{(2)} &\geq \left((1 - \delta/4)\mu_W N + \sum_{i=k}^N O_i\right)^2 - NK^2 - \sum_{i=k}^N O_i^2 \\ &\geq (1 - \delta/4)^2 \mu_W^2 N^2 - NK^2 = \left(1 - \frac{\delta}{2} + \frac{\delta^2}{16} - \frac{K^2}{\mu_W^2 N}\right) \mu_W^2 N^2 \geq (1 - \delta) \mu_W^2 N^2 \end{aligned}$$

for N large enough. Thus, also $\{(S_{k,N})^2 - S_{k,N}^{(2)} < (1 - \delta)\mu_W^2 N^2\} \subset \{\sum_{i=k}^N U_i < (1 - \delta/4)\mu_W N\}$ and we have that (3.29) follows from (3.31).

On the event $\{1 - \delta/4 \leq (\mu_W N)^{-1} \sum_{i=k}^N U_i \leq 1 + \delta/4\}$ we can estimate similarly to the above for all $\delta \in (0, 1)$:

$$\begin{aligned} (S_{k,N})^2 - S_{k,N}^{(2)} &\geq \left((1 - \delta/4) \mu_W N + \sum_{i=k}^N O_i \right)^2 - NK^2 - \sum_{i=k}^N O_i^2 \\ &= (1 - \delta/4)^2 \mu_W^2 N^2 + 2(1 - \delta/4) \mu_W N \sum_{i=k}^N O_i + \left(\sum_{i=k}^N O_i \right)^2 - NK^2 - \sum_{i=k}^N O_i^2 \\ &\geq \frac{\mu_W N}{2} \left((1 + \delta/4) \mu_W N + \sum_{i=k}^N O_i \right) \geq \frac{\mu_W N}{2} \left(\sum_{i=k}^N W_i \right) = \frac{\mu_W N}{2} S_{k,N}. \end{aligned}$$

Thus, (3.30) follows from (3.31). □

For ease of reference we recall here a classical fact about multinomial distributions (see, e.g., Formula (35.5) in [13]).

Lemma 3.11. For $Y = (Y_1, \dots, Y_m) \sim \text{Multinomial}(N, p_1, p_2, \dots, p_m)$ and $n_1, n_2, \dots, n_m \in \mathbb{N}_0$ we have

$$\mathbb{E}[(Y_1)_{n_1} (Y_2)_{n_2} \dots (Y_m)_{n_m}] = (N)_{n_1+n_2+\dots+n_m} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m} \tag{3.32}$$

with $n = n_1 + n_2 + \dots + n_m$.

Proof of Lemma 2.2. To verify (2.7) observe that from (2.4) with Q_N from (2.6)

$$\mathcal{L}(V_1 | (W_i)) = \text{Bin}(N, Q_N)$$

hence using (3.32) from Lemma 3.11

$$\mathbb{E}[V_1(V_1 - 1) | (W_i)] = N(N - 1)Q_N^2.$$

(2.7) follows from this via the formula $c_N = \mathbb{E}[V_1(V_1 - 1)]/(8(N - 1))$, see Lemma 3.1.

We now assume that (2.5) holds. In order to prove (2.9) we first verify that

$$\limsup_{N \rightarrow \infty} N^\alpha \mathbb{E}[Q_N^2] \leq 8C_{\text{pair}}^{(\text{Beta})}. \tag{3.33}$$

Note that using (3.27) we can re-write

$$Q_N = \frac{W_1 S_{2,N}}{W_1 S_{2,N} + \frac{1}{2}(S_{2,N})^2 - \frac{1}{2}S_{2,N}^{(2)}} = \frac{W_1}{W_1 + \frac{(S_{2,N})^2 - S_{2,N}^{(2)}}{2S_{2,N}}}. \tag{3.34}$$

Put

$$\tilde{A}_{N,\delta} := \left\{ \frac{1 - \delta}{2} \mu_W N < \frac{(S_{2,N})^2 - S_{2,N}^{(2)}}{2S_{2,N}} < \frac{1 + \delta}{2} \mu_W N \right\}.$$

We have

$$\mathbb{P}(\tilde{A}_{N,\delta}) \xrightarrow[N \rightarrow \infty]{} 1, \tag{3.35}$$

noting that $S_{2,N}/N \rightarrow \mu_W$ and $S_{2,N}^{(2)}/N^2 \rightarrow 0$ as $N \rightarrow \infty$ in probability (for the latter use that W_i^2 have regularly varying tails of index $\alpha/2 \in (1/2, 1)$, so in particular $S_{2,N}^{(2)}/N^{2/\alpha}$ is tight; this follows e.g. from [11, Thm. 2 in Section XVII.5]). Furthermore, consider

$$B_N := \left\{ \frac{(S_{2,N})^2 - S_{2,N}^{(2)}}{2S_{2,N}} < \mu_W N/5 \right\}$$

and note that $\mathbb{P}(B_N) \leq e^{-rN}$ for N large enough with $r > 0$ due to (3.30) from Lemma 3.10 for $k = 2$. Since for small enough $\delta > 0$ we have $B_N \subset \tilde{A}_{N,\delta}^c$ it then follows for those δ that

$$\begin{aligned} \mathbb{E}[Q_N^2] &\leq e^{-rN} + \mathbb{E}\left[\frac{W_1^2}{(W_1 + \mu_W N/5)^2} \mathbb{1}_{\tilde{A}_{N,\delta}^c}\right] + \mathbb{E}\left[\frac{W_1^2}{(W_1 + (1 - \delta)\mu_W N/2)^2}\right] \\ &= e^{-rN} + \mathbb{E}\left[\frac{W_1^2}{(W_1 + \mu_W N/5)^2}\right] \mathbb{P}(\tilde{A}_{N,\delta}^c) + \mathbb{E}\left[\frac{W_1^2}{(W_1 + (1 - \delta)\mu_W N/2)^2}\right] \end{aligned} \quad (3.36)$$

and Lemma 3.9 together with (3.35) implies

$$\limsup_{N \rightarrow \infty} N^\alpha \mathbb{E}[Q_N^2] \leq c_W \left(\frac{2}{(1-\delta)\mu_W}\right)^\alpha \alpha B(2 - \alpha, \alpha),$$

(3.33) follows by taking $\delta \downarrow 0$.

Analogous, in fact a little easier, arguments can be used to show that

$$\liminf_{N \rightarrow \infty} N^\alpha \mathbb{E}[Q_N^2] \geq 8C_{\text{pair}}^{(\text{Beta})}. \quad (3.37)$$

Again using (3.34) we get

$$\mathbb{E}[Q_N^2] \geq \mathbb{P}(\tilde{A}_{N,\delta}) \mathbb{E}\left[\frac{W_1^2}{(W_1 + (1 + \delta)\mu_W N/2)^2}\right], \quad (3.38)$$

now combine (3.35) with Lemma 3.9 as above and then let $\delta \downarrow 0$ to conclude (3.37). (3.33) and (3.37) combined with (2.7) yield (2.9).

We now assume $\mu_W^{(2)} < \infty$. The proof of (2.8) is similar, in fact simpler: Instead of using Lemma 3.9, we simply observe that in this case

$$\lim_{M \rightarrow \infty} M^2 \mathbb{E}\left[\frac{W_1^2}{(W_1 + M)^2}\right] = \lim_{M \rightarrow \infty} \mathbb{E}\left[W_1^2 \frac{M^2}{(W_1 + M)^2}\right] = \mathbb{E}[W_1^2] = \mu_W^{(2)}$$

by dominated convergence. Thus, (3.36) implies that $\limsup_{N \rightarrow \infty} N^2 \mathbb{E}[Q_N^2] \leq 4(1 - \delta)^{-2} \mu_W^{(2)} / \mu_W^2$ and (3.38) implies $\limsup_{N \rightarrow \infty} N^2 \mathbb{E}[Q_N^2] \geq 4(1 - \delta)^2 \mu_W^{(2)} / \mu_W^2$. Taking $\delta \downarrow 0$, this combined with (2.7) yields (2.8). \square

Proof of Lemma 3.7. 1. Assume $\mu_W^{(2)} < \infty$ and recall $S_{2,N}$ and $S_{2,N}^{(2)}$ from (3.26). Since $\mathcal{L}(V_1 | (W_i)) = \text{Bin}(N, Q_N)$ we have by (3.32)

$$\mathbb{E}[(V_1)_3 | (W_i)] = (N)_3 \frac{W_1^3 (S_{2,N})^3}{Z_N^3}.$$

We re-write $Z_N = W_1 S_{2,N} + \frac{1}{2}((S_{2,N})^2 - S_{2,N}^{(2)})$ as in (3.27). Lemma 3.10 shows that $(S_{2,N})^2 - S_{2,N}^{(2)} \geq (\mu_W/2)N S_{2,N}$ holds with probability $\geq 1 - 2e^{-rN}$. Thus,

$$\mathbb{E}[(V_1)_3] \leq N^3 \mathbb{E}\left[\frac{W_1^3}{(W_1 + \mu_W N/2)^3}\right] + 2N^3 e^{-rN},$$

and (3.18) follows as in the proof of [29, Proposition 7] with the help of (2.8) in Lemma 2.2.

2. Write

$$V' := \sum_{j=3}^N V_{1,j}, \quad V'' := \sum_{j=3}^N V_{2,j},$$

hence $V_1 = V_{1,2} + V'$, $V_2 = V_{1,2} + V''$ and a straightforward calculation yields

$$\begin{aligned} (V_1)_2(V_2)_2 &= (V_{1,2})_4 + 4(V_{1,2})_3 + 2(V_{1,2})_2 + 2(V_{1,2})_3(V' + V'') + 4(V_{1,2})_2(V' + V'') \\ &\quad + (V_{1,2})_2(V'')_2 + (V_{1,2})_2(V')_2 + 4(V_{1,2})_2V'V'' + 4V_{1,2}V'V'' \\ &\quad + 2V_{1,2}V'(V'')_2 + 2V_{1,2}V''(V')_2 + (V')_2(V'')_2. \end{aligned}$$

Instead of spelling out the details of this computation, note that there is a combinatorial interpretation: Consider $V_{1,2} + V' + V''$ numbered balls, of which $V_{1,2}$ are white, V' are red and V'' are blue. Then $(V_1)_2(V_2)_2$ counts the number of ordered pairs we can form where the first pair consists of two distinct balls which are either white or red and the second pair consists of two distinct balls which are either white or blue (and the same ball(s) might possibly appear in both pairs). The right-hand side decomposes this number: There are $(V_{1,2})_4$ pairs where all balls are white and all are distinct, $4(V_{1,2})_3$ pairs where all balls are white and exactly one ball appears twice, etc.

Since the law of $(V_{1,2}, V', V'', N - V_{1,2} - V' - V'')$ given the (W_i) is

$$\text{Multinomial}\left(N, \frac{W_1W_2}{Z_N}, \frac{W_1S_{3,N}}{Z_N}, \frac{W_2S_{3,N}}{Z_N}, \frac{\sum_{3 \leq j < k \leq N} W_jW_k}{Z_N}\right)$$

we find using Lemma 3.11 in the first equality that

$$\begin{aligned} \mathbb{E}[(V_1)_2(V_2)_2 \mid (W_i)] & \tag{3.39} \\ &= (N)_4 \left(\frac{W_1W_2}{Z_N}\right)^4 + 4(N)_3 \left(\frac{W_1W_2}{Z_N}\right)^3 + 2(N)_2 \left(\frac{W_1W_2}{Z_N}\right)^2 \\ &\quad + 2(N)_4 \left(\frac{W_1W_2}{Z_N}\right)^3 \frac{(W_1+W_2)S_{3,N}}{Z_N} + 4(N)_3 \left(\frac{W_1W_2}{Z_N}\right)^2 \frac{(W_1+W_2)S_{3,N}}{Z_N} \\ &\quad + (N)_4 \left(\frac{W_1W_2}{Z_N}\right)^2 \left(\frac{W_2S_{3,N}}{Z_N}\right)^2 + (N)_4 \left(\frac{W_1W_2}{Z_N}\right)^2 \left(\frac{W_1S_{3,N}}{Z_N}\right)^2 \\ &\quad + 4(N)_4 \frac{W_1W_2}{Z_N} \frac{W_1S_{3,N}W_2S_{3,N}}{Z_N^2} + 4(N)_3 \frac{W_1W_2}{Z_N} \frac{W_1S_{3,N}W_2S_{3,N}}{Z_N^2} \\ &\quad + 2(N)_4 \frac{W_1W_2}{Z_N} \frac{W_1S_{3,N}}{Z_N} \left(\frac{W_2S_{3,N}}{Z_N}\right)^2 + 2(N)_4 \frac{W_1W_2}{Z_N} \frac{W_2S_{3,N}}{Z_N} \left(\frac{W_1S_{3,N}}{Z_N}\right)^2 \\ &\quad + (N)_4 \left(\frac{W_1S_{3,N}}{Z_N}\right)^2 \left(\frac{W_2S_{3,N}}{Z_N}\right)^2 \\ &= \frac{(N)_4}{Z_N^4} \left(W_1^4W_2^4 + 2W_1^3W_2^3(W_1 + W_2)S_{3,N} + W_1^2W_2^4(S_{3,N})^2 + W_1^4W_2^2(S_{3,N})^2 \right. \\ &\quad \left. + 4W_1^3W_2^3(S_{3,N})^2 + 2W_1^2W_2^3(S_{3,N})^3 + 2W_1^3W_2^2(S_{3,N})^3 + W_1^2W_2^2(S_{3,N})^4 \right) \\ &\quad + \frac{4(N)_3}{Z_N^3} \left(W_1^3W_2^3 + W_1^2W_2^2(W_1 + W_2)S_{3,N} + W_1^2W_2^2(S_{3,N})^2 \right) + \frac{2(N)_2W_1^2W_2^2}{Z_N^2}. \end{aligned}$$

Recall

$$Z_N = W_1W_2 + (W_1 + W_2)S_{3,N} + \frac{1}{2}((S_{3,N})^2 - S_{3,N}^{(2)})$$

from (3.28) and that we can bound

$$(S_{3,N})^2 - S_{3,N}^{(2)} \geq (1 - \delta)\mu_W^2 N^2 \vee \frac{\mu_W}{2} NS_{3,N}$$

except on an event with exponentially small probability, cf. (3.29) and (3.30) from Lemma 3.10.

We will not treat all the terms on the right-hand side of (3.39) in detail (but see Remark 3.12 below) since the computations are long but otherwise relatively straightforward. Consider for example the term

$$\mathbb{E} \left[\frac{W_1^2W_2^2(S_{3,N})^4}{Z_N^4} \right] = \mathbb{E} \left[\frac{W_1^2W_2^2(S_{3,N})^4}{\left(W_1W_2 + (W_1 + W_2)S_{3,N} + \frac{1}{2}((S_{3,N})^2 - S_{3,N}^{(2)}) \right)^4} \right].$$

If $(1 - \delta)\mu_W N \leq S_{3,N} \leq 2\mu_W N$, say, we can estimate

$$\begin{aligned} \frac{W_1^2 W_2^2 (S_{3,N})^4}{Z_N^4} &\leq 16\mu_W^4 N^4 \frac{W_1^2 W_2^2}{((W_1 + W_2)(1 - \delta)\mu_W N + (1 - \delta)\mu_W^2 N^2/2)^4} \\ &\leq 16(1 - \delta)^{-4} \frac{W_1^2}{(W_1 + \mu_W N/2)^2} \frac{W_2^2}{(W_2 + \mu_W N/2)^2}, \end{aligned}$$

hence

$$\begin{aligned} \mathbb{E} \left[\frac{W_1^2 W_2^2 (S_{3,N})^4}{Z_N^4} \mathbb{I}(S_{3,N} \leq 2\mu_W N) \right] &\leq e^{-rN} + 16(1 - \delta)^{-4} \left(\mathbb{E} \left[\frac{W_1^2}{(W_1 + \mu_W N/2)^2} \right] \right)^2 \\ &\leq CN^{-2\alpha} \end{aligned}$$

for N large enough where we used Lemma 3.9 in the last inequality.

On $\{S_{3,N} \geq 2\mu_W N\}$ we also have $S_{3,N}^2 - S_{3,N}^{(2)} \geq (\mu_W/2)NS_{3,N}$ with high probability. Then

$$\begin{aligned} &\mathbb{E} \left[\frac{W_1^2 W_2^2 (S_{3,N})^4}{Z_N^4} \mathbb{I}(S_{3,N} > 2\mu_W N) \right] \\ &\leq \mathbb{E} \left[\frac{W_1^2 W_2^2 (S_{3,N})^4}{((W_1 + W_2)S_{3,N} + (\mu_W/4)NS_{3,N})^4} \mathbb{I}(S_{3,N} > 2\mu_W N) \right] + e^{-rN} \\ &\leq \mathbb{E} \left[\frac{W_1^2 W_2^2}{((W_1 + W_2) + \mu_W N/4)^4} \right] \mathbb{P}(S_{3,N} > 2\mu_W N) + e^{-rN} \\ &\leq \left(\mathbb{E} \left[\frac{W_1^2}{(W_1 + \mu_W N/2)^2} \right] \right)^2 \mathbb{P}(S_{3,N} > 2\mu_W N) + e^{-rN} = O(N^{-2\alpha}) \end{aligned}$$

and we obtain

$$\limsup_{N \rightarrow \infty} N^{2\alpha} \mathbb{E} \left[\frac{W_1^2 W_2^2 (S_{3,N})^4}{Z_N^4} \right] < \infty.$$

Similarly, using Lemma 3.10

$$\begin{aligned} \mathbb{E} \left[\frac{W_1^4 W_2^4}{Z_N^4} \right] &\leq \mathbb{E} \left[\frac{W_1^4 W_2^4}{(W_1 W_2 + (W_1 + W_2)(1 - \delta)\mu_W N + (1 - \delta)\mu_W^2 N^2/2)^4} \right] + e^{-rN} \\ &\leq \frac{2^4}{(1 - \delta)^4} \mathbb{E} \left[\frac{W_1^4 W_2^4}{(W_1 W_2 + (W_1 + W_2)\mu_W N + \mu_W^2 N^2)^4} \right] + e^{-rN} \\ &= \frac{2^4}{(1 - \delta)^4} \mathbb{E} \left[\frac{W_1^4}{(W_1 + \mu_W N)^4} \frac{W_2^4}{(W_2 + \mu_W N)^4} \right] + e^{-rN} \end{aligned}$$

and we obtain again

$$\limsup_{N \rightarrow \infty} N^{2\alpha} \mathbb{E} \left[\frac{W_1^4 W_2^4}{Z_N^4} \right] < \infty$$

from Lemma 3.9.

The other terms in (3.39) can be treated analogously (see Remark 3.12) to yield

$$\limsup_{N \rightarrow \infty} N^{2\alpha-4} \mathbb{E}[(V_1)_2(V_2)_2] < \infty. \tag{3.40}$$

Since $N^2 c_N \sim C_{\text{pair}}^{(\text{Beta})} N^{3-\alpha}$ by (2.9) of Lemma 2.2 and $4 - 2\alpha < 3 - \alpha$ this proves the claim (3.19). \square

Remark 3.12. For a rough idea of the size of the terms on the right-hand side of (3.39) we can argue as follows: Consider the “typical event” $S_{3,N} \approx \mu N, (S_{3,N})^2 - S_{3,N}^{(2)} \approx \mu^2 N^2$. When W_1 and W_2 are both bounded, the right-hand side of (3.39) is then $O(1)$, the contribution of the case $W_1 \approx N^{\beta_1}, W_2 \approx N^{\beta_2}$ is then $\approx N^\gamma$ with

$$\begin{aligned} \gamma = & -\alpha(\beta_1 + \beta_2) + 4 \\ & + [(4\beta_1 + 4\beta_2) \vee (4\beta_1 + 3\beta_2 + 1) \vee (3\beta_1 + 4\beta_2 + 1) \vee (2\beta_1 + 4\beta_2 + 2) \vee (4\beta_1 + 2\beta_2 + 2) \\ & \vee (3\beta_1 + 3\beta_2 + 2) \vee (2\beta_1 + 3\beta_2 + 3) \vee (3\beta_1 + 2\beta_2 + 2) \vee (2\beta_1 + 2\beta_2 + 4)] \\ & - 4[(\beta_1 + \beta_2) \vee (\beta_1 + 1) \vee (\beta_2 + 1) \vee 2]. \end{aligned}$$

Observe that when $\beta_1, \beta_2 \leq 1$ this is $= -\alpha(\beta_1 + \beta_2) + 4 + (2\beta_1 + 2\beta_2 + 4) - 4 \cdot 2 = (2 - \alpha)(\beta_1 + \beta_2)$ ($< 3 - \alpha$, note that we divide in (3.19) by $c_N N^2 \approx N^{3-\alpha}$); when $\beta_1 < 1 \leq \beta_2$, say, this is $= -\alpha(\beta_1 + \beta_2) + 4 + (2\beta_1 + 4\beta_2 + 2) - 4(\beta_2 + 1) = 2 + (2 - \alpha)\beta_1 - \alpha\beta_2$ ($< 3 - \alpha$); when $\beta_1, \beta_2 > 1$ this is $= -\alpha(\beta_1 + \beta_2) + 4 + (4\beta_1 + 4\beta_2) - 4(\beta_1 + \beta_2) = 4 - \alpha(\beta_1 + \beta_2)$ ($< 3 - \alpha$). This confirms (3.40) at least on an intuitive level.

Proof of Lemma 3.8. We start by arguing that in order to show (3.20) it suffices to check that for $x \in (0, 1)$

$$\lim_{N \rightarrow \infty} \frac{N}{c_N} \mathbb{P}(Q_N > x) \tag{3.41}$$

exists and is given by the right-hand side of (3.20) (note that this expression is continuous in x). Indeed, since $\mathcal{L}(V_1 | (W_i)) = \text{Bin}(N, Q_N)$ and

$$\text{Bin}(N, p)(\{0, 1, \dots, [(1 - \epsilon)pN]\} \cup \{(1 + \epsilon)pN, \dots, N\}) \leq e^{-rN}$$

for any $p, \epsilon \in (0, 1)$ with $r = r(p, \epsilon) > 0$ by classical large deviation estimates for the binomial distribution (e.g. [7], Thm. 2.2.3 and Ex. 2.2.23 (b)) we may write for $\epsilon > 0$

$$\begin{aligned} \mathbb{P}\left(\frac{V_1}{N} > x\right) &= \mathbb{P}\left(\frac{V_1}{N} > x | Q_N > x(1 + \epsilon)\right) \mathbb{P}(Q_N > x(1 + \epsilon)) \\ &+ \mathbb{P}\left(\frac{V_1}{N} > x | x(1 - \epsilon) < Q_N \leq x(1 + \epsilon)\right) \times \mathbb{P}(x(1 - \epsilon) < Q_N \leq x(1 + \epsilon)) \\ &+ \mathbb{P}\left(\frac{V_1}{N} > x | Q_N \leq x(1 - \epsilon)\right) \mathbb{P}(Q_N \leq x(1 - \epsilon)). \end{aligned}$$

Due to (3.41) and the continuity of the limit we can choose $\epsilon > 0$ small and then N large to make the second probability in the second term multiplied by N/c_N and thus the second term multiplied by N/c_N arbitrarily small. By choosing N potentially larger and by using the above large deviations result the conditional probabilities in the first and third line are arbitrarily close to 1 respectively arbitrarily close to 0 even when multiplied by $N/c_N \sim 1/C_{\text{pair}}^{(\text{Beta})} N^\alpha$ by (2.9). Thus, the claim (3.20) now follows from (3.41).

As in the proof of Lemma 2.2 we re-express (see (3.34) and recall $S_{2,N}$ and $S_{2,N}^{(2)}$ from (3.26))

$$Q_N = \frac{W_1}{W_1 + \frac{(S_{2,N})^2 - S_{2,N}^{(2)}}{2S_{2,N}}}.$$

Recall that $\tilde{A}_{N,\delta} := \left\{ \frac{1-\delta}{2} \mu_W N < \frac{(S_{2,N})^2 - S_{2,N}^{(2)}}{2S_{2,N}} < \frac{1+\delta}{2} \mu_W N \right\}$ satisfies $\mathbb{P}(\tilde{A}_{N,\delta}) \rightarrow 1$ as $N \rightarrow \infty$ for every $\delta > 0$ (see (3.35) in the proof of Lemma 2.2), furthermore

$$\mathbb{P}\left(\tilde{A}_{N,\delta}^c \cap \left\{ \frac{(S_{2,N})^2 - S_{2,N}^{(2)}}{2S_{2,N}} < \mu_W N/5 \right\}\right) \leq 3e^{-rN}$$

for N large enough with $r = r(\delta) > 0$ (combine (3.29) and (3.30) from Lemma 3.10). Thus

$$\limsup_{N \rightarrow \infty} \frac{N}{c_N} \mathbb{P} \left(\{Q_N > x\} \cap \tilde{A}_{N,\delta}^c \right) \leq \limsup_{N \rightarrow \infty} \frac{N}{c_N} \mathbb{P}(W_1 > N\mu_W x/5) \mathbb{P}(\tilde{A}_{N,\delta}^c) = 0$$

where we use $N/c_N \sim N^\alpha / C_{\text{pair}}^{(\text{Beta})}$ by Lemma 2.2, combined with (2.5) and (3.35).

Now

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{N}{c_N} \mathbb{P} \left(\{Q_N > x\} \cap \tilde{A}_{N,\delta} \right) &\leq \limsup_{N \rightarrow \infty} \frac{N}{c_N} \mathbb{P} \left(\frac{W_1}{W_1 + (1 - \delta)\mu_W N/2} > x \right) \mathbb{P} \left(\tilde{A}_{N,\delta} \right) \\ &= \limsup_{N \rightarrow \infty} \frac{N^\alpha}{C_{\text{pair}}^{(\text{Beta})}} \mathbb{P} \left(W_1 > \frac{x}{1-x} (1 - \delta)\mu_W N/2 \right) = \frac{(1 - \delta)^{-\alpha}}{C_{\text{pair}}^{(\text{Beta})}} c_W \frac{2^\alpha}{\mu_W^\alpha} \left(\frac{1-x}{x} \right)^\alpha \end{aligned}$$

by the tail assumption (2.5), analogous to the proof of Lemma 14 in [29], and similarly for the lim inf. Finally note that from (2.9) in Lemma 2.2 we have

$$\begin{aligned} \frac{c_W (2/\mu_W)^\alpha}{C_{\text{pair}}^{(\text{Beta})}} \left(\frac{1-x}{x} \right)^\alpha &= \frac{8}{B(2-\alpha, \alpha)} \frac{1}{\alpha} \left(\frac{1-x}{x} \right)^\alpha \\ &= \frac{8}{B(2-\alpha, \alpha)} \int_x^1 \frac{1}{y^2} y^{1-\alpha} (1-y)^{\alpha-1} dy. \quad \square \end{aligned}$$

3.3 Proof of Proposition 2.5

In this section we prove Lemma 2.4 and Proposition 2.5, the main convergence result for the diploid population model of Section 2.2, where potential offspring to parental couples are generated as in a supercritical Galton-Watson process and then pruned. In order to prove the main Proposition 2.5 we need two lemmas in addition to Lemma 2.4. The proof of these auxiliary lemmas is postponed to Section 3.3.1.

Our proofs in this section are to some extent parallel to those in [29], especially those of Lemmas 2.4 and 3.13. We note however that the arguments are somewhat more involved, in particular due to the fact that $X_i^{(N)}, 1 \leq i \leq N$ are not independent. Our proof of Proposition 2.5, 2. follows a slightly different route than that of its analogue, [29, Thm. 4 (c)] in that we verify condition (1.5) on the law of the ranked offspring frequencies directly without recourse to the moment criterion (1.6). One can alternatively prove Proposition 2.5 by verifying (1.6) but this route appears more cumbersome here because the X_i are not independent in our set-up.

Recall that the offspring $V_{i,j}^{(N)}$ of individual i and j are sampled from the ‘‘potential offspring’’ $X_{i,j}^{(N)}$, which are i.i.d.. We write $V_i^{(N)} = \sum_{j \neq i}^N V_{i,j}^{(N)}$ as before. Our first observation concerns large deviations of the total number of potential offspring $S_N = \sum_{1 \leq i < j \leq N} X_{i,j}^{(N)}$ of (2.11) which by (2.15) has $\mathbb{E}[S_N] \sim N\mu$ with $\mu > 1$.

Lemma 3.13. *For any $\varepsilon > 0$ there exist constants $c = c(\varepsilon) > 0$ and $C = C(\varepsilon) < \infty$ such that*

$$\mathbb{P}(S_N \leq (1 - \varepsilon)N\mu) \leq e^{-cN} \quad \text{for all } N \text{ large enough} \tag{3.42}$$

and under Assumption (2.17) we have

$$\mathbb{P}(S_N \geq (1 + \varepsilon)N\mu) \leq CN^{1-\alpha} \quad \text{for all } N \text{ large enough} \tag{3.43}$$

(under Assumption (2.16), (3.43) holds with the right-hand side replaced by CN^{-1} , i.e., by formally setting $\alpha = 2$.)

The same bounds hold for $S'_N := S_N - X_{1,2}^{(N)}$ and for $S''_N := S_N - X_1^{(N)}$.

We note that on $\{S_N \geq N\}$, conditioned on the $X_{k,\ell}^{(N)}$'s, each $V_i^{(N)}$ is hypergeometric. More precisely, if $\text{hypergeom}(n, m, x)$ is the number of white balls in n draws without replacement from an urn that contains m balls of which x are white balls and $m - x$ are black balls then $V_i^{(N)} \sim \text{hypergeom}(N, S_N, X_i^{(N)})$ conditionally on the $X_{k,\ell}^{(N)}$'s. However, the $V_i^{(N)}$'s for different i are not independent.

We now consider $A_{i,j}^{(N)}, 1 \leq i < j \leq N$ to be i.i.d. $\sim \text{Ber}(p_N)$ (for other $i \neq j \in [N]$ put $A_{j,i}^{(N)} = A_{i,j}^{(N)}, A_{i,i}^{(N)} = 0$), as well as $Y_{i,j}^{(N)}, i, j \in \mathbb{N}$ to be independent copies of X from (2.13), independent of the $A_{i,j}^{(N)}$'s. A convenient parametrisation of the set-up from (2.12) and (2.13) is to set

$$X_{i,j}^{(N)} = A_{i,j}^{(N)} Y_{i,j}^{(N)}. \tag{3.44}$$

Note that

$$A_i^{(N)} := \sum_{j=1}^N A_{i,j}^{(N)}, \tag{3.45}$$

the number of different ‘‘potential partners’’ of individual i , is $\text{Bin}(N - 1, p_N)$ -distributed,

$$A^{(N)} := \sum_{1 \leq i < j \leq N} A_{i,j}^{(N)}, \tag{3.46}$$

the total number of ‘‘potential offspring-generating’’ pairs, is $\text{Bin}(N(N - 1)/2, p_N)$ -distributed.

Lemma 3.14. *Let $A_{i,j}^{(N)}, 1 \leq i < j \leq N$ be i.i.d. $\sim \text{Ber}(p_N)$ with $p_N \sim c_{X,1}/N$ and $A_{j,i}^{(N)} = A_{i,j}^{(N)}$ for $j > i$ as defined above. Then*

$$\mathbb{P}\left(\exists k \leq N : \sum_{i \neq k}^N A_{i,k}^{(N)} \geq \log N\right) = O(N^{-b}) \tag{3.47}$$

for every $b > 0$.

For $A^{(N)} = \sum_{1 \leq i < j \leq N} A_{i,j}^{(N)}$ as in (3.46) we have $\mathbb{E}[A^{(N)}] = \binom{N}{2} p_N \sim c_{X,1} N/2$ and

$$\mathbb{P}\left(|A^{(N)} - \mathbb{E}[A^{(N)}]| > \varepsilon \mathbb{E}[A^{(N)}]\right) \leq 2 \exp\left(-\frac{\varepsilon^2}{3} \mathbb{E}[A^{(N)}]\right). \tag{3.48}$$

for all N and $0 < \varepsilon \leq 1/2$.

In the following we will drop the superscript (N) often for notational simplicity.

Proof of Proposition 2.5. 1. In view of Lemma 2.4 and Condition (1.7) it suffices to verify that $\mathbb{E}[(V_1)_3] = o(N)$ for then $\phi_1(3) = \lim_N \mathbb{E}[(V_1)_3]/(c_N N^2) = 0$ and this implies (1.5) and (1.7) with $\Xi' = \delta_0$ (see e.g. [20, Thm. 4 (b)]). This can be checked along the lines of the proof of [29, Proposition 7]. We simply need that for some $0 < a < 1$

$$\limsup_{N \rightarrow \infty} \mathbb{E}[X_1^2 \mathbb{I}_{\{X_1 \geq N^a\}}] = 0. \tag{3.49}$$

To see that this is true we use (2.13) and (2.16) to estimate

$$\begin{aligned} \mathbb{E}[X_1^2 \mathbb{I}_{\{X_1 \geq N^a\}}] &= (N - 1) \mathbb{E}\left[X_{1,2}^2 \left(\mathbb{I}_{\{X_{1,2} \geq N^a/2\}} + \mathbb{I}_{\{\sum_{j=3}^N X_{1,j} \geq N^a/2\}}\right)\right] \\ &\leq (N - 1) p_N \mathbb{E}[X^2 \mathbb{I}_{\{X \geq N^a/2\}}] + (N - 1) p_N \left(\frac{2}{N^a}\right)^2 \mathbb{E}[X^2] \rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$. This shows (3.49) and thus the claim.

2. Let $V_{(1)} \geq V_{(2)} \geq \dots \geq V_{(N)}$ be the ranked V_i 's. We will verify that (1.5) holds with

$$\Xi'(dx) = \int_{[0,1]} \delta_{(x/2, x/2, 0, 0, \dots)} \text{Beta}(2 - \alpha, \alpha)(dx)$$

(and then Proposition 2.5, 2. follows from our main result, Theorem 1.1, via Condition (1.5)).

This in turn will follow if we show that for every $k \in \mathbb{N}$, $1 > x_1 \geq x_2 \geq x_3 \geq \dots \geq x_k > 0$ we have

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P} \left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_2, \dots, \frac{V_{(k)}}{2N} > x_k \right) \\ &= \int_{\Delta} \mathbb{I}(y_1 \geq x_1, y_2 \geq x_2, \dots, y_k \geq x_k) \frac{1}{(y, y)} \Xi'(dy) \\ &= \mathbb{I}(x_1 \leq 1/2, k \leq 2) \int_{2x_1}^1 \frac{2}{y^2} \text{Beta}(2 - \alpha, \alpha)(dy) \\ &= \mathbb{I}(x_1 \leq 1/2, k \leq 2) \frac{2}{\alpha \Gamma(2 - \alpha) \Gamma(\alpha)} \left(\frac{1 - 2x_1}{2x_1} \right)^\alpha. \end{aligned} \tag{3.50}$$

Here we have used in the last line that with the substitution of $z = (1 - y)/y$

$$\begin{aligned} & \Gamma(2 - \alpha) \Gamma(\alpha) \int_{2x_1}^1 \frac{1}{y^2} \text{Beta}(2 - \alpha, \alpha)(dy) \\ &= \int_{2x_1}^1 \left(\frac{1 - y}{y} \right)^{\alpha - 1} \frac{dy}{y^2} = \int_0^{(1 - 2x_1)/2x_1} z^{\alpha - 1} dz = \alpha^{-1} \left(\frac{1 - 2x_1}{2x_1} \right)^\alpha. \end{aligned} \tag{3.51}$$

The intuition behind this result is the following: $V_{i,j} > Ny$ with $y > 0$ is possible (only) if $X_{i,j}$ is of order N (and then with overwhelming probability the sum of all other $X_{k,\ell}$ with $\{k, \ell\} \neq \{i, j\}$ is $\approx N\mu$ and both V_i and V_j are $\approx V_{i,j}$ up to terms which are $o(N)$); we then have

$$V_{i,j} \approx N \frac{X_{i,j}}{X_{i,j} + N\mu}$$

(using fluctuation bounds for hypergeometric distributions), thus

$$V_{i,j} > Ny \iff \frac{X_{i,j}}{X_{i,j} + N\mu} > y \iff X_{i,j} > N\mu \frac{y}{1 - y}.$$

Furthermore

$$\begin{aligned} & \mathbb{P} \left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_2 \right) \\ & \approx \mathbb{P} (\exists \text{ exactly one pair } i < j \leq N \text{ with } V_{(1)} \approx V_{(2)} \approx V_{i,j} > 2Nx_1) + o(N^{1-\alpha}). \end{aligned}$$

Recall $A^{(N)}$ from (3.45). By Lemma 3.14, $A^{(N)} \approx c_{X,1}N/2$, then

$$\begin{aligned} \mathbb{P} \left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_2 \right) & \sim \frac{c_{X,1}N}{2} \mathbb{P} \left(X > N\mu \frac{2x_1}{1 - 2x_1} \right) \\ & \sim \frac{c_{X,1}N}{2} c_{X,2} N^{-\alpha} \mu^{-\alpha} \left(\frac{2x_1}{1 - 2x_1} \right)^{-\alpha} = \frac{c_{X,1}c_{X,2}}{2\mu^\alpha} \left(\frac{1 - 2x_1}{2x_1} \right)^\alpha N^{1-\alpha} \end{aligned}$$

(note that then because $x_1 \geq x_2$, the event $\{V_{(1)} > 2Nx_1, V_{(2)} < 2Nx_2\}$ is very unlikely).

Combining this with Lemma 2.4 suggests (3.50) for $k = 2$, namely

$$\begin{aligned} \frac{1}{2c_N} \mathbb{P} \left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_2 \right) & \approx \frac{4\mu^\alpha}{c_{X,1}c_{X,2}\alpha B(2 - \alpha, \alpha)} N^{\alpha-1} \frac{c_{X,1}c_{X,2}}{2\mu^\alpha} \left(\frac{1 - 2x_1}{2x_1} \right)^\alpha N^{1-\alpha} \\ & = \frac{2}{\alpha B(2 - \alpha, \alpha)} \left(\frac{1 - 2x_1}{2x_1} \right)^\alpha. \end{aligned}$$

For $\varepsilon > 0$,

$$\mathbb{P}(V_{(3)} > \varepsilon N) = o(N^{1-\alpha}) \text{ and } \mathbb{P}(V_{(1)} > Ny, N(1-\varepsilon)y \geq V_{(2)}) = o(N^{1-\alpha})$$

(both events require essentially that there are at least two distinct pairs $\{i, j\} \neq \{k, \ell\}$ with $X_{i,j}, X_{k,\ell} \geq (\varepsilon/2)N$, say; observe that $V_{i,j} \leq X_{i,j}$ by definition).

A more detailed argument runs as follows: Let

$$B_{i,j}(y, \varepsilon) = B_{i,j}^{(N)}(y, \varepsilon) := \left\{ X_{i,j}^{(N)} > N\mu \frac{2y}{1-2y}, S_N - X_{i,j}^{(N)} \leq N\mu(1+\varepsilon) \right\}$$

for $y \in (0, 1/2)$, $\varepsilon > 0$. In the following we again drop the index N and note that

$$\mathbb{P}(B_{i,j}(y, \varepsilon)) \sim \mathbb{P}\left(X_{1,2} > N\mu \frac{2y}{1-2y}\right) \sim \frac{c_{X,1}}{N} c_{X,2} \frac{(1-2y)^\alpha}{(2y)^\alpha \mu^\alpha} N^{-\alpha} \text{ as } N \rightarrow \infty \quad (3.52)$$

because

$$\begin{aligned} & \mathbb{P}\left(X_{i,j} > N\mu \frac{2y}{1-2y}, S_N - X_{i,j} > N\mu(1+\varepsilon)\right) \\ &= \mathbb{P}\left(X_{1,2} > N\mu \frac{2y}{1-2y}\right) \mathbb{P}(S_N - X_{1,2} > N\mu(1+\varepsilon)) \leq C \frac{1}{N} \frac{(1-2y)^\alpha}{(2y)^\alpha \mu^\alpha} N^{-\alpha} \cdot N^{1-\alpha} = C' N^{-2\alpha} \end{aligned}$$

by Assumption (2.17) and Lemma 3.13 using the fact that $\alpha > 1$.

Since the $V_{k,\ell}$'s are hypergeometric conditional on the $X_{k,\ell}$'s with

$$\mathbb{E}[V_{i,j} | X_{k,\ell}\text{'s}] = N \frac{X_{i,j}}{S_N} = N \frac{X_{i,j}}{X_{i,j} + (S_N - X_{i,j})} \text{ on } \{S_N \geq N\}$$

and the right-hand side is on $B_{i,j}(y, \varepsilon) \cap \{S_N \geq N\}$ bounded below by

$$N \frac{X_{i,j}}{X_{i,j} + N\mu(1+\varepsilon)} \geq N \frac{2y/(1-2y)}{2y/(1-2y) + (1+\varepsilon)} \geq N \frac{2y}{1+\varepsilon},$$

we have by fluctuation bounds for hypergeometric laws (e.g. as recalled in [29, Lemma 19]; see [6, 12]) and by conditioning on $X_{k,\ell}$'s that

$$\mathbb{P}\left(\{V_{i,j} \leq 2Ny/(1+2\varepsilon)\} \cap B_{i,j}(y, \varepsilon) \cap \{S_N \geq N\}\right) \leq e^{-c(\varepsilon)N} \quad (3.53)$$

with $c(\varepsilon) > 0$. Using (3.53) as well as the fact that $\{V_{i,j} \geq m\} \subset \{V_i \geq m, V_j \geq m\} \subset \{V_{(1)} \geq m, V_{(2)} \geq m\}$ for every m and all $i < j \leq N$ we obtain that

$$\begin{aligned} & \limsup_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P}\left(\bigcup_{i < j \leq N} B_{i,j}(x_1(1+2\varepsilon), \varepsilon) \cap \{S_N \geq N\}\right) \\ &= \limsup_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P}\left(\bigcup_{i < j \leq N} \{V_{i,j} > 2Nx_1\} \cap B_{i,j}(x_1(1+2\varepsilon), \varepsilon) \cap \{S_N \geq N\}\right) \\ &\leq \liminf_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P}\left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_1\right). \end{aligned} \quad (3.54)$$

Furthermore $\mathbb{P}(S_N < N) \leq e^{-cN}$ for N large enough by Lemma 3.13 and for $\{i, j\} \neq \{k, \ell\}$ we have that

$$\begin{aligned} \mathbb{P}\left(B_{i,j}(y, \varepsilon) \cap B_{k,\ell}(y, \varepsilon)\right) &\leq \mathbb{P}\left(X_{i,j} > N\mu \frac{2y}{1-2y}, X_{k,\ell} > N\mu \frac{2y}{1-2y}\right) \\ &\leq \left(C \frac{1}{N} c_{X,2} \frac{(1-2y)^\alpha}{(2y)^\alpha \mu^\alpha} N^{-\alpha}\right)^2. \end{aligned}$$

This gives

$$\begin{aligned} & \mathbb{P} \left(\bigcup_{i < j \leq N} B_{i,j}(x_1(1 + 2\varepsilon), \varepsilon) \cap \{S_N \geq N\} \right) \\ &= \sum_{i < j \leq N} \mathbb{P} \left(B_{i,j}(x_1(1 + 2\varepsilon), \varepsilon) \right) + O \left(N^2 \cdot N^{-2-2\alpha} + e^{-cN} \right) \\ &= \binom{N}{2} \mathbb{P} \left(B_{1,2}(x_1(1 + 2\varepsilon), \varepsilon) \right) + o(N^{1-\alpha}). \end{aligned} \tag{3.55}$$

Combining (3.52), (3.54), (3.55) with the calculation in (3.51) and Lemma 2.4 yields

$$\liminf_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P} \left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_1 \right) \geq \int_{2x_1(1+2\varepsilon)}^1 \frac{2}{u^2} \text{Beta}(2 - \alpha, \alpha)(du). \tag{3.56}$$

For the matching upper bound we let

$$D^{(N)} := \bigcap_{i < j \leq N} \{S_N - X_{i,j}^{(N)} \geq (1 - \varepsilon)\mu N\} \cap \bigcap_{k \leq N} \{A_k^{(N)} < \log N\}.$$

Note that if we choose ε small such that $(1 - \varepsilon)\mu > 1$ we in particular have $S_N \geq N$ on $D^{(N)}$. Also,

$$\mathbb{P}((D^{(N)})^c) \leq N^2 e^{-cN} + N^{-b} \tag{3.57}$$

by Lemmas 3.13 and 3.14 where $c > 0$ and the constant $b > 0$ can be chosen (much) larger than $\alpha - 1$. Choose $\delta \in (\frac{\alpha-1}{2}, \alpha - 1)$. Then

$$E^{(N)} := \left\{ \exists i < j \leq N, k < \ell \leq N \text{ with } \{i, j\} \neq \{k, \ell\} \text{ and } X_{i,j} \geq N^{(1+\delta)/\alpha}, X_{k,\ell} \geq N^{(1+\delta)/\alpha} \right\}$$

due to (2.13) and (2.17) that

$$\mathbb{P}(E^{(N)}) \leq (N^2)^2 \left(\frac{1}{N} \frac{C}{(N^{(1+\delta)/\alpha})^\alpha} \right)^2 = O(N^{-2\delta}) = o(N^{1-\alpha}) \tag{3.58}$$

because $\delta > (\alpha - 1)/2$. We also have

$$D^{(N)} \cap (E^{(N)})^c \cap \{V_{(1)} \geq 2Ny\} \subset \bigcup_{i < j \leq N} \{V_{i,j} \geq (1 - \varepsilon)2Ny\} \tag{3.59}$$

for N large enough since on $D^{(N)} \cap (E^{(N)})^c$ every V_i consists of the sum of at most $\log N$ many nonzero $V_{i,j}$ (since this is true for X_i and $X_{i,j}$) of which only one can be larger than $N^{(1+\delta)/\alpha}$ (note that $(1 + \delta)/\alpha < 1$ so that the largest $V_{i,j}$ needs to be of the same order as $V_{(1)}$).

We now set

$$\tilde{B}_{i,j}(y, \varepsilon) = \tilde{B}_{i,j}^{(N)}(y, \varepsilon) := \{X_{i,j}^{(N)} < N\mu \frac{2y}{1 - 2y}, S_N - X_{i,j}^{(N)} \geq N\mu(1 - \varepsilon)\}.$$

Then we have by fluctuation bounds for hypergeometric distributions, analogous to the argument for (3.53), that

$$\mathbb{P} \left(\{V_{i,j} \geq (1 - \varepsilon)2Ny\} \cap \tilde{B}_{i,j}((1 - 2\varepsilon)y, \varepsilon) \cap \{S_N \geq N\} \right) \leq e^{-c(\varepsilon)N}. \tag{3.60}$$

Combining (3.57), (3.58) and then (3.59) we see that

$$\begin{aligned} \mathbb{P}(V_{(1)} \geq 2Ny) &\leq \mathbb{P}((D^{(N)})^c) + \mathbb{P}(E^{(N)}) + \mathbb{P}\left(\{V_{(1)} \geq 2Ny\} \cap D^{(N)} \cap (E^{(N)})^c\right) \\ &\leq \mathbb{P}\left(\{V_{(1)} \geq 2Ny\} \cap D^{(N)} \cap (E^{(N)})^c \cap \bigcap_{i < j \leq N} \left\{X_{i,j} < \frac{2y(1-2\varepsilon)}{1-2y(1-2\varepsilon)}\mu N\right\}\right) \\ &\quad + \mathbb{P}\left(\bigcup_{i < j \leq N} \left\{X_{i,j} \geq \frac{2y(1-2\varepsilon)}{1-2y(1-2\varepsilon)}\mu N\right\}\right) + o(N^{1-\alpha}) \\ &\leq \mathbb{P}\left(\bigcup_{i < j \leq N} \{V_{i,j} \geq (1-\varepsilon)2Ny\} \cap \bigcap_{i < j \leq N} \tilde{B}_{i,j}((1-2\varepsilon)y, \varepsilon) \cap \{S_N \geq N\}\right) \\ &\quad + \binom{N}{2} \mathbb{P}\left(X_{1,2} \geq \frac{2y(1-2\varepsilon)}{1-2y(1-2\varepsilon)}\mu N\right) + o(N^{1-\alpha}), \end{aligned}$$

hence, by (3.60),

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P}\left(\frac{V_{(1)}}{2N} > x_1, \frac{V_{(2)}}{2N} > x_2\right) &\leq \limsup_{N \rightarrow \infty} \frac{1}{2c_N} \mathbb{P}\left(\frac{V_{(1)}}{2N} > x_1\right) \\ &\leq \limsup_{N \rightarrow \infty} \frac{1}{2c_N} \binom{N}{2} \mathbb{P}\left(X_{1,2} \geq \frac{2x_1(1-2\varepsilon)}{1-2x_1(1-2\varepsilon)}\mu N\right) \\ &\leq \int_{2x_1(1-2\varepsilon)}^1 \frac{2}{u^2} \text{Beta}(2-\alpha, \alpha)(du), \end{aligned} \tag{3.61}$$

where we have (as in the lower bound (3.56)) used (3.51) and (3.52) as well as Lemma 2.4. Now let $\varepsilon \downarrow 0$ in (3.56) and (3.61) to obtain (3.50) for $k \leq 2$.

Finally, for $\varepsilon > 0$ and N sufficiently large, we have

$$\begin{aligned} \mathbb{P}(V_{(3)} \geq \varepsilon N) &\leq \mathbb{P}(\exists \text{ distinct } i, j, k \leq N : X_i, X_j, X_k \geq \varepsilon N) \\ &\leq \mathbb{P}((D^{(N)})^c) + \mathbb{P}(E^{(N)}) = o(N^{1-\alpha}) \end{aligned}$$

as above. Combining again with Lemma 2.4, this completes the proof of (3.50) for $k \geq 3$. \square

3.3.1 Proofs of auxiliary results

In this section we prove Lemma 2.4, 3.13 and 3.14. To start with we need one more lemma.

Lemma 3.15. *Let Y_1, Y_2, \dots be independent copies of $Y > 0$ with*

$$\mathbb{P}(Y > y) \sim c_Y y^{-\alpha},$$

where $c_Y \in (0, \infty)$, $\alpha \in (1, 2)$. Let $B_N \sim \text{Bin}(N, p_N)$ where $p_N \sim c_p/N$ with $c_p \in (0, \infty)$ and set

$$X_N := \sum_{i=1}^{B_N} Y_i.$$

Then we have that

$$\forall \varepsilon > 0 \exists N_0, x_0 < \infty : \sup_{N \geq N_0} \sup_{x \geq x_0} \left| \frac{\mathbb{P}(X_N > x)}{c_p c_Y x^{-\alpha}} - 1 \right| \leq \varepsilon.$$

Proof. Write $\mu := \mathbb{E}[Y_1]$ ($< \infty$ because $\alpha > 1$), $S_m := Y_1 + \dots + Y_m$. Inspection of the proof of [24, Thm. 1] gives

$$\forall c > \mu, \varepsilon > 0 \exists x_* = x_*(c, \varepsilon) \text{ such that } \sup_{m \in \mathbb{N}} \sup_{x \geq cm \vee x_*} \left| \frac{\mathbb{P}(S_m > x)}{m c_Y (x - m\mu)^{-\alpha}} - 1 \right| \leq \varepsilon \quad (3.62)$$

(note that [24] considers centred summands, we apply the results from [24] to $\mathbb{P}(S_m > x) = \mathbb{P}(S_m - m\mu > x - m\mu)$.)

By assumption, $\bar{c}_p := \sup_{N \in \mathbb{N}} N p_N < \infty$. For $y > \bar{c}_p$ we have the elementary large deviations bound

$$\mathbb{P}(B_N \geq y) \leq e^{-\lambda y} \mathbb{E}[e^{\lambda B_N}] \leq e^{-\lambda y} \left(\frac{\bar{c}_p}{N} (e^\lambda - 1) + 1 \right)^N \leq \exp \left(-\lambda y + \bar{c}_p (e^\lambda - 1) \right)$$

which holds for all $N \in \mathbb{N}, \lambda > 0$. Choosing $\lambda = \log(y/\bar{c}_p) > 0$ yields

$$\mathbb{P}(B_N \geq y) \leq \exp \left(-y \tilde{I}(y) \right) \quad \text{for all } N \in \mathbb{N}, y > \bar{c}_p \quad (3.63)$$

where $\tilde{I}(y) = \log(y/\bar{c}_p) - 1$.

Fix $\varepsilon > 0$, choose x_* so that the bound in (3.62) holds for some $c > \mu$ and potentially enlarge x_* further such that for all $x \geq x_*$ we have $x \geq cm > \mu m$ for all $m \leq \log x$. Let $x_{**} > e^{\bar{c}_p}$ be so large that

$$\sup_{x \geq x_{**}} \frac{x^\alpha}{c_p c_Y} \exp \left(-\tilde{I}(\log x) \cdot \log x \right) < \varepsilon$$

and

$$\sup_{x \geq x_{**}} \sup_{1 \leq m \leq \lfloor \log x \rfloor} \left(\frac{x}{x - m\mu} \right)^\alpha < 1 + \varepsilon.$$

For $x \geq x_* \vee x_{**}$ we have

$$\begin{aligned} \frac{\mathbb{P}(X_N > x)}{c_p c_Y x^{-\alpha}} - 1 &= \sum_{m=0}^N \binom{N}{m} p_N^m (1 - p_N)^{N-m} \left(\frac{\mathbb{P}(S_m > x)}{c_p c_Y x^{-\alpha}} - 1 \right) \\ &\leq \sum_{m=0}^{\lfloor \log x \rfloor \wedge N} \binom{N}{m} p_N^m (1 - p_N)^{N-m} \left(\frac{\mathbb{P}(S_m > x)}{c_p c_Y (x - m\mu)^{-\alpha}} \frac{(x - m\mu)^{-\alpha}}{x^{-\alpha}} - 1 \right) \\ &\quad + \frac{x^\alpha}{c_p c_Y} \sum_{m=\lfloor \log x \rfloor \wedge N}^N \binom{N}{m} p_N^m (1 - p_N)^{N-m} \\ &\leq \sum_{m=0}^{\lfloor \log x \rfloor \wedge N} \binom{N}{m} p_N^m (1 - p_N)^{N-m} \left(\frac{m}{c_p} (1 + \varepsilon)^2 - 1 \right) + \frac{x^\alpha}{c_p c_Y} \mathbb{P}(B_N \geq \log x) \\ &\leq \left(\frac{(1 + \varepsilon)^2}{c_p} \mathbb{E}[B_N] - 1 \right) + \frac{x^\alpha}{c_p c_Y} \exp \left(-\tilde{I}(\log x) \cdot \log x \right). \end{aligned}$$

The right-hand side is smaller than 8ε when $N \geq N_0$ where N_0 is chosen so that

$$\mathbb{E}[B_N] < c_p (1 + \varepsilon) \quad \text{for all } N \geq N_0.$$

The lower bound can be proved analogously: it suffices to keep in the computation above the sum over m up to $\lfloor \log x \rfloor \wedge N$ where $x \geq \max\{x_*, x_{**}, x_\dagger\}$ with x_\dagger chosen so large that $\sum_{m=0}^{\lfloor \log x \rfloor \wedge N} \binom{N}{m} p_N^m (1 - p_N)^{N-m} \geq (1 - \varepsilon) \mathbb{E}[B_N]$ uniformly in $N \geq N_0$, with a suitably enlarged N_0 . \square

Proof of Lemma 3.14. Put $\bar{c}_p := \sup_{n \in \mathbb{N}} N p_N < \infty$. $A_k^{(N)} := \sum_{i \neq k}^N A_{i,k} \sim \text{Bin}(N-1, p_N)$, we have (see (3.63))

$$\mathbb{P}(A_k^{(N)} \geq \log N) \leq \exp\left(-\log N(\log \log(N) - \log \bar{c}_p - 1)\right) = O\left(\exp\left(- (b+1) \log N\right)\right)$$

for any $b > 0$. The probability on the left-hand side of (3.47) is bounded by $\sum_{k=1}^N \mathbb{P}(A_k^{(N)} \geq \log N) = O(N \cdot N^{-b-1})$.

The probability bound in (3.48) is a classical large deviation bound for sums of Bernoulli random variables (the claimed constant is not sharp).

(For completeness, here are some details:

$$\mathbb{E}[e^{\beta A^{(N)}}] = (p_N(e^\beta - 1) + 1)^{N(N-1)/2} \leq \exp\left(p_N \frac{N(N-1)}{2} (e^\beta - 1)\right) = \exp\left(\mathbb{E}[A^{(N)}](e^\beta - 1)\right)$$

for any $\beta \in \mathbb{R}$, hence (we parametrise $\lambda \geq 0$ in the formulas below)

$$\mathbb{P}(A^{(N)} > (1 + \varepsilon)\mathbb{E}[A^{(N)}]) \leq e^{-\lambda(1+\varepsilon)\mathbb{E}[A^{(N)}]} \mathbb{E}[e^{\lambda A^{(N)}}] \leq \exp\left(\mathbb{E}[A^{(N)}](e^\lambda - 1 - \lambda - \varepsilon\lambda)\right)$$

and

$$\mathbb{P}(A^{(N)} < (1 - \varepsilon)\mathbb{E}[A^{(N)}]) \leq e^{\lambda(1-\varepsilon)\mathbb{E}[A^{(N)}]} \mathbb{E}[e^{-\lambda A^{(N)}}] \leq \exp\left(\mathbb{E}[A^{(N)}](e^{-\lambda} - 1 + \lambda - \varepsilon\lambda)\right).$$

Now put $\lambda = \varepsilon$ and use that $0 \leq e^\beta - 1 - \beta \leq (2/3)\beta^2$ for $|\beta| \leq 1/2$. □

Proof of Lemma 3.13. (3.42) can be proved by classical large deviation bound arguments, e.g. along the lines of the proof of [29, Lemma 5]. Alternatively, we see from (3.48) in Lemma 3.14 by conditioning on the $A_{i,j}^{(N)}$'s in (3.44) that S_N is indeed a sum of $\approx c_{X,1}N/2$ i.i.d. summands (up to an exponentially small error term) and then we can literally apply [29, Lemma 5].

For (3.43) under Assumption (2.17) we represent S_N via (3.44) and Lemma 3.14 as a sum of at most $(1 + \varepsilon/2)c_{X,1}N/2$ i.i.d. copies of X (up to an exponentially small error term) and then apply [24, Thm. 1].

Under Assumption (2.16) we have $c := \sup_N \mathbb{E}[(X_1^{(N)})^2] < \infty$ (cf. e.g. (3.68) below), so $\mathbb{P}(X_1^{(N)} \geq x) \leq c/x^2$ by Markov's inequality. [24, Thm. 2] then gives a bound of the form CN^{-1} for the probability in (3.43).

Finally, note that $S_N'' = {}^d S_{N-1}$ and $S_N'' \leq S_N' \leq S_N$. □

With the help of Lemma 3.13 (and Lemma 3.15) we can now prove Lemma 2.4. For this let us recall that the factorial moments for a hypergeometric random variable $H \sim \text{hypergeom}(n, m, x)$ are given by (see e.g. Formula (39.6) in [13]),

$$\mathbb{E}[(H)_k] = (n)_k \frac{(x)_k}{(m)_k}. \tag{3.64}$$

Proof of Lemma 2.4. Recall that $X_1 = X_1^{(N)}$. We have by (3.64) and (3.42) of Lemma 3.13 that

$$c_N = \frac{\mathbb{E}[(V_1)_2]}{8(N-1)} = \frac{1}{8(N-1)} N(N-1) \mathbb{E}\left[\frac{(X_1)_2}{(S_N)_2} \mathbb{1}(S_N \geq N)\right] + O(e^{-cN}). \tag{3.65}$$

In the case of 1. we now observe that for every $\varepsilon > 0$

$$\mathbb{E}\left[\frac{(X_1)_2}{(S_N)_2} \mathbb{1}(S_N \geq N)\right] \leq \frac{\mathbb{E}[(X_1)_2]}{((1-\varepsilon)\mu N)_2} + O(e^{-cN}), \tag{3.66}$$

$$\mathbb{E}\left[\frac{(X_1)_2}{(S_N)_2} \mathbb{1}(S_N \geq N)\right] \geq \frac{\mathbb{E}[(X_1)_2 \mathbb{1}(S_N \leq (1+\varepsilon)\mu N)]}{((1+\varepsilon)\mu N)_2} + O(N^{1-\alpha}) \tag{3.67}$$

by (3.42) and (3.43) of Lemma 3.13.

$$\begin{aligned} \mathbb{E}[(X_1)_2] &= \mathbb{E}[X_1^2] - \mathbb{E}[X_1] = (N - 1)\mathbb{E}[X_{1,2}^2] + (N - 1)(N - 2)\mathbb{E}[X_{1,2}]^2 - (N - 1)\mathbb{E}[X_{1,2}] \\ &\sim c_{X,1}(\mathbb{E}[X^2] - \mathbb{E}[X]) + (c_{X,1}\mathbb{E}[X])^2. \end{aligned} \tag{3.68}$$

Furthermore,

$$\begin{aligned} \mathbb{E}[(X_1)_2 \mathbb{I}(S_N > (1 + \varepsilon)\mu N)] &\leq \mathbb{E}[(X_1)_2 \mathbb{I}(X_1 > \varepsilon\mu N/2)] \\ &\quad + \mathbb{E}[(X_1)_2 \mathbb{I}(S_N - X_1 > (1 + \varepsilon/2)\mu N)] \xrightarrow[N \rightarrow \infty]{} 0 \end{aligned}$$

due to the fact that by (3.68) and (3.43) of Lemma 3.13

$$\mathbb{E}[(X_1)_2 \mathbb{I}(S_N - X_1 > (1 + \varepsilon/2)\mu N)] \leq \mathbb{E}[(X_1)_2] \cdot \mathbb{P}[S_N - X_1 > (1 + \varepsilon/2)\mu N] \leq CN^{1-\alpha}$$

with $\alpha = 2$ and so as $N \rightarrow \infty$,

$$\mathbb{E}[(X_1)_2 \mathbb{I}(S_N \leq (1 + \varepsilon)\mu N)] \sim \mathbb{E}[(X_1)_2].$$

Combining this and (3.66), (3.67), and (3.68) with $\mu = c_{X,1}c_{X,2}/2$ gives (2.18).

For showing 2. we first use (3.65), write $S_N = X_1 + (S_N - X_1)$ and use that by Lemma 3.13 for $\varepsilon > 0$,

$$\mathbb{P}[(1 - \varepsilon)\mu N \leq S_N - X_1 \leq (1 + \varepsilon)\mu N] \rightarrow 1.$$

We may then formally follow the argument in the proof of [29, Lemma 13] in order to obtain

$$c_N \sim \frac{N}{8} \mathbb{E}\left[\frac{X_1^2}{(X_1 + N\mu)^2}\right].$$

Now note that $\mathbb{P}(X_1 > x) = \mathbb{P}(X_1^{(N)} > x) \sim c_{X,1}c_{X,2}x^{-\alpha}$ as $x \rightarrow \infty$ uniformly in N by Lemma 3.15 and so the statement of Lemma 3.9 also holds in the current setting implying that

$$c_N \sim \frac{N}{8} (N\mu)^{-\alpha} c_{X,1}c_{X,2}\alpha B(2 - \alpha, \alpha]$$

A The weak convergence criterion for the total offspring numbers

Here, for ease of reference, we briefly recall some of the main results from Möhle and Sagitov [22], Sagitov [27], Schweinsberg [28] in the notation of our Theorem 1.1 and its assumptions. Note that our offspring numbers (V_1, \dots, V_N) by (1.2) are analogous to the family size in each generation for haploid Cannings models. The only difference comes from the fact that the V_i here sum to $2N$ whereas theirs sum to N .

Lemma A.1. *Assume (1.6). For any $j \in \mathbb{N}$, there exists a symmetric measure F_j , uniquely determined on the simplex $\Delta_j := \{(x_1, \dots, x_j) : x_1 \geq x_2 \geq \dots \geq x_j \geq 0, \sum_{i=1}^j x_i \leq 1\}$ via its moments*

$$\int_{\Delta_j} x_1^{k_1-2} \dots x_j^{k_j-2} F_j(dx_1, \dots, dx_j) = \phi_j(k_1, \dots, k_j) \text{ for } k_1, \dots, k_j \geq 2.$$

Proof. We refer to the analog proof for Lemma 3.1 in [22]. □

Lemma A.2. *The recursive formula for $\psi_{j,s}(k_1, \dots, k_j)$ by (1.8) over s holds as follows:*

$$\begin{aligned} \psi_{j,s+1}(k_1, \dots, k_j) = & \psi_{j,s}(k_1, \dots, k_j) - \sum_{i=1}^j \psi_{j,s}(k_1, \dots, k_{i-1}, k_i + 1, k_{i+1}, \dots, k_j) \\ & - s\psi_{j+1,s-1}(k_1, \dots, k_j, 2). \end{aligned} \tag{A.1}$$

Proof. We refer to the analog proof for Lemma 3.3 in [22]. □

Proposition A.3. *Assume (1.6).*

1. *For any $j \geq \ell$, $k_1 \geq m_1, \dots, k_\ell \geq m_\ell$, we have $\phi_j(k_1, \dots, k_j) \leq \phi_\ell(m_1, \dots, m_\ell)$.*
2. *For all $j \in \mathbb{N}$ and $k_1, \dots, k_j \geq 2$, $\phi_j(k_1, \dots, k_j)$ are uniformly bounded.*

Proof. Note that these ϕ_j 's are exactly the particular case of $\psi_{j,s}$'s when $s = 0$. By recursive formula (A.1), the monotonicity is true for ϕ_j 's. This result has also been proposed in the Remark on Page 39 of Möhle [21].

The uniform bounded property follows from monotonicity as

$$\frac{\mathbb{E} \left((V_1)_{k_1} \cdots (V_j)_{k_j} \right)}{c_N N^{k_1 + \dots + k_j - j} 2^{k_1 + \dots + k_j}} \leq \frac{\mathbb{E}((V_1)_2)}{4c_N N} \leq 2,$$

where we have used $c_N = \mathbb{E}((V_1)_2) / [8(N - 1)]$. □

In order to represent those $\{\psi_{j,s} : j \in \mathbb{N}, s \in \mathbb{N}_0\}$ by integrations with respect to symmetric measures $\{F_j : j \in \mathbb{N}\}$, we define analogous polynomials

$$T_{j,s}^{(r)}(x_1, \dots, x_r), \quad 1 \leq j \leq r, r \in \mathbb{N} \text{ and } s \in \mathbb{N}_0$$

as

$$T_{r,s}^{(r)}(x_1, \dots, x_r) = (1 - x_1 - \dots - x_r)^s$$

and

$$T_{r-j,s}^{(r)}(x_1, \dots, x_r) = (-1)^{j+1} \sum_{i_j=2j-1}^{i_{j+1}-2} \cdots \sum_{i_1=1}^{i_2-2} \prod_{k=0}^j i_k \left(1 - \sum_{i=1}^{r-k} x_i \right)^{i_{k+1}-i_k-2}, \quad j = 1, \dots, r,$$

where $i_0 = -1$ and $i_{j+1} = s + 1$. Note that this implies $T_{r-j,s}^{(r)}(x_1, \dots, x_r) = 0$ for $s < 2j$.

Lemma A.4. *Assume (1.6). Then we have*

$$\psi_{j,s}(k_1, \dots, k_j) = \sum_{r \geq j} \int_{\Delta_r} x_1^{k_1-2} \cdots x_j^{k_j-2} T_{j,s}^{(r)}(x_1, \dots, x_r) F_r(dx_1 \cdots dx_r) \tag{A.2}$$

for all $j \in \mathbb{N}$, $k_1, \dots, k_j \geq 2$ and all $s \in \mathbb{N}_0$.

Proof. The proof is analogous to Lemma 3.5 and Lemma 3.14 in [22]. □

Theorem A.5. *Assume (1.6). $\{\psi_{j,s}(k_1, k_2, \dots, k_j) : s \in \mathbb{N}_0, j \in \mathbb{N}, k_1, \dots, k_j \geq 2\}$ is the collection of real numbers given by (A.2). It is clearly that these $\psi_{j,s}(k_1, k_2, \dots, k_j)$ are nonnegative. The sequence of measures $(F_j)_{j \in \mathbb{N}}$ in Lemma A.1 satisfies*

1. *each F_j is concentrated on $\Delta_j = \{(x_1, \dots, x_j) : x_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^j x_i \leq 1\}$;*
2. *each F_j is symmetric with respect to the j coordinates of Δ_j ;*

3. $F_1(\Delta_1) \geq F_2(\Delta_2) \geq \dots$

Then there exists an \mathcal{E}_∞ -valued coalescent process $R_\infty := (R_\infty(t))_{t \geq 0}$ satisfying

1. $R_\infty(0)$ is the partition of \mathbb{N} into singletons;
2. for each n , $(R_n(t))_{t \geq 0}$ is the restriction of $(R_\infty(t))_{t \geq 0}$ on $\{1, 2, \dots, n\}$. When R_n has $b := k_1 + \dots + k_j + s$ blocks, $k_1, \dots, k_j \geq 2$, each $(b; k_1, \dots, k_j; s)$ -collision is occurring at rate $\psi_{j,s}(k_1, \dots, k_j)$ given by (A.2). For such a simultaneous multiple coalescent process R_∞ , there is a finite measure Ξ' on the infinite simplex Δ of the form $\Xi' = \Xi'_0 + a\delta_0$, where Ξ'_0 has no atom at zero and δ_0 is a unit mass at zero, such that

$$\begin{aligned} \psi_{j,s}(k_1, k_2, \dots, k_j) &= \int_{\Delta} \sum_{\ell=0}^s \binom{s}{\ell} \sum_{\substack{i_1, \dots, i_{j+\ell} \\ \text{distinct}}} x_{i_1}^{k_1} \dots x_{i_j}^{k_j} x_{i_{j+1}} \dots x_{i_{j+\ell}} (1 - |x|)^{s-\ell} \frac{2\Xi'_0(dx)}{(x, x)} \\ &\quad + 2a\mathbb{I}_{\{j=1, k_1=2\}}. \end{aligned}$$

The connection between the probability measure Ξ' on the infinite dimensional simplex Δ and the sequence of measures $(F_j)_{j \in \mathbb{N}}$ is given by

$$F_j(S) = \int_{\Delta} \sum_{\substack{i_1, \dots, i_j \\ \text{distinct}}} y_{i_1}^2 \dots y_{i_j}^2 \mathbb{I}_{\{(y_{i_1}, \dots, y_{i_r}) \in S\}} \frac{2\Xi'_0(dy)}{(y, y)} + 2a\mathbb{I}_{\{r=1, (0, \dots, 0) \in S\}}$$

with $S \subseteq \Delta_r$.

Proof. We refer to Theorem 2, Proposition 8 and Proposition 11 in Schweinsberg [28] as well as Theorem 2.1 in Möhle and Sagitov [22] for details. \square

Condition I: The limits of

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E} \left((V_1 - 2)^{k_1} \dots (V_j - 2)^{k_j} \right)}{c_N N^{k_1 + \dots + k_j - j} 2^{k_1 + \dots + k_j}}$$

exist for all $j \in \mathbb{N}$ and $k_1, \dots, k_j \geq 2$, which is also equal to $\phi_j(k_1, \dots, k_j)$.

Condition II: The limits of

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E} \left((V_1)_2 \dots (V_j)_2 \right)}{c_N N^j 2^{2j}} = F_j(\Delta_j)$$

exist for all $j \in \mathbb{N}$ and

$$\lim_{N \rightarrow \infty} \frac{N^j}{c_N} \mathbb{P}(V_1 > 2Nx_1, \dots, V_j > 2Nx_j) = \int_{x_1}^1 \dots \int_{x_j}^1 \frac{F_j(dy_1 \dots dy_j)}{y_1^2 \dots y_j^2}$$

holding for all points (x_1, \dots, x_j) of continuity for the measure F_j .

For any $\epsilon > 0$, let $\Delta_{r,\epsilon} := \Delta_r \cap \{x_1, \dots, x_r > \epsilon\}$. Denote by $\Xi_{r,N}$ the symmetric measure on Δ_r giving the joint distribution of the vector $(V_1/2N, \dots, V_r/2N)$ of unranked offspring frequencies. Let Ξ_r be a symmetric measure on the r -dimensional simplex Δ_r .

Condition III: The weak convergence condition

$$\frac{1}{2c_N} N^r \Xi_{r,N} \rightarrow \Xi_r \text{ as } N \rightarrow \infty$$

over $\Delta_{r,\epsilon}$, where the connection between Ξ_r and Ξ' is given by

$$\Xi_r(S) := \int_{\Delta} \sum_{\substack{i_1, \dots, i_r \\ \text{distinct}}} \mathbb{I}_{\{(y_{i_1}, \dots, y_{i_r}) \in S\}} \frac{\Xi'(dy)}{(y, y)}$$

for all $S \subseteq \Delta_r$.

It is clear that the probability measure Ξ' is uniquely determined by the sequence of measures $(\Xi_r)_{r \in \mathbb{N}}$.

Lemma A.6. *Conditions (1.5), (1.6), I, II and III are equivalent.*

Proof. (1.6) \Leftrightarrow Condition I and (1.6) \Leftrightarrow Condition II can be proved similarly as Page 1557 of [22]. Condition II \Leftrightarrow Condition III and (1.5) \Leftrightarrow Condition III can be proved following Pages 849-852 of [27]. Here we omit the details. \square

References

- [1] Austin, T.: On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.* **5**, (2008), 80–145. MR-2426176
- [2] Berestycki, J.; Berestycki, N.; Schweinsberg, J.: Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35**, (2007), 1835–1887. MR-2349577
- [3] Birkner, M.; Blath, J.; Capaldo, M.; Etheridge, A.; Möhle, M.; Schweinsberg J.; Wakolbinger, A.: Alpha-stable Branching and Beta-Coalescents. *Electron. J. Probab.* **10**, (2005), 303–325. MR-2120246
- [4] Birkner, M.; Blath, J.; Eldon, B.: An ancestral recombination graph for diploid populations with skewed offspring distribution. *Genetics* **193**, (2013), 255–290. arXiv:1203.4950
- [5] Billingsley, P.: *Convergence of Probability Measures*, 2nd edition. *John Wiley & Sons, Inc.*, Chicago, 1999.
- [6] Chvátal, V.: The tail of the hypergeometric distribution. *Discrete Math.* **25**, (1979), 285–287. MR-0534946
- [7] Dembo, A. and Zeitouni, O.: *Large deviations techniques and applications*. 2nd corrected edition. *Springer-Verlag*, Berlin, 2010. xvi+396 pp. MR-2571413
- [8] Donnelly, P. and Kurtz, T. G.: Particle representations for measure-valued population models. *Ann. Probab.* **27**, (1999), 166–205. MR-1681126
- [9] Eldon, B. and Wakeley, J.: Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172**, (2006), 2621–2633.
- [10] Ethier, S. N. and Kurtz, T. G.: *Markov processes: Characterization and convergence*. *John Wiley & Sons, Inc.*, New Jersey, 1986. MR-1567615
- [11] Feller, W.: *An Introduction to Probability Theory and Its Applications, II*, 2nd edition. *John Wiley & Sons, Inc.*, New York-London-Sydney, 1971. xxiv+669 pp. MR-0270403
- [12] Hoeffding, W.: Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, (1963), 13–30. MR-0144363
- [13] Johnson, N. L.; Kotz, S.; Balakrishnan, N.: *Discrete Multivariate Distributions*. *John Wiley & Sons, Inc.*, New York, 1997. xxii+299 pp. MR-1429617
- [14] Kallenberg, O.: *Probabilistic symmetries and invariance principles*. *Springer*, New York, 2005. xii+510 pp. MR-2161313
- [15] Kingman, J. F. C.: On the genealogy of large populations. *J. Appl. Probab.* **19**, (1982), 27–43. MR-0633178
- [16] Kingman, J. F. C.: The coalescent. *Stochastic Process. Appl.* **13**, (1982), 235–248. MR-0671034
- [17] Möhle, M.: A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. in Appl. Probab.* **30**, (1998), 493–512. MR-1642850
- [18] Möhle, M.: Coalescent results for two-sex population models. *Adv. in Appl. Probab.* **30**, (1998), 513–520. MR-1642851

Coalescent results for diploid exchangeable population models

- [19] Möhle, M.: Weak convergence to the coalescent in neutral population models. *J. Appl. Probab.* **36**, (1999), 446–460. MR-1724816
- [20] Möhle, M.: Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. in Appl. Probab.* **32**, (2000), 983–993. MR-1808909
- [21] Möhle, M.: On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12**, (2006), 35–53. MR-2202319
- [22] Möhle, M. and Sagitov, S.: A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, (2001), 1547–1562. MR-1880231
- [23] Möhle, M. and Sagitov, S.: Coalescent patterns in diploid exchangeable population models. *J. Math. Biol.* **47**, (2003), 337–352. MR-2024501
- [24] Nagaev, S. V.: On the asymptotic behavior of probabilities of one-sided large deviations. *Teor. Veroyatnost. i Primenen* **26**, (1981), 369–372. MR-0616627
- [25] Pitman, J.: Coalescents with multiple collisions. *Ann. Probab.* **27**, (1999), 1870–1902. MR-1742892
- [26] Sagitov, S.: The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, (1999), 1116–1125. MR-1742154
- [27] Sagitov, S.: Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40**, (2003), 839–854. MR-2012671
- [28] Schweinsberg, J.: Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, (2000), 1–50. MR-1781024
- [29] Schweinsberg, J.: Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.* **106**, (2003), 107–139. MR-1983046
- [30] van der Hofstad, R.: Random Graphs and Complex Networks, Vol. I and II. *Cambridge University Press*, 2017. <http://www.win.tue.nl/~rhofstad/NotesRGCN.html>

Acknowledgments. We would like to thank two anonymous referees for their careful reading and useful comments which helped to improve the presentation.