Submitted to Bernoulli
arXiv: arXiv:1406.1432

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

ASER CORTINES[1,*]

[1] *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's 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 $(\nu_1, \ldots, \nu_N)$ is given by a Multinomial law with N trials and probability outcomes η_i's. We then show that the average coalescence times scale 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: Coalescence, Traveling waves, Ancestral processes, Bolthausen-Sznitman coalescent.

1. Introduction

An important question in the study of populations' evolution is to understand the effects of selection and mutation in the genealogy of a population. For a given population we would like to know how individuals are related and how many generations do we have to trace back in time in order to find a common ancestor. Kingman [14] was one of the first mathematicians to give a mathematical formulation to this problem and study the ancestral history of a population. He shows that in the absence of selection (neutral models) the populations' genealogical structure satisfies universal features, *see also* [15, 16, 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. The mutations are represented by real numbers, indicating how fit an individual is to the environment. 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 nothing but a stochastic model in front propagation. The selection mechanism constrains the particles to stay together. And 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 [4, 7, 8].

Recent results suggest that in rapidly adapting population models the genealogical correlations between individuals have universal features. It is conjecture [4, 7, 8] that the genealogical trees of these populations converge to those of a Bolthausen-Sznitman coalescent and that the average coalescence times scale like the logarithmic of the populations' size. The conjectures contrast with classical results in neutral population models, such as Wright-Fisher and Moran's models. It is proved that in neutral population models the genealogical trees converge to those of a Kingman's coalescent and that the average coalescence times scale like N, the size of the population [14, 15, 16, 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" [4, 7, 8] is an example of constant size population dynamics, for which a complete mathematical treatment is possible. Each individual i at 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. J. Berestycki, N. Berestycki, J. 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, thus in the near-critical regime 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. 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 (Λ-coalescent). 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's dynamics derived from the following model in front propagation [5]. It consists in a constant number N of evolving particles on the real line initially at positions $X_1(0), \ldots, X_N(0)$. 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:

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

where $\big\{\xi_{ij}(s);\, 1 \le i, j \le N,\, s \in \mathbb{N}\big\}$ are i.i.d. real random variables. The model may 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 samples 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 j-th point on the L-th plane [10, 11]. The optimal path starting at the same point but arriving at different points give 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 [6].

If the distribution of $\xi_{ij}(t+1)$ in (1.1) has no atoms, *i.e.* 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^*

$$X_j(t+1) = X_{i^*}(t) + \xi_{i^*j}(t+1). \tag{1.2}$$

In this sense, we may say that $X_j(t+1)$ is an offspring or a descendant of $X_{i^*}(t)$ and we denote by $\nu_{i^*}(t)$ the number of descendants of $X_{i^*}(t)$ at generation $t+1$. The fitnesses of the individuals are given by their positions $X_1(t), \dots, X_N(t)$ and conditionally on $\mathcal{F}_t := \sigma\{\xi_{ij}(s), s \le t, 1 \le i, j \le N\}$ the probability that $X_j(t+1)$ descends from $X_{i^*}(t)$ is given by

$$\eta_{i^*}(t) := \mathbb{P}\Big(\xi_{i^*j}(t+1) + X_{i^*}(t) \ge \xi_{kj}(t+1) + X_k(t)\,;\ \text{for every } 1 \le k \le N \,\Big|\, \mathcal{F}_t\Big). \tag{1.3}$$

Since $\{\xi_{ij}(t+1);\, 1 \le i, j \le N\}$ are independent, it is easy to see that for $j_1, \dots, j_m$ pairwise distinct

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

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

We analyze the genealogical tree of the population by observing the ancestral partition process, *i.e.* 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, *i.e.*

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

the microscopic dynamics can be solved allowing precise calculations, see also [5] where Brunet and Derrida use a similar technique to compute the exact asymptotic for the speed of the front. We then obtain the following weak limit for $\Pi_t^{N,n}$.

Theorem 1.1. *Assume that ξ_{ij} in (1.1) are Gumbel $G(\rho, \beta)$-distributed and that the initial position of particles $\big(X_1(0), \dots, X_N(0)\big)$ are distributed according to μ a probability distribution on $\mathbb{R}^N$. Choose n particles uniformly at random from the N particles at generation $\lfloor T(\log N)\rfloor$ and label those particles $e_1, \dots, e_n$. For $0 \le T_0 < T$ let $\left(\Pi^{N,n}_{\lfloor t(\log N)\rfloor}; t \in [0, T_0]\right)$ 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 at generation $\lfloor (T-t)(\log N)\rfloor$.*

Then, the processes $\left(\Pi^{N,n}_{\lfloor t(\log N)\rfloor}; t \in [0, T_0]\right)$ converge weakly as $N \to \infty$ to a continuous time process $\big(\Pi_t^{\infty,n}; t \in [0, T_0]\big)$ that has the same law as the restriction to $[n]$ of the Bolthausen-Sznitman coalescent (up to time T_0). 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 $T_0 = T$.

We draw the reader's attention to the differences between the population dynamics in (1.1) and the exponential model in [7]. 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's dynamics in the presence of selection (or directed polymer in random medium) that validates the conjectures in [4, 6, 7, 8].

Furthermore, we also study the evolution of an asexual (haploid) population inspired by (1.1), in which the individuals at time t have a (random) genetic fitness $\eta_i(t)$, that determines their average reproductive success. The total genetic fitness at time t is a.s. constant and equal to one $\sum \eta_i(t) = 1$, given $\eta(t)$ the offspring vectors $\big(\nu_1(t), \dots, \nu_N(t)\big)$ are distributed according to a N-class Multinomial with N trials and success probabilities $\eta_i(t)$'s. We focus on a "toy model" and we will assume that $\nu(t)$ is N-exchangeable and independent from generation to generation. Therefore it consists in a generalization of the classical Wright-Fisher model, in which the offspring vectors are i.i.d. copies of a N-class Multinomial random variable with N trials and success probabilities $1/N$.

We also make two assumptions on the fitness $\eta(t)$. First, we assume that each $\eta_i(t)$ is of the form

$$\eta_i(t) = Y_i(t) \Big/ \sum_{j=1}^{N} Y_j(t), \tag{1.4}$$

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

$$\lim_{y\to\infty} \mathbb{P}\left(Y_i(t) \geq y\right)/y^{-\alpha} = C, \tag{1.5}$$

where α and C are positive constants. To simplify the notation, the time parameter t is often omitted. Moreover, $\eta_i(t)$ in (1.4) 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 in (1.5). 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 backwards in time defined by the vectors $\nu(t) = \big(\nu_1(t), \dots, \nu_N(t)\big)$, $t \in \mathbb{Z}$ 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 $\eta = \big(\eta_1, \dots, \eta_N\big)$:*

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

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 (1.4) with i.i.d. Y_i's. Then, the following holds.

a. If $\mathbb{E}[Y_1^2] < \infty$ (in particular, if (1.5) holds and $\alpha > 2$), then the processes $\big(\Pi^{N,n}_{\lfloor t/c_N \rfloor}; t \geq 0\big)$ converge weakly as $N \to \infty$ to the Kingman's n-coalescent. The scaling factor c_N is asymptotically equivalent to N, precisely

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

b. If the Y_i's satisfy (1.5) with $\alpha = 2$, then the processes $\Big(\Pi^{N,n}_{\lfloor t/c_N \rfloor}; t \geq 0\Big)$ converge in the Skorokhod sense as $N \to \infty$ to the Kingman's n-coalescent. The scaling factor c_N is asymptotically equivalent to $N/\log N$

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

c. *When (1.5) holds with $1 \le \alpha < 2$, then the processes $\left(\Pi^{N,n}_{\lfloor t/c_N \rfloor}; t \ge 0\right)$ converge in the Skorokhod sense as $N \to \infty$ to a continuous-time process $(\Pi^{\infty,n}_t; t \ge 0)$ 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*

$$\lambda_{b;k} = \frac{B(k-\alpha; b-k+\alpha)}{B(2-\alpha;\alpha)}, \tag{1.6}$$

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

$$p_{b;b_1;\ldots;b_a;s} = \frac{\alpha^{a+s-1}(a+s-1)!}{(b-1)!} \cdot \prod_{i=1}^{a} \frac{\Gamma(b_i-\alpha)}{\Gamma(1-\alpha)}. \tag{1.7}$$

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 $\zeta_i(t)$ 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 $\big(\nu_1(t),\ldots,\nu_N(t)\big)$ 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. On the other hand, if we consider the same urn model, but ν_i is the number of i-balls sampled after N draws with replacement. Then, $\big(\nu_1(t),\ldots,\nu_N(t)\big)$ is distributed according to a Multinomial with N trials and probability outcomes $\zeta_i(t)/\big(\zeta_1(t)+\cdots+\zeta_N(t)\big)$ and we are under the hypothesis of Theorem 1.2.

The paper is organized as follows: in Section 2 we recall some necessary definition 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. In the end of the paper we include two Appendix, 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 $\big(\Pi^n_t; t \ge 0\big)$ a n-coalescent if it has right-continuous step function paths and if Π^n_s is a refinement of Π^n_t, 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 $\mathcal{D}\big([0,\infty);\mathscr{P}_n\big)$ the space of càdlàg functions on $[0,\infty)$ taking values in $\mathscr{P}_n$, obviously $\big(\Pi^n_t; t\geq 0\big)\in\mathcal{D}\big([0,\infty);\mathscr{P}_n\big)$. Since $\mathscr{P}_n$ endowed with the discrete metric is a separable complete metric space, the space $\mathcal{D}\big([0,\infty);\mathscr{P}_n\big)$ 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 $\mathcal{D}\big([0,\infty);\mathscr{P}_n\big)$ equipped with this metric.

2.1. Λ-coalescent.

In [18], Pitman studies 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 $\big(\lambda_{b,k}\big)_{2\leq k\leq b}$ 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 nonnegative 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. An other 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 $b,b_1,\ldots,b_a,s$ be nonnegative integers such that $b_1\geq\ldots\geq b_a\geq 2$ and $b=s+\sum b_i$. Then, Ξ-coalescent are $\mathscr{P}_\infty$-Markov processes characterized by the rates $\lambda_{b;b_1,\ldots,b_a;s}$ 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 $b_1,\ldots,b_a$ 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:

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

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

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, 16, 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 $\Pi^{N,\cdot}_t$ converge in the Skorokhod sense to a coalescent process. As usual denote by $\nu_i(t)$ the number of children of the i-th 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 $\big(\nu_1, \dots, \nu_N\big)$ 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

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

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

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

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).

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

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

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

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

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

then the time-rescaled ancestral processes $\Big(\Pi^{N,n}_{\lfloor t/c_N \rfloor},\, t \geq 0\Big)$ converge weakly as $N \to \infty$ to a process $\Big(\Pi^{\infty,n}_t,\, t \geq 0\Big)$ 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 $\Big(\Pi^{N,n}_t,\, t \in \mathbb{N}\Big)$ converge weakly as $N \to \infty$ to a process $\Big(\Pi^{\infty,n}_t,\, t \in \mathbb{N}\Big)$, 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

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

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 hypothesis of Theorem 1.2 with $\alpha = 1$, which implies Theorem 1.1. We bring the reader's attention to two important details.

First detail is that the time restriction in the statement of Theorem 1.1 is a necessary condition. One immediate reason for it 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 $X_1(0), \dots, X_N(0)$. For example, if the initial distribution of points is deterministic and $X_1(0) \gg X_i(0)$ for every $i \geq 2$, then with a large probability

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

implying that $\Pi^{\infty,n}_T$ is not distributed as a n-Bolthausen-Sznitman coalescent at time T. On the other hand, if we suppose that the initial distribution of particles satisfies certain conditions, then the convergence in Theorem 1.1 holds for $T_0 \leq T$.

Secondly, we emphasize that in the general case the family sizes in (1.1) may not be independent. We refer to [11] to provide a picture of a situation, in which the positions of particles are highly related to the positions of their ancestors. It is considered a slight different case, in which the distribution of ξ_{ij} depends on N

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

In this case, there exists a stopping time τ, such that if $t < \tau$ and $X_i(t)$ is at leading position (*i.e.* $X_i(t) = \max\{X_j(t)\}$) then its ancestor is also a leader at generation $t-1$. Therefore the genetic advantages are transmitted between generations, implying correlation between the family sizes (hence (i) in page 7 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

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

It is then clear that there is no invariant measure for $X(t) := \big(X_1(t), \dots, X_N(t)\big)$. On the other hand, if we consider the process $X^0(t) := X(t) - \Phi\big(X(t)\big)$ (*the process seen from the leading edge*), then there exists a unique invariant measure (depending on $\Phi(\cdot)$) for it. In the Gumbel case an appropriate measure of the front's location is

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

If ξ_{ij} are Gumbel $G(\rho, \beta)$-distributed and Φ is as in (3.2), then the invariant measure for the process $X^0(t)$ has the law of a shifted vector $V^0 := V - \Phi(V)$ of a vector V obtained from a N-sample from a Gumbel $G(0, \beta)$. Hence the model is completely soluble, allowing exact computations [5, 9].

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 $X_i(t)$ at generation $t + 1$.*

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

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

where $\big\{\mathcal{E}_i;\ 1 \leq i \leq N\big\}$ 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 $X_j^0(t)$ be the process seen from the leading edge: $X_j^0(t) = X_j(t) - \Phi\big(X(t)\big)$, where the front position is given by (3.2). Then, for $t \geq 1$ we may write $X_j(t)$ as follows, *see Theorem 3.1 in [9]*

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

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

$$\exp\big(- \beta(\xi_{ij}(t) - \rho) - \beta X_j^0(t-1)\big); \qquad 1 \leq j \leq N$$

are independent and $\exp\big(- \beta(\xi_{ij}(t) - \rho) - \beta X_j^0(t-1)\big)$ is distributed according to an exponential random variable with parameter $\exp\big(\beta X_j^0(t-1)\big)$. 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 $\mathcal{E}(t) := \big(\mathcal{E}_i(t), i \leq N\big)$ 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 we get that

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

Then, from (3.4) we obtain that

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

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

$$\mathbb{E}\left[f_1\big(\nu(1)\big)\dots f_{t+1}\big(\nu(t+1)\big)\right] = \mathbb{E}\left[f_1\big(\nu(1)\big)\dots f_t\big(\nu(t)\big)\right]\mathbb{E}\big[f_{t+1}\big(\nu(t+1)\big)\big]\,, \tag{3.7}$$

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

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

that $X_i(t+1)$ descends from $X_j(t)$. Denote by $\mathcal{G}_t$ the sigma 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 $\big\{\mathcal{E}_k(t+1),\, 1 \leq k \leq N\big\}$ and $\big\{\xi_{kl}(t+2),\, 1 \leq k, l \leq N\big\}$ 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

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

it suffices to show that $A_{i,j;t}$ is independent from $\sigma\big\{\mathcal{E}_k(t+1),\, 1 \leq k \leq N\big\}$. 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 $\exp\big(\beta X_k^0(t-1)\big)$ and the set $A_{i,j;t}$ is the event that the minimum is attained by $\exp\big(-\beta(\xi_{ji}(t) - \rho) - \beta X_j^0(t)\big)$. Then using standard properties of exponential distributions we obtain

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

We used that X^0 is the process seen from the leading edge, 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 $\mathcal{E}_i^{-1}(t)$

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

where "$\sim$" means that the ratio of the sides approaches to one as $x \to \infty$, so (1.5) 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 (1.5). 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 in the cases where $\alpha \leq 2$. 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 $\nu = \big(\nu_1, \dots, \nu_N\big)$ be a doubly stochastic Multinomial random variable with probability outcomes $\eta_1, \dots, \eta_N$. Let also $b_1 \geq \cdots \geq b_a \geq 1$ and $b = b_1 + \cdots + b_a$ (we also assume that $b \leq N$). Then,*

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

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.

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

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

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

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 $\mathbb{E}[\mathbf{Y_1^2}] < \infty$.

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*

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

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 backwards in time.

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

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

Proof. From Lemma 4.1, we obtain that

$$Nc_N = N^2\mathbb{E}\left[\eta_1^2\right].$$

Let $\delta_1 > 0$, then from the definition of η_1

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

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 bigger 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

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

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

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

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

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

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

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

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

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

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

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

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)

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

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

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

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

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

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

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

that finishes the proof. □

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

The strategy to prove Theorem 1.2 in the case $\alpha \le 2$ 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 $b_1 \ge b_2 \ge \ldots \ge b_a \ge 2$ be positive integers, $b = b_1 + \cdots + b_a$ and for $1 \le i \le N$*

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

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

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

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

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

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

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

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

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

In the last equality in (4.15) 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 its 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 of the proof of 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 $\mathbb{E}\big[\eta_1^{b_1} \dots \eta_a^{b_a}\big]$.
4. We prove Theorem 1.2 using Lemma 4.1 that relates (2.4) with $\mathbb{E}\big[\eta_1^{b_1} \dots \eta_a^{b_a}\big]$.

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

a. If Y_i satisfies (1.5) with $\alpha = 2$ and $C = 1$. Then,

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

b. When Y_i satisfies (1.5) with $1 < \alpha < 2$ and $C = 1$. Then,

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

c. If (1.5) holds with $\alpha = 1$ and $C = 1$. Then,

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

d. Assume that Y_i satisfies (1.5) with $0 < \alpha < 1$ and $C = 1$. Then,

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

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 (1.5) with $\alpha \leq 2$ and $C = 1$. Then, for every $K \in \mathbb{N}$*

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

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

Proof. Since I_0 is a decreasing function

$$\begin{aligned}
&\int_{\kappa_N}^{\infty} u^{b-1} I_0(u)^{N-a} I_{b_1}(u) \dots I_{b_a}(u)\mathrm{d}u \\
&\quad \leq I_0(\kappa_N)^{N-a} \int_{\kappa_N}^{\infty} u^{b-1} I_{b_1}(u) \dots I_{b_a}(u)\mathrm{d}u \\
&\quad \leq I_0(\kappa_N)^{N-a} \int_{0}^{\infty} u^{b-1} \mathbb{E}\Big[Y_1^{b_1} e^{-uY_1}\Big] \dots \mathbb{E}\Big[Y_a^{b_a} e^{-uY_a}\Big]\mathrm{d}u \\
&\quad = I_0(\kappa_N)^{N-a}\, \Gamma(b) \mathbb{E}\left[\frac{Y_1^{b_1} \dots Y_a^{b_a}}{\left(\sum_{i=1}^{a} Y_i\right)^b}\right]. \qquad (4.17)
\end{aligned}$$

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 $N \to \infty$

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

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 $\mathbb{E}\left[\eta_1^{b_1} \dots \eta_a^{b_a}\right]$.

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

a. Suppose Y_i satisfies (1.5) with $\alpha = 2$ and $C = 1$. Let $g := \max\{i;\, b_i \geq 3\}$, we adopt the convention that $\max\{\emptyset\} = 0$. Then,

$$\lim_{N\to\infty} \mathbb{E}\left[\eta_1^{b_1} \dots \eta_a^{b_a}\right] \cdot \frac{N^{2a}}{(\log N)^{a-g}} = \Gamma(2a) \cdot \frac{2^a \prod_{i=1}^{g} \Gamma(b_i - 2)}{\Gamma(b)\mathbb{E}[Y_1]^{2a}}. \qquad (4.18)$$

b. If (1.5) holds with $1 < \alpha < 2$ and $C = 1$. Then,

$$\lim_{N\to\infty} \mathbb{E}\left[\eta_1^{b_1} \dots \eta_a^{b_a}\right] N^{a\alpha} = \Gamma(a\alpha) \cdot \frac{\prod_{i=1}^{a} \alpha\Gamma(b_i - \alpha)}{\Gamma(b)\mathbb{E}[Y_1]^{a\alpha}}. \qquad (4.19)$$

c. If we assume that Y_i satisfies (1.5) with $\alpha = 1$ and $C = 1$. Then,

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

d. If (1.5) holds with $0 < \alpha < 1$ and $C = 1$. Then,

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

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 hypothesis of Theorem 1.2 hold and let c_N be as in (2.3). Assume also that the Y_i's satisfy (1.5) with $\alpha \le 2$ and $C = 1$. Then,*

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

Finally, if Y_i satisfies (1.5) 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

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

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 $N \to \infty$

$$\begin{aligned} &\frac{\mathbb{E}\Big[(\nu_1)_{b_1}\dots(\nu_a)_{b_a}\Big]}{N^{b-a}c_N} \\ &= \frac{(N)_b}{N^{b-a}c_N}\cdot \mathbb{E}\Big[\eta^{b_1}\dots\eta^{b_a}\Big] && \text{(Lemma 4.1)} \\ &\sim N^a \frac{N}{\log N}\cdot\frac{\mathbb{E}[Y_1]^2}{2}\cdot\mathbb{E}\Big[\eta^{b_1}\dots\eta^{b_a}\Big] && \text{(Corollary 4.8)} \\ &\sim \frac{N^{a+1}}{\log N}\cdot\frac{\mathbb{E}[Y_1]^2}{2}\cdot\frac{(\log N)^{a-g}}{N^{2a}}\cdot\Gamma(2a)\cdot\frac{2^a\prod_{i=1}^g\Gamma(b_i-2)}{\Gamma(b)\mathbb{E}[Y_1]^{2a}} && \text{(Proposition 4.7)} \\ &= \frac{(\log N)^{a-g-1}}{N^{a-1}}\cdot\Gamma(2a)\cdot\frac{2^{a-1}\prod_{i=1}^g\Gamma(b_i-2)}{\Gamma(b)\mathbb{E}[Y_1]^{2(a-1)}}, \end{aligned}$$

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

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

On the other hand, if $a = 1$ and $g = 0$, *i.e.* $b = 2$, then

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

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

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

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

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

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

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*

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

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* $b_a=b$

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

Hence using the recursive formula (2.1) for $\lambda_{b;k}$ we may conclude the convergence to the Bolthausen-Sznitman coalescent.

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

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

where $[x]_{m,y} := x(x+y)\dots(x+(m-1)y)$. 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 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 proof's idea 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $u \to 0^+$

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

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 Appendix B: Proof of Proposition 4.7.

In this appendix we prove Proposition 4.7. Once more, the proof's main idea 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

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

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

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

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 bigger than N_0 and $u \leq \kappa_N$

$$(1-\varepsilon)\big(2\Gamma(b_i-2)\big) \leq I_{b_i}(u)/u^{2-b_i} \leq (1+\varepsilon)\big(2\Gamma(b_i-2)\big), \quad \text{if} \quad b_i \geq 3;$$

$$2(1-\varepsilon) \leq I_2(u)/\log(u^{-1}) \leq 2(1+\varepsilon), \qquad \text{if} \quad b_i = 2.$$

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

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

where we used $b = b_1 + \cdots + b_a = b_1 + \cdots + b_g + 2(a-g)$ (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 $N_1 \geq N_0$)

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

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

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

where $\gamma_N = N(1+\varepsilon)\mathbb{E}[Y_1]\kappa_N$. We have that

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

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

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

Moreover, (B.3) decays uniformly to zero for $v \leq \gamma_N$. *We direct the reader's attention to 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 $N_2 \geq N_1$)

$$(1-\varepsilon)\log N \leq -\log\Big(v/\left(N(1+\varepsilon)\mathbb{E}[Y_1]\right)\Big) \leq (1+\varepsilon)\log N, \quad \text{for every } v \leq \gamma_N.$$

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

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

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

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

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

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

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*:

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

By Lemma 4.5, $I_0(u)^{N-a} \cong (1+u \log u)^{N-a}$. Then we make the changing of variables $v = uN \log N$, obtaining 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 $v \leq \kappa_N N \log N = (\log N)^3$, the equation inside of the parenthesis 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, Prof. Francis Comets, for suggesting this problem, for the helpful discussions and for the patient guidance.

References

[1] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers I.* Springer-Verlag, 1999.

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

[3] E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.

[4] E. Brunet and B. Derrida. Genealogies in simple models of evolution. *J. Stat. Mech. Theory Exp.*, 2013:P01006.

[5] E. Brunet and B. Derrida. Exactly soluble noisy traveling-wave equation appearing in the problem of directed polymers in a random medium. *Phys. Rev. E*, 70(1):016106, 2004.

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

[7] E. Brunet, B. Derrida, A.H. Mueller, and S. Munier. Noisy traveling waves: effect of selection on genealogies. *EPL (Europhysics Letters)*, 76(1):1–7, 2006.

[8] E. Brunet, B. Derrida, A.H. Mueller, and S. Munier. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E*, 76(4):041104, 2007.

[9] F. Comets, J. Quastel, and A.F. Ramírez. Last passage percolation and traveling fronts. *J. Stat. Phys.*, 152(3):419–451, 2013.

[10] J. Cook and B. Derrida. Directed polymers in a random medium: 1/d expansion and the n-tree approximation. *J. Phys. A*, 23(9):1523–1554, 1990.

[11] A. Cortines. Front velocity and directed polymers in random medium. *Stochastic Process. Appl.* to appear.

[12] T. Huillet and M. Möhle. Population genetics models with skewed fertilities: a forward and backward analysis. *Stoch. Models*, 27(3):521–554, 2011.

[13] T. Huillet and M. Möhle. On the extended Moran model and its relation to coalescents with multiple collisions. *Theoretical Population Biology*, 87:5–14, 2013.

[14] J.F.C. Kingman. On the genealogy of large population. *J. Appl. Probab.*, 19:27–43, 1982.

[15] M. Möhle. Weak convergence to the coalescent in neutral population models. *J. Appl. Probab.*, 36(2):446–460, 1999.

[16] M. Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. in Appl. Probab.*, 32(4):983–993, 2000.

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

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

[19] J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, 5:1–50, 2000.

[20] J. Schweinsberg. Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.*, 106(1):107–139, 2003.