How genealogies are affected by the speed of evolution

Éric Brunet[*] and Bernard Derrida[†]

Laboratoire de Physique Statistique, École Normale Supérieure,
UPMC, Université Paris Diderot, CNRS,
24 rue Lhomond, 75231 Paris cedex 05, France

July 13, 2011

Abstract

In a series of recent works it has been shown that a class of simple models of evolving populations under selection leads to genealogical trees whose statistics are given by the Bolthausen-Sznitman coalescent rather than by the well known Kingman coalescent in the case of neutral evolution. Here we show that when conditioning the genealogies on the speed of evolution, one finds a one parameter family of tree statistics which interpolates between the Bolthausen-Sznitman and Kingman's coalescents. This interpolation can be calculated explicitly for one specific version of the model, the exponential model. Numerical simulations of another version of the model and a phenomenological theory indicate that this one-parameter family of tree statistics could be universal. We compare this tree structure with those appearing in other contexts, in particular in the mean field theory of spin glasses.

1 Introduction

An important question in the study of evolving populations is to understand the effect of selection on the ancestry and on the genealogies [3, 2, 4, 5, 1]. In absence of selection, for a well mixed population such as in the Wright-Fisher model, the statistical properties of the genealogy of a large population of constant size is described by Kingman's coalescent [9, 8, 7, 6]. Recent attempts to modify the Wright-Fisher model in order to introduce selection lead to a change of the statistical properties of genealogies: in [10, 11], the study of a whole class of models indicates that the genealogies of populations evolving under selection are given by Bolthausen-Sznitman's coalescent [12] rather than by Kingman's (this has been shown analytically only for one specific version of the model, the exponential model (see below), but it has also been checked in numerical simulations and proved for a modified version of the model where the effect of selection is represented by a moving absorbing wall along the fitness axis [13]). In the present paper, we further study these simple models of evolution with selection and we calculate how the statistical properties of the genealogies are correlated to the speed of evolution.

---

[*]Email: `Eric.Brunet@lps.ens.fr`

[†]Email: `Bernard.Derrida@lps.ens.fr`

The models of evolution with selection we consider here have been introduced in [10, 11] (see also [15, 14]). They can be defined as follows: at each generation t the population consists of a fixed number N of individuals and each individual i is characterized by a single number $x_i(t)$ representing its adaptation in the environment. So $x_i(t)$ is the position of individual i on a fitness or adaptation axis (very much like in the Bak-Sneppen model [16]). This individual has several offspring at positions $x_i(t) + \epsilon_{i,1}(t)$, $x_i(t) + \epsilon_{i,2}(t)$, $x_i(t) + \epsilon_{i,3}(t)$, etc., where the $\epsilon_{i,j}(t)$ are random numbers representing the change of adaptation due to mutations between parent i and child j. The total number of offspring produced this way by all individuals at a given generation t exceeds N; the population at generation $t + 1$ is then obtained by keeping the N most adapted children (i.e. the N rightmost points along the axis) among all these offspring, see figure 1. The model is fully specified when the distribution of the number offspring of each individual and the distribution of the random shifts ϵ are given.
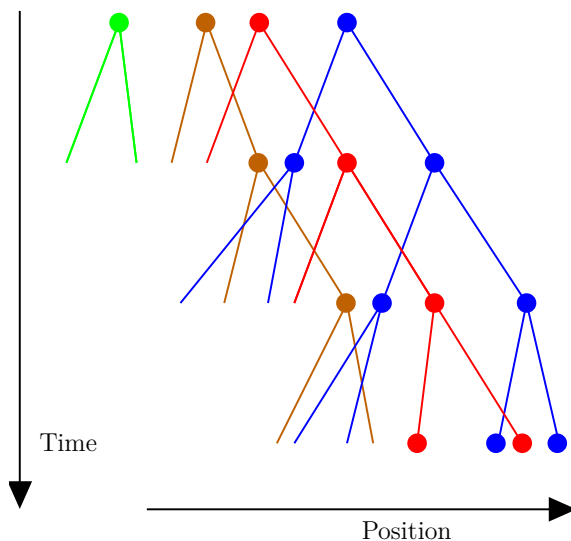

Figure 1: Three time steps for a population of size $N = 4$ in a model where each individual has two offspring.

After letting such a model evolve for a large number t of generations, the positions $x_i(t)$ of the individuals on the adaptation axis from a cloud of points grouped around a position X_t which grows linearly with time with some velocity v_N. There is some arbitrariness in the way this position X_t can be defined (one could choose for example X_t to be the position of the rightmost individual, or of the leftmost individual or the center of mass of the population) but, as the positions of all the individuals remain grouped, a change of definition modifies the value of X_t by an amount which does not grow with time and therefore does not affect the velocity v_N. In addition to the velocity, the position X_t has fluctuations: in particular it diffuses with a variance which grows linearly in time [17].

At each generation, one can study the genealogy of the population of this model by considering the matrix $\tau_{i,j}(t)$ of the ages of the most recent common ancestors, or *coalescence times*, of all pairs of individuals i and j living at generation t; see figure 2. As the genealogy is a tree, the whole ancestry of the population can be deduced from the knowledge of this matrix. In particular the age
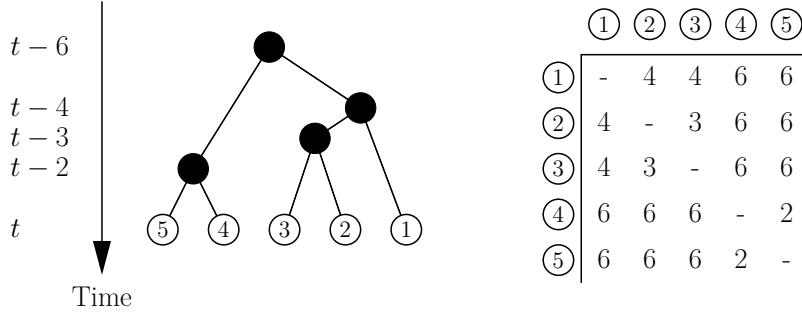
2

Figure 2: *Left:* an exemple of the genealogical tree of $N = 5$ individuals. *Right:* the coalescence times $\tau_{i,j}(t)$ corresponding to the tree.

of the most recent common ancestor of any subset of the population can be expressed in terms of this matrix: for example the age $\tau_{i,j,k}(t)$ of the most recent common ancestor of three individuals i, j and k at generation t is simply

$$\tau_{i,j,k}(t) = \max\left(\tau_{i,j}(t), \tau_{i,k}(t), \tau_{j,k}(t)\right). \tag{1}$$

One property common to the models of evolution under selection studied here and to the neutral models of evolution described by Kingman's coalescent is that the heights and the shapes of the genealogical trees fluctuate with t even when the size N of the population becomes large. The statistical properties of these trees and their time scales are however different: for instance, the typical age of the most recent common ancestors of k individuals grows logarithmically with the population size N in presence of selection (as in the models studied here) while it grows linearly with N in the neutral case.

There are several ways of describing the statistical properties of these trees (see section 5). In [10, 11] we chose to characterize them by the average coalescence times $\langle T_k \rangle$ of k individuals chosen at random in the population:

$$\langle T_k \rangle = \left\langle \tau_{i_1,\ldots,i_k}(t) \right\rangle. \tag{2}$$

(here $\langle \cdot \rangle$ denotes an average over the individuals $i_1, \ldots, i_k$ and over the generation t). The N dependence of $\langle T_2 \rangle$ gives the time scale over which coalescence events occur, while the ratios $\langle T_k \rangle / \langle T_2 \rangle$ are a signature of the statistical properties of the shape of the trees. We found in [10, 11] that these ratios in presence of selection converge when $N \to \infty$ to those of a Bolthausen-Snitzmann coalescent:

$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{5}{4}, \qquad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{25}{18}, \qquad \text{(Bolthausen-Snitzmann)} \tag{3}$$

in contrast to the neutral case where they converge to those of Kingman's coalescent:

$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4}{3}, \qquad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{3}{2}. \qquad \text{(Kingman)} \tag{4}$$

Our goal here is to calculate how these ratios are correlated to the speed of evolution, by weighting all the events during a long time interval τ by a factor $e^{-\beta X_\tau}$. ($\beta < 0$ favors events with

a speed of evolution faster than average, while $\beta > 0$ correponds to events with a slower speed of evolution.) Our main result, derived below for the exponential model, is that the above ratios become for large N

$$\frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{5 + 4\beta}{4 + 3\beta}, \qquad \frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{100 + 204\beta + 133\beta^2 + 27\beta^3}{72 + 142\beta + 90\beta^2 + 18\beta^3} \tag{5}$$

It is remarkable that these expressions interpolate between the neutral case (Kingman) for $\beta \to +\infty$ (low speed limit) and the selection case (Bolthausen-Snitzmann) for $\beta = 0$. When $\beta \to -1$ (high speed limit), all the ratios become 1 indicating a "star-shaped" coalescent.

This paper is organized as follows: in section 2, we explain how the weighting by the factor $e^{-\beta X_\tau}$ is done. In section 3, we show that in presence of the bias, one version of the model (the exponential model) can be solved exactly by analyzing a coalescent model, the rates of which depend on β. This leads to (5). In section 4 we argue using the phenomenological theory developed in [11] that (5) should remain valid for other versions of the model up to a change of scale of β. Lastly in section 5 we compare the β-random tree structure which leads to (5) to the statistics of the partitions in mean field spin glasses and in the Poisson-Dirichlet distribution.

2 How to condition on the velocity

If one performs a simulation of the model described in the introduction, one can measure at each generation t the position X_t of the population (defined in any reasonable way: as explained in the introduction, the precise definition does not matter) and the ages $T_2(t), \ldots, T_k(t)$ of the most recent common ancestor of $2, \ldots, k$ individuals chosen at random in the population at time t (for more efficiency one can average these times $T_k(t)$ over all the choices of the k individuals in the population at time t). Then we choose a long time interval τ and we want to determine

$$\langle T_k \rangle_\beta = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \frac{\langle T_k(t) e^{-\beta X_\tau} \rangle}{\langle e^{-\beta X_\tau} \rangle}. \tag{6}$$

2.1 Theoretical considerations

As the model is a Markov process and correlations decay fast enough in time, we expect that, for large t,

$$\left\langle e^{-\beta(X_t - X_0)} \right\rangle \sim e^{tG(\beta)}. \tag{7}$$

The knowledge of $G(\beta)$ determines all the cumulants of the position X_t

$$\lim_{t \to \infty} \frac{\langle X_t^n \rangle_c}{t} = (-)^n \left. \frac{\mathrm{d}^n G(\beta)}{\mathrm{d}\beta^n} \right|_{\beta=0}. \tag{8}$$

It is also related to the large deviation function $F(v)$ of the velocity defined by

$$\mathrm{Proba}(X_t = vt) \sim e^{tF(v)} \tag{9}$$

through a Legendre transform

$$G(\beta) = \max_v [-\beta v + F(v)]. \tag{10}$$

4

For large t, the value of v which dominates the weighted averages in (6) and (7) is given by

$$v = -\frac{\mathrm{d}G(\beta)}{\mathrm{d}\beta},\tag{11}$$

with fluctuations of order $t^{-1/2}$. Therefore the weighted averages (6) become equivalent in the $t \to \infty$ limit to conditioning on the velocity v given by (11). This is, in the present context, the analog of the well known equivalence of ensembles in statistical physics.

2.2 In numerical simulations

Numerically it is difficult to perform averages such as (6) because the events which dominate both the numerator and the denominator of (6) are rare events. In order to overcome this difficulty, we use an importance sampling method. We consider a sample periodic in time of period τ where τ is chosen large enough. This means that the random shifts $\epsilon_{i,j}(t)$ are periodic in time ($\epsilon_{i,j}(t+\tau) = \epsilon_{i,j}(t)$ for all i, j and t). With these periodic $\epsilon_{i,j}(t)$, the evolution of the system becomes also periodic in time: the shift X_τ of the position of the population after one period τ can therefore be unambigously defined and depends on all the $\epsilon_{i,j}(t)$. Then, we perform a standard Monte-Carlo simulation: at each step we try a new sample by changing some of the $\epsilon_{i,j}(t)$ and we let the system evolve till it becomes periodic (here we change all the $\epsilon_{i,j}(t)$ at a random time t uniformly distributed between 1 and τ). The outcome of this change is to modify the X_τ to a new value X_τ^{new}. Then, as always with a Metropolis algorithm, we accept the change with a probability $\max\left[1, \exp[-\beta(X_\tau^{\mathrm{new}} - X_\tau)]\right]$. With this procedure samples are produced with a weight $\exp[-\beta X_\tau]$, so that by averaging quantities such as the T_k over many samples one gets an estimate of (6).

We have simulated the exponential model (see section 3 where we give the precise definition of the exponential model and its analytic solution in the $N \to \infty$ limit) for $N = 100$ and a value of $\tau \approx 30 \ln N$ which is much larger than $\langle T_2 \rangle$ [11]. For each value of β, we measured $\langle T_2 \rangle$, $\langle T_3 \rangle$, $\langle T_4 \rangle$ averaged on 10^6 Monte-Carlo steps. We have also simulated a more generic model where each individual has exactly two offspring with independent random shifts $\epsilon_{i,j}(t)$ uniformly distributed between 0 and 1. This model cannot be solved exactly, but a phenomenological theory (see [11] and section 4) predicts when $N \to \infty$ the same statistics of the genealogical trees (5) as in the exponential model with β replaced by β/γ, where $\gamma \simeq 5.262$ is the value which minimizes the function $\ln[2(e^\gamma - 1)/\gamma]/\gamma$, see section 4. We simulated the sizes $N = 30$, $N = 100$ and $N = 300$ with values of $\tau \approx 8 \ln^3 N$ which is much larger than $\langle T_2 \rangle$, and again averaged over 10^6 Monte-Carlo steps. We also checked for several values of β that our results remain unchanged by choosing a value of the period τ twice as big (results not shown), indicating that our Monte-Carlo results would be the same for an infinite time-window.

The results for $\langle T_3 \rangle / \langle T_2 \rangle$ and $\langle T_4 \rangle / \langle T_2 \rangle$ are presented in figures 3 and 4. As in [11], we observe that for the exponential model, the results are already very close to their asymptotic limits even for $N = 100$; only for $\beta \lesssim -1$ there is a discrepancy with the theoretical prediction (5). As in [11], the convergence is however slower in the generic case but the curves seem in both cases to converge to the prediction.

3 Exponential model

In this section we consider a version of the model, the exponential model, which can be solved exactly [11]. In the exponential model, the shifts of the offspring of each individual are generated
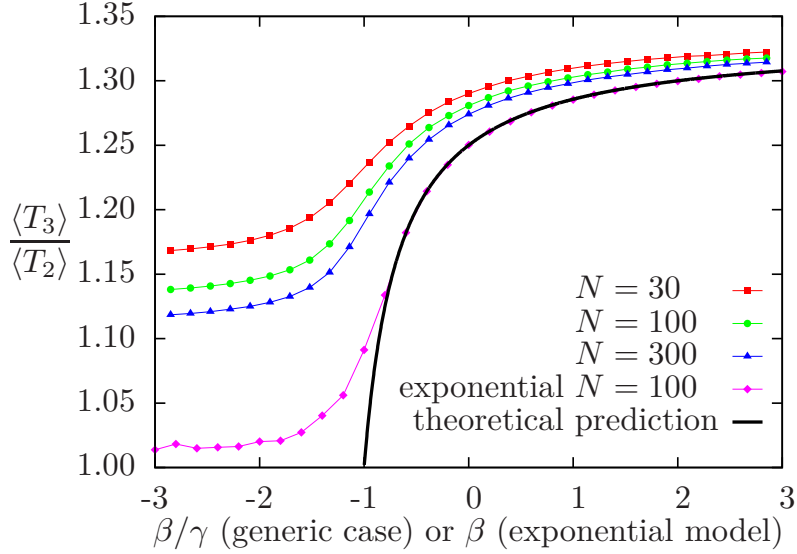
Figure 3: $\langle T_3 \rangle / \langle T_2 \rangle$ as a function of β for the exponential model and the prediction (5), and as a function of β/γ for the generic model described in the text.

by a Poisson process of density $\rho(\epsilon) = e^{-\epsilon}$. This means that an individual at position $x_i(t)$ has a probability $e^{-\epsilon} d\epsilon$ of having an offspring in the infinitesimal interval $(x_i(t) + \epsilon, x_i(t) + \epsilon + d\epsilon)$. Then the population at the next generation is obtained by selecting the N rightmost points among all the offspring produced by generation t. (Note that in the exponential model the number of offspring produced by each generation is infinite but their number at the right of any position y is finite. There is therefore no problem to select the N survivors at generation $t + 1$).

In the exponential model, there is a convenient way of defining the position X_t of the population

$$X_t = \ln \left[\sum_{i=1}^{N} e^{x_i(t)} \right]. \tag{12}$$

The simplicity of the exponential model comes from the fact that, with this definition of X_t, one has

$$\sum_i e^{-[x - x_i(t)]} = e^{-(x - X_t)}, \tag{13}$$

which means that one can generate the offspring of the whole population at time t by replacing the N Poisson processes centered at the positions $x_i(t)$ by a single Poisson process centered at position X_t. Therefore, with definition (12) of X_t, the N points $x_i(t + 1)$ at generation $t + 1$ are the N rightmost points of a Poisson point process with a density $\exp[-(x - X_t)]$.

As explained in [11] a way of drawing these N points is to choose a number z with a density of probability $\text{Proba}(z) = \exp[-(N + 1)z - e^{-z}]/N!$ and, independently, N numbers y_i with an exponential density $\text{Proba}(y) = e^{-y}\theta(y)$; the points $x_i(t + 1)$ are then given (in an arbitrary order) by

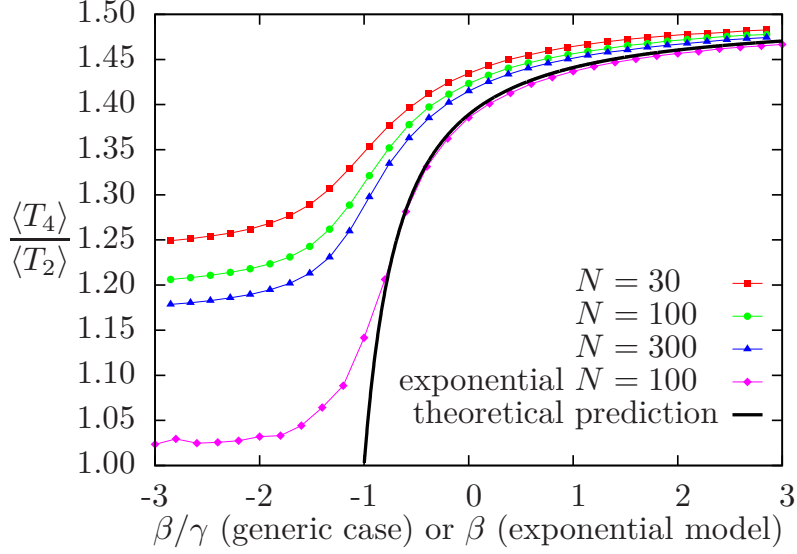$$x_i(t + 1) = X_t + z + y_i, \tag{14}$$

Figure 4: $\langle T_4 \rangle / \langle T_2 \rangle$ as a function of β for the exponential model and the prediction (5), and as a function of β/γ for the generic model described in the text.

and one gets from (12)

$$X_{t+1} = X_t + z + \ln\left[\sum_{i=1}^{N} e^{y_i}\right]. \tag{15}$$

We see that, with the definition (12) of X_t, the differences $X_{t+1} - X_t$ are independent variables. Therefore from (7)

$$e^{tG(\beta)} = \left\langle e^{-\beta(X_t - X_0)} \right\rangle = \left\langle e^{-\beta(X_{t+1} - X_t)} \right\rangle^t, \tag{16}$$

and $G(\beta)$ can be computed by averaging over a single generation

$$
\begin{aligned}
e^{G(\beta)} &= \langle e^{-\beta z} \rangle \left\langle \left[\sum_{i=1}^{N} e^{y_i}\right]^{-\beta} \right\rangle \\
&= \frac{\Gamma(N+1+\beta)}{\Gamma(N+1)} \int_0^\infty dy_1 \cdots \int_0^\infty dy_n \ e^{-y_1 - \cdots - y_N} \left[\sum_{i=1}^{N} e^{y_i}\right]^{-\beta}.
\end{aligned}
\tag{17}
$$

Using the integral representation (valid for $\beta > 0$)

$$A^{-\beta} = \frac{1}{\Gamma(\beta)} \int_0^\infty d\lambda\, \lambda^{\beta-1} e^{-\lambda A} \qquad \text{(for } \beta > 0\text{)}, \tag{18}$$

one obtains

$$e^{G(\beta)} = \frac{\Gamma(N+1+\beta)}{\Gamma(N+1)\Gamma(\beta)} \int_0^\infty d\lambda\, \lambda^{\beta-1} I_0(\lambda)^N, \tag{19}$$

7

where the integral $I_0(\lambda)$ and more general integrals $I_p(\lambda)$ are defined as

$$I_p(\lambda) = \int_0^\infty \mathrm{d}y\, e^{(p-1)y - \lambda e^y} = \lambda^{1-p} \int_\lambda^\infty \mathrm{d}u\, u^{p-2} e^{-u}. \tag{20}$$

(These integrals are in fact, up to a simple change of variables, incomplete gamma functions). For large N, the expression (19) is dominated by small values of λ where the integrals $I_p(\lambda)$ for non-negative integers p can be approximated by [11]:

$$I_0(\lambda) = 1 + \lambda(\ln\lambda + \gamma_E - 1) + \mathcal{O}(\lambda^2), \qquad I_1(\lambda) = -(\ln\lambda + \gamma_E) + \mathcal{O}(\lambda),$$
$$I_{p\geq 2}(\lambda) = \frac{(p-2)!}{\lambda^{p-1}} + \mathcal{O}(\lambda^{2-p}), \tag{21}$$

where $\gamma_E = -\Gamma'(1) \simeq .577$ is Euler's constant.

Given the small λ expansion of $I_0(\lambda)$, the integral in (19) is dominated, for N large and β of order 1, by λ of order $1/(N \ln N)$. Making the change of variable $\lambda = \mu/(N \ln N)$ one has

$$I_0(\lambda)^N = e^{-\mu} \left(1 + \mu \frac{\ln\mu - \ln\ln N + \gamma_E - 1}{\ln N} + \mathcal{O}\left(\frac{\mu \ln\mu}{\ln N} \right)^2 \right), \tag{22}$$

We can now evaluate (19) for N large and β of order 1; using (22) and $\Gamma(N+1+\beta)/\Gamma(N+1) \simeq N^\beta$, one gets [11]

$$e^{G(\beta)} = \frac{1}{\ln^\beta N} \left[1 + \frac{\beta}{\ln N} \left(\frac{\Gamma'(\beta+1)}{\Gamma(\beta+1)} - \ln\ln N + \gamma_E - 1 \right) + \cdots \right] \tag{23}$$

or

$$G(\beta) = -\beta \ln\ln N + \frac{\beta}{\ln N} \left(\frac{\Gamma'(\beta+1)}{\Gamma(\beta+1)} - \ln\ln N + \gamma_E - 1 \right) + \cdots \tag{24}$$

We see that, as $v = -G'(\beta)$, see (11), varying β does not change the leading N dependence $v \simeq \ln\ln N$ of the velocity but only shifts it by a small amount of order $\ln\ln N/\ln N$ which vanishes in the $N \to \infty$ limit. We are now going to show that, on the contrary, β does change the statistical properties of the trees even in the $N \to \infty$ limit.

3.1 Trees

We have already seen that all the offspring produced by generation t are distributed according to a Poisson point process of density $\exp[-(x - X_t)]$. On the other hand, the offspring of individual $x_i(t)$ are distributed as a Poisson point process of density $\exp\left[-[x - x_i(t)] \right]$. This implies that, given that there is an offspring in an interval $\mathrm{d}x$ around x, its probability of being an offspring of $x_i(t)$ is

$$W_i = e^{x_i(t) - X_t} = \frac{e^{x_i(t)}}{\sum_{j=1}^N e^{x_j(t)}}. \tag{25}$$

This probability is independent of x. Therefore the probability $Q_p(t)$ that p individuals at generation $t+1$ have the same ancestor at generation t is

$$Q_p(t) = \sum_{i=1}^N W_i^p = \sum_{i=1}^N e^{px_i(t) - pX_t}. \tag{26}$$

8

If one weights these coalescence rates with the factor $e^{-\beta X_t}$, then using (14) and (15) with t replaced by $t-1$ and using the fact that X_{t-1}, z and the y_i are independent, one gets

$$\langle Q_p \rangle_\beta = N \frac{\left\langle e^{px_1(t)-(\beta+p)X_t} \right\rangle}{\left\langle e^{-\beta X_t} \right\rangle} = N \frac{\left\langle e^{py_1-(\beta+p)\ln\left[\sum_{i=1}^N e^{y_i}\right]} \right\rangle}{\left\langle e^{-\beta \ln\left[\sum_{i=1}^N e^{y_i}\right]} \right\rangle} \tag{27}$$

with the y_i independent exponential variables. The numerator and the denominator can be computed in the same way as in (19) and one obtains

$$\langle Q_p \rangle_\beta = N \frac{\Gamma(\beta)}{\Gamma(\beta+p)} \times \frac{\int_0^\infty d\lambda \, \lambda^{\beta+p-1} I_p(\lambda) I_0(\lambda)^{N-1}}{\int_0^\infty d\lambda \, \lambda^{\beta-1} I_0(\lambda)^N}. \tag{28}$$

We take $p \geq 2$ and β of order 1. The integrals are dominated by $\lambda = \mu/(N \ln N)$ and μ of order 1. To the leading order, $I_0(\lambda)^N \approx e^{-\mu}$, see (22), and using (21) for $I_p(\lambda)$ one easily gets, to leading order

$$\langle Q_p \rangle_\beta \simeq \frac{1}{\ln N} \frac{(p-2)! \, \Gamma(\beta+1)}{\Gamma(\beta+p)} = \frac{1}{\ln N} \frac{(p-2)!}{(1+\beta)(2+\beta)\cdots(p-1+\beta)}. \tag{29}$$

After rescaling time by a factor $\ln N$, one gets a coalescent with transition rates $q_p = \langle Q_p \rangle_\beta \ln N$. One can check that

$$q_p = \frac{(p-2)! \, \Gamma(\beta+1)}{\Gamma(\beta+p)} = \int_0^1 x^{p-2} \Lambda(dx) \qquad \text{with } \Lambda(dx) = (1-x)^\beta \, dx, \tag{30}$$

Using the expressions (55) of the appendix, where the ratios $\langle T_3 \rangle/\langle T_2 \rangle$ and $\langle T_4 \rangle/\langle T_2 \rangle$ have been obtained for a general coalescent, one finally gets (5).

4   The phenomenological theory

In this section, we show that the phenomenological theory developed in [17, 11] in the context of the noisy Fisher-KPP equation predicts that (5) remains valid for other versions of the model described in the introduction. When the number of offspring of each individual is bounded and when the shifts $\epsilon_{i,j}$ are also bounded, one can describe the evolution of the population by a noisy traveling wave equation of the Fisher-KPP type. In [17, 11], a phenomenological theory was proposed to describe the large N behavior of these noisy equations. In the $N \to \infty$ limit, the effect of noise vanishes and the traveling wave has a finite velocity v_∞ (in contrast to the exponential model where the velocity diverges as $N \to \infty$). The first correction when N is large can be understood by considering the cutoff introduced by the discrete number of particles: this leads to $v_{\text{cutoff}} = v_\infty - A/\ln^2 N$. The next order correction leads to a positive term of order $\ln \ln N/\ln^3 N$ which can be understood, as well as the fluctuations of X_t, by the following phenomenological theory: the front has, most of the time, the shape and the velocity predicted by the cutoff theory. However, every (typically) $\ln^3 N$ time steps, a rare event occurs where some particles escape significantly ahead of the front. When this happens, the shape of the front is at first deformed, but it relaxes to its cutoff shape after $\sim \ln^2 N$ time steps. The end result is a finite increase of the position of the front [11]. It has been shown that this phenomenological theory predicts genealogies described by the Bolthausen-Snitzmann coalescent. We are now going to show that when we condition on the velocity by using the weight $\exp(-\beta X_t)$, this leads to (5).

We consider a time interval Δt which is large compared to $\ln^2 N$ but small compared to $\ln^3 N$. During this interval, there is a small probability $p(\delta)\,\mathrm{d}\delta\,\Delta t$ that an event of size δ occurs. When this happens, the front position increases (after relaxation) by $R(\delta)$. The time interval $\ln^2 N \ll \Delta t \ll \ln^3 N$ is such that each event has the time to relax during Δt and that the probability that two events occur during the same time interval is negligible.

With these notations, the position X_t of the front evolves according to

$$X_{t+\Delta t} - X_t = \begin{cases} v_{\text{cutoff}}\Delta t + R(\delta) & \text{proba. } p(\delta)\,\mathrm{d}\delta\,\Delta t, \\ v_{\text{cutoff}}\Delta t + 0 & \text{proba. } 1 - \Delta t \int \mathrm{d}\delta\, p(\delta). \end{cases} \tag{31}$$

We argued in [11] that, for large δ,

$$p(\delta) \approx C_1 e^{-\gamma\delta}, \quad R(\delta) \approx \frac{1}{\gamma}\ln\left(1 + C_2\frac{\gamma^3 e^{\gamma\delta}}{\ln^3 N}\right), \quad C_1 C_2 \approx \pi^2 \gamma v''(\gamma). \tag{32}$$

The number γ and the function $v(r)$ depend on the details of the model. (For $N \to \infty$, $v(r)$ gives the velocity of a front starting with the initial condition e^{-rx}. For a step initial condition, the system moves at the velocity $v(\gamma)$ where γ is the value at which $v(r)$ reaches its minimum.)

Using (31) to compute $G(\beta)$ given by (7), one gets

$$G(\beta) = -\beta v_{\text{cutoff}} + \int \mathrm{d}\delta\, p(\delta)\left[e^{-\beta R(\delta)} - 1\right]. \tag{33}$$

Let us now weight all the events by the factor $\exp(-\beta X_t)$. (31) becomes

$$X_{t+\Delta t} - X_t = \begin{cases} v_{\text{cutoff}}\Delta t + R(\delta) & \text{proba. } \frac{1}{Z(\beta)}e^{-\beta[v_{\text{cutoff}}\Delta t + R(\delta)]}p(\delta)\,\mathrm{d}\delta\,\Delta t, \\ v_{\text{cutoff}}\Delta t + 0 & \text{proba. } \frac{1}{Z(\beta)}e^{-\beta v_{\text{cutoff}}\Delta t}\left[1 - \Delta t \int \mathrm{d}\delta\, p(\delta)\right]. \end{cases} \tag{34}$$

$Z(\beta)$ is such that the probabilities are normalized; clearly $Z(\beta) = e^{\Delta t G(\beta)}$.

Now, we can try to determine the probability $Q_p \Delta t$ that the p particles coalesce into one during the time interval Δt. We argued in [11] that when a rare event of size δ occurs, a fraction $f = 1 - e^{-\gamma R(\delta)}$ of the population is replaced by the offspring of the single particle that originated the event (γ is the model specific number appearing in (32)). When this happens, there is a probability f^p that the p particles coalesce during that interval of time Δt. This leads to

$$Q_p = \int \mathrm{d}\delta\, p(\delta)e^{-\beta R(\delta)}\left[1 - e^{-\gamma R(\delta)}\right]^p \tag{35}$$

We can now use (32) in (35). Rewriting the integral in term of the variable $f = 1 - e^{-\gamma R(\delta)}$ (the fraction of the population replaced by the offspring of an individual), one gets $p(\delta)\,\mathrm{d}\delta = C_1 C_2 \gamma^2 / \ln^3 N \,\mathrm{d}f/f^2$, so that

$$Q_p = \frac{C_1 C_2 \gamma^2}{\ln^3 N} \int_0^1 f^{p-2}(1-f)^{\beta/\gamma}\,\mathrm{d}f, \tag{36}$$

which is the same as (30) up to a prefactor (which only changes the time scale) and the fact that β is replace by β/γ. Therefore the phenomenological theory leads to the same coalescent as in the exponential model but with a different time scale of order $\ln^3 N$ instead of $\ln N$.

10

5 Comparison with mean-field spin glasses and the Poisson-Dirichlet distribution

5.1 Various ways of characterizing random trees

There are several ways of characterizing the statistical properties of the trees generated by some given coalescence rates. (We only consider here the cases where the coalescence rates do not vary in time, where the particles play symmetrical roles and where at most one coalescence event can occur during an interval of time dt.)

- One can specify the coalescence rates q_p (which take the values (30) for the models of evolution with selection that we consider in this paper). In terms of these coalescence rates, Kingman's coalescent corresponds to

$$q_2 \neq 0, \qquad q_p = 0 \quad \text{for } p \geq 3, \tag{37}$$

  while the Bolthausen-Sznitman coalescent corresponds to

$$q_p = \frac{q_2}{p-1}. \tag{38}$$

  It is easy to see that the rates (30) interpolate between (37) for $\beta = \infty$ and (38) for $\beta = 0$.

- One can alternatively specify all the ratios $\langle T_p \rangle / \langle T_2 \rangle$. It is clear (see (52,55) in the appendix) that the knowledge of the q_p determines all these time ratios and conversely that the knowledge of the time ratios allows one to calculate all the ratios q_p/q_2.

- One can also characterize the trees by the partition of the population they induce at a given time in the past: the population at a generation t can be decomposed into several τ-families where, by definition of these families, two individuals i and j belong to the same τ-family if the age of their most recent common ancestor is less than τ (i.e. $\tau_{i,j}(t) < \tau$). One can then associate to this τ-partition of the population at generation t the following numbers

$$Y_k^{(\tau)}(t) = \left\langle \theta(\tau - \tau_{i_1,\ldots,i_k}) \right\rangle_t \tag{39}$$

  where $\langle \cdot \rangle_t$ means an average over all the possible choices of the k individuals $i_1, \ldots, i_k$ at generation t.

  One can interpret these $Y_k^{(\tau)}(t)$ as the probability that k individuals chosen at random in the population at generation t belong to the same τ-family. These $Y_k^{(\tau)}(t)$ fluctuate from generation to generation and the expressions $\langle Y_k^{(\tau)} \rangle$ of their averages over t can be computed in terms of the coalescence rates q_p. In the appendix, they are given for $k = 2, 3, 4$ by the quantities $Z_{k \to 1}(\tau) \equiv \langle Y_k^{(\tau)} \rangle$.

  Knowing all the $\langle Y_k^{(\tau)} \rangle$ (even for a single τ) determines also in principle all the coalescence rates and therefore all the statistical properties of the trees.

5.2 Comparison with mean-field spin glasses

Very much like in the coalescence problems discussed above, where all the individuals at a generation t can be grouped into τ-families, one can group the spin configurations of a spin glass model

according to their distances d (or to their overlap= $1 - d$) in phase space. One can then define [18], for a given sample, the probability $Y_k(d)$ that k configurations, at thermal equilibrium, have all their $k(k-1)/2$ mutual distances in phase space less than d.

One of the predictions [18, 20, 19] of the Parisi solution [23, 21, 22] of the Sherrington Kirkpatrick model [25, 24] is that this $Y_k(d)$ fluctuates with the spin glass sample even when the system size becomes large. The Parisi theory predicts also all the statistical properties of these $Y_k(d)$. For example

$$\langle Y_k(d) \rangle = \lim_{n \to 0} y_k(n, \mu) = \frac{\Gamma(k - \mu)}{\Gamma(k) \, \Gamma(1 - \mu)} \tag{40}$$

where according to the broken replica symmetry

$$y_k(n, \mu) = \frac{\Gamma(1 - n) \, \Gamma(k - \mu)}{\Gamma(k - n) \, \Gamma(1 - \mu)}. \tag{41}$$

In (40) all the dependence on the distance d, on the details of the model, and on the parameters such as the temperature or the magnetic field is through the parameter μ. Formula (41) follows from a very simple replica calculation: assume that one has n replicas grouped into n/μ families of μ replicas, $y_k(n, \mu)$ is simply the probability that k replicas chosen at random among the n replicas belong to the same family.

For typical samples one has to take the $n \to 0$ limit as in (40) and the statistics of the Y_k coincide with those of the Bolthausen-Sznitman coalescent: one can check that the $Z_{k \to 1}(\tau) \equiv \langle Y_k^{(\tau)} \rangle$ obtained in (57) coincides with (40) by choosing $\mu = e^{-q_2 \tau}$ and the q_p given by (38).

In the spin glass case, one can also weight the samples according to their free energy (by weighting them by a factor Z^n where Z is the partition function [27, 26]). One then expects from the replica theory [26] that the statistics of the $Y_k(d)$ be modified and that

$$\frac{\langle Z^n \, Y_k(d) \rangle}{\langle Z^n \rangle} = y_k(n, \mu). \tag{42}$$

One can check easily from (57) with q_p given by (30) that there exists no choice of n and μ as functions of β and τ such that $y_k(n, \mu) = Z_{k \to 1}(\tau)$. Therefore, although the statistical properties of the trees in the model of evolution with selection and in the spin glass problem are the same for typical samples, they become different when one introduces a bias (related to the free energy in the spin glass problem and to the speed of adaptation in the models of evolution with selection).

5.3 The Poisson-Dirichlet distribution

The Poisson-Dirichlet distribution [29, 28] is a probability distribution of the partitions of a unit interval into infinitely many subintervals. It is parametrized by two parameters α and θ. One way of defining the Poisson-Dirichlet distribution is to consider an infinite sequence $z_1, z_2, \ldots, z_n, \ldots$ of independent numbers, each z_n being distributed according to a distribution

$$P_n(z_n) = \frac{\Gamma(1 + \theta + n\alpha - \alpha)}{\Gamma(1 - \alpha)\Gamma(\theta + n\alpha)} z_n^{-\alpha}(1 - z_n)^{\theta + n\alpha - 1} \tag{43}$$

(which is a β distribution). Then one considers a partition of the unit interval into subintervals of lengths $W_1, W_2, \ldots W_n, \ldots$ with

$$W_1 + W_2 + \cdots + W_n + \cdots = 1, \tag{44}$$

where the W_i are given by

$$
\begin{aligned}
W_1 &= z_1, \\
W_2 &= (1 - z_1)z_2, \\
&\ldots \\
W_n &= (1 - z_1)(1 - z_2)\cdots(1 - z_{n-1})z_n, \\
&\ldots
\end{aligned}
\tag{45}
$$

For such a partition one can introduce the quantities

$$
Y_k = \sum_i W_i^k,
\tag{46}
$$

which represents the probability that k points chosen at random on the unit interval fall in the same subinterval. It is easy to check that when one averages over the z_i, one gets

$$
\langle Y_k \rangle_{\alpha,\theta} = \langle z_1^k \rangle + \langle (1 - z_1)^k \rangle \langle Y_k \rangle_{\alpha,\alpha+\theta},
\tag{47}
$$

The solution of this recursion is

$$
\langle Y_k \rangle_{\alpha,\theta} = \frac{\Gamma(1 + \theta)\Gamma(k - \alpha)}{\Gamma(k + \theta)\Gamma(1 - \alpha)}.
\tag{48}
$$

One can notice [30] that these expressions are identical to the replica expressions (41) of $y_k(n,\mu)$ when one chooses $\theta = -n$ and $\alpha = \mu$. Therefore as soon as one introduces the bias $\beta \neq 0$ the statistical properties τ-families of our models of evolution with selection differ from those of the Poisson-Dirichlet distribution.

6   Conclusion

In this paper we have seen that, for a family of simple models of evolution under selection, the statistics of the genalogies are modified when conditioning on the speed of evolution (5). For one particular version of these models, the exponential model, the trees can be generated by a coalescent with modified rates (30). Numerical simulations (figures 3 and 4) and a phenomenological theory (section 4) indicate a similar behavior of more generic versions of the model.

Despite their simplicity, there is not yet a full theoretical understanding of the models of evolution with selection we consider here. The introduction of the bias opens new questions which would be interesting to consider. For exemple, what is the effect of the bias on the steady state density profile of the population along the fitness axis, or on the distances between the rightmost points in the population? Numerically, the Monte-Carlo approach we developed here should give a rather powerful tool to study these questions accurately and to test more precisely the phenomenological theory developped in [17, 11].

It would also be interesting to study the genealogies of other models of evolution with selection[31] to test the genericity of our results.

We are very happy to dedicate this work to David Sherrington, on the occasion of his 70th birthday.

A   Coalescence times and sizes of families in the Λ coalescent

In this appendix we calculate a few simple properties of a continuous time coalescent defined as follows: one starts with N points and during every infinitesimal time interval $dt \ll 1$, every subset of k points has a probability $q_k\, dt$ of coalescing into one point. It is assumed that there is at most one coalescence event during a time interval dt.

This model is called the Λ-coalescent [33, 32, 34, 13, 28]. The coalescence rates can be written [33] in terms of a positive measure Λ on the interval $(0,1)$

$$q_k = \int_0^1 x^{k-2} \Lambda(dx), \tag{49}$$

and, more generally [33], the rate $\lambda_{b,k}$ at which the $k \geq 2$ first points out of b coalesce into one point (while the other $b - k$ points remain single) is given by

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k}\, \Lambda(dx) = \sum_{n=0}^{b-k} \frac{(b-k)!}{n!(b-k-n)!}(-1)^n q_{n+k}. \tag{50}$$

It is more convenient in the following to use of the quantities $r_b(b')$, defined as the rate at which a set of b points coalesce into a set of $b' < b$ points. Clearly

$$r_b(b') = \frac{b!}{(b'-1)!(b-b'+1)!}\lambda_{b,b-b'+1} \tag{51}$$

(This simply means that the total number of distinct points jumps from b to b' with probability $r_b(b')dt$ during an infinitesimal time interval dt. The binomial factor in (51) comes from the number of ways of choosing the $b - b' + 1$ points which coalesce.)

As already noticed, the rates (30) correspond to $\Lambda(dx) = (1-x)^\beta dx$.

By analyzing what happens during a time interval dt one can then see that the age T_b of the most recent common ancestor of b individuals chosen at generation $t + dt$ satisfies

$$T_b(t+dt) = \begin{cases} dt & \text{with probability} & r_b(1)\,dt, \\ dt + T_2(t) & & r_b(2)\,dt, \\ \ldots & & \\ dt + T_{b-1}(t) & & r_b(b-1)\,dt, \\ dt + T_b(t) & & 1 - \sum_{b'=1}^{b-1} r_b(b')\,dt. \end{cases} \tag{52}$$

Therefore one can determine recursively the average coalescence times by writing that $\langle T_b(t+dt)\rangle = \langle T_b(t)\rangle$, which leads to

$$\left[\sum_{b'=1}^{b-1} r_b(b')\right]\langle T_b\rangle = 1 + \sum_{b'=2}^{b-1} r_b(b')\langle T_{b'}\rangle \tag{53}$$

As (50, 51) imply that

$$\begin{aligned} r_2(1) &= q_2, \\ r_3(1) &= q_3, & r_3(2) &= 3(q_2 - q_3), \\ r_4(1) &= q_4, & r_4(2) &= 4(q_3 - q_4), & r_4(3) &= 6(q_2 - 2q_3 + q_4), \end{aligned} \tag{54}$$

14

one gets

$$\langle T_2 \rangle = \frac{1}{q_2}, \qquad \frac{\langle T_3 \rangle}{\langle T_2 \rangle} = \frac{4q_2 - 3q_3}{3q_2 - 2q_3},$$

$$\frac{\langle T_4 \rangle}{\langle T_2 \rangle} = \frac{27q_2^2 - 56q_2 q_3 + 28q_3^2 + 12q_2 q_4 - 10q_3 q_4}{(3q_2 - 2q_3)(6q_2 - 8q_3 + 3q_4)}. \tag{55}$$

If one defines $Z_{b \to b'}(\tau)$ as the probability that b points have coalesced into b' points during some time τ, one can easily see that it evolves according to

$$\frac{\mathrm{d}Z_{b \to b'}}{\mathrm{d}\tau} = \sum_{b'' > b'} r_{b''}(b')\, Z_{b \to b''} - \sum_{b'' < b'} r_{b'}(b'')\, Z_{b \to b'}, \tag{56}$$

with the initial condition that $Z_{b \to b'}(0) = \delta_{b,b'}$. Then, using the expressions (54) one gets

$$Z_{2 \to 2} = e^{-q_2 \tau},$$
$$Z_{2 \to 1} = 1 - e^{-q_2 \tau},$$

$$Z_{3 \to 3} = e^{-(3q_2 - 2q_3)\tau},$$
$$Z_{3 \to 2} = \frac{3}{2} e^{-q_2 \tau} - \frac{3}{2} e^{-(3q_2 - 2q_3)\tau},$$
$$Z_{3 \to 1} = 1 - \frac{3}{2} e^{-q_2 \tau} + \frac{1}{2} e^{-(3q_2 - 2q_3)\tau}, \tag{57}$$

$$Z_{4 \to 4} = e^{-(6q_2 - 8q_3 + 3q_4)\tau},$$
$$Z_{4 \to 3} = 2 e^{-(3q_2 - 2q_3)\tau} - 2 e^{-(6q_2 - 8q_3 + 3q_4)\tau},$$
$$Z_{4 \to 2} = \frac{9q_2 - 14q_3 + 5q_4}{5q_2 - 8q_3 + 3q_4} e^{-q_2 \tau} - 2 e^{-(3q_2 - 2q_3)\tau} + \frac{6q_2 - 10q_3 + 4q_4}{5q_2 - 8q_3 + 3q_4} e^{-(6q_2 - 8q_3 + 3q_4)\tau},$$
$$Z_{4 \to 1} = 1 - \frac{9q_2 - 14q_3 + 5q_4}{5q_2 - 8q_3 + 3q_4} e^{-q_2 \tau} + e^{-(3q_2 - 2q_3)\tau} - \frac{q_2 - 2q_3 + q_4}{5q_2 - 8q_3 + 3q_4} e^{-(6q_2 - 8q_3 + 3q_4)\tau}.$$

References

[1] J. Degnan and L. Salter, Evolution 59 (2005) p.24–37.

[2] R.R. Hudson, Oxford Surveys in Evolutionary Biology 7 (1991) p.1–44.

[3] R.R. Hudson and N.L. Kaplan, Genetics 120 (1988) p.831–840.

[4] C. Neuhauser and S. Krone, Genetics 145 (1997) p.519–534.

[5] M. Nordborg, Handbook of Statistical Genetics (2001) p.179–212.

[6] B. Derrida and L. Peliti, Bull. Math. Biol. 53 (1991) p.355–382.

[7] P. Donnelly and S. Tavare, Annual Review of Genetics 29 (1995) p.401–421.

[8] J.F.C. Kingman, Stoch. Proc. Appl. 13 (1982) p.235–248.

[9] J.F.C. Kingman, J. Appl. Probab. 19 (1982) p.27–43.

[10] É. Brunet, B. Derrida, A.H. Mueller and S. Munier, Europhys. Lett. 76 (2006) p.1–7.

[11] É. Brunet, B. Derrida, A.H. Mueller and S. Munier, Phys. Rev. E 76 (2007) p.041104.

[12] E. Bolthausen and A.S. Sznitman, Comm. Math. Phys. 197 (1998) p.247–276.

[13] J. Berestycki, N. Berestycki and J. Schweinsberg, arXiv:1001.2337 math.PR (2010).

[14] R. Durrett and D. Remenik, arXiv:0907.5180 math.PR (2009).

[15] R.E. Snyder, Ecol. 84 (2003) p.1333–1339.

[16] P. Bak and K. Sneppen, Phys. Rev. Lett. 71 (1993) p.4083–4086.

[17] É. Brunet, B. Derrida, A.H. Mueller and S. Munier, Phys. Rev. E 73 (2006) p.056126.

[18] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse and M.A. Virasoro, Journal de Physique 45 (1984) p.843–854.

[19] B. Derrida and H. Flyvbjerg, J. Phys. A 20 (1987) p.5273–5288.

[20] M. Mézard, G. Parisi and M.A. Virasoro *Spin glass theory and beyond*, Vol. 9, World Scientific Lecture Notes in Physics, 1987.

[21] G. Parisi, J. Phys. A 13 (1980) p.1101–1112.

[22] G. Parisi, Phys. Rev. Lett. 50 (1983) p.1946–1948.

[23] G. Parisi, Phys. Rev. Lett. 43 (1979) p.1754–1756.

[24] S. Kirkpatrick and D. Sherrington, Phys. Rev. B 17 (1978) p.4384–4403.

[25] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. 35 (1975) p.1792–1796.

[26] A.C.C. Coolen, R.W. Penney and D. Sherrington, Phys. Rev. B 48 (1993) p.16116–16118.

[27] I. Kondor, J. Phys. A 16 (1983) p.L127–L131.

[28] N. Berestycki, Ensaios Matematicos 16 (2009) p.1–193.

[29] J. Pitman and M. Yor, Ann. Probab. 25 (1997) p.855–900.

[30] B. Derrida, Physica D 107 (1997) p.186–198.

[31] D.A. Kessler, H. Levine, D. Ridgway and L. Tsimring, J. Stat. Phys. 87 (1997) p.519–544.

[32] M. Möhle and S. Sagitov, Ann. Probab. 29 (2001) p.1547–1562.

[33] J. Pitman, Ann. Probab. 27 (1999) p.1870–1902.

[34] J. Schweinsberg, Elect. Journ. Prob. 5 (2000) p.1–50.