Bernoulli **22**(4), 2016, 2209–2236
DOI: 10.3150/15-BEJ726

The genealogy of a solvable population model under selection with dynamics related to directed polymers

ASER CORTINES

Université Paris Diderot, Mathématiques, case 7012, F-75 205 Paris Cedex 13, France.
E-mail: cortines@math.univ-paris-diderot.fr

We consider a stochastic model describing a constant size N population that may be seen as a directed polymer in random medium with N sites in the transverse direction. The population dynamics is governed by a noisy traveling wave equation describing the evolution of the individual fitnesses. We show that under suitable conditions the generations are independent and the model is characterized by an extended Wright–Fisher model, in which the individual i has a random fitness η_i and the joint distribution of offspring $(\nu_1, \dots, \nu_N)$ is given by a multinomial law with N trials and probability outcomes η_i's. We then show that the average coalescence times scales like $\log N$ and that the limit genealogical trees are governed by the Bolthausen–Sznitman coalescent, which validates the predictions by Brunet, Derrida, Mueller and Munier for this class of models. We also study the extended Wright–Fisher model, and show that, under certain conditions on η_i, the limit may be Kingman's coalescent, a coalescent with multiple collisions, or a coalescent with simultaneous multiple collisions.

Keywords: ancestral processes; Bolthausen–Sznitman coalescent; coalescence; travelling waves

1. Introduction

An important question in the study of populations evolution is to understand the effect of selection and mutation on its genealogy. For a given population, we would like to know how individuals are related and how many generations do we have to go back in time in order to find a common ancestor. Kingman [14] was one of the first to give a mathematical formulation for this problem and study the ancestral history of a population. He showed that in the absence of selection (neutral models) the populations genealogical structure satisfies universal features; *see also* [15–17].

In this paper, we focus on the evolution of a fixed size population model with N individuals subjected to the effects of mutation and selection. We assign to each individual a real number, which represents the fitness of this individual. This fitness is transmitted to the offspring, up to variations due to mutations. Individuals with large fitness spawn a considerable fraction of the population, whereas the children of low fitness individuals tend to be eliminated. Therefore, these population models are sometimes referred to as "rapidly adapting." If we consider the evolution of the fitnesses along the real axis, it is simply a stochastic model of front propagation. The selection mechanism constrains the particles to stay together. Since individuals with large fitness quickly overrun the whole population, the front is essentially pulled by the leading edge. These models are then related to noisy traveling wave equations of the Fisher–Kolmogorov–Petrovsky–Piscounov (FKPP) type [5,7,8].

Recent results suggest that in rapidly adapting population models the genealogical correlations between individuals have universal features. It is conjectured [5,7,8] that the genealogical trees of these populations converge to the Bolthausen–Sznitman coalescent and that the average coalescence times scales like the logarithmic of the populations size. The conjectures contrast with classical results in neutral population models, such as Wright–Fisher and Moran models. It is known that in neutral population models the genealogical trees converge to those of a Kingman's coalescent and that the average coalescence times scales like N, the size of the population [14–17]. In Section 2, we will give a general introduction and present some relevant results about coalescent processes.

We now mention some models, for which the conjectures have been proved. The "exponential model" [5,7,8] is an example of constant size population dynamics, for which a complete mathematical treatment is possible. Each individual i in generation t carries a value $x_i(t)$, which represents the fitness. The offspring of the individuals are generated by independent Poisson point process of densities $e^{-y+x_i(t)}\,\mathrm{d}y$. One then selects the N right-most individuals to form the next generation $t+1$. The authors show that, after rescaling time by a factor $\log N$, one obtains the convergence to the Bolthausen–Sznitman coalescent. Berestycki, Berestycki and Schweinsberg [2] consider a system of particles, performing branching Brownian motion with negative drift and killed upon reaching zero. The authors choose the appropriate drift such that the model is in the near-critical regime and the initial population size N is roughly preserved. They show that the expected time to observe a merge is of order $(\log N)^3$ and that the genealogy of the particles is also governed by the Bolthausen–Sznitman coalescent.

We also mention other related models, for which the genealogical trees do not converge to those of a Kingman's coalescent. Schweinsberg [20] considers a model, in which the numbers of offspring for the individuals are i.i.d. (Galton–Watson processes), but in each generation only N of the offspring are chosen at random for survival (selection mechanism). He proves that depending on the tail probabilities of the reproductive law, the limit may be Kingman's coalescent, a coalescent with multiple collisions (Λ-coalescent), or a coalescent with simultaneous multiple collisions (Ξ-coalescent). The authors in [13] study the asymptotic of the extended Moran model as the total population size N diverges, and show that the ancestral process of the population may be approximated by a coalescent process with multiple collisions. Discrete population models with unequal (skewed) fertilities, such as the skewed Wright–Fisher model and the Kimura model, are not necessarily in the domain of attraction of the Kingman's coalescent [12].

In the present paper, we consider a population dynamics derived from the following model of front propagation [4]. It consists in a constant number N of evolving particles on the real line initially at positions $X_1(0),\dots,X_N(0)$. Then, given the positions $X_i(t)$ of the N particles at time $t\in\mathbb{N}$, we define the positions at time $t+1$ by

$$X_j(t+1) := \max_{1\le i\le N}\bigl\{X_i(t)+\xi_{ij}(t+1)\bigr\}, \tag{1.1}$$

where $\{\xi_{ij}(s); 1\le i, j\le N, s\in\mathbb{N}\}$ are i.i.d. real random variables. The model can be seen as a directed polymer in random medium at zero temperature. The lattice consists in L planes in the transversal direction. In every plane, there are N points that are connected to all points of the previous plane and the next one and for each edge ij, connecting the planes t and $t+1$, a random energy $-\xi_{ij}(t+1)$ is sampled from a common probability distribution. At zero temperature, the directed polymer chooses the path which minimizes its energy (the optimal path) and $-X_j(L)$ is equal to minimal energy among all paths connecting the origin to the jth point on the Lth plane [10,11]. The optimal path starting at the same point but arriving at different points gives rise to a tree structure. It is well known that population dynamics in presence of selection may be related to directed polymers in random medium at zero temperature and it is expected that they belong to the same universality class [5–8].

If the distribution of $\xi_{ij}(t+1)$ in (1.1) has no atoms, *that is,* for every $x \in \mathbb{R}$ the probability $\mathbb{P}(\xi_{ij}(t+1) = x) = 0$, then for all j the following equation has a.s. a unique solution i:

$$X_j(t+1) = X_i(t) + \xi_{ij}(t+1). \tag{1.2}$$

In this sense, we may say that $X_j(t+1)$ is an offspring or a descendant of $X_i(t)$ and denote by $\nu_i(t)$ the number of descendants of $X_i(t)$ in generation $t+1$. The fitnesses of the individuals are given by their positions $X_1(t), \dots, X_N(t)$ and conditionally on

$$\mathcal{F}_t := \sigma\big\{\xi_{ij}(s) \text{ and } X_i(0); 0 \le s \le t, 1 \le i, j \le N\big\},$$

the probability that $X_j(t+1)$ descends from $X_i(t)$ is given by

$$\eta_i(t) := \mathbb{P}\big(\xi_{ij}(t+1) + X_i(t) \ge \xi_{kj}(t+1) + X_k(t);\ \text{for every } 1 \le k \le N | \mathcal{F}_t\big). \tag{1.3}$$

Since $\{\xi_{ij}(t+1); 1 \le i, j \le N\}$ are independent, it is easy to see that, for $j_1, \dots, j_m$ distinct and $i_1, \dots, i_m$ (not necessarily distinct),

$$\begin{aligned}&\mathbb{P}\big(X_{j_k}(t+1) \text{ descends from } X_{i_k}(t),\ \text{for } 1 \le k \le m | \mathcal{F}_t\big)\\ &\quad = \eta_{i_1}(t)\eta_{i_2}(t)\cdots\eta_{i_m}(t).\end{aligned}$$

If $i_k = i_l$, the individuals j_k and j_l have a common ancestor in generation t. As a consequence, given $\mathcal{F}_t$ the offspring vector $\nu(t) := (\nu_1(t), \dots, \nu_N(t))$ is distributed according to a N-class multinomial with N trials and probabilities outcomes $\eta(t) := (\eta_1(t), \dots, \eta_N(t))$.

We analyse the genealogical tree of the population by observing the ancestral partition process, that is, we sample without replacement $n \ll N$ individuals from a given generation T, say $e_1, \dots, e_n$ and for $0 \le t \le T$ we consider $\Pi_t^{N,n}$ the random partition of $[n] := \{1, \dots, n\}$ such that i and j belong to the same equivalent class if and only if e_i and e_j share the same ancestor at time $T - t$. It is very important to realize that the direction of time for the ancestral process is the opposite of the direction of time for the "natural" evolution of the population.

If ξ_{ij} in (1.1) is Gumbel $G(\rho, \beta)$-distributed, *that is,*

$$\mathbb{P}(\xi_{ij} \le x) = \exp\big(-e^{-\beta(x-\rho)}\big), \qquad x \in \mathbb{R},$$

the microscopic dynamics can be solved allowing precise calculations; see also [4] where Brunet and Derrida use a similar technique to compute the exact asymptotic for the speed of the front. In this case, *see Proposition* 3.1 *in Section* 3, the positions of the particles in generation $t+1$ can be obtained by a $\mathcal{F}_t$-measurable function $\Phi(X(t))$ (that may be interpreted as the front position at time t) and a $\mathcal{F}_t$-independent family of i.i.d. random variables $(\mathcal{E}_i(t+1))_{1\le i\le N}$

$$X_i(t+1) = \rho + \Phi\big(X(t)\big) - \beta^{-1}\log\mathcal{E}_i(t+1). \tag{1.4}$$

Hence, one only needs the information $\Phi(X(t))$ from $\mathcal{F}_t$ to generate the particle position $X_i(t+1)$. In this case, one obtains the following weak limit for the ancestral partition process.

Theorem 1.1. *Assume that ξ_{ij} in (1.1) are Gumbel $G(\rho,\beta)$-distributed and that the initial position of particles $(X_1(0),\dots,X_N(0))$ are distributed according to a probability distribution μ on $\mathbb{R}^N$. Choose n particles $e_1,\dots,e_n$ uniformly at random from the N particles in generation $\lfloor T(\log N)\rfloor$. Let $(\Pi^{N,n}_{\lfloor t(\log N)\rfloor}; t\in[0,T[)$ be the random partition of $[n]$ such that i and j are in the same block if and only if e_i and e_j have the same ancestor in generation $\lfloor (T-t)(\log N)\rfloor$.*

Then the processes $(\Pi^{N,n}_{\lfloor t(\log N)\rfloor}; t\in[0,T[)$ converge weakly as $N\to\infty$ to a continuous time process $(\Pi^{\infty,n}_t; t\in[0,T[)$ that has the same law as the restriction to $[n]$ of the Bolthausen–Sznitman coalescent (up to time T^-).

The exponential model [5,7] and the population dynamics in (1.1) share the common property that the only information one needs from generation t in order to obtain generation $t+1$ is contained on a single function of the particles positions at time t. Using this property and shifting the particles positions appropriately, one can prove, for example, the independence between generations. Yet, it is important to point out that the techniques used to prove independence and the population models are different. In the exponential model, each individual has infinitely many offspring, but only the N right-most are selected to form the next generation. On the other hand, in (1.1) each individual has only N offspring and the selection mechanism is of a different nature. Indeed, we may label the offspring of $X_i(t)$ according to the $\xi_{ij}(t+1)$'s, so the child labeled $j\in\{1,\dots,N\}$ is at position $X_i(t)+\xi_{ij}(t+1)$. The selection is then made among individuals having the same label: $X_j(t+1)=\max_{1\le i\le N}\{X_i(t)+\xi_{ij}(t+1)\}$, and in generation $t+1$ we keep the right-most individual of each label j and not the N right-most individuals, as in the exponential model. Hence, Theorem 1.1 provides an other example of population dynamics in the presence of selection (or directed polymer in random medium) that validates the conjectures in [5–8].

The population dynamics obtained from (1.1) can be described as follows. The individuals in generation t have a (random) genetic fitness $\eta_i(t)$, that determines their average reproductive success. The total genetic fitness is a.s. constant and equal to one, $\sum\eta_i(t)=1$, then given $\eta(t)$ the offspring vectors $(\nu_1(t),\dots,\nu_N(t))$ are distributed according to a N-class multinomial with N trials and probabilities outcomes $\eta_i(t)$'s. If we assume that the offspring vectors $(\nu(t))_{t\in\mathbb{N}}$ are identically distributed and independent from generation to generation, then we obtain a "toy model," in which generations are not correlated. In this paper, we also study the ancestral history of this population. We make two additional assumptions on the fitness $\eta(t)$. First, we assume that each $\eta_i(t)$ is of the form

$$\eta_i(t) = Y_i(t)\Big/\sum_{j=1}^{N} Y_j(t), \tag{1.5}$$

where $Y_j(t)$ are i.i.d. positive random variables. Secondly, for some of our results, we assume that the tail distribution of $Y_i(t)$ satisfies

$$\lim_{y\to\infty} \mathbb{P}\big(Y_i(t) \geq y\big)/y^{-\alpha} = C, \tag{1.6}$$

where α and C are positive constants. To simplify the notation, the time parameter t is often omitted. Moreover, $\eta_i(t)$ in (1.5) does not change if we replace $Y_j(t)$ by $Y_j(t)C^{-1/\alpha}$, for this reason we may always assume that $C = 1$. Then we show that the ancestral processes converge weakly and that the limit distribution depends on α. Our result resembles Theorem 4 in [20], where Schweinsberg studies coalescent processes obtained from supercritical Galton–Watson processes.

Theorem 1.2. *Consider the dynamics of a constant size N population with infinitely many generations backward in time defined by the vectors $\nu(t) = (\nu_1(t), \dots, \nu_N(t))$, $t \in \mathbb{Z}$ of family sizes and denote by $\Pi_t^{N,n}$ the ancestral partition process. Suppose that the family sizes $\nu(t)$ are i.i.d. copies of ν a doubly stochastic multinomial random variable with N trials and probability outcomes $\eta = (\eta_1, \dots, \eta_N)$:*

$$\mathbb{P}\big(\nu = (i_1, \dots, i_N)|\eta\big) = \frac{N!}{i_1! \cdots i_N!} \eta_1^{i_1} \cdots \eta_N^{i_N},$$

where $i_1, \dots, i_N \in \mathbb{N}$ and $i_1 + \cdots + i_N = N$. Suppose also that η_i is of the form (1.5) with i.i.d. Y_i's. Then the following holds.

(a) *If $\mathbb{E}[Y_1^2] < \infty$ (in particular, if (1.6) holds and $\alpha > 2$), then the processes $(\Pi_{\lfloor t/c_N \rfloor}^{N,n}; t \geq 0)$ converge weakly as $N \to \infty$ to the Kingman's n-coalescent. The scaling factor c_N is asymptotically equivalent to N, precisely*

$$\lim_{N\to\infty} Nc_N = \frac{\mathbb{E}[Y_i^2]}{\mathbb{E}[Y_i]^2}.$$

(b) *If the Y_i's satisfy (1.6) with $\alpha = 2$, then the processes $(\Pi_{\lfloor t/c_N \rfloor}^{N,n}; t \geq 0)$ converge in the Skorokhod sense as $N \to \infty$ to the Kingman's n-coalescent. The scaling factor c_N is asymptotically equivalent to $N/\log N$*

$$\lim_{N\to\infty} \frac{Nc_N}{\log N} = \frac{2}{\mathbb{E}[Y_i]^2}.$$

(c) *When (1.6) holds with $1 \leq \alpha < 2$, then the processes $(\Pi_{\lfloor t/c_N \rfloor}^{N,n}; t \geq 0)$ converge in the Skorokhod sense as $N \to \infty$ to a continuous-time process $(\Pi_t^{\infty,n}; t \geq 0)$ that has the same law as the restriction to $[n]$ of the Λ-coalescent, where Λ is the probability measure associated with the* Beta$(2-\alpha; \alpha)$ *distribution. The transition rates are given by*

$$\lambda_{b;k} = \frac{B(k-\alpha; b-k+\alpha)}{B(2-\alpha; \alpha)}, \tag{1.7}$$

where $B(c,d) = \Gamma(c)\Gamma(d)/\Gamma(c+d)$ *is the beta function. The scaling factor* c_N *satisfies*

$$\lim_{N\to\infty} N^{\alpha-1}c_N = \frac{\alpha\Gamma(\alpha)\Gamma(2-\alpha)}{\mathbb{E}[Y_i]^\alpha} \qquad \text{if } 1<\alpha<2,$$
$$\lim_{N\to\infty} c_N \log N = 1 \qquad \text{if } \alpha = 1.$$

(d) *When* (1.6) *holds with* $0<\alpha<1$, *then the processes* $(\Pi_t^{N,n}; t\in\mathbb{N})$ *converge as* $N\to\infty$ *to a discrete-time Markov chain* $(\Pi_t^{\infty,n}; t\in\mathbb{N})$ *that has the same law as the restriction to* $[n]$ *of a discrete-time* Ξ_α*-coalescent. The transition probabilities are given by*

$$p_{b;b_1;\ldots;b_a;s} = \frac{\alpha^{a+s-1}(a+s-1)!}{(b-1)!}\cdot\prod_{i=1}^{a}\frac{\Gamma(b_i-\alpha)}{\Gamma(1-\alpha)}. \tag{1.8}$$

Despite the similarities between Theorem 1.2 and Theorem 4 in [20], we consider a population dynamics that is different from the one studied by Schweinsberg. In [20], each individual gives birth to $\zeta_i(t)\in\mathbb{N}$ children, but only N among the $\zeta_1(t)+\cdots+\zeta_N(t)$ survive. The survivors are chosen uniformly without replacement and $\nu_i(t)$ is the number of descendants that remain after the selection step. The distribution of $(\nu_1(t),\ldots,\nu_N(t))$ is then characterized by an urn model. Indeed, if $\zeta_i(t)$, $1\le i\le N$ is the number of balls in the urn which are labeled i, so ν_i is the number of i-balls sampled after N draws without replacement. If we suppose that the Y_i in Theorem 1.2 are integer valued, we may also compare the population dynamics with an urn model. In this case, though, ν_i is the number of i-balls sampled after N draws with replacement. Then $(\nu_1(t),\ldots,\nu_N(t))$ is distributed according to a multinomial with N trials and probability outcomes $Y_i(t)/(Y_1(t)+\cdots+Y_N(t))$ and we are under the hypotheses of Theorem 1.2. On the other hand, the Y_i are not necessarily integer valued and may be distributed according to any distribution satisfying (1.6). In fact, as the reader will see in the proof of Theorem 1.1, a relevant case is when the Y_i are distributed according to the inverse of an exponential distribution. Whereas in [20], $\zeta_i(t)$ must be an integer valued random variable, since it represents the number of offspring of the ith individual in generation t.

The paper is organized as follows: in Section 2, we recall some necessary definitions and results about coalescent processes. Then, in Section 3, we study the case where the disorder ξ_{ij} is Gumbel distributed and we obtain Theorem 1.1 as an application of Theorem 1.2, that will be proved later in Section 4. At the end of the paper, we include two Appendices, in which we prove some technical results.

2. Coalescent processes

Let $\mathscr{P}_n$ be the finite set of all partitions of $[n]$ and $\mathscr{P}_\infty$ the set of partitions of $\mathbb{N}^*$. For $\pi,\pi'\in\mathscr{P}_n$ we say that π' is a refinement of π if every equivalent class of π is either a union of several equivalence classes of π' or coincides with an equivalence class of π', we denote it by $\pi'\subset\pi$.

We call a $\mathscr{P}_n$-valued process $(\Pi_t^n; t\ge 0)$ a n-coalescent if it has right-continuous step function paths and if Π_s^n is a refinement of Π_t^n, whenever $s\le t$. We call a $\mathscr{P}_\infty$-valued process $(\Pi_t; t\ge 0)$

a coalescent if it has càdlàg paths and if Π_s is a refinement of Π_t for all $s < t$. In this paper, we use the notation $\Pi^{N,\cdot}$ to denote the ancestral partition of a constant size population with N individuals, while the notation $\Pi^{\infty,\cdot}$, or simply Π, stands for a coalescent process.

We denote by $\mathcal{D}([0,\infty);\mathscr{P}_n)$ the space of càdlàg functions on $[0,\infty)$ taking values in $\mathscr{P}_n$, obviously $(\Pi_t^n; t \geq 0) \in \mathcal{D}([0,\infty);\mathscr{P}_n)$. Since $\mathscr{P}_n$ endowed with the discrete metric is a separable complete metric space, the space $\mathcal{D}([0,\infty);\mathscr{P}_n)$ is also separable and complete in the Skorokhod distance. We say that a process converges in the Skorokhod sense if the distribution of the process converges weakly in $\mathcal{D}([0,\infty);\mathscr{P}_n)$ equipped with this metric.

2.1. Λ-coalescent

In [18], Pitman studied the so-called Λ-coalescent. It consists in "coalescents with multiple collisions" that are continuous time Markov chains taking value in $\mathscr{P}_\infty$. Λ-coalescents have the property that the rate at which blocks are merging does not depend on the size of the blocks nor on the integers that are in the blocks, moreover simultaneous collisions do not happen. Let $\lambda_{b,k}$ be the rate that k blocks merge into a single one when there are b blocks in total. The array $(\lambda_{b,k})_{2\leq k\leq b}$ determines the distribution of Π^n's and, consequently, the distribution of Π. As Pitman shows in [18], there exists a coalescent process with transition rates $\lambda_{b,k}$ if and only if the consistency condition

$$\lambda_{b,k} = \lambda_{b+1,k} + \lambda_{b+1,k+1}$$

holds. In this case, there exists a non-negative and finite measure on the Borel subsets of $[0,1]$ such that

$$\lambda_{b,k} = \int_{[0,1]} u^{k-2}(1-u)^{b-k}\Lambda(\mathrm{d}u).$$

The process is then called the Λ-coalescent. When Λ is a unit mass at zero, we obtain the Kingman's coalescent. Another notorious case is when Λ is the uniform distribution on $[0,1]$; this process was studied by Bolthausen and Sznitman in [3] and is named after the authors.

One can further generalize these processes and obtain $\mathscr{P}_\infty$-Markov processes that may undergo "simultaneous multiple collisions," the Ξ-coalescent, *see Möhle and Sagitov [17] and Schweinsberg [19].* Let $b, b_1, \dots, b_a, s$ be non-negative integers such that $b_1 \geq \dots \geq b_a \geq 2$ and $b = s + \sum b_i$. Then, Ξ-coalescent are $\mathscr{P}_\infty$-Markov processes characterized by the rates $\lambda_{b;b_1,\dots,b_a;s}$ at which b blocks merge into $a+s$ blocks, with s blocks that remain unchanged and a blocks that are obtained by the union of $b_1, \dots, b_a$ blocks before the merging. As Möhle and Sagitov observe in Lemma 3.3 of [17] (see also Schweinsberg [19]) the transition rates satisfy the following recursion:

$$\lambda_{b;b_1,\dots,b_a;s+1} = \lambda_{b;b_1,\dots,b_a;s} - \sum_{j=1}^{a} \lambda_{b+1;b_1,\dots,b_j+1,\dots,b_a;s} - s\lambda_{b+1;b_1,\dots,b_a,2;s-1}. \tag{2.1}$$

Hence, the distribution of a Ξ-coalescent is completely determined by the rates $\lambda_{b;b_1,\dots,b_a}$.

2.2. Weak convergence of ancestral processes

It is well known that coalescent processes may be obtained as the weak limit of ancestral processes [15–17]. Mölhe and Sagitov study a wide class of constant size population models, which have "been living forever" (so we may trace back the individuals genealogical tree indefinitely). They obtain general conditions under which the ancestral processes $\Pi_t^{N,\cdot}$ converge in the Skorokhod sense to a coalescent process. As usual denote by $\nu_i(t)$ the number of children of the ith individual in generation t

$$\nu_1(t) + \nu_2(t) + \cdots + \nu_N(t) = N, \qquad t \in \mathbb{Z}.$$

They assume that generations do not overlap and that the family sizes in different generations are i.i.d. Generally, it is also assumed that individuals in a given generation have the same propensity to reproduce:

(i) The offspring vectors $\nu(t), t \in \mathbb{Z}$ are i.i.d. copies of ν.
(ii) The offspring vector $(\nu_1, \dots, \nu_N)$ is N-exchangeable.

The first assumption is necessary since it ensures the Markov property of the ancestral partition process. Under the above assumptions, it is easy to compute the transition probability of $\Pi^{N,n}$. Let $\pi' \subset \pi$ be two partitions of $\mathscr{P}_n$ and denote by a and b the number of equivalent classes of π and π', respectively. Then b may be decomposed as follows: $b = b_1 + \cdots + b_a$, where b_i's are ordered positive integers denoting the number of equivalent classes of π' that we have to merge in order to obtain one equivalent class of π. By a combinatorial "putting balls into boxes" argument, we obtain that the transition probability from π' to π is

$$\begin{aligned} p_N(\pi', \pi) &= \mathbb{P}\big(\Pi_{t+1}^{N,n} = \pi \,|\, \Pi_t^{N,n} = \pi'\big) \\ &= \frac{1}{(N)_b} \sum_{\substack{i_1,\dots,i_a=1 \\ \text{all distinct}}}^{N} \mathbb{E}\big[(\nu_{i_1})_{b_1} \cdots (\nu_{i_a})_{b_a}\big], \end{aligned} \tag{2.2}$$

where $(N)_b := N(N-1)\cdots(N-b+1)$. If the offspring vector is N-exchangeable we may further simplify (2.2) obtaining

$$p_N(\pi', \pi) = \frac{(N)_a}{(N)_b} \mathbb{E}\big[(\nu_1)_{b_1} \cdots (\nu_a)_{b_a}\big].$$

We now state Mölhe and Sagitov result, we keep their notation and let c_N be the probability that two individuals, chosen randomly without replacement from some generation, have a common ancestor one generation backward in time (it is the same c_N appearing in the statement of Theorem 1.2).

$$c_N := \frac{1}{N(N-1)} \sum_i^N \mathbb{E}\big[\nu_i(t)\big(\nu_i(t) - 1\big)\big] = \frac{1}{(N-1)} \mathbb{E}\big[\nu_1(t)\big(\nu_1(t) - 1\big)\big]. \tag{2.3}$$

Theorem 2.1 (Mölhe and Sagitov [17]). *Suppose that for all $a \geq 1$ and $b_1 \geq \cdots \geq b_a \geq 2$, the limits*

$$\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_{b_1} \cdots (\nu_a)_{b_a}]}{N^{b_1+\cdots+b_a-a} c_N} \tag{2.4}$$

exist, and let $b := b_1 + \cdots + b_a$. If

$$\lim_{N\to\infty} c_N = 0,$$

then the time-rescaled ancestral processes $(\Pi^{N,n}_{\lfloor t/c_N \rfloor}, t \geq 0)$ converge weakly as $N \to \infty$ to a process $(\Pi^{\infty,n}_t, t \geq 0)$ that has the same law as the restriction to $[n]$ of a Ξ-coalescent. Furthermore, the transition rates $\lambda_{b;b_1,\dots,b_a}$, that characterize the distribution of $\Pi^{\infty,n}_t$, are equal to the limits in (2.4). On the other hand, if

$$\lim_{N\to\infty} c_N = c > 0,$$

then the processes $(\Pi^{N,n}_t, t \in \mathbb{N})$ converge weakly as $N \to \infty$ to a process $(\Pi^{\infty,n}_t, t \in \mathbb{N})$, which has the same law as the restriction to $[n]$ of a discrete-time Ξ-coalescent. The transition probabilities $p_{b;b_1,\dots,b_a}$ satisfy

$$p_{b;b_1,\dots,b_a} = \lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_{b_1} \cdots (\nu_a)_{b_a}]}{N^{b_1+\cdots+b_a-a}}. \tag{2.5}$$

The existence of the limits in (2.4) implies that the finite-dimensional distributions of $\Pi^{N,n}_{\lfloor t/c_N \rfloor}$ converge to those of the coalescent Π^n_t, as proved in [17]. The authors in [15,17] prove that when $c_N \to 0$ the sequence of processes $\Pi^{N,n}_{\lfloor t/c_N \rfloor}$ is tight, which implies the weak convergence in the Skorokhod sense.

3. Relation with Brunet and Derrida's model

In this section, we will assume that Theorem 1.2 holds and we show that when the ξ_{ij}'s are Gumbel distributed, then the family sizes $\nu(t)$ of the model (1.1) are i.i.d. and the distribution satisfies the hypotheses of Theorem 1.2 with $\alpha = 1$, which implies Theorem 1.1. We bring to the reader's attention two important details.

The first one is that the time restriction in the statement of Theorem 1.1 is a necessary condition. One immediate reason is that the ancestral process is not even defined for $t > T$. A more subtle reason is that the partition $\Pi^{N,n}_{\lfloor T(\log N) \rfloor}$ depends on the initial distribution $X_1(0), \dots, X_N(0)$. This dependence can be easily illustrated by the following example. One chooses an initial position of points: $X_1(0), \dots, X_N(0)$, for which $X_1(0) \gg X_i(0)$. Then, with an overwhelming probability, every individual in generation one descends from $X_1(0)$ and

$$\Pi^{N,n}_{\lfloor T(\log N) \rfloor} = \{(1, \dots, n)\},$$

in particular, as $N \to \infty$ the partition $\Pi^{N,n}_{\lfloor T(\log N)\rfloor}$ does not converge in distribution to the n-Bolthausen–Sznitman coalescent at time T.

Secondly, we emphasize that, in the general case, the offspring vectors $\nu(t)$ obtained from (1.1) may not be independent from generation to generation. We refer to [11] to provide a picture of a situation, in which the positions of the particles are highly related to the positions of their ancestors. It is considered the case, in which the distribution of ξ_{ij} depends on N

$$\mathbb{P}(\xi_{ij} = 0) = 1 - \mathbb{P}(\xi_{ij} = -1) = 1/N^{1+r}.$$

In this model, the number of leaders $\sharp\{i; X_i(t) = \max\{X_j(t)\}\}$ in generation t has a strong correlation with the number of leaders in generation $t-1$. Therefore, the fitness vectors $(\eta(t))_{t\in\mathbb{N}}$ between successive generations are correlated, and hence the offspring vectors $\nu(t)$ are not independent (in particular (i) in page 2216 does not hold).

Before proving Theorem 1.1, let us present some preliminary results and explain why the Gumbel case is particular. In [9], it is shown that the particles remain grouped as t increases and that the position of the front at time t may be described by any numerical function $\Phi : \mathbb{R}^N \to \mathbb{R}$ that is increasing for the partial order on $\mathbb{R}^N$ and that commutes to space translations by constant vectors

$$\Phi(x + r\mathbf{1}) = r + \Phi(x), \tag{3.1}$$

where $\mathbf{1}$ is the vector $(1, 1, \dots, 1) \in \mathbb{R}^N$. For a given function Φ, we denote by x^0 the vector $x \in \mathbb{R}^N$ shifted by $\Phi(x)$.

$$x^0 = x - \Phi(x).$$

The authors also prove that there exists a non-random constant v_N (not depending on $\Phi(\cdot)$) called speed of the front such that

$$\lim_{t\to\infty} \frac{\Phi(X(t))}{t} = v_N \qquad \text{a.s.}$$

It is then clear that there is no invariant measure for $X(t) := (X_1(t), \dots, X_N(t))$. On the other hand, if we consider the shifted process $X^0(t) := X(t) - \Phi(X(t))$, then there exists a unique invariant measure (depending on $\Phi(\cdot)$) for it. In the Gumbel case, an appropriate measure of the front location is

$$\Phi(x) = \beta^{-1} \log \sum_{i=1}^{N} \exp(\beta x_i). \tag{3.2}$$

In the proof of Proposition 3.1, we show that if the ξ_{ij} are Gumbel $G(\rho, \beta)$-distributed, then $\Phi(X(t))$ has all information needed to construct the next generation. The technique that we will present has been used in [4] to calculate the velocity and diffusion constant of the N-particles system. In [9], the authors use the same argument to calculate explicitly the invariant measure for the process $X^0(t)$. It has the law of a shifted vector $V^0 := V - \Phi(V)$ of a vector V obtained from a N-sample from a Gumbel $G(0, \beta)$. Summing up, when the disorder is Gumbel distributed the model is completely soluble, allowing exact computations.

Proposition 3.1. *Assume that ξ_{ij} in (1.1) are Gumbel $G(\rho,\beta)$-distributed and denote by $\nu_i(t)$ the number of descendants of $X_i(t)$ in generation $t+1$.*

Then, for every starting configuration μ the family sizes $\nu(t)=(\nu_1(t),\dots,\nu_N(t))$, $t\geq 1$ are i.i.d. copies of ν a doubly stochastic multinomial random variable with N trials and probability outcomes η_i given by

$$\eta_i=\mathcal{E}_i^{-1}\Big/\left(\sum_{k=1}^{N}\mathcal{E}_k^{-1}\right), \tag{3.3}$$

where $\{\mathcal{E}_i;1\leq i\leq N\}$ are independent and exponentially distributed with parameter 1. If μ has the law of a shifted vector $V^0:=V-\Phi(V)$ of a vector V obtained from a N-sample from a Gumbel $G(0,\beta)$, then we may take $t\geq 0$.

Proof. Let $\Phi(x)$ be given by (3.2), then $\Phi(x)$ has all information one needs to construct the next generation and the process shifted by Φ: $X_j^0(t)=X_j(t)-\Phi(X(t))$, are independent from generation to generation. Indeed, for $t\geq 1$ we may write $X_j(t)$ as follows *(see [9] (Theorem 3.1) and [4]):*

$$X_j(t)=\rho+\Phi\big(X(t-1)\big)-\beta^{-1}\log\mathcal{E}_j(t), \tag{3.4}$$

where $\mathcal{E}_j(t):=\min_{1\leq i\leq N}\{\exp(-\beta(\xi_{ij}(t)-\rho)-\beta X_i^0(t-1))\}$. Since $\xi_{ij}(t)$ are Gumbel $G(\rho,\beta)$-distributed, $\exp(-\beta(\xi_{ij}(t)-\rho))$ are exponentially distributed with parameter one. Hence, conditionally on $\mathcal{F}_{t-1}$,

$$\exp\big(-\beta\big(\xi_{ij}(t)-\rho\big)-\beta X_i^0(t-1)\big),\qquad 1\leq i\leq N$$

are independent and $\exp(-\beta(\xi_{ij}(t)-\rho)-\beta X_i^0(t-1))$ is distributed according to an exponential random variable with parameter $\exp(\beta X_i^0(t-1))$. Applying the stability property of the exponential law under independent minimum, we obtain that conditionally on $\mathcal{F}_{t-1}$ each variable $\mathcal{E}_i(t)$ is exponentially distributed with parameter one and, moreover, that the whole vector $\mathcal{E}(t):=(\mathcal{E}_i(t),i\leq N)$ is conditionally independent. Therefore, the vector $\mathcal{E}(t)$ is independent from $\mathcal{F}_{t-1}$ and its coordinates $\mathcal{E}_i(t)$, $1\leq i\leq N$ are i.i.d. having an exponential law with parameter one. Using once again the stability property of the exponential law under independent minimum,

$$\begin{aligned}\eta_i(t)&:=\mathbb{P}\big(\xi_{ij}(t+1)+X_i(t)>\xi_{kj}+X_k(t),\ \text{for every } k\neq i|\mathcal{F}_t\big)\\&=\mathbb{P}\Big(e^{-\beta(\xi_{ij}(t+1)-\rho)}e^{-\beta X_i(t)}<\min_{k\neq i}e^{-\beta(\xi_{kj}(t+1)-\rho)}e^{-\beta X_k(t)}|\mathcal{F}_t\Big)\\&=\exp\big(\beta X_i(t)\big)\Big/\left(\sum_{k=1}^{N}\exp\big(\beta X_k(t)\big)\right).\end{aligned} \tag{3.5}$$

Then, from (3.4) we obtain that

$$\eta_i(t)=\mathcal{E}_i^{-1}(t)\Big/\left(\sum_{k=1}^{N}\mathcal{E}_k^{-1}(t)\right), \tag{3.6}$$

which proves (3.3). In particular, the family sizes $\nu(1), \nu(2), \dots$ have the same distribution. If at $t=0$ the particles are distributed according to the invariant measure the same argument holds and $\nu(t), t \geq 0$ have the same distribution.

We now prove that the $\nu(t)$*'s are independent*. It suffices to show that

$$\mathbb{E}\bigl[f_1\bigl(\nu(1)\bigr)\cdots f_{t+1}\bigl(\nu(t+1)\bigr)\bigr] = \mathbb{E}\bigl[f_1\bigl(\nu(1)\bigr)\cdots f_t\bigl(\nu(t)\bigr)\bigr]\mathbb{E}\bigl[f_{t+1}\bigl(\nu(t+1)\bigr)\bigr], \tag{3.7}$$

for all continuous bounded functions $f_1(\cdot), \dots, f_t(\cdot), f_{t+1}(\cdot)$. Let $A_{i,j;t}$ be the event

$$A_{i,j;t} = \Bigl\{\xi_{ji}(t+1) + X_j(t) > \max_{k \neq i}\bigl\{\xi_{ki}(t+1) + X_k(t)\bigr\}\Bigr\}$$

that $X_i(t+1)$ descends from $X_j(t)$. Denote by $\mathcal{G}_t$ the σ-algebra generated by $\mathcal{F}_t$ and $A_{i,j;t}$ for every $1 \leq i, j \leq N$, then $\nu(1), \dots, \nu(t)$ are $\mathcal{G}_t$-measurable. We claim that $\nu(t+1)$ is independent from $\mathcal{G}_t$, which proves (3.7). Since $\nu(t+1)$ is completely determined by $\{\mathcal{E}_k(t+1), 1 \leq k \leq N\}$ and $\{\xi_{kl}(t+2), 1 \leq k, l \leq N\}$, it is immediate that it is independent from $\mathcal{F}_t$. Hence, we prove the claim once we show that $\nu(t+1)$ and $A_{i,j;t}$ are independent for every $1 \leq i, j \leq N$. Since

$$A_{i,j;t} \in \sigma\bigl\{\mathcal{F}_t; \bigl\{\xi_{ki}(t+1); 1 \leq k \leq N\bigr\}\bigr\} \subset \mathcal{F}_{t+1},$$

it suffices to show that $A_{i,j;t}$ is independent from $\sigma\{\mathcal{E}_k(t+1), 1 \leq k \leq N\}$. It is not hard to show that $\mathcal{E}_k(t+1)$ and $A_{i,j;t}$ are independent, whenever $k \neq i$ and we leave the details to the reader. Let $g(\cdot)$ be a bounded continuous function. Conditionally, on $\mathcal{F}_t$, $\mathcal{E}_i(t+1)$ is the minimum of N independent random variables exponentially distributed with parameters $\exp(\beta X_k^0(t-1))$ and the set $A_{i,j;t}$ is the event that the minimum is attained by $\exp(-\beta(\xi_{ji}(t) - \rho) - \beta X_j^0(t))$. Then, using standard properties of exponential distributions, we obtain

$$\begin{aligned}
&\mathbb{E}\bigl[g\bigl(\mathcal{E}_i(t+1)\bigr)\mathbf{1}_{A_{i,j;t}}|\mathcal{F}_t\bigr] \\
&\quad= \mathbb{P}(A_{i,j;t}|\mathcal{F}_t)\int_{\mathbb{R}_+} g(y)\cdot\frac{\exp(-y\sum e^{\beta X_k^0(t-1)})}{\sum e^{\beta X_k^0(t-1)}}\cdot \mathrm{d}y \\
&\quad= \mathbb{P}(A_{i,j;t}|\mathcal{F}_t)\int_{\mathbb{R}_+} \mathrm{d}y\, g(y)\exp(-y).
\end{aligned}$$

We used that X^0 is the process shifted by Φ, which satisfies $\sum e^{\beta X_k^0(t-1)} = 1$. Then $\mathcal{E}_i(t+1)$ and $A_{i,j;t}$ are independent, which proves the claim and, therefore, the proposition. □

Proof of Theorem 1.1. By Proposition 3.1, the family sizes $\nu(t)$ are independent and identically distributed for $t \geq 1$ (and $t \geq 0$ if the initial position of particles is distributed according to the invariant measure). Furthermore, it is easy to compute the tail distribution of $\mathcal{E}_i^{-1}(t)$

$$\mathbb{P}\bigl(\mathcal{E}_i^{-1}(t) \geq x\bigr) = 1 - e^{-x^{-1}} \sim 1/x, \qquad x \to \infty,$$

where "$\sim$" means that the ratio of the sides approaches to one as $x \to \infty$, so (1.6) holds with $\alpha = 1$.

If $T_0 < T$ and N is sufficient large such that $(T - T_0)(\log N) \geq 1$, then the family sizes $\nu(t), t \in \{\lfloor (T - T_0)(\log N)\rfloor, \dots, \lfloor T(\log N)\rfloor\}$ are i.i.d. It is then possible to apply Theorem 1.2 with $\alpha = 1$, which concludes the proof. □

4. Proof of Theorem 1.2

The proof of Theorem 1.2 will be divided in two main parts. In the first one, we focus on the case where Y_1 has finite second moment, which generalize $\alpha > 2$ in (1.6). The proof of the first part of Theorem 1.2 is an adaptation of the proof of part (a) of Theorem 4 in [20]. In the second part, we prove Theorem 1.2 for $\alpha \leq 2$. We do so by studying the Laplace transform of Y_i and its derivatives.

Before proving Theorem 1.2, we prove a general statement about multinomial distributions. In the next lemma, we will denote by ν a N-class multinomial random variable with N trials and by η_i the probability outcomes, that are not necessarily N-exchangeable.

Lemma 4.1. *Let $\nu = (\nu_1, \dots, \nu_N)$ be a doubly stochastic multinomial random variable with probability outcomes $\eta_1, \dots, \eta_N$. Let also $b_1 \geq \dots \geq b_a \geq 1$ and $b = b_1 + \dots + b_a$ (we also assume that $b \leq N$). Then*

$$\mathbb{E}\big[(\nu_1)_{b_1} \cdots (\nu_a)_{b_a}\big] = (N)_b \mathbb{E}\big[\eta_1^{b_1} \cdots \eta_a^{b_a}\big]. \tag{4.1}$$

Proof. To simplify the notation, we assume that $\eta_1, \dots, \eta_N$ are non-random. Then, ν is distributed according to a standard multinomial distribution.

$$\begin{aligned} &\mathbb{E}\big[(\nu_1)_{b_1} \cdots (\nu_a)_{b_a}\big] \\ &\quad= \sum_{\substack{i_j \geq b_j \\ i_1+\dots+i_a \leq N}} \frac{N! \eta_1^{i_1} \cdots \eta_a^{i_a} (1 - \eta_{1,\dots,a})^{N - i_{1,\dots,a}}}{i_1! \cdots i_a! (N - i_{1,\dots,a})!} \cdot \frac{i_1!}{(i_1 - b_1)!} \cdots \frac{i_a!}{(i_a - b_a)!}, \end{aligned} \tag{4.2}$$

where $i_{1,\dots,a} := i_1 + \dots + i_a$ and $\eta_{1,\dots,a} := \eta_1 + \dots + \eta_a$. By a change of variables $k_j = i_j - b_j$ we rewrite (4.2)

$$\begin{aligned} &\sum_{k_1+\dots+k_a \leq N-b} \frac{N!}{k_1! \cdots k_a! (N - b - k_{1,\dots,a})!} \cdot \eta_1^{k_1+b_1} \cdots \eta_a^{k_a+b_a} (1 - \eta_{1,\dots,a})^{N-b-k_{1,\dots,a}} \\ &\quad= (N)_b \eta_1^{b_1} \cdots \eta_a^{b_a} \sum \frac{(N-b)!}{k_1! \cdots k_a! (N - b - k_{1,\dots,a})!} \cdot \eta_1^{k_1} \cdots \eta_a^{k_a} (1 - \eta_{1,\dots,a})^{N-b-k_{1,\dots,a}} \\ &\quad= (N)_b \eta_1^{b_1} \cdots \eta_a^{b_a} \big(\eta_1 + \dots + \eta_a + (1 - \eta_{1,\dots,a})\big)^{N-b}, \end{aligned}$$

proving the result in the non-random case. The random case is obtained by conditioning on $\sigma\{\eta_1, \dots, \eta_N\}$. □

4.1. Convergence to Kingman's coalescent $\mathbb{E}[Y_1^2] < \infty$

In [16], Möhle shows that if the family sizes are not "too large" the processes $\Pi^{N,n}_{\lfloor t/c_N \rfloor}$ converge to the Kingman's n-coalescent.

Proposition 4.2 (Möhle [16]). *Suppose that*

$$\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_i)_3]}{N^2 c_N} = 0. \tag{4.3}$$

Then, as $N \to \infty$, the processes $\Pi^{N,n}_{\lfloor t/c_N \rfloor}$ converge to the Kingman's n-coalescent.

We will use Proposition 4.2 to prove Theorem 1.2 in the case where the Y_i's are square integrable. We first estimate c_N, the probability that two individuals have a common ancestor one generation backward in time.

Lemma 4.3. *Assume that the hypotheses of Theorem 1.2 hold with $\mathbb{E}[Y_1^2] < \infty$ and let c_N be as in (2.3). Then*

$$\lim_{N\to\infty} N c_N = \frac{\mathbb{E}[Y_1^2]}{\mathbb{E}[Y_1]^2}. \tag{4.4}$$

Proof. From Lemma 4.1, we obtain that

$$N c_N = N^2 \mathbb{E}\big[\eta_1^2\big].$$

Let $\delta_1 > 0$, then by definition of η_1

$$N^2 \mathbb{E}\big[\eta_1^2\big] = \mathbb{E}\left[\frac{Y_1^2}{(N^{-1}\sum_{j=1}^N Y_j)^2}\right] \geq \mathbb{E}\left[\frac{Y_1^2}{\delta_1 + (N^{-1}\sum_{j=1}^N Y_j)^2}\right]. \tag{4.5}$$

Since $Y_1 > 0$, we use dominated convergence in (4.5) to obtain that

$$\liminf_{N\to\infty} N c_N \geq \frac{\mathbb{E}[Y_1^2]}{\delta_1 + (\mathbb{E}[Y_1])^2}.$$

The inequality holds for every δ_1 positive, which implies that the above lim inf is larger than $\mathbb{E}[Y_1^2]/\mathbb{E}[Y_1]^2$. We now obtain an upper bound for the lim sup. We use the Markov inequality to obtain that for all $c > 0$

$$\lim_{x\to\infty} x^2 \mathbb{P}(Y_1 \geq cx) = 0. \tag{4.6}$$

Let $S_{2,N} = \sum_{i=2}^N Y_i$ and take $0 < \delta_2 < \mathbb{E}[Y_1]$ sufficiently small such that

$$\frac{\mathbb{E}[Y_1^2]}{(\mathbb{E}[Y_1] - \delta_2)^2} \leq \frac{\mathbb{E}[Y_1^2]}{\mathbb{E}[Y_1]^2} + \varepsilon/3, \tag{4.7}$$

for a fixed $\varepsilon > 0$. Then we write

$$
\begin{aligned}
N^2\mathbb{E}\big[\eta_1^2\big] &= \mathbb{E}\left[\frac{Y_1^2}{(N^{-1}Y_1 + N^{-1}S_{2,N})^2}; S_{2,N} \geq N\big(\mathbb{E}[Y_1] - \delta_2\big)\right] \\
&\quad + \mathbb{E}\left[\frac{Y_1^2}{(N^{-1}Y_1 + N^{-1}S_{2,N})^2}; S_{2,N} \leq N\big(\mathbb{E}[Y_1] - \delta_2\big)\right] \\
&= (\mathrm{I}) + (\mathrm{II}).
\end{aligned} \tag{4.8}
$$

Since $Y_i > 0$, we may bound (II) in (4.8) as follows:

$$
\begin{aligned}
(\mathrm{II}) &\leq \mathbb{E}\left[\frac{Y_1^2}{(N^{-1}Y_1)^2}; S_{2,N} \leq N\big(\mathbb{E}[Y_1] - \delta_2\big)\right] \\
&= N^2\mathbb{P}\big(S_{2,N} \leq N\big(\mathbb{E}[Y_1] - \delta_2\big)\big).
\end{aligned}
$$

So we apply the Chernoff inequality to conclude that if δ_2 is fixed and N sufficiently large, then (II) is smaller than $\varepsilon/3$.

$$
\begin{aligned}
(\mathrm{I}) &\leq \mathbb{E}\left[\frac{Y_1^2}{(\mathbb{E}[Y_1] - \delta_2)^2}; Y_1 \leq N\big(\mathbb{E}[Y_1] - \delta_2\big)\right] \\
&\quad + N^2\mathbb{P}\big(Y_1 \geq N\big(\mathbb{E}[Y_1] - \delta_2\big)\big) \\
&\leq \mathbb{E}\left[\frac{Y_1^2}{(\mathbb{E}[Y_1] - \delta_2)^2}\right] + N^2\mathbb{P}\big(Y_1 \geq N\big(\mathbb{E}[Y_1] - \delta_2\big)\big).
\end{aligned}
$$

From (4.6) with $c = \mathbb{E}[Y_1] - \delta_2$, the second term in the right-hand side converges to zero as $N \to \infty$, and we may choose N conveniently such that it is smaller than $\varepsilon/3$. It is implied that N is taken such that (II) is also smaller than $\varepsilon/3$. Then, applying the upper bounds in (4.8) we obtain

$$
N^2\mathbb{E}\big[\eta_1^2\big] \leq \frac{\mathbb{E}[Y_1^2]}{(\mathbb{E}[Y_1] - \delta_2)^2} + \frac{2}{3} \cdot \varepsilon < \frac{\mathbb{E}[Y_1^2]}{\mathbb{E}[Y_1]^2} + \varepsilon.
$$

Since the inequality holds for every $\varepsilon > 0$ and N large enough, we conclude that $\limsup Nc_N \leq \mathbb{E}[Y_1^2]/\mathbb{E}[Y_1]^2$ proving the lemma. □

Proof of Theorem 1.2 in the case $\mathbb{E}[Y_1^2] < \infty$. In order to prove Theorem 1.2, it suffices to show that (4.3) holds and apply Proposition 4.2. From Lemma 4.3, there exists a constant $c < 1$ such that for N sufficiently large $Nc_N > c\mathbb{E}[Y_1^2]/\mathbb{E}[Y_1]^2$, hence

$$
0 \leq \frac{\mathbb{E}[(\nu_1)_3]}{N^2 c_N} \leq \frac{\mathbb{E}[(\nu_1)_3]}{N} \cdot \frac{\mathbb{E}[Y_1]^2}{c\mathbb{E}[Y_1^2]}.
$$

Then, to prove the convergence in (4.3), it suffices to show that $N^{-1}\mathbb{E}[(\nu_1)_3] \to 0$. From (4.1), it is equivalent to $N^2\mathbb{E}[\eta_1^3] \to 0$ as $N \to \infty$. We proceed as in (4.8) and obtain

$$\begin{aligned} N^2\mathbb{E}\big[\eta_1^3\big] &= N^2\mathbb{E}\left[\frac{Y_1^3}{(Y_1+S_{2,N})^3}; S_{2,N} \geq N\big(\mathbb{E}[Y_1]-\delta_2\big)\right] \\ &\quad + N^2\mathbb{E}\left[\frac{Y_1^3}{(Y_1+S_{2,N})^3}; S_{2,N} \leq N\big(\mathbb{E}[Y_1]-\delta_2\big)\right] \\ &= (\mathrm{I}) + (\mathrm{II}). \end{aligned} \tag{4.9}$$

Applying the same argument of Lemma 4.3, we conclude that (II) converges to zero as N diverges and we also obtain the following upper bound to (I)

$$(\mathrm{I}) \leq N^2\mathbb{E}\left[\frac{Y_1^3}{(N(\mathbb{E}[Y_1]-\delta_2))^3}; Y_1 \leq N\big(\mathbb{E}[Y_1]-\delta_2\big)\right] + N^2\mathbb{P}\big(Y_1 \geq N\big(\mathbb{E}[Y_1]-\delta_2\big)\big). \tag{4.10}$$

We use the Markov inequality to show that the second term in the right-hand side of (4.10) converges to zero as $N \to \infty$. As a consequence, to finish the proof it suffices to show that the first term in the right-hand side of (4.10) converges to zero as $N \to \infty$. For $\varepsilon > 0$ let $L \in \mathbb{R}_+$ be such that

$$\mathbb{E}\big[Y_1^2; Y_1 \geq L\big]/\big(\mathbb{E}[Y_1]-\delta_2\big)^2 < \varepsilon/2.$$

Since L, δ_2 and ε are fixed we may choose N sufficiently large such that

$$\frac{L\mathbb{E}[Y_1^2]}{N(\mathbb{E}[Y_1]-\delta_2)^3} < \varepsilon/2,$$

and we bound the first term in the right-hand side of (4.10)

$$\begin{aligned} & N^2\mathbb{E}\left[\frac{Y_1^3}{(N\mathbb{E}[Y_1]-\delta_2)^3}; Y_1 \leq N\big(\mathbb{E}[Y_1]-\delta_2\big)\right] \\ &\quad \leq \frac{L}{N(\mathbb{E}[Y_1]-\delta_2)^3} \cdot \mathbb{E}\big[Y_1^2; Y_1 \leq L\big] + \frac{\mathbb{E}[Y_1^2; L \leq Y_1 \leq N(\mathbb{E}[Y_1]-\delta_2)]}{(\mathbb{E}[Y_1]-\delta_2)^2} \\ &\quad \leq \frac{L}{N(\mathbb{E}[Y_1]-\delta_2)^3} \cdot \mathbb{E}\big[Y_1^2\big] + \frac{\mathbb{E}[Y_1^2\mathbf{1}_{\{Y_1 \geq L\}}]}{(\mathbb{E}[Y_1]-\delta_2)^2} < \varepsilon, \end{aligned} \tag{4.11}$$

that finishes the proof. $\square$

4.2. Proof of Theorem 1.2 when $\alpha \leq 2$

The strategy to prove Theorem 1.2 in the case $\alpha \leq 2$ is to compute the limits (2.4) and apply Theorem 2.1. In the next proposition, we show how the moments of η_i's are related to the Laplace transform of Y_i.

Proposition 4.4. *Let $b_1 \geq b_2 \geq \cdots \geq b_a \geq 2$ be positive integers, $b = b_1 + \cdots + b_a$ and for $1 \leq i \leq N$*

$$\eta_i := \frac{Y_i}{\sum_{i=1}^N Y_j},$$

where $Y_1, \dots, Y_N$ are i.i.d. random variables. Then

$$\mathbb{E}\big[\eta_1^{b_1} \cdots \eta_a^{b_a}\big] = \frac{1}{\Gamma(b)} \int_0^\infty u^{b-1} I_0(u)^{N-a} I_{b_1}(u) \cdots I_{b_a}(u)\,\mathrm{d}u, \tag{4.12}$$

where $\Gamma(\cdot)$ is the gamma function and

$$I_p(u) = \mathbb{E}\big[Y_1^p e^{-uY_1}\big], \qquad p \in \mathbb{N}. \tag{4.13}$$

Proof. For every $z \in \mathbb{R}_+^*$ we have the following integral representation

$$z^{-b} = \frac{1}{\Gamma(b)} \int_0^\infty u^{b-1} e^{-uz}\,\mathrm{d}u, \tag{4.14}$$

then applying (4.14) with $z = \sum_{i=1}^N Y_i$ we obtain

$$\begin{aligned}
\mathbb{E}\big[\eta_1^{b_1} \cdots \eta_a^{b_a}\big] &= \mathbb{E}\left[Y_1^{b_1} \cdots Y_a^{b_a} \frac{1}{\Gamma(b)} \int_0^\infty u^{b-1} e^{-u\sum_{i=1}^N Y_i}\,\mathrm{d}u\right] \\
&= \int_0^\infty \frac{u^{b-1}}{\Gamma(b)} \mathbb{E}\big[Y_1^{b_1} \cdots Y_a^{b_a} e^{-u\sum_{i=1}^N Y_i}\,\mathrm{d}u\big] \qquad \text{(Fubini)} \qquad (4.15) \\
&= \int_0^\infty \frac{u^{b-1}}{\Gamma(b)} \mathbb{E}\big[\exp(-uY_1)\big]^{N-a} \prod_{i=1}^a \mathbb{E}\big[Y_1^{b_i} \exp(-uY_1)\big]\,\mathrm{d}u.
\end{aligned}$$

In the last equality, we used the fact that Y_i are i.i.d. Hence, from the definition of I_{b_i} we obtain that (4.15) and (4.12) are equal, proving the result. □

It is clear that the functions $I_p(u)$ are decreasing and attain their maximum at zero. Moreover, the following relation can be easily deduced

$$\frac{\mathrm{d}^p}{\mathrm{d}u^p} I_0(u) = (-1)^p I_p(u).$$

We now outline the strategy to prove Theorem 1.2.

1. We first obtain a precise asymptotic of $I_p(u)$ in the neighborhood of zero, where $I_p(u)$ attains its maximum. As the reader will see, the behavior of $I_p(u)$ depends on α and each case will be studied separately.

2. We show that the integral in the right-hand side of (4.12) is essentially determined by the immediate neighborhood of zero.
3. We estimate $\mathbb{E}[\eta_1^{b_1} \cdots \eta_a^{b_a}]$.
4. We prove Theorem 1.2 using Lemma 4.1 that relates (2.4) with $\mathbb{E}[\eta_1^{b_1} \cdots \eta_a^{b_a}]$.

Lemma 4.5. *Let $I_{\cdot}(u)$ be given by (4.13).*

(a) *If Y_i satisfies (1.6) with $\alpha = 2$ and $C = 1$. Then*

$$\begin{aligned} I_0(u) &= 1 - u\mathbb{E}[Y_1] + o(u) \qquad \text{when } u \to 0^+; \\ I_2(u) &= (-2\log u) + o\big(\log\big(u^{-1}\big)\big) \qquad \text{when } u \to 0^+; \\ I_p(u) &= u^{2-p}\big(2\Gamma(p-2)\big) + o\big(u^{2-p}\big) \qquad \text{when } p \geq 3 \text{ and } u \to 0^+. \end{aligned}$$

(b) *When Y_i satisfies (1.6) with $1 < \alpha < 2$ and $C = 1$. Then*

$$\begin{aligned} I_0(u) &= 1 - u\mathbb{E}[Y_1] + o(u) \qquad \text{when } u \to 0^+; \\ I_p(u) &= u^{\alpha-p}\big(\alpha\Gamma(p-\alpha)\big) + o\big(u^{\alpha-p}\big) \qquad \text{when } p \geq 2 \text{ and } u \to 0^+. \end{aligned}$$

(c) *If (1.6) holds with $\alpha = 1$ and $C = 1$. Then*

$$\begin{aligned} I_0(u) &= 1 + (u\log u) + o(u\log u), \qquad \text{when } u \to 0^+; \\ I_p(u) &= u^{1-p}\Gamma(p-1) + o\big(u^{1-p}\big), \qquad \text{when } p \geq 2 \text{ and } u \to 0^+. \end{aligned}$$

(d) *Assume that Y_i satisfies (1.6) with $0 < \alpha < 1$ and $C = 1$. Then*

$$\begin{aligned} I_0(u) &= 1 - u^{\alpha}\Gamma(1-\alpha) + o\big(u^{\alpha}\big) \qquad \text{when } u \to 0^+; \\ I_p(u) &= u^{\alpha-p}\big(\alpha\Gamma(p-\alpha)\big) + o\big(u^{\alpha-p}\big) \qquad \text{when } p \geq 2 \text{ and } u \to 0^+. \end{aligned}$$

Proof. See Appendix A. □

In the next lemma, we show that only the immediate neighborhood of zero contributes to the integral in (4.12) of Proposition 4.4.

Lemma 4.6. *Let $I_{\cdot}(u)$ be given by (4.13) and $\kappa_N := (\log N)^2/N$, assume also that Y_i satisfies (1.6) with $\alpha \leq 2$ and $C = 1$. Then, for every $K \in \mathbb{N}$*

$$\lim_{N\to\infty} N^K \int_{\kappa_N}^{\infty} u^{b-1} I_0(u)^{N-a} I_{b_1}(u) \cdots I_{b_a}(u)\,\mathrm{d}u = 0, \tag{4.16}$$

where $b_1 \geq \cdots \geq b_a$ are fixed integers and $b = b_1 + \cdots + b_a$. Hence, the integral in (4.16) decreases faster than any polynomial in N.

Proof. Since I_0 is a decreasing function

$$\begin{aligned}
&\int_{\kappa_N}^{\infty} u^{b-1} I_0(u)^{N-a} I_{b_1}(u)\cdots I_{b_a}(u)\,\mathrm{d}u \\
&\le I_0(\kappa_N)^{N-a}\int_{\kappa_N}^{\infty} u^{b-1} I_{b_1}(u)\cdots I_{b_a}(u)\,\mathrm{d}u \\
&\le I_0(\kappa_N)^{N-a}\int_{0}^{\infty} u^{b-1}\mathbb{E}\big[Y_1^{b_1}e^{-uY_1}\big]\cdots\mathbb{E}\big[Y_a^{b_a}e^{-uY_a}\big]\,\mathrm{d}u \\
&= I_0(\kappa_N)^{N-a}\Gamma(b)\mathbb{E}\bigg[\frac{Y_1^{b_1}\cdots Y_a^{b_a}}{(\sum_{i=1}^a Y_i)^b}\bigg].
\end{aligned} \tag{4.17}$$

In the last equality, we proceed as in Proposition 4.4 and use the integral representation (4.14) with $z=\sum_{i=1}^a Y_i$. The expected value in the right-hand side of (4.17) is bounded from above by one. Applying Lemma 4.5 with $u=\kappa_N\to 0^+$ as $N\to\infty$

$$\begin{aligned}
&I_0(\kappa_N)^{N-a} = \exp\big\{-\mathbb{E}[Y_i](\log N)^2 + o\big(\log^2 N\big)\big\} \qquad \text{if } 1<\alpha\le 2; \\
&I_0(\kappa_N)^{N-a} = \exp\big\{-(\log N)^3 + (\log N)^2(\log 2\log N) + o\big(\log^3 N\big)\big\} \qquad \text{if } \alpha=1; \\
&I_0(\kappa_N)^{N-a} = \exp\big\{-\Gamma(1-\alpha)N^{1-\alpha}(\log N)^{2\alpha} + o\big(N^{1-\alpha}(\log N)^{2\alpha}\big)\big\} \qquad \text{if } 0<\alpha<1;
\end{aligned}$$

that decreases faster than any polynomial in N. □

The κ_N in Lemma 4.6 is not optimal. The reason we have chosen such κ_N will be clear in the proof of Proposition 4.7 below, where we estimate $\mathbb{E}[\eta_1^{b_1}\cdots\eta_a^{b_a}]$.

Proposition 4.7. *Let $b_1\ge b_2\ge\cdots\ge b_a\ge 2$ be positive integers, $b=b_1+\cdots+b_a$, and η_i be as in Proposition 4.4.*

(a) *Suppose Y_i satisfies (1.6) with $\alpha=2$ and $C=1$. Let $g:=\max\{i; b_i\ge 3\}$, we adopt the convention that $\max\{\varnothing\}=0$. Then*

$$\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\big[\eta_1^{b_1}\cdots\eta_a^{b_a}\big]\cdot\frac{N^{2a}}{(\log N)^{a-g}} \\
&\quad= \Gamma(2a)\cdot\frac{2^a\prod_{i=1}^g\Gamma(b_i-2)}{\Gamma(b)\mathbb{E}[Y_1]^{2a}}.
\end{aligned} \tag{4.18}$$

(b) *If (1.6) holds with $1<\alpha<2$ and $C=1$. Then*

$$\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\big[\eta_1^{b_1}\cdots\eta_a^{b_a}\big]N^{a\alpha} \\
&\quad= \Gamma(a\alpha)\cdot\frac{\prod_{i=1}^a\alpha\Gamma(b_i-\alpha)}{\Gamma(b)\mathbb{E}[Y_1]^{a\alpha}}.
\end{aligned} \tag{4.19}$$

(c) *If we assume that Y_i satisfies* (1.6) *with $\alpha = 1$ and $C = 1$. Then*

$$\lim_{N\to\infty} \mathbb{E}\left[\eta_1^{b_1}\cdots\eta_a^{b_a}\right](N\log N)^a = \Gamma(a)\cdot\frac{\prod_{i=1}^a\Gamma(b_i-1)}{\Gamma(b)}. \tag{4.20}$$

(d) *If* (1.6) *holds with $0<\alpha<1$ and $C=1$. Then*

$$\lim_{N\to\infty} \mathbb{E}\left[\eta_1^{b_1}\cdots\eta_a^{b_a}\right]N^a = \Gamma(a)\cdot\frac{\alpha^{a-1}\prod_{i=1}^a\Gamma(b_i-\alpha)}{\Gamma(1-\alpha)^a\Gamma(b)}. \tag{4.21}$$

Proof. See Appendix B. □

We now compute c_N the probability that two individuals randomly chosen have the same ancestor.

Corollary 4.8. *Assume that the hypotheses of Theorem* 1.2 *hold and let c_N be as in* (2.3). *Assume also that the Y_i's satisfy* (1.6) *with $\alpha \le 2$ and $C = 1$. Then*

$$\begin{aligned}
\lim_{N\to\infty}\frac{Nc_N}{\log N} &= \frac{2}{\mathbb{E}[Y_1]^2} && \text{if } \alpha=2;\\
\lim_{N\to\infty}\frac{c_N}{N^{1-\alpha}} &= \frac{\alpha\Gamma(\alpha)\Gamma(2-\alpha)}{\mathbb{E}[Y_1]^\alpha} && \text{if } 1<\alpha<2;\\
\lim_{N\to\infty}(\log N)c_N &= 1 && \text{if } \alpha=1.
\end{aligned} \tag{4.22}$$

Finally, if Y_i satisfies (1.6) *with $0<\alpha<1$ and $C=1$, then*

$$\lim_{N\to\infty} c_N = \frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}. \tag{4.23}$$

Proof. It is a direct application of Lemma 4.1 and Proposition 4.7. □

Proof of Theorem 1.2 in the cases $\alpha \le 2$. We analyze each case separately and compute the limits

$$\lim_{N\to\infty}\frac{\mathbb{E}[(\nu_1)_{b_1}\cdots(\nu_a)_{b_a}]}{N^{b-a}c_N}.$$

If $\mathbb{P}(Y_i \ge x) \sim x^{-2}$ as $x\to\infty$, denote by $g = \max\{i; b_i \ge 3\}$ (as in Proposition 4.7). Then, as $N\to\infty$

$$\begin{aligned}
&\frac{\mathbb{E}[(\nu_1)_{b_1}\cdots(\nu_a)_{b_a}]}{N^{b-a}c_N}\\
&\quad=\frac{(N)_b}{N^{b-a}c_N}\cdot\mathbb{E}\left[\eta^{b_1}\cdots\eta^{b_a}\right] \qquad \text{(Lemma 4.1)}
\end{aligned}$$

$$\sim N^a \frac{N}{\log N} \cdot \frac{\mathbb{E}[Y_1]^2}{2} \cdot \mathbb{E}\big[\eta^{b_1} \cdots \eta^{b_a}\big] \qquad \text{(Corollary 4.8)}$$

$$\sim \frac{N^{a+1}}{\log N} \cdot \frac{\mathbb{E}[Y_1]^2}{2} \cdot \frac{(\log N)^{a-g}}{N^{2a}} \cdot \Gamma(2a) \cdot \frac{2^a \prod_{i=1}^{g} \Gamma(b_i - 2)}{\Gamma(b)\mathbb{E}[Y_1]^{2a}} \qquad \text{(Proposition 4.7)}$$

$$= \frac{(\log N)^{a-g-1}}{N^{a-1}} \cdot \Gamma(2a) \cdot \frac{2^{a-1} \prod_{i=1}^{g} \Gamma(b_i - 2)}{\Gamma(b)\mathbb{E}[Y_1]^{2(a-1)}},$$

which converges to zero whenever $a \geq 2$. If $a = 1 = g$, which implies $b_a = b \geq 3$, then

$$\frac{\mathbb{E}[(\nu_1)_{b_1}]}{N^{b-1}c_N} \sim \frac{1}{\log N} \cdot \frac{\Gamma(b-2)}{\Gamma(b)\mathbb{E}[Y_1]} \to 0 \qquad \text{as } N \to \infty.$$

On the other hand, if $a = 1$ and $g = 0$, that is, $b = 2$, then

$$\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_2]}{N^{2-1}c_N} = 1.$$

Hence, in the scaling limit we may only observe collisions of two distinct blocks that do not occur simultaneously, *that is,* Kingman's coalescent.

In the case $1 < \alpha < 2$ *we proceed as above obtaining*

$$\frac{\mathbb{E}[(\nu_1)_{b_1} \cdots (\nu_a)_{b_a}]}{N^{b-a}c_N} \sim \frac{\Gamma(\alpha a)}{N^{(a-1)(\alpha-1)}} \cdot \frac{\mathbb{E}[Y_1]^{\alpha}}{\alpha\Gamma(\alpha)\Gamma(2-\alpha)} \cdot \frac{\prod_{i=1}^{a} \alpha\Gamma(b_i - \alpha)}{\Gamma(b)\mathbb{E}[Y_1]^{\alpha a}} \qquad \text{as } N \to \infty,$$

that converges to zero whenever $a \geq 2$. If $a = 1$ and *a fortiori* $b_a = b$

$$\begin{aligned}
\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_b]}{N^{b-1}c_N} &= \frac{\Gamma(b-\alpha)}{\Gamma(b)\Gamma(2-\alpha)} \\
&= \frac{(b-1-\alpha)\cdots(2-\alpha)}{(b-1)!} \\
&= \frac{B(b-\alpha,\alpha)}{B(2-\alpha,\alpha)} = \lambda_{b;b},
\end{aligned}$$

where $B(c,d) = \Gamma(c)\Gamma(d)/\Gamma(c+d)$, as defined in Theorem 1.2. Hence, using the recursive formula (2.1) for $\lambda_{b;k}$

$$\begin{aligned}
\lambda_{b;b-1;1} &= \lambda_{b-1,b-1} - \lambda_{b,b} \\
&= \frac{\Gamma(b-1-\alpha)}{\Gamma(b-1)\Gamma(2-\alpha)} - \frac{\Gamma(b-\alpha)}{\Gamma(b)\Gamma(2-\alpha)} \\
&= \frac{\alpha}{b-1} \cdot \frac{\Gamma(b-1-\alpha)}{\Gamma(b-1)\Gamma(2-\alpha)} \\
&= \frac{B(b-1-\alpha,1+\alpha)}{B(2-\alpha,\alpha)} = \lambda_{b;b-1}.
\end{aligned}$$

We may proceed by recurrence and conclude the convergence to the Beta-coalescent.

In the case $\alpha = 1$, we have that

$$\frac{\mathbb{E}[(\nu_1)_{b_1}\cdots(\nu_a)_{b_a}]}{N^{b-a}c_N} \sim \frac{\Gamma(a)}{(\log N)^{a-1}} \cdot \frac{\prod_{i=1}^{a}\Gamma(b_i - 1)}{\Gamma(b)} \qquad \text{as } N \to \infty,$$

that converges to zero whenever $a \geq 2$, implying that we do not observe simultaneous collisions in the time scale. If $a = 1$ and *a fortiori* $b_a = b$

$$\lim_{N\to\infty} \frac{\mathbb{E}[(\nu_1)_b]}{N^{b-1}c_N} = \frac{\Gamma(b-1)}{\Gamma(b)} = \frac{1}{b-1} = \int_{[0,1]} x^{b-2}\,\mathrm{d}x.$$

Hence, using the recursive formula (2.1) for $\lambda_{b;k}$, we can conclude the convergence in distribution to the Bolthausen–Sznitman coalescent.

When $\alpha < 1$, by Corollary 4.8 $\lim c_N > 0$. Then, as $N \to \infty$

$$\begin{aligned}
\frac{\mathbb{E}[(\nu_1)_{b_1}\cdots(\nu_a)_{b_a}]}{N^{b-a}} &= \frac{(N)_b}{N^{b-a}} \cdot \mathbb{E}\bigl[\eta^{b_1}\cdots\eta^{b_a}\bigr] \qquad \text{(Lemma 4.1)} \\
&\sim \Gamma(a) \cdot \frac{\alpha^{a-1}\prod_{i=1}^{a}\Gamma(b_i - \alpha)}{\Gamma(1-\alpha)^a\Gamma(b)} \qquad \text{(Proposition 4.7)} \\
&= \frac{\alpha^{a-1}(a-1)!}{(b-1)!} \cdot \prod \frac{\Gamma(b_i - \alpha)}{\Gamma(1-\alpha)} \\
&= \frac{\alpha^{a-1}(a-1)!}{(b-1)!} \cdot \prod [1-\alpha]_{b_i - 1;1},
\end{aligned} \tag{4.24}$$

where $[x]_{m,y} := x(x+y)\cdots(x+(m-1)y)$. We finish the proof by observing that the limit in (4.24) is exactly the same limit that Schweinsberg obtains when studying coalescent processes that govern the genealogical trees of supercritical Galton–Watson processes with selection; see Section 4 of [20]. □

Appendix A: Proof of Lemma 4.5

In this appendix, we present the proof of Lemma 4.5. We first prove the expansion of $I_0(u)$ and then of $I_p(u)$ for $p \geq 2$. The idea of the proof is more or less the same for every $0 < \alpha \leq 2$, but some technical adaptations are required in specific cases.

The Laplace transform I_0 of Y_i is differentiable, when $1 < \alpha \leq 2$ and $I_0'(0) = \mathbb{E}[Y_i]$, then in this case, the expansion of $I_0(u)$ is obtained by a simple Taylor development at zero. For $\alpha \leq 1$, the Laplace transform of Y_1 is no longer differentiable at zero. On the other hand, we have that

$$\begin{aligned}
\mathbb{E}\bigl[e^{-uY_1}\bigr] &= \int_0^\infty e^{-x}\mathbb{P}(Y_1 \leq x/u)\,\mathrm{d}x \\
&= 1 - \int_0^{c(u)} e^{-x}\mathbb{P}(Y_1 \geq x/u)\,\mathrm{d}x - \int_{c(u)}^\infty e^{-x}\mathbb{P}(Y_1 \geq x/u)\,\mathrm{d}x,
\end{aligned} \tag{A.1}$$

where $c(u)$ is a function depending on u to be chosen. Let $c(u) = u \log\log(u^{-1})$, then

$$\frac{x}{u} \geq \log\log\bigl(u^{-1}\bigr) \qquad \text{if } x \geq c(u);$$

that diverges if $u \to 0^+$. It is also trivial that $c(u) = o(u^\alpha)$ (in the case $\alpha < 1$) and $c(u) = o(u \log u)$ (in the case $\alpha = 1$) as $u \to 0^+$. Hence, we can easily bound the first term in (A.1) by

$$\int_0^{c(u)} e^{-x} \mathbb{P}(Y_1 \geq x/u) \, \mathrm{d}x \leq c(u),$$

that it is negligible as $u \to 0^+$. We study the second term in (A.1), since x/u diverges if $x \geq c(u)$, we can replace $\mathbb{P}(Y_i \geq x/u)$ by its asymptotic equivalent u^α / x^α

$$\int_{c(u)}^{\infty} e^{-x} \mathbb{P}(Y_1 \geq x/u) \, \mathrm{d}x \sim u^\alpha \int_{c(u)}^{\infty} \frac{e^{-x}}{x^\alpha} \, \mathrm{d}x \qquad \text{as } u \to 0^+.$$

When $\alpha < 1$, we have that $\int_{c(u)}^{\infty} \frac{e^{-x}}{x^\alpha} \, \mathrm{d}x \to \Gamma(1-\alpha) < \infty$, that proves the statement in this case. For $\alpha = 1$, we use the following result, that may be found in [1] (Section 6.2, Example 4):

$$\int_z^{\infty} \frac{e^{-x}}{x} \, \mathrm{d}x = -\gamma - \log z - \sum_{m \geq 1} (-1)^m \frac{z^m}{m(m!)}, \qquad z \to 0^+, \tag{A.2}$$

where γ stands for the Euler–Mascheroni constant. Taking $z = c(u)$, we obtain that

$$\begin{aligned} \int_{c(u)}^{\infty} \frac{e^{-x}}{x} \, \mathrm{d}x &= -\gamma - \log\bigl(u \log\log\bigl(u^{-1}\bigr)\bigr) - \sum_{m \geq 1} (-1)^m \frac{(u \log\log(u^{-1}))^m}{m(m!)} \\ &= -\log u + o(\log u) \qquad \text{as } u \to 0^+, \end{aligned}$$

finishing the proof.

We now focus on the case $p \geq 2$. We start with the following relation:

$$\begin{aligned} I_p(u) &= \int_0^{\infty} \bigl(p x^{p-1} e^{-ux} - u x^p e^{-ux}\bigr) \mathbb{P}(Y_i \geq x) \, \mathrm{d}x \\ &= \int_0^{c(u)} \bigl(p u^{-p} x^{p-1} e^{-x} - u^{-p} x^p e^{-x}\bigr) \mathbb{P}(Y_i \geq x/u) \, \mathrm{d}x \qquad \text{(A.3)} \\ &\quad + \int_{c(u)}^{\infty} \bigl(p u^{-p} x^{p-1} e^{-x} - u^{-p} x^p e^{-x}\bigr) \mathbb{P}(Y_i \geq x/u) \, \mathrm{d}x, \qquad \text{(A.4)} \end{aligned}$$

where $c(u)$ is a function depending on u to be chosen. As we did above, we will choose $c(u)$ such that it is negligible in comparison to $u^{\alpha-p}$, but x/u diverges if $x \geq c(u)$.

Suppose that $\alpha < 2$ *or* $\alpha = 2$ *and* $p \geq 3$. Let $\beta \in]0,1[$ such that $\beta p > \alpha$ and choose $c(u) = u^\beta$ (it is trivial that such β does not exist if $p = \alpha = 2$). We bound (A.3) by

$$\begin{aligned}
&\left| \int_0^{c(u)} \left(pu^{-p}x^{p-1}e^{-x} - u^{-p}x^pe^{-x}\right)\mathbb{P}(Y_i \geq x/u)\,\mathrm{d}x \right| \\
&\quad \leq u^p \int_0^{c(u)} pu^{-p}x^{p-1} + u^{-p}x^p \mathrm{d}x \\
&\quad = u^{(\beta+1)p} + \frac{u^{(\beta+1)p+1}}{p+1},
\end{aligned}$$

that is negligible in comparison to $u^{\alpha-p}$ as $u \to 0^+$. We now turn our attention to (A.4), where x/u diverges as $u \to 0^+$. We may replace $\mathbb{P}(Y_i \geq x/u)$ by its asymptotic equivalent u^α/x^α, then as $u \to 0^+$

$$\begin{aligned}
&\int_{c(u)}^{\infty} \left(pu^{-p}x^{p-1}e^{-x} - u^{-p}x^pe^{-x}\right)\mathbb{P}(Y_i \geq x/u)\,\mathrm{d}x \\
&\quad \sim u^{\alpha-p} \int_{c(u)}^{\infty} \left(px^{p-\alpha-1}e^{-x} - x^{p-\alpha}e^{-x}\right)\mathrm{d}x \qquad \text{(A.5)} \\
&\quad = u^{\alpha-p}\alpha\Gamma(p-\alpha) - u^{\alpha-p}\int_0^{c(u)} \left(px^{p-\alpha-1}e^{-x} - x^{p-\alpha}e^{-x}\right)\mathrm{d}x.
\end{aligned}$$

Finally, the second term in the right-hand side of (A.5) is $o(u^{\alpha-p})$ as $u \to 0^+$, concluding the proof in the cases $\alpha < 2$ and $\alpha = 2$, with $p \geq 2$.

The case $p = 2$ *and* $\alpha = 2$ *is obtained as above, choosing* $c(u) = u \log\log(u^{-1})$ *and using the asymptotic development* (A.2). *We leave the details to the reader.*

Appendix B: Proof of Proposition 4.7

In this appendix, we prove Proposition 4.7. Once more, the main idea of the proof is roughly the same for every $0 < \alpha \leq 2$, but some technical adaptations are required in specific cases. For this reason, we will present a detailed proof of the case $\alpha = 2$ and only sketch the proofs of the other cases.

Let $\kappa_N = (\log N)^2/N$ be as in Lemma 4.6. By (4.12) and Lemma 4.6, we have that

$$\mathbb{E}\left[\eta_1^{b_1} \cdots \eta_a^{b_a}\right] = \frac{1}{\Gamma(b)} \int_0^{\kappa_N} u^{b-1} I_0(u)^{N-a} I_{b_1}(u) \cdots I_{b_a}(u)\,\mathrm{d}u + \varepsilon_N,$$

where ε_N decreases to zero faster than any polynomial in N. Hence, it suffices to show that

$$\lim_{N\to\infty} \frac{N^{2a}}{(\log N)^{a-g}} \cdot \int_0^{\kappa_N} u^{b-1} I_0(u)^{N-a} I_{b_1}(u) \cdots I_{b_a}(u)\,\mathrm{d}u = \frac{2^a \prod_{i=1}^{g} \Gamma(b_i - 2)}{\mathbb{E}[Y_1]^{2a}} \cdot \Gamma(2a). \quad \text{(B.1)}$$

Let $\varepsilon > 0$, since $\lim_{N\to\infty} \kappa_N = 0$ we apply Lemma 4.5 to conclude that there exists a N_0 such that for N larger than N_0 and $u \leq \kappa_N$

$$(1-\varepsilon)\big(2\Gamma(b_i-2)\big) \leq I_{b_i}(u)/u^{2-b_i} \leq (1+\varepsilon)\big(2\Gamma(b_i-2)\big) \qquad \text{if } b_i \geq 3;$$
$$2(1-\varepsilon) \leq I_2(u)/\log\big(u^{-1}\big) \leq 2(1+\varepsilon) \qquad \text{if } b_i = 2.$$

Since there are finitely many b_i's, we may take N_0 such that the inequalities hold for every $i \in \{1, 2, \dots, a\}$. As a consequence, for $N > N_0$

$$\begin{aligned}
&\int_0^{\kappa_N} u^{b-1} I_0(u)^{N-a} I_{b_1}(u)\cdots I_{b_a}(u)\,\mathrm{d}u \\
&\quad\geq (1-\varepsilon)^a 2^a \prod_{i=1}^{g} \Gamma(b_i-2) \int_0^{\kappa_N} u^{b-b_1-\dots-b_g-1+2g}\big(\log\big(u^{-1}\big)\big)^{a-g} I_0(u)^{N-a}\,\mathrm{d}u \qquad \text{(B.2)} \\
&\quad= (1-\varepsilon)^a 2^a \prod_{i=1}^{g} \Gamma(b_i-2) \int_0^{\kappa_N} u^{2a-1}\big(\log\big(u^{-1}\big)\big)^{a-g} I_0(u)^{N-a}\,\mathrm{d}u,
\end{aligned}$$

where we used $b = b_1 + \dots + b_a = b_1 + \dots + b_g + 2(a-g)$ (a similar argument may be used to obtain a similar upper bound). Applying Lemma 4.5 for I_0, we get that

$$\lim_{u\to 0^+} \frac{I_0(u)-1}{-u\mathbb{E}[Y_1]} = 1.$$

Hence, there exists a N_1 such that for $N \geq N_1$ and $u \leq \kappa_N$ (we assume that $N_1 \geq N_0$)

$$\big(1-u(1+\varepsilon)\mathbb{E}[Y_1]\big)^{N-a} \leq I_0(u)^{N-a} \leq \big(1-u(1-\varepsilon)\mathbb{E}[Y_1]\big)^{N-a}.$$

Applying the above inequality in (B.2) to obtain a lower bound, and by the change of variables $v = u(1+\varepsilon)\mathbb{E}[Y_1]N$ we get

$$\begin{aligned}
&(1-\varepsilon)^a 2^a \prod_{i=1}^{g} \Gamma(b_i-2) \int_0^{\kappa_N} u^{2a-1}\big(\log\big(u^{-1}\big)\big)^{a-g} I_0(u)^{N-a}\,\mathrm{d}u \\
&\quad\geq \frac{(1-\varepsilon)^a}{(1+\varepsilon)^{2a}} \cdot \frac{1}{N^{2a}} \cdot \frac{2^a \cdot \prod_{i=1}^{g} \Gamma(b_i-2)}{\mathbb{E}[Y_1]^{2a}} \\
&\qquad\times \int_0^{\gamma_N} v^{2a-1}\left(-\log\left(\frac{v}{N(1+\varepsilon)\mathbb{E}[Y_1]}\right)\right)^{a-g}\left(1-\frac{v}{N}\right)^{N-a}\mathrm{d}v,
\end{aligned}$$

where $\gamma_N = N(1+\varepsilon)\mathbb{E}[Y_1]\kappa_N$, then

$$-\log\big(v/\big(N(1+\varepsilon)\mathbb{E}[Y_1]\big)\big) = \log N\left(1 + \frac{\log((1+\varepsilon)\mathbb{E}[Y_1]) - \log v}{\log N}\right),$$

and for $v \leq (1+\varepsilon)\mathbb{E}[Y_1](\log N)^2 = \gamma_N$

$$\frac{|\log((1+\varepsilon)\mathbb{E}[Y_1]) - \log v|}{\log N} \to 0 \qquad \text{as } N \to \infty. \tag{B.3}$$

Moreover, (B.3) decays uniformly to zero for $v \leq \gamma_N$. *We bring to the reader's attention the choice of κ_N in Lemma 4.6, because it was chosen such that* (B.3) *decays to zero uniformly.* Then there exists a N_2 such that for $N \geq N_2$ (we assume that $N_2 \geq N_1$)

$$(1-\varepsilon)\log N \leq -\log\bigl(v/\bigl(N(1+\varepsilon)\mathbb{E}[Y_1]\bigr)\bigr) \leq (1+\varepsilon)\log N \qquad \text{for every } v \leq \gamma_N.$$

Then, for $N \geq N_2$ we may further bound (B.2) and obtain

$$\begin{aligned}\int_0^{\kappa_N} & u^{b-1} I_0(u)^{N-a} I_{b_1}(u) \cdots I_{b_a}(u)\,\mathrm{d}u \\ &\geq \frac{(1-\varepsilon)^{2a-g}}{(1+\varepsilon)^{2a}} \cdot \frac{(\log N)^{a-g}}{N^{2a}} \cdot \frac{2^a \prod_{i=1}^g \Gamma(b_i-2)}{\mathbb{E}[Y_1]^{2a}} \cdot \int_0^{\gamma_N} v^{2a-1}\left(1-\frac{v}{N}\right)^{N-a} \mathrm{d}v.\end{aligned} \tag{B.4}$$

Since $v \leq \gamma_N$, both v/N and v^2/N decay to zero as $N \to \infty$. We also have that

$$\left(1-\frac{v}{N}\right)^{N-a} = \exp\bigl(-v + \mathcal{O}\bigl(v^2/N\bigr)\bigr) \qquad \text{as } N \to \infty.$$

As a consequence, the following limit holds:

$$\lim_{N\to\infty} \int_0^{\gamma_N} v^{2a-1}\left(1-\frac{v}{N}\right)^{N-a} \mathrm{d}v = \Gamma(2a).$$

Since ε in (B.4) is arbitrary, we have that

$$\liminf_{N\to\infty} \mathbb{E}\bigl[\eta_1^{b_1} \cdots \eta_a^{b_a}\bigr] \cdot \frac{N^{2a}}{(\log N)^{a-g}} \geq \frac{2^a \prod_{i=1}^g \Gamma(b_i-2)}{\mathbb{E}[Y_1]^{2a}} \cdot \Gamma(2a).$$

We obtain an upper bound for the lim sup using a similar argument with the obvious changes, and we leave the details to the reader. Hence, the limit in (B.1) holds, which proves the statement.

We now sketch the proof of Proposition 4.7 in the remaining cases ($\alpha < 2$), and we explain briefly how to overcome possible difficulties. *The case $1 < \alpha < 2$ has no further difficulties and we leave the details of the proof to the reader. In the case $\alpha = 1$, the relevant term to estimate is of the form*:

$$\Gamma(b_1-1)\cdots\Gamma(b_a-1) \cdot \int_0^{\kappa_N} u^{b-1} I_0(u)^{N-a} u^{1-b_i} \cdots u^{1-b_a}\,\mathrm{d}u.$$

By Lemma 4.5, $I_0(u)^{N-a} \cong (1+u\log u)^{N-a}$. Then, by the change of variables $v = uN\log N$, we obtain an expression of the form:

$$\frac{\prod \Gamma(b_i-1)}{(N\log N)^a} \cdot \int_0^{\kappa_N N\log N} v^{a-1}\left(1+\frac{v}{N\log N}\log\frac{v}{N\log N}\right)^{N-a} \mathrm{d}v.$$

Since $v \le \kappa_N N \log N = (\log N)^3$, the equation inside of the parentheses has the following asymptotic behavior as $N \to \infty$:

$$\begin{aligned} 1 + \frac{v}{N \log N} \log\left(\frac{v}{N \log N}\right) &= 1 - \frac{v}{N} \cdot \left(1 + \frac{\log\log N - \log v}{\log N}\right) \\ &\cong 1 - \frac{v}{N}, \end{aligned}$$

then we may proceed as in the case $\alpha = 2$ to prove the statement. *In the case $\alpha < 1$, we will arrive to an equation of the form*

$$\prod \alpha \Gamma(b_i - \alpha) \int_0^{\kappa_N} u^{a\alpha - 1} I_0(u)^{N-a} \, \mathrm{d}u.$$

We then use the development of $I_0(u)$ in a neighborhood of zero and the change of variables $v = u^\alpha \Gamma(1-\alpha) N$ to obtain

$$\frac{\prod \alpha \Gamma(b_i - \alpha)}{\alpha \Gamma(1-\alpha)^a N^a} \int_0^{\kappa_N^\alpha \Gamma(1-\alpha) N} v^{a-1} \left(1 - \frac{v}{N}\right)^{N-a} \mathrm{d}v,$$

that finishes the proof.

Acknowledgments

I would like to thank my supervisor, Professor Francis Comets, for suggesting this problem, for the helpful discussions and for his patient guidance.

References

[1] Bender, C.M. and Orszag, S.A. (1999). *Advanced Mathematical Methods for Scientists and Engineers. I: Asymptotic Methods and Perturbation Theory*. New York: Springer. MR1721985

[2] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2013). The genealogy of branching Brownian motion with absorption. *Ann. Probab.* **41** 527–618. MR3077519

[3] Bolthausen, E. and Sznitman, A.-S. (1998). On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197** 247–276. MR1652734

[4] Brunet, É. and Derrida, B. (2004). Exactly soluble noisy traveling-wave equation appearing in the problem of directed polymers in a random medium. *Phys. Rev. E* (3) **70** 016106, 5. MR2125704

[5] Brunet, É. and Derrida, B. (2013). Genealogies in simple models of evolution. *J. Stat. Mech. Theory Exp.* **1** P01006, 20. MR3036206

[6] Brunet, E., Derrida, B. and Damien, S. (2008). Universal tree structures in directed polymers and models of evolving populations. *Phys. Rev. E* **78** 061102.

[7] Brunet, E., Derrida, B., Mueller, A.H. and Munier, S. (2006). Noisy traveling waves: Effect of selection on genealogies. *Europhys. Lett.* **76** 1–7. MR2299937

[8] Brunet, É., Derrida, B., Mueller, A.H. and Munier, S. (2007). Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Phys. Rev. E* (3) **76** 041104, 20. MR2365627

[9] Comets, F., Quastel, J. and Ramírez, A.F. (2013). Last passage percolation and traveling fronts. *J. Stat. Phys.* **152** 419–451. MR3082639
[10] Cook, J. and Derrida, B. (1990). Directed polymers in a random medium: $1/d$ expansion and the n-tree approximation. *J. Phys. A* **23** 1523–1554. MR1048783
[11] Cortines, A. (2014). Front velocity and directed polymers in random medium. *Stochastic Process. Appl.* **124** 3698–3723. MR3249352
[12] Huillet, T. and Möhle, M. (2011). Population genetics models with skewed fertilities: A forward and backward analysis. *Stoch. Models* **27** 521–554. MR2827443
[13] Huillet, T. and Möhle, M. (2013). On the extended Moran model and its relation to coalescents with multiple collisions. *Theoretical Population Biology* **87** 5–14.
[14] Kingman, J.F.C. (1982). On the genealogy of large populations. *J. Appl. Probab.* **19A** 27–43. MR0633178
[15] Möhle, M. (1999). Weak convergence to the coalescent in neutral population models. *J. Appl. Probab.* **36** 446–460. MR1724816
[16] Möhle, M. (2000). Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. in Appl. Probab.* **32** 983–993. MR1808909
[17] Möhle, M. and Sagitov, S. (2001). A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29** 1547–1562. MR1880231
[18] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.* **27** 1870–1902. MR1742892
[19] Schweinsberg, J. (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5** 50 pp. (electronic). MR1781024
[20] Schweinsberg, J. (2003). Coalescent processes obtained from supercritical Galton–Watson processes. *Stochastic Process. Appl.* **106** 107–139. MR1983046

Received June 2014 and revised March 2015