DOI: 10.1214/EJP.v15-771

Electronic Journal of Probability

Vol. 15 (2010), Paper no. 26, pages 776–800.

Journal URL
http://www.math.washington.edu/~ejpecp/

On the two oldest families for the Wright-Fisher process

Jean-François Delmas *
delmas@cermics.enpc.fr

Jean-Stéphane Dhersin †
dhersin@math.univ-paris13.fr

Arno Siri-Jegousse ‡
Arno.Jegousse@mi.parisdescartes.fr

Abstract

We extend some of the results of Pfaffelhuber and Wakolbinger on the process of the most recent common ancestors in evolving coalescent by taking into account the size of one of the two oldest families or the oldest family which contains the immortal line of descent. For example we give an explicit formula for the Laplace transform of the extinction time for the Wright-Fisher diffusion. We give also an interpretation of the quasi-stationary distribution of the Wright-Fisher diffusion using the process of the relative size of one of the two oldest families, which can be seen as a resurrected Wright-Fisher diffusion .

Key words: Wright-Fisher diffusion, MRCA, Kingman coalescent tree, resurrected process, quasi-stationary distribution.

AMS 2000 Subject Classification: Primary 60J70, 60K35, 92D25.

Submitted to EJP on June 2, 2009, final version accepted April 30, 2010.

*Université Paris-Est, CERMICS, 6-8 av. Blaise Pascal, Champs-sur-Marne, 77455 Marne La Vallée, France.
†Département de Mathématiques, Université Paris 13, 99 av. B-B. Clément, F-93430 Villetaneuse, France
‡MAP5, Université Paris Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France

1 Introduction

Many models have been introduced to describe population dynamics in population genetics. Fisher [14], Wright [34] and Moran [25] have introduced two models for exchangeable haploid populations of constant size. A generalization has been given by Cannings [2]. Looking backward in time at the genealogical tree leads to coalescent processes, see Griffiths [17] for one of the first papers with coalescent ideas. For a large class of exchangeable haploid population models of constant size, when the size N tends to infinity and time is measured in units of "N generations", the associated coalescent process is Kingman's coalescent [22] (see also [24; 27; 28; 29] for general coalescent processes associated with Cannings' model). One of the associated object of interest is the time to the most recent common ancestor (TMRCA), say A_t, of the population living at time t. This is also the depth of their genealogical tree (see [9; 12]). In the case of Kingman's coalescent, each couple of individual merges at rate one, which gives a TMRCA with expectation 2, or an expected time equivalent to $2N$ generations in the discrete case (see [12] for more results on this approximation, [15] for the exact coalescent in the Wright-Fisher model and [24] for the statement for the convergence to the Kingman coalescent). When time t evolves forward from a fixed time t_0, the TMRCA at time t is $A_t = A_{t_0} + (t - t_0)$ until the most recent common ancestor (MRCA) of the population changes, and A_t jumps down. We say that at this time a new MRCA is established. Recent papers give an exhaustive study of times when MRCAs live and times when new MRCAs are established, see [26] and also [30] (see also [11] for genealogies of continuous state branching processes). In particular, for the Wright-Fisher (WF) model with infinite population size, the times when MRCAs live as well as the times when new MRCAs are established, are distributed according to a Poisson process, see [6] and [26].

In the Moran model (with finite population size) and in WF model with infinite population size, only two lineages can merge at a time. The population is divided in two "oldest" families each one born from one of the two children of the MRCA. Let X_t and $1-X_t$ denote the relative proportion of those two oldest families. One of these two oldest families will fixate (in the future); this one contains the immortal line of descent. Let Y_t be its relative size. We have: $Y_t = X_t$ with probability X_t and $Y_t = 1-X_t$ with probability $1-X_t$. At time $\tau_t = \inf\{s; X_{t+s} \in \{0,1\}\} = \inf\{s; Y_{t+s} = 1\}$, one of the two oldest families fixate, a new MRCA is established and two new oldest families appear. This corresponds to a jump of the processes $\mathbf{X} = (X_t, t \in \mathbb{R})$ and $\mathbf{Y} = (Y_t, t \in \mathbb{R})$. Processes $\mathbf{X}$ and $\mathbf{Y}$ are functionals of the genealogical trees, and could be studied by the approach developed in [16] on martingale problems for tree-valued process. In between two jumps the process $\mathbf{X}$ is a Wright-Fisher (WF) diffusion on $[0,1]$: $dX_t = \sqrt{X_t(1-X_t)}dB_t$, where B is a standard Brownian motion, with absorbing states 0 and 1. Similarly, in between two jumps the process $\mathbf{Y}$ is a WF diffusion on $[0,1]$ conditioned not to hit 0: $dY_t = \sqrt{Y_t(1-Y_t)}dB_t + (1-Y_t)dt$, with absorbing state 1. The WF diffusion and its conditioned version have been largely used to model allelic frequencies in a neutral two-types population, see [9; 12; 18]. A key tool to study the proportion of the two oldest families is the look-down representation for the genealogy introduced by Donnelly and Kurtz [7; 8] and a direct connection between the tree topology generated by a Pólya's urns and the Kingman's coalescent, see Theorem 2.1. In fact, we consider a biased Pólya's urn (because of the special role played by the immortal line of descendants).

Following Pfaffelhuber and Wakolbinger [26], we are interested in the distribution of the following quantities:

- A_t: the TMRCA for the population at time t.
- $\tau_t \geq 0$: the time to wait before a new MRCA is established (that is the next hitting time of $\{0,1\}$ for $\mathbf{X}$).
- $L_t \in \mathbb{N}^* = \{1,2,\ldots\}$: the number of living individuals which will have descendants at time τ.
- $Z_t \in \{0,\ldots,L_t\}$: the number of individuals present in the genealogy which will become MRCA of the population in the future.
- $Y_t \in (0,1)$ the relative size of the oldest family to which belongs the immortal line of descent.
- $X_t \in (0,1)$ the relative size of one of the two oldest families taken at random (with probability one half it has the immortal individual).

The distribution of (A_t, τ_t, L_t, Z_t) is given in [26] with t either a fixed time or a time when a new MRCA is established. We complete this result by giving, see Lemma 1.1 and Theorem 1.2 below, the joint distribution of $(A_t, \tau_t, L_t, Z_t, X_t, Y_t)$.

Let $(E_k, k \in \mathbb{N}^*)$ be independent exponential random variables with mean 1. We introduce

$$T_K = \sum_{k\geq 1} \frac{2}{k(k+1)} E_k \quad \text{and} \quad T_T = \sum_{k\geq 2} \frac{2}{k(k+1)} E_k.$$

Notice that T_K is distributed as the lifetime of a Kingman's coalescent process. The first part of the next Lemma is well known, and can be deduced from the look-down construction recalled in Section 2.1 and from [32]. The second part is proved in Section 4.6.

Lemma 1.1. *If t is a fixed time (resp. a time when a new MRCA is established) then A_t is distributed as T_K (resp. T_T). If t is a fixed time or a time when a new MRCA is established, then A_t is independent of $(\tau_t, L_t, Z_t, X_t, Y_t)$.*

By stationarity, the distribution of $(\tau_t, L_t, Z_t, X_t, Y_t)$ does not depend on t for fixed t. It does not depend on t either if t is a time when a new MRCA is established (the argument is the same as in the proof of Theorem 2 in [26]). This property is the analogue of the so-called PASTA (Poisson Arrivals See Time Average) property in queuing theory. For this reason, we shall write (τ, L, Z, X, Y) instead of $(\tau_t, L_t, Z_t, X_t, Y_t)$. We now state the main result of this paper, whose proof is given in Section 4.7.

Theorem 1.2. *At a fixed time t or at a time when a new MRCA is established, we have:*

i) Y is distributed as a beta $(2,1)$.

ii) Conditionally on Y, we have $X = \varepsilon Y + (1-\varepsilon)(1-Y)$ where ε is an independent random variable (of Y) such that $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = 0) = 1/2$. And X is uniform on $[0,1]$.

iii) Conditionally on Y, L is geometric with parameter $1-Y$.

iv) Conditionally on (Y, L), $\tau \stackrel{(d)}{=} \sum_{k=L}^{\infty} \frac{2}{k(k+1)} E_k$, where $(E_k, k \in \mathbb{N}^*)$ are independent exponential random variables with mean 1 and independent of (Y, L).

v) *Conditionally on* (Y, L)*,* $Z \overset{(d)}{=} \sum_{k=2}^{L} B_k$ *(with the convention* $\sum_{\emptyset} = 0$*), where* $(B_k, k \geq 2)$ *are independent Bernoulli random variables independent of* (Y, L) *and such that* $\mathbb{P}(B_k = 1) = 1/\binom{k}{2}$*.*

vi) *Conditionally on* (Y, L)*,* τ *and* Z *are independent.*

vii) *Conditionally on* Y*,* X *and* (τ, L, Z) *are independent.*

In Section 2, we also give formulas for the Laplace transform and the first two moments of Z conditionally on (Y, L), Y or X, see Corollaries 2.10, 2.11 and 2.12 (see also Remark 2.9 and (17) for a direct representation of the distribution of Z). Notice that results iii), iv), v) and vi) imply that given L the random variables Y, τ and Z are jointly independent. Those results also give a detailed proof of the heuristic arguments of Remarks 3.2 and 7.3 in [26]. From the conditional distribution of τ given in iv), we give its first two moments, see (10), and we recover the formula from Kimura and Ohta [20; 21] of its conditional expectation and second moment, see (13) and (14). See also (15) and (16) for the first and second moment of τ conditionally on X. The conditional distribution of τ given X is well known. Its Laplace transform is the solution of the ODE: $\mathscr{L}^X f = \lambda f$, with boundary condition $f(0) = f(1) = 1$, where $\mathscr{L}^X$ is the generator of the WF diffusion: $\mathscr{L}^X h(x) = x(1-x)h''(x)$ in $(0, 1)$. This Laplace transform is explicitly given by (12). We also recover (Corollary 2.6) that τ is an exponential random variable with mean 1, see [26] or [6].

We then give a new formula linking Z and τ, which is a consequence of Theorem 1.2 iv) and v) (see also (18)).

Corollary 1.3. *We have for all* $\lambda \geq 0$*:*

$$\mathbb{E}\left[\mathrm{e}^{-\lambda\tau} |Y, L\right] = \mathbb{E}\left[\mathrm{e}^{-\lambda T_K}\right] \mathbb{E}\left[(1+\lambda)^Z |Y, L\right]. \tag{1}$$

In particular, we deduce that

$$\mathbb{E}[\mathrm{e}^{-\lambda\tau} |X] = \mathbb{E}[\mathrm{e}^{-\lambda T_K}]\mathbb{E}[(1+\lambda)^Z |X]. \tag{2}$$

Notice that we also immediately get the following relations for the first moments:

$$\mathbb{E}[\tau|Y, L] = 2 - \mathbb{E}[Z|Y, L], \tag{3}$$

$$\mathbb{E}[\tau^2|Y, L] = \mathbb{E}[Z^2|Y, L] - 5\mathbb{E}[Z|Y, L] + \frac{4\pi^2}{3} - 8, \tag{4}$$

using that $\mathbb{E}[T_K] = 2$ for the first equality and that $\mathbb{E}[T_K^2] = \frac{4\pi^2}{3} - 8$ for the last.

Detailed results on the distribution of X, Y, L, Z, τ using the look-down process and ideas of [26] are stated in Section 2.

In Section 3.1, we recall that a probability measure μ is a quasi-stationary distribution (QSD) for a diffusion killed at an hitting time if and only if this is a stationary measure for the associated resurrected diffusion with resurrection measure μ, see Lemma 2.1 in [4] and also the pioneer work of [13] in a discrete setting. Notice that Theorem 1.2 ii) states that μ_0, the uniform distribution on $[0, 1]$, is a stationary measure for the resurrected WF diffusion with resurrection measure μ_0. We thus recover that μ_0 is a QSD of the WF diffusion, see [12; 18] and also [3]. The only QSD distribution of the WF diffusion is the uniform distribution, see [12, p. 161], or [18] for an explicit computation. Similarly, Theorem 1.2 i) states that μ_1, the beta $(2,1)$ distribution, is a stationary measure for the resurrected WF diffusion conditioned not to hit 0 with resurrection measure μ_1 and thus is a QSD of the WF diffusion conditioned not to hit 0, see also [18]. We check in Proposition 3.1 that μ_1 is indeed its only QSD.

In those two examples, the QSD distribution can be seen as the stationary distribution of the size of one of the two oldest families (either taken at random, or the one that fixates). A similar result is also true for the Moran model, see Section 3.4. But there is no such interpretation for the WF model for finite population, see Remark 3.2.

The proofs are postponed to Section 4.

2 Presentation of the main results on the conditional distribution

2.1 The look-down process and notations

The look-down process and the modified look-down process have been introduced by Donnelly and Kurtz [7; 8] to give the genealogical process associated to a diffusion model of population evolution (see also [10] for a detailed construction for the Fleming-Viot process). This powerful representation is now currently used. We briefly recall the definition of the modified look-down process, without taking into account any spatial motion or mutation for the individuals.

2.1.1 The set of individuals

Consider an infinite size population evolving forward in time. Let $E=\mathbb{R}\times\mathbb{N}^*$. Each (s,i) in E denotes the (unique) individual living at time s and level i. This level is affected according to the persistence of each individual: the higher the level is, the faster the individual will die. Let $(N_{i,j}, 0\leq i<j)$ be independent Poisson processes with rate 1. At a jumping time t of N_{ij}, the individual $(t-,i)$ reproduces and its unique child appears at level j. At the same time every individual having level at least j is pushed one level up (see Figure 1). These reproduction events involving levels i and j are called look-down events (as j looks down at i).

2.1.2 Partition of the set of individuals in lines

We can construct a partition of E in lines associated to the processes $N_{i,j}$ as follows. This partition contains the immortal line $\iota=\mathbb{R}\times\{1\}$. All the individuals which belong to the immortal line are called immortal individuals. The other lines of the partition start at look-down events: if an individual is born at level $j\geq 2$ at time s_0 by a look-down event (which means that s_0 is a jumping time of $N_{i,j}$ for some i), it initiates a new line

$$G=\bigcup_{k\in\mathbb{N}}[s_k,s_{k+1})\times\{j+k\},$$

where for $k\in\mathbb{N}^*$, s_k is the first birth time after s_{k-1} of an individual with level less than $j+k+1$. We shall write $b_G=s_0$ for the birth time of the line G. We say that $d_G=\lim_{k\to\infty}s_k$ is the death time of this line. We say that a line is alive at time t if $b_G \le t < d_G$, and k is the level of G at time t if the individual $(t,k) \in G$. We write $\mathscr{G}_t$ for the set of all lines alive at time t. A line at level j is pushed at rate $\binom{j}{2}$ to level $j+1$ (since there are $\binom{j}{2}$ possible independent look-down events which arrive at rate 1 and which push a line living at level j). Since $\sum_{j\ge 2} 1/\binom{j}{2} < \infty$, we get that any line but the immortal one dies in finite time.

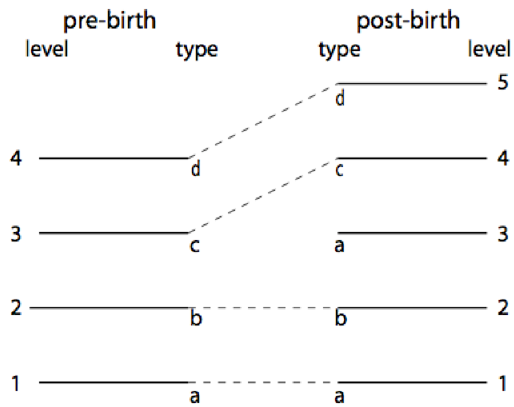

Figure 1: A look-down event between levels 1 and 3. Each line living at level at least 3 before the look-down event is pushed one level up after it.

2.1.3 The genealogy

Let (t,j) be an individual in a line G. An individual (s,i) is an ancestor of the individual (t,j) if $s \le t$ and either $(s,i) \in G$, or there is a finite sequence of lines $G_0, G_1, G_2, \dots, G_n = G$ such that each line G_k is initiated by a child of an individual in G_{k-1}, $k = 1, 2, \dots, n$ and $(s,i) \in G_0$.

For any fixed time t_0, we can introduce the following family of equivalence relations $\mathscr{R}^{(t_0)} = (\mathscr{R}_s^{(t_0)}, s \ge 0)$: $i\mathscr{R}_s^{(t_0)}j$ if the two individuals (t_0,i) and (t_0,j) have a common ancestor at time $t_0 - s$. It is then easy to show that the coalescent process on $\mathbb{N}^*$ defined by $\mathscr{R}^{(t_0)}$ is the Kingman's coalescent. See Figure 2 for a graphical representation.

2.1.4 Fixation curves

In the study of MRCA, some lines will play a particular role. We say that a line G is a fixation curve if $(b_G, 2) \in G$: the initial look-down event was from 2 to 1.

For a fixed time t, let G_t be the highest fixation curve. It has been initiated by the MRCA of the whole population living at time t. Notice that $t - A_t = \inf\{b_G; G \in \mathscr{G}_t, G \neq \iota\} = b_{G_t}$. Let $Z_t + 1$ denote the number of fixation curves living at time t: $Z_t \ge 0$ is the number of individuals present in the genealogy which will become MRCA of the population in the future. We denote by $L_0(t) > L_1(t) > \cdots > L_{Z_t}(t)$ the decreasing levels of the fixation curves alive at time t. Notice $L(t) = L_0(t) - 1$ is the number of living individuals at time t which will have descendants at the next MRCA change. The

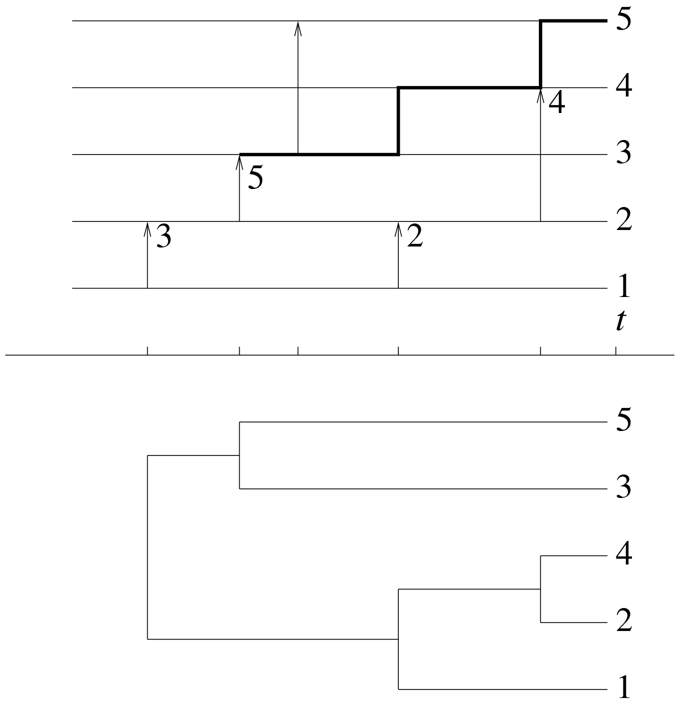

Figure 2: The look down process and its associated coalescent tree, started at time t for the 5 first levels. At each look-down event, a new line is born. We indicate at which level this line is at time t. The line of the individual $(t, 5)$ is bold.

joint distribution of $(Z_t, L_0(t), L_1(t), \ldots, L_{Z_t}(t))$ is given in Theorem 2 of [26], and the distribution of Z_t is given in Theorem 3 of [26].

2.1.5 The two oldest families

We consider the partition of the population into the two oldest families given by the equivalence relation $\mathscr{R}^{(t)}_{t-A_t}$. This corresponds to the partition of individuals alive at time t whose ancestor at time b_{G_t} is either $(b_{G_t}, 2)$ or the immortal individual $(b_{G_t}, 1)$. We shall denote by Y_t the relative proportion of the sub-population (i.e. the oldest family) whose ancestor at time b_{G_t} is the immortal individual, that is the oldest family which contains the immortal individual. Let X_t be the relative proportion of an oldest family picked at random: with probability $1/2$ it is the one which contains the immortal individual and with probability $1/2$ the other one.

2.1.6 Stationarity and PASTA property

We set $H_t = (X_t, Y_t, Z_t, L(t), L_1(t), \dots, L_{Z_t}(t))$. We are interested in the distribution of H_t at as well as the distribution of the labels of the individuals of the same oldest family. By stationarity, those distributions does not depend of t for fixed t. Arguing as in the proof of Theorem 2 of [26] they are also the same if t is a time when a new MRCA is established. This is the so-called PASTA (Poisson Arrivals See Time Average) property, see [1] for a review on this subject, where the Poisson process considered corresponds to the times when the MRCA changes. For this reason, we shall omit the subscript and write H, and carry out the proofs at a time when a new MRCA is established.

2.2 Size of the new two oldest families

We are interested in the description of the population, and more precisely in the relative size of the two oldest families at the time when a new MRCA is established. Let G_* be a fixation curve and G be the next fixation curve: they have been initiated by two successive present or future MRCAs. Let $s_0 = b_{G_*}$ be the birth time of G_* and $(s_k, k \in \mathbb{N}^*)$ be the jumping times of G_*. Notice that $s_1 = b_G$ corresponds to the birth time of G. Let $N \geq 2$. Notice that at time s_{N-1}, only the individuals with level 1 to N will have descendants at the death time d_{G^*} of G^*. They correspond to the ancestors at time s_{N-1} of the population living at time d_{G^*}. We consider the partition into 2 subsets given by $\mathscr{R}^{(s_{N-1})}_{s_{N-1}-s_0}$ which corresponds to the partition of individuals alive at time s_{N-1} with labels $k \in \{1, \dots, N\}$ whose ancestor at time s_1 is either the individual $(s_1, 2)$ which has initiated G or the immortal individual. We set $\sigma_N(k) = 1$ if this ancestor is the immortal individual and $\sigma_N(k) = 0$ if it is $(s_1, 2)$. Let $V_N = \sum_{k=1}^{N} \sigma_N(k)$ be the number of individuals at time s_{N-1} whose ancestor at time s_1 is the immortal individual, see Figure 3 for an example. Notice that $\lim_{N\to\infty} V_N/N$ will be the proportion of the oldest family which contains the immortal individual when $(b_{G^*}, 2)$ becomes the MRCA of the population. By construction the process $(\sigma_N, N \in \mathbb{N}^*)$ is Markov. Notice Theorem 2.1 below gives that $(V_N, N \in \mathbb{N}^*)$ is also Markov.

In order to give the law of (V_N, σ_N) we first recall some facts on Pólya's urns, see [19]. Let $S_N^{(i,j)}$ be the number of green balls in an urn after N drawing, when initially there was i green balls and j of some other color in the urn, and where at each drawing, the chosen ball is returned together with one ball of the same color. The process $(S_N^{(i,j)}, N \in \mathbb{N})$ is a Markov chain, and for $\ell \in \{0, \dots, N\}$

$$\mathbb{P}\left(S_N^{(i,j)} = i + \ell\right) = \binom{N}{\ell} \frac{(i+\ell-1)!(j+N-\ell-1)!(i+j-1)!}{(i-1)!(j-1)!(i+j+N-1)!}.$$

In particular, for $i = 2$, $j = 1$ and $k \in \{1, N+1\}$, we have

$$\mathbb{P}(S_N^{(2,1)} = k+1) = \frac{2k}{(N+2)(N+1)}. \tag{5}$$

The next Theorem is proved in Section 4.1.

Theorem 2.1. *Let $N \geq 2$.*

i) The process $(1+V_{N+2}, N \in \mathbb{N})$ is a Pólya's urn starting at $(2,1)$. In particular, V_N has a size-biased uniform distribution on $\{1, \dots, N-1\}$, i.e.

$$\mathbb{P}(V_N = k) = \frac{2k}{N(N-1)}.$$

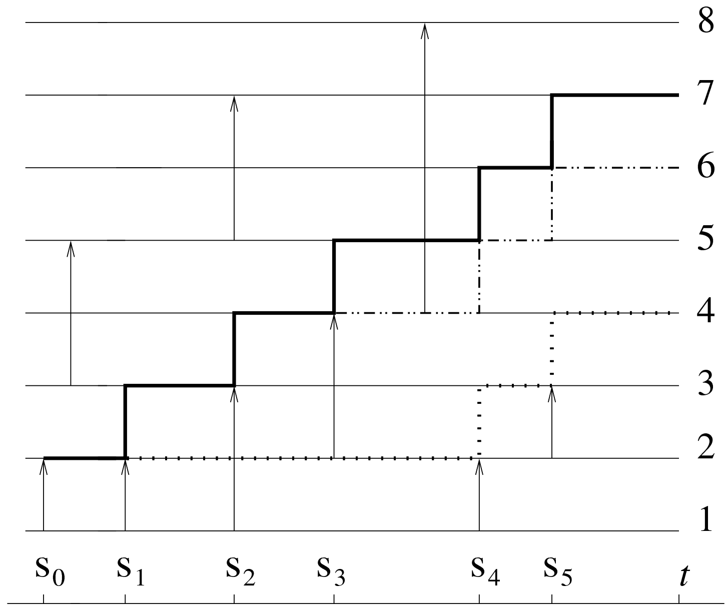

Figure 3: In this example, the fixation curve G_* is bold whereas the next fixation curve G is dotted. At time s_1, we have $\sigma_2 = (1,0)$ and $V_2 = 1$. At time s_2, we have $\sigma_3 = (1,0,1)$ and $V_3 = 2$. At time s_3, we have $\sigma_4 = (1,0,1,0)$ and $V_4 = 2$. At time s_4, we have $\sigma_5 = (1,1,0,1,0)$ and $V_5 = 3$. At time s_5, we have $\sigma_6 = (1,1,1,0,1,0)$ and $V_6 = 4$.

ii) Conditionally on $(V_1, \dots, V_N)$, σ_N is uniformly distributed on the possible configurations: $\{\sigma \in \{0,1\}^N; \sigma(1) = 1$ and $\sum_{k=1}^N \sigma(k) = V_N\}$.

Remark 2.2. We now give an informal proof of Theorem 2.1-i). If one forgets about the levels of the individuals but for the immortal one, one gets that when there are N lines, the immortal line gives birth to a new line at rate N, whereas one line taken at random (different from the immortal one) gives birth to a new line at rate $N/2$. Among those $N-1$ lines, $V_N - 1$ have a common ancestor with the immortal line at time s_1, $N - V_N$ do not. Let us say the former are of type 1 and the other are of type 0. The lines of type 0 are increased by 1 at rate $(N - V_N)N/2$. Taking into account that the immortal line gives birth to lines of type 1, we get that the lines of type 1 are increased by 1 at rate $N + (V_N - 1)N/2$. The probability to add a line of type 1 is then $(V_N + 1)/(N+1)$. Since $V_2 = 1$, we recover that $(1 + V_N, N \geq 2)$ is a Pólya's urn starting at $(2,1)$.

Notice that, in general, if $N_0 \geq 3$, the process $(V_{N_0+N}, N \in \mathbb{N})$ conditionally on σ_{N_0} can not be described using Pólya's urns.

Results on Pólya's urns, see Section 6.3.3 of [19], give that $(V_N/N, N \in \mathbb{N}^*)$ converges a.s. to a random variable Y with a beta distribution with parameters $(2,1)$. This gives the following result.

Corollary 2.3. *When a new MRCA is established, the relative proportion Y of the new oldest family which contains the immortal line of descent is distributed as a beta $(2,1)$.*

If one chooses a new oldest family at random (with probability $1/2$ the one which contains the immortal individual and with probability $1/2$ the other one), then its relative proportion X is uniform on $(0,1)$. This is coherent with the Remark 3.2 given in [26]. Notice that Y has the size biased distribution of X, which corresponds to the fact that the immortal individual is taken at random from the two oldest families with probability proportional to their size.

2.3 Level of the next fixation curve

We keep notations from the previous section. Let $L^{(N)}+1$ be the level of the fixation curve G when the fixation curve G_* reaches level $N+1$, that is at time s_{N-1}. Notice that $L^{(N)}$ belongs to $\{1,\dots,V_N\}$. The law of $(L^{(N)},V_N)$ will be useful to give the joint distribution of (Z,Y), see Section 2.5. It also implies (7) which was already given by Lemma 7.1 of [26]. The process $L^{(N)}$ is an inhomogeneous Markov chain, see Lemma 6.1 of [26]. By construction, the sequence $(L^{(N)},N\geq 2)$ is non-decreasing and converges a.s. to L defined in Section 2.1. The next Proposition is proved in Section 4.2.

Proposition 2.4. *Let $N\geq 2$.*

i) For $1\leq i\leq k\leq N-1$, we have

$$\mathbb{P}(L^{(N)}=i,V_N=k)=2\frac{(N-i-1)!}{N!}\frac{k!}{(k-i)!}\frac{N-k}{N-1}, \tag{6}$$

and for all $i\in\{1,\dots,N-1\}$,

$$\mathbb{P}(L^{(N)}=i)=\frac{N+1}{N-1}\frac{2}{(i+1)(i+2)}. \tag{7}$$

ii) The sequence $((L^{(N)},V_N/N),N\in\mathbb{N}^*)$ converges a.s. to a random variable (L,Y), where Y has a beta $(2,1)$ distribution and conditionally on Y, L is geometric with parameter $1-Y$.

A straightforward computation gives that for $i\in\mathbb{N}^*$

$$\mathbb{P}(L=i)=\frac{2}{(i+1)(i+2)}.$$

This result was already in Proposition 3.1 of [26].

The level $L+1$ corresponds to the level of the line of the current MRCA, when the MRCA is newly established. Recall $L_1(t)$ is the level at time t of the second fixation curve. We use the convention $L_1(t)=1$ if there is only one fixation curve i.e. $Z(t)=0$. Just before the random time d_{G_*} of the death of the fixation curve G_*, we have $L_1(d_{G_*}-)=L_0(d_{G^*})=L+1$. At a fixed time t, by stationarity, the distribution of $L_1(t)$ does not depend on t, and equation (3.4) from [26] gives that $L_1(t)$ is distributed as L. In view of Remark 4.1 in [26], notice the result is also similar for $M/M/k$ queue where the invariant distribution for the queue process and the queue process just before arrivals time are the same, thanks to the PASTA property.

2.4 Next fixation time

We consider the time d_{G_*} of death of the line G_* (which corresponds to a time when a new MRCA is established). At this time, Y is the proportion of the oldest family which contains the immortal individuals. We denote by τ the time we have to wait for the next fixation time. It is the time needed by the highest fixation curve alive at time d_{G_*} to reach ∞. Hence, by the look-down construction, we get that

$$\tau \stackrel{(d)}{=} \sum_{k=L}^{\infty} \frac{2}{k(k+1)} E_k \tag{8}$$

where E_k are independent exponential random variables with parameter 1 and independent of Y and L. See also Theorem 1 in [26].

Proposition 2.5. *Let $a \in \mathbb{N}^*$. The distribution of the waiting time for the next fixation time is given by: for $\lambda \in \mathbb{R}_+$,*

$$\mathbb{E}[\mathrm{e}^{-\lambda\tau} | Y, L = a] = \prod_{k=a}^{\infty} \left(\frac{k(k+1)}{k(k+1) + 2\lambda} \right). \tag{9}$$

Its first two moments are given by:

$$\mathbb{E}[\tau | Y, L = a] = \frac{2}{a} \quad \text{and} \quad \mathbb{E}[\tau^2 | Y, L = a] = -\frac{8}{a} + 8 \sum_{k \geq a} \frac{1}{k^2}. \tag{10}$$

We also have: for $y, x \in (0,1)$ and $\lambda \in \mathbb{R}_+$,

$$\mathbb{E}[\mathrm{e}^{-\lambda\tau} | Y = y] = (1-y) \sum_{\ell=1}^{\infty} y^{\ell-1} \prod_{k=\ell}^{\infty} \left(\frac{k(k+1)}{k(k+1) + 2\lambda} \right), \tag{11}$$

$$\mathbb{E}[\mathrm{e}^{-\lambda\tau} | X = x] = x(1-x) \sum_{\ell=1}^{\infty} \left[x^{\ell-1} + (1-x)^{\ell-1} \right] \prod_{k=\ell}^{\infty} \left(\frac{k(k+1)}{k(k+1) + 2\lambda} \right). \tag{12}$$

We deduce from (10) that $\mathbb{E}[\tau | L = a] = \frac{2}{a}$, which was already in Theorem 1 in [26]. Notice that using (11), we recover the following result.

Corollary 2.6. *The random variable τ is exponential with mean* 1.

Using (10) and the fact that L is geometric with parameter $1 - Y$, we recover the well known results from Kimura and Ohta [20; 21] (see also [12]):

$$\mathbb{E}[\tau | Y = y] = -2 \frac{(1-y)\log(1-y)}{y}, \tag{13}$$

$$\mathbb{E}[\tau^2 | Y = y] = 8 \left(\frac{(1-y)\log(1-y)}{y} - \int_y^1 \frac{\log(1-z)}{z} \, dz \right). \tag{14}$$

The following Lemma is elementary.

Lemma 2.7. *Let Y be a beta (2,1) random variable and $X = \varepsilon Y + (1-\varepsilon)(1-Y)$ where ε is independent of Y and such that $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. Then X is uniform on $[0,1]$. Furthermore, if W is integrable and independent of ε, then we have $\mathbb{E}[W|X] = Xg(X) + (1-X)g(1-X)$ where $g(y) = \mathbb{E}[W|Y = y]$.*

We also get, thanks to the above Lemma that:

$$\mathbb{E}[\tau|X = x] = -2\left(x\log(x) + (1-x)\log(1-x)\right), \quad \text{and} \tag{15}$$

$$\mathbb{E}[\tau^2|X = x] = 8\left(x\log(x) + (1-x)\log(1-x) - x\int_x^1 \frac{\log(1-z)}{z}\,dz \right. \tag{16}$$
$$\left. -(1-x)\int_{1-x}^1 \frac{\log(1-z)}{z}\,dz\right).$$

2.5 Number of individuals present which will become MRCA

We keep notations from Sections 2.1 and 2.3. We set $Z = Z_{d_{G_*}}$ the number of individuals living at time d_{G_*} which will become MRCA of the population in the future. Let $L_0 = L(d_{G_*}) + 1$ and $(L_0, L_1, \ldots, L_Z) = (L_0(d_{G_*}), \ldots, L_Z(d_{G_*}))$ be the levels of the fixation curves at the death time of G_*. Recall notations from Section 2.2. The following Lemma and Proposition 2.4 characterize the joint distribution of $(Y, Z, L, L_1, \ldots, L_Z)$.

Lemma 2.8. *Conditionally on (L, Y) the distribution of $(Z, L_1, \ldots, L_Z)$ does not depend on Y. Conditionally on $\{L = N\}$, $(Z, L_1, \ldots, L_Z)$ is distributed as follows:*

1. *$Z = 0$ if $N = 1$;*
2. *Conditionally on $\{Z \geq 1\}$, L_1 is distributed as $L^{(N)} + 1$.*
3. *For $N' \in \{1, \ldots, N-1\}$, conditionally on $\{Z \geq 1, L_1 = N' + 1\}$, $(Z - 1, L_2, \ldots, L_Z)$ is distributed as $(Z, L_1, \ldots, L_Z)$ conditionally on $\{L = N'\}$.*

Remark 2.9. If one is interested only in the distribution of $(Z, L_0, \ldots, L_Z, L)$, one gets that $\{L_Z, \ldots, L_0\}$ is distributed as $\{k; B_k = 1\}$ where $(B_n, n \geq 2)$ are independent Bernoulli r.v. such that $\mathbb{P}(B_k = 1) = 1/\binom{k}{2}$. In particular we have

$$Z \overset{(d)}{=} \sum_{k \geq 2} B_k - 1. \tag{17}$$

Indeed, set $B_k = 1$ if the individual (k, d_{G_*}) at level k belongs to a fixation curve and $B_k = 0$ otherwise. Notice that $B_k = 1$ if none of the $k-2$ look-down events which pushed the line of (k, d_{G_*}) between its birth time and d_{G_*} involved the line of (k, d_{G_*}). This happens with probability

$$\mathbb{P}(B_k = 1) = \frac{\binom{k-1}{2}}{\binom{k}{2}} \cdots \frac{\binom{2}{2}}{\binom{3}{2}} = \frac{1}{\binom{k}{2}}.$$

Moreover B_k is independent of $B_2, \ldots, B_{k-1}$ which depends on the lines below the line of (k, d_{G_*}) from the look-down construction. This gives the announced result. Notice that $L = L_0 - 1 =$

$\sup\{k; B_k = 1\} - 1$. We deduce that conditionally on $L = a$, $Z = \sum_{k=2}^{a} B_k$ (with the convention $Z = 0$ if $a = 1$). In particular, we get

$$\mathbb{E}[(1+\lambda)^Z | L = a] = \prod_{k=2}^{a} \mathbb{E}[(1+\lambda)^{B_k}] = \prod_{k=2}^{a} \frac{k(k-1)+2\lambda}{k(k-1)} = \prod_{k=1}^{a-1} \frac{k(k+1)+2\lambda}{k(k+1)}.$$

The result does not change if one considers a fixed time t instead of d_{G_*}.

We deduce the following Corollary from the previous Remark and Lemma 2.8 and for the first two moments (20) we use (10) and Proposition 1.3.

Corollary 2.10. *Let $a \geq 1$. Conditionally on (Y, L), $Z \overset{(d)}{=} \sum_{k=2}^{L} B_k$ (with the convention $\sum_{\emptyset} = 0$), where $(B_k, k \geq 2)$ are independent Bernoulli random variables independent of (Y, L) and such that $\mathbb{P}(B_k = 1) = 1/\binom{k}{2}$. We have for all $\lambda \geq 0$,*

$$\mathbb{E}[(1+\lambda)^Z | Y, L = a] = \prod_{k=1}^{a-1} \frac{k(k+1)+2\lambda}{k(k+1)}, \tag{18}$$

with the convention $\prod_{\emptyset} = 1$. We have $\mathbb{P}(Z = 0 | Y, L = 1) = 1$ and for $k \geq 1$,

$$\mathbb{P}(Z = k | Y, L = a) = \frac{2^{k-1}}{3} \frac{a+1}{a-1} \sum_{1 < a_k < \cdots < a_2 < a} \prod_{i=2}^{k} \frac{1}{(a_i - 1)(a_i + 2)}. \tag{19}$$

We also have

$$\mathbb{E}[Z | Y, L = a] = 2 - \frac{2}{a} \quad \text{and} \quad \mathbb{E}[Z^2 | Y, L = a] = 18 - \frac{4\pi^2}{3} - \frac{18}{a} + 8 \sum_{k \geq a} \frac{1}{k^2}. \tag{20}$$

We are now able to give the distribution of Z conditionally on Y or X. We deduce from ii) of Proposition 2.4 and from Corollary 2.10 the next result.

Corollary 2.11. *Let $y \in [0, 1]$. We have, for all $\lambda \geq 0$,*

$$\mathbb{E}[(1+\lambda)^Z | Y = y] = (1-y) \sum_{a=1}^{+\infty} y^{a-1} \prod_{k=1}^{a-1} \frac{k(k+1)+2\lambda}{k(k+1)}, \tag{21}$$

with the convention $\prod_{\emptyset} = 1$. We have $\mathbb{P}(Z = 0 | Y = y) = 1 - y$, and, for all $k \in \mathbb{N}^*$,

$$\mathbb{P}(Z = k | Y = y) = \frac{2^{k-1}}{3}(1-y) \sum_{1 < a_k < \cdots < a_1 < \infty} (a_1+1)(a_1+2) y^{a_1 - 1} \prod_{i=1}^{k} \frac{1}{(a_i - 1)(a_i + 2)}. \tag{22}$$

We also have

$$\mathbb{E}[Z | Y = y] = 2\left(1 + \frac{1-y}{y} \log(1-y)\right). \tag{23}$$

The next Corollary is a direct consequence of Lemma 2.7 and Corollary 2.10 .

Corollary 2.12. *Let $x \in [0,1]$. We have, for all $\lambda \geq 0$,*

$$\mathbb{E}[(1+\lambda)^Z | X = x] = x(1-x) \sum_{a=2}^{\infty} \left(x^{a-1} + (1-x)^{a-1} \right) \sum_{a=1}^{+\infty} \prod_{k=1}^{a-1} \frac{k(k+1)+2\lambda}{k(k+1)}, \tag{24}$$

with the convention $\prod_{\emptyset} = 1$. We have $\mathbb{P}(Z=0|X=x) = 2x(1-x)$, and, for all $k \in \mathbb{N}^*$,

$$\mathbb{P}(Z=k|X=x)$$
$$= \frac{2^{k-1}}{3} x(1-x) \sum_{1<a_k<\cdots<a_1<\infty} (a_1+1)(a_1+2) \left(x^{a_1-2} + (1-x)^{a_1-2} \right) \prod_{i=1}^{k} \frac{1}{(a_i-1)(a_i+2)}. \tag{25}$$

We also have

$$\mathbb{E}[Z|X=x] = 2\left(1 + x\log(x) + (1-x)\log(1-x)\right). \tag{26}$$

The second moment of Z conditionally on Y (resp. X) can be deduced from (21) (resp. (24)) or from (4) and (14) (resp. (16)).

Some elementary computations give:

$$\begin{aligned}
\mathbb{P}(Z=0|X=x) &= 2x(1-x), \\
\mathbb{P}(Z=1|X=x) &= \frac{1}{3}\left[x^2 + (1-x)^2 - 2x(1-x)\ln(x(1-x)) \right], \\
\mathbb{P}(Z=2|X=x) &= \frac{2}{3}\left[\frac{11}{6}(x^2+(1-x)^2) - (1-x)\ln(1-x) - x\ln(x) \right] \\
&\qquad + \frac{2}{3}x(1-x)\left[2 - \frac{\pi^2}{3} + 2\ln(x)\ln(1-x) - \frac{1}{3}\ln(x(1-x)) \right].
\end{aligned}$$

We recover by integration of the previous equations the following results from [26]:

$$\mathbb{P}(Z=0) = \frac{1}{3}, \quad \mathbb{P}(Z=1) = \frac{11}{27} \quad \text{and} \quad \mathbb{P}(Z=2) = \frac{107}{243} - \frac{2}{81}\pi^2.$$

3 Stationary distribution of the relative size for the two oldest families

3.1 Resurrected process and quasi-stationary distribution

Let E be a subset of $\mathbb{R}$. We recall that if $U = (U_t, t \geq 0)$ is an E-valued diffusion with absorbing states Δ, we say that a distribution ν is a quasi-stationary distribution (QSD) of U if for any Borel set $A \subset \mathbb{R}$,

$$\mathbb{P}_\nu(U_t \in A | U_t \notin \Delta) = \nu(A) \quad t \geq 0,$$

where we write $\mathbb{P}_\nu$ when the distribution of U_0 is ν. See also [31] for QSD for diffusions with killing.

Let μ and ν be two distributions on $E \backslash \Delta$. We define U^μ the resurrected process associated to U, with resurrection distribution μ, under $\mathbb{P}_\nu$ as follows:

1. U_0 is distributed according to ν and $U_t^\mu = U_t$ for $t \in [0, \tau_1)$, where $\tau_1 = \inf\{s \geq 0; U_s \in \Delta\}$.

2. Conditionally on $(\tau_1, \{\tau_1 < \infty\}, (U_t^\mu, t \in [0, \tau_1)))$, $(U_{t+\tau_1}^\mu, t \geq 0)$ is distributed as U^μ under $\mathbb{P}_\mu$.

According to Lemma 2.1 of [4], the distribution μ is a QSD of U if and only if μ is a stationary distribution of U^μ. See also the pioneer work of [13] in a discrete setting.

The uniqueness of quasi-stationary distributions is an open question in general. We will give a genealogical representation of the QSD for the Wright-Fisher diffusion and the Wright-Fisher diffusion conditioned not to hit 0, as well as for the Moran model for the discrete case.

We also recall that the so-called Yaglom limit μ is defined by

$$\lim_{t\to\infty} \mathbb{P}_x(U_t \in A | U_t \notin \Delta) = \mu(A) \quad \forall A \in \mathscr{B}(\mathbb{R}),$$

provided the limit exists and is independent of $x \in E \backslash \Delta$.

3.2 The resurrected Wright-Fisher diffusion

From Corollary 2.3 and comments below it, we get that the relative proportion of one of the two oldest families at a time when a new MRCA is established is distributed according to the uniform distribution over $[0, 1]$. Then the relative proportion evolves according to a Wright-Fisher (WF) diffusion. In particular it hits the absorbing state of the WF diffusion, $\{0, 1\}$, in finite time. At this time one of the two oldest families dies out and there a new MRCA is (again) established.

The QSD distribution of the WF diffusion exists and is the uniform distribution, see [12, p. 161], or [18] for an explicit computation. From Section 3.1, we get that in stationary regime, for fixed t (and of course at time when a new MRCA is established) the relative size, X_t, of one of the two oldest families taken at random is uniform over $(0, 1)$.

Similar arguments as those developed in the proof of Proposition 3.1 yield that the uniform distribution is the only QSD of the WF diffusion. Lemma 2.1 in [4] implies there is no other resurrection distribution which is also the stationary distribution of the resurrected process.

3.3 The oldest family with the immortal line of descent

Recall that $Y = (Y_t, t \in \mathbb{R})$ is the process of relative size for the oldest family containing the immortal individual. From Corollary 2.3, we get that, at a time when a new MRCA is established, Y is distributed according to the beta $(2, 1)$ distribution. Then Y evolves according to a WF diffusion conditioned not to hit 0; its generator is given by $\mathscr{L} = \frac{1}{2} x(1-x)\partial_x^2 + (1-x)\partial_x$, see [9; 18]. Therefore Y is a resurrected Wright-Fisher diffusion conditioned not to hit 0, with beta $(2, 1)$ resurrection distribution.

The Yaglom distribution of the Wright-Fisher diffusion conditioned not to hit 0 exists and is the beta $(2, 1)$ distribution, see [18] for an explicit computation. In fact the Yaglom distribution is the only QSD according to the next proposition.

Proposition 3.1. *The only quasi-stationary distribution of the Wright-Fisher diffusion conditioned not to hit* 0 *is the beta* $(2, 1)$ *distribution.*

Lemma 2.1 in [4] implies that the beta $(2, 1)$ distribution is therefore the stationary distribution of Y. Furthermore, the resurrected Wright-Fisher diffusion conditioned not to hit 0, with resurrection distribution μ has stationary distribution μ if and only if μ is the beta $(2, 1)$ distribution.

3.4 Resurrected process in the Moran model

The Moran model has been introduced in [25]. This mathematical model represents the neutral evolution of a haploid population of fixed size, say N. Each individual gives, at rate 1, birth to a child, which replaces an individual taken at random among the N individuals. Notice the population size is constant. Let ξ_t denote the size of the descendants at time t of a given initial group. The process $\xi = (\xi_t, t \geq 0)$ goes from state k to state $k + \varepsilon$, where $\varepsilon \in \{-1, 1\}$, at rate $k(N-k)/N$. Notice that 0 and N are absorbing states. They correspond respectively to the extinction of the descendants of the initial group or its fixation. The Yaglom distribution of the process ξ is uniform over $\{1, \ldots, N-1\}$ (see [12, p. 106]). Since the state is finite, the Yaglom distribution is the only QSD.

Let μ be a distribution on $\{1, \ldots, N-1\}$. We consider the resurrected process $(\xi_t^\mu, t \geq 0)$ with resurrection distribution μ. The resurrected process has the same evolution as ξ until it reaches 0 or N, and it immediately jumps according to μ when it hits 0 or N. The process ξ^μ is a continuous time Markov process on $\{1, \ldots, N-1\}$ with transition rates matrix Λ^μ given by:

$$\begin{aligned}
\Lambda^\mu(1,k) &= \left(\mu(k) + \mathbf{1}_{\{k=2\}}\right) \frac{N-1}{N} \quad \text{for } k \in \{2, \ldots, N-1\}, \\
\Lambda^\mu(k, k+\varepsilon) &= \frac{k(N-k)}{N} \quad \text{for } \varepsilon \in \{-1, 1\} \text{ and } k \in \{2, \ldots, N-2\}, \\
\Lambda^\mu(N-1, k) &= \left(\mu(k) + \mathbf{1}_{\{k=N-2\}}\right) \frac{N-1}{N} \quad \text{for } k \in \{1, \ldots, N-2\}.
\end{aligned}$$

We deduce from [13], that μ is a stationary distribution for ξ^μ (i.e. $\mu \Lambda^\mu = 0$) if and only if μ is a QSD for ξ, hence if and only if μ is uniform over $\{1, \ldots, N-1\}$.

Using the genealogy of the Moran model, we can give a natural representation of the resurrected process ξ^μ when the resurrection distribution is the Yaglom distribution. Since the genealogy of the Moran model can be described by the restriction of the look-down process to $E^{(N)} = \mathbb{R} \times \{1, \ldots, N\}$, we get from Theorem 2.1 that the size of the oldest family which contains the immortal individual is distributed as the size-biased uniform distribution on $\{1, \ldots, N-1\}$ at a time when a new MRCA is established. The PASTA property also implies that this is the stationary distribution. If, at a time when a new MRCA is established, we consider at random one of the two oldest families (with probability $1/2$ the one with the immortal individual and with probability $1/2$ the other one), then the size process is distributed as $(\xi_t^\mu, t \in \mathbb{R})$ under its stationary distribution, with μ the uniform distribution.

Remark 3.2. We can also consider the Wright-Fisher model (see e.g. [9]) in discrete time with a population of fixed finite size N, $\zeta = (\zeta_k, k \in \mathbb{N})$. This is a Markov chain with state space $\{0, \ldots, N\}$ and transition probabilities

$$P(i,j) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}.$$

There exists a unique quasi-stationary distribution, μ_N (which is not the uniform distribution), see [5]. We deduce that the resurrected process ζ^μ has stationary distribution μ if and only if $\mu = \mu_N$. Notice, that in this example there is no biological interpretation of μ_N as the size of one of the oldest family at a time when a new MRCA is established.

4 Proofs

4.1 Proof of Theorem 2.1

We consider the set

$$A_N = \left\{(k_1,\dots,k_N); k_1 = 1, \text{ for } i \in \{1,\dots,N-1\}, k_{i+1} \in \left\{k_i, k_i+1\right\}\right\}.$$

Notice that $\mathbb{P}(V_1 = k_1,\dots,V_N = k_N) > 0$ if and only if $(k_1,\dots,k_N) \in A_N$. To prove the first part of Theorem 2.1, it is enough to show that, for $N \geq 2$ and $(k_1,\dots,k_{N+1}) \in A_{N+1}$,

$$\mathbb{P}(V_{N+1} = k_{N+1} | V_N = k_N,\dots,V_1 = k_1) = \begin{cases} 1 - \frac{1+k_N}{N+1} & \text{if } k_{N+1} = k_N, \\ \frac{1+k_N}{N+1} & \text{if } k_{N+1} = 1 + k_N. \end{cases} \tag{27}$$

For p and q in $\mathbb{N}^*$ such that $q < p$, we introduce the set:

$$\Delta_{p,q} = \{\alpha = (\alpha_1,\dots,\alpha_p) \in \{0,1\}^p, \alpha_1 = 1, \sum_{i=1}^{p} \alpha_i = q\}.$$

Notice that Card $(\Delta_{p,q}) = \binom{p-1}{q-1}$. Hence to prove the second part of Theorem 2.1, it is enough to show that: for all $(k_1,\dots,k_N) \in A_N$, and all $\alpha \in \Delta_{N,k_N}$,

$$\mathbb{P}(\sigma_N = \alpha | V_N = k_N,\dots,V_1 = k_1) = \frac{1}{\binom{N-1}{k_N-1}}. \tag{28}$$

We proceed by induction on N for the proof of (27) and (28). The result is obvious for $N = 2$. We suppose that (27) and (28) are true for a fixed N. We denote by I_N and J_N, $1 \leq I_N < J_N \leq N+1$, the two levels involved in the look-down event at time s_N. Notice that (I_N, J_N) and σ_N are independent. This pair is chosen uniformly so that, for $1 \leq i < j \leq N+1$,

$$\mathbb{P}(I_N = i, J_N = j) = \frac{2}{(N+1)N},$$
$$\mathbb{P}(I_N = i) = \frac{2(N-i+1)}{(N+1)N},$$
$$\mathbb{P}(J_N = j) = \frac{2(j-1)}{(N+1)N}.$$

For $\alpha = (\alpha_1,\dots,\alpha_{N+1}) \in \{0,1\}^{N+1}$ and $j \in \{1,\dots,N+1\}$, we set $\alpha_\times^j = (\alpha_1,\dots,\alpha_{j-1}, \alpha_{j+1},\dots,\alpha_{N+1}) \in \{0,1\}^N$.

Let us fix $(k_1,\dots,k_{N+1}) \in A_{N+1}$, and $\alpha = (\alpha_1,\dots,\alpha_{N+1}) \in \Delta_{N+1,k_{N+1}}$. Notice that $\{\sigma_{N+1} = \alpha\} \subset \{V_{N+1} = k_{N+1}\}$. We first compute

$$\mathbb{P}(\sigma_{N+1} = \alpha | V_N = k_N,\dots,V_1 = k_1).$$

1st case: $k_{N+1} = k_N + 1$. We have:

$$
\begin{aligned}
\mathbb{P}(\sigma_{N+1} = \alpha | V_N = k_N, \ldots, V_1 = k_1) \\
&= \sum_{1 \le i < j \le N+1} \mathbb{P}(I_N = i, J_N = j, \sigma_{N+1} = \alpha | V_N = k_N, \ldots, V_1 = k_1) \\
&= \sum_{1 \le i < j \le N+1, \alpha_i = \alpha_j = 1} \mathbb{P}(I_N = i, J_N = j, \sigma_N = \alpha^j_\times | V_N = k_N, \ldots, V_1 = k_1) \\
&= \sum_{1 \le i < j \le N+1, \alpha_i = \alpha_j = 1} \mathbb{P}(I_N = i, J_N = j) \mathbb{P}(\sigma_N = \alpha^j_\times | V_N = k_N, \ldots, V_1 = k_1) \\
&= \sum_{1 \le i < j \le N+1, \alpha_i = \alpha_j = 1} \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N - 1}} \\
&= \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N - 1}} \frac{k_{N+1}(k_{N+1} - 1)}{2} \\
&= \frac{(k_N + 1)!(N - k_N)!}{(N+1)!},
\end{aligned} \tag{29}
$$

where we used the independence of (I_N, J_N) and σ_N for the third equality, the uniform distribution of σ_N conditionally on V_N for the fourth, and that $k_{N+1} = k_N + 1$ for the sixth. Hence, we get

$$
\begin{aligned}
\mathbb{P}(V_{N+1} = k_N + 1 | V_N = k_N, \ldots, V_1 = k_1) &= \sum_{\alpha \in \Delta_{N+1, k_{N+1}}} \mathbb{P}(\sigma_{N+1} = \alpha | V_N = k_N, \ldots, V_1 = k_1) \\
&= \binom{N}{k_{N+1} - 1} \frac{(k_N + 1)!(N - k_N)!}{(N+1)!} \\
&= \frac{1 + k_N}{N+1}.
\end{aligned} \tag{30}
$$

2nd case: $k_{N+1} = k_N$. Similarly, we have:

$$
\begin{aligned}
\mathbb{P}(\sigma_{N+1} = \alpha | V_N = k_N, \ldots, V_1 = k_1) &= \sum_{1 \le i < j \le N+1, \alpha_i = \alpha_j = 0} \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N - 1}} \\
&= \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N - 1}} \frac{(N + 1 - k_N)(N - k_N)}{2} \\
&= \frac{(N - k_N)(k_N - 1)!(N - k_N + 1)!}{(N+1)!}.
\end{aligned} \tag{31}
$$

Hence, we get

$$
\begin{aligned}
\mathbb{P}(V_{N+1} = k_N | V_N = k_N, \ldots, V_1 = k_1) &= \sum_{\alpha \in \Delta_{N+1, k_{N+1}}} \mathbb{P}(\sigma_{N+1} = \alpha | V_N = k_N, \ldots, V_1 = k_1) \\
&= \binom{N}{k_{N+1} - 1} \frac{(N - k_N)(k_N - 1)!(N - k_N + 1)!}{(N+1)!} \\
&= 1 - \frac{1 + k_N}{N+1}.
\end{aligned} \tag{32}
$$

Equalities (30) and (32) imply (27). Moreover, we deduce from (29) and (31) that, for $k_{N+1} \in \{k_N, k_N+1\}$,

$$\mathbb{P}(\sigma_{N+1}=\alpha|V_{N+1}=k_{N+1},\dots,V_1=k_1) = \frac{\mathbb{P}(\sigma_{N+1}=\alpha, V_{N+1}=k_{N+1}|V_N=k_N,\dots,V_1=k_1)}{\mathbb{P}(V_{N+1}=k_{N+1}|V_N=k_N,\dots,V_1=k_1)} = \frac{1}{\binom{N}{k_{N+1}-1}},$$

which proves that (28) with N replaced by $N+1$ holds. This ends the proof.

4.2 Proof of Proposition 2.4

Theorem 2.1 shows that the distribution of σ_N conditionally on V_N is uniform. Then, if $V_N=k$, we can see $L^{(N)}$ as the number of draws (without replacement) we have to do in a two-colored urn of size $N-1$ with $k-1$ black balls until we obtain a white ball. Hence, for $k \in \{1,\dots,N-1\}$ and $i \in \{1,\dots,k\}$,

$$\begin{aligned}\mathbb{P}(L^{(N)}=i|V_N=k) &= \frac{k-1}{N-1}\frac{k-2}{N-2}\cdots\frac{k-i+1}{N-i+1}\frac{N-k}{N-i} \\ &= \frac{(N-i-1)!}{(N-1)!}\frac{(k-1)!}{(k-i)!}(N-k).\end{aligned}$$

This and Theorem 2.1 give (6).

It is easy to prove by induction on j that for all $j \in \mathbb{N}$,

$$\sum_{k=i}^{i+j}\frac{k!}{(k-i)!} = \frac{(i+j+1)!}{j!(i+1)}. \tag{33}$$

Summing (6) over $k \in \{i,\dots,N-1\}$ gives:

$$\begin{aligned}\mathbb{P}(L^{(N)}=i) &= \frac{2(N-i-1)!}{N!(N-1)}\sum_{k=i}^{N-1}\frac{k!}{(k-i)!}(N-k) \\ &= \frac{2(N-i-1)!}{N!(N-1)}\left[(N+1)\sum_{k=i}^{N-1}\frac{k!}{(k-i)!} - \sum_{k=i}^{N-1}\frac{(k+1)!}{((k+1)-(i+1))!}\right] \\ &= \frac{2(N-i-1)!}{N!(N-1)}\left[\frac{(N+1)!}{(N-i-1)!(i+1)} - \frac{(N+1)!}{(N-i-1)!(i+2)}\right] \\ &= 2\frac{N+1}{N-1}\frac{1}{(i+1)(i+2)},\end{aligned}$$

where we used (33) twice in the third equality.

Since $(L^{(N)}, n \in \mathbb{N}^*)$ is non-decreasing, we deduce from Theorem 2.1 that the sequence $((L^{(N)}, V^{(N)}/N), N \in \mathbb{N}^*)$ converges a.s. to a limit (L, Y). Let $i \geq 1$ and $v \in [0,1)$. We have:

$$\begin{aligned}
\mathbb{P}\left(L^{(N)} = i, \frac{V_N}{N} \leq v\right) &= \sum_{k=i}^{\lfloor Nv \rfloor} \mathbb{P}\left(L^{(N)} = i, V_N = k\right) \\
&= \sum_{k=i}^{\lfloor Nv \rfloor} 2\frac{(N-i-1)!}{N!} \frac{k!}{(k-i)!} \frac{N-k}{N-1} \\
&= \frac{2}{N} \sum_{k=i}^{\lfloor Nv \rfloor} \frac{k}{N-1} \frac{k-1}{N-2} \cdots \frac{k-i+1}{N-i} \left(1 - \frac{k-1}{N-1}\right),
\end{aligned}$$

which converges to $2\displaystyle\int_0^v y^i(1-y)dy$ as N goes to infinity. We deduce that $\mathbb{P}(L = i, Y \leq v) = 2\int_0^v y^i(1-y)dy$ for $i \in \mathbb{N}^*$ and $v \in [0,1)$. Thus Y has a beta $(2,1)$ distribution and conditionally on Y, L is geometric with parameter $1 - Y$.

4.3 Proof of Proposition 2.5

The Laplace transform (9) comes from (8). We deduce from (8) that

$$\mathbb{E}[\tau | Y, L = a] = \sum_{k \geq a} \frac{2}{k(k+1)} = \frac{2}{a},$$

and that

$$\begin{aligned}
\mathbb{E}[\tau^2 | Y, L = a] &= 8 \sum_{k \geq a} \frac{1}{k(k+1)} \sum_{\ell \geq k} \frac{1}{\ell(\ell+1)} \\
&= 8 \sum_{k \geq a} \frac{1}{k^2(k+1)} \\
&= 8 \sum_{k \geq a} \frac{1}{k^2} - 8 \sum_{k \geq a} \frac{1}{k(k+1)} \\
&= 8 \sum_{k \geq a} \frac{1}{k^2} - \frac{8}{a}.
\end{aligned}$$

We get (11) from (9) and Proposition 2.4. We get (12) from (11) and Lemma 2.7.

4.4 Proof of Corollary 2.6

We give a direct proof. We set $c_k = k(k+1)$ and $b_k = c_k - 2 = (k-1)(k+2)$. Notice that $c_k + 2\lambda = b_k + 2(1+\lambda)$. We have from (11)

$$
\begin{aligned}
\mathbb{E}[\mathrm{e}^{-\lambda\tau}] &= \int_0^1 2y\,dy \left(\sum_{a\geq 1} (1-y)y^{a-1} \prod_{k\geq a} \frac{c_k}{c_k+2\lambda} \right) \\
&= 2\sum_{a\geq 1} \frac{1}{(a+1)(a+2)} \prod_{k\geq a} \frac{c_k}{c_k+2\lambda} \\
&= 2\sum_{a\geq 1} \frac{1}{b_a+2(\lambda+1)} \prod_{k\geq a+1} \frac{b_k}{b_k+2(1+\lambda)} \\
&= \frac{1}{1+\lambda} \lim_{N\to+\infty} \sum_{a=1}^{N} \left(1 - \frac{b_a}{b_a+2(1+\lambda)}\right) \prod_{k\geq a+1} \frac{b_k}{b_k+2(1+\lambda)} \\
&= \frac{1}{1+\lambda} \lim_{N\to+\infty} \prod_{k\geq N+1} \frac{b_k}{b_k+2(1+\lambda)} \\
&= \frac{1}{1+\lambda},
\end{aligned}
$$

where for the last equality we used that $\lim_{N\to\infty} \prod_{k\geq N+1} \frac{b_k}{b_k+2(1+\lambda)} = 1$.

4.5 Proof of Lemma 2.8

Let us fix $N \geq 2$. We have introduced $L^{(N)}+1$ as the level of the fixation curve G when the fixation curve G_* reaches level $N+1$, that is at time s_{N-1}. We denote by Z_N the number of other fixation curves alive at this time, and $L_1^{(N)} > L_2^{(N)} > \cdots > L_{Z_N}^{(N)} = 2$ their levels. By construction of the fixation curves, the result given by Lemma 2.8 is straightforward for $(V_N/N, Z_N, L^{(N)}, L_1^{(N)}, L_2^{(N)}, \ldots, L_{Z_N}^{(N)})$ instead of $(Y, Z, L, L_1, \ldots, L_Z)$. Now, using similar arguments as for the proof of the second part of Proposition 2.4, we get that $\big((V_N/N, Z_N, L^{(N)}, L_1^{(N)}, L_2^{(N)}, \ldots, L_{Z_N}^{(N)}), N \geq 2\big)$ converges a.s. to $(Y, Z, L, L_1, \ldots, L_Z)$ which ends the proof.

4.6 Proof of Lemma 1.1

Only the second part of this Lemma has to be proved. Assume t is either fixed or a time when a new MRCA is established. The fact that the coalescent times A_t (and thus the TMRCA) does not depend on the coalescent tree shape can be deduced from [33], Section 3, see also [6]. In particular, A_t does not depend on (X_t, Y_t, L_t, Z_t) neither on τ_t which conditionally on the past depends only on the coalescent tree shape (see Section 2.4).

4.7 Proof of Theorem 1.2

The properties i)-vii) are proved at time d_{G_*}, but arguments as in the proof of Theorem 2 in [26] yield that these results also hold at fixed time.

The distribution of Y is given by Corollary 2.3. Properties ii) and vii) are straightforward by construction of X. Proposition 2.4 implies iii). Proposition 2.5 implies iv) and Corollary 2.10 implies v). We deduce vi) from (8), as the exponential random variables are independent of the past before d_{G_*}.

4.8 Proof of Proposition 3.1

Let μ_1 be the beta $(2,1)$ distribution. Using [4], it is enough to prove that μ_1 is the only probability distribution μ on $[0,1)$ such that μ is invariant for Y^μ. Since $x \mapsto \mathbb{E}_x[\tau]$ is bounded (see (15)), we get that $\mathbb{E}_\mu[\tau] < \infty$. For a measure μ and a function f, we set $\langle \mu, f \rangle = \int f \, d\mu$ when this is well defined. As $\mathbb{E}_\mu[\tau] < \infty$, it is straightforward to deduce from standard results on Markov chains having one atom with finite mean return time (see e.g. [23] for discrete time Markov chains) that Y^μ has a unique invariant probability measure π which is defined by $\langle \pi, f \rangle = \mathbb{E}_\mu\left[\int_0^\tau f(Y_s)\, ds\right] / \mathbb{E}_\mu[\tau]$. Hence we have

$$\mathbb{E}_\mu\left[\int_0^\tau f(Y_s) ds\right] = \langle \pi, f \rangle \mathbb{E}_\mu[\tau]. \tag{34}$$

Let τ_n be the n-th resurrection time (i.e. n-th hitting time of 1) after 0 of the resurrected process Y^μ: $\tau_1 = \tau$ and for $n \in \mathbb{N}^*$, $\tau_{n+1} = \inf\{t > \tau_n; Y^\mu_{t-} = 1\}$. The strong law of large numbers implies that for any real measurable bounded function f on $[0,1)$,

$$\mathbb{P}_\mu - a.s. \quad \frac{1}{\tau_n} \int_0^{\tau_n} f(Y_s) ds \xrightarrow[n\to\infty]{} \langle \pi, f \rangle.$$

Recall $\mathscr{L}$ is the infinitesimal generator of Y. For g any C^2 function defined on $[0,1]$, the process $M_t = g(Y_t) - \int_0^t \mathscr{L} g(Y_s)\, ds$ is a martingale. Since $|M_t| \le \|g\|_\infty + t(\|g'\|_\infty + \|g''\|_\infty)$ and $\mathbb{E}_\mu[\tau] < \infty$, we can apply the optional stopping theorem for $(M_t, t \ge 0)$ at time τ to get that

$$g(1) - \mathbb{E}_\mu\left[\int_0^\tau \mathscr{L} g(Y_s)\, ds\right] = \langle \mu, g \rangle \cdot$$

If a C^2 function g_λ is an eigenvector with eigenvalue $-\lambda$ (with $\lambda > 0$) such that $g_\lambda(1) = 0$, we deduce from (34) that $\langle \mu, g_\lambda \rangle = \lambda \mathbb{E}_\mu[\tau] \langle \pi, g_\lambda \rangle$. Therefore, if the resurrection measure is the invariant measure, we get:

$$\langle \mu, g_\lambda \rangle = \lambda \mathbb{E}_\mu[\tau] \langle \mu, g_\lambda \rangle. \tag{35}$$

Let $(a_n^\lambda, n \ge 0)$ be defined by $a_0^\lambda = 1$ and, for $n \ge 0$,

$$a_{n+1}^\lambda = \frac{n(n+1) - 2\lambda}{(n+1)(n+2)} a_n^\lambda.$$

The function $\sum_{n=0}^\infty a_n^\lambda x^n$ solves $\mathscr{L} f = -\lambda f$ on $[0,1)$. For $N \in \mathbb{N}^*$ and $\lambda = \dfrac{N(N+1)}{2}$, notice that $P_N(x) = \sum_{n=0}^\infty a_n^\lambda x^n$ is a polynomial function of degree N. By continuity at 1, P_N is an eigenvector of $\mathscr{L}$ with eigenvalue $-N(N+1)/2$, and such that $P_N(1) = 0$ (as $\mathscr{L} f(1) = 0$ for any C^2 function

f). Notice that $P_1(x) = 1 - x$, which implies that $\langle \mu, P_1 \rangle > 0$. We deduce from (35) that $\mathbb{E}_\mu[\tau] = 1$ and $\langle \mu, P_N \rangle = 0$ for $N \geq 2$. As $P_N(1) = 0$ for all $N \geq 1$, we can write $P_N(x) = (1-x)Q_{N-1}(x)$, where Q_{N-1} is a polynomial function of degree $N-1$. For the probability distribution $\bar{\mu}(dx) = \dfrac{1-x}{\langle \mu, P_1 \rangle}\mu(dx)$, as $\frac{\langle \mu, P_{N+1} \rangle}{\langle \mu, P_1 \rangle} = 0$, we get that:

$$\langle \bar{\mu}, Q_N \rangle = 0, \quad \text{for all } N \geq 1. \tag{36}$$

Since $\bar{\mu}$ is a probability distribution on $[0,1]$, it is characterized by (36). To conclude, we just have to check that $\bar{\mu}_1$ satisfies (36). In fact, we shall check that $\langle \mu_1, g_\lambda \rangle = 0$ for any C^2 function g_λ eigenvector of $\mathscr{L}$ with eigenvalue $-\lambda$ such that $g_\lambda(1) = 0$ and $\lambda \neq 1$. Indeed, we have

$$\begin{aligned}
-\lambda \langle \mu_1, g_\lambda \rangle &= -\lambda \int_0^1 2x g_\lambda(x) dx \\
&= \int_0^1 x^2(1-x) g_\lambda''(x) dx + \int_0^1 2x(1-x) g_\lambda'(x) dx \\
&= \left[x^2(1-x) g_\lambda'(x)\right]_0^1 - \int_0^1 (2x(1-x) - x^2) g_\lambda'(x) dx + \int_0^1 2x(1-x) g_\lambda'(x) dx \\
&= \int_0^1 x^2 g_\lambda'(x) dx \\
&= \left[x^2 g_\lambda(x)\right]_0^1 - \int_0^1 2x g_\lambda(x) dx = -\langle \mu_1, g_\lambda \rangle,
\end{aligned}$$

which implies $\langle \mu_1, g_\lambda \rangle = 0$ unless $\lambda = 1$.

Acknowledgments

The authors would like to thank the referees and the associate editor for their comments and in particular for suggesting Remark 2.2 and also Remark 2.9 which entails the representation for the distribution of Z in Theorem 1.2-v).

References

[1] P. Brémaud, R. Kannurpatti, and R. Mazumdar. Event and time averages: a review. *Adv. in Appl. Probab.*, 24(2):377–411, 1992. MR1167264

[2] C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Advances in Appl. Probability*, 6:260–290, 1974. MR0343949

[3] P. Cattiaux, P. Collet, A. Lambert, S. Martinez, S. Méléard, and J. S. Martin. Quasi-stationarity distributions and diffusion models in population dynamics, *Ann. Probab.*, 37(5):1926–1969, 2009. MR2561437

[4] P. Collet, S. Martínez, and V. Maume-Deschamps. On the existence of conditionally invariant probability measures in dynamical systems. *Nonlinearity*, 13(4):1263–1274, 2000. MR1767958

[5] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probability*, 2:88–100, 1965. MR0179842

[6] P. Donnelly and T. G. Kurtz. The Eve process. Manuscript, personal communication.

[7] P. Donnelly and T. G. Kurtz. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, 24(2):698–742, 1996. MR1404525

[8] P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27(1):166–205, 1999. MR1681126

[9] R. Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer, New York, second edition, 2008. MR2439767

[10] A. M. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2000. MR1779100

[11] S. N. Evans and P. L. Ralph. Dynamics of the time to the most recent common ancestor in a large branching population. *Ann. Appl. Probab.*, 20(1):1–25, 2010. MR2582640

[12] W. J. Ewens. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. Theoretical introduction. MR2026891

[13] P. A. Ferrari, H. Kesten, S. Martinez, and P. Picco. Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Probab.*, 23(2):501–521, 1995. MR1334159

[14] R. Fisher. *The genetical theory of natural selection/RA Fisher.* Clarendon Press, Oxford, 1930.

[15] Y. Fu. Exact coalescent for the Wright–Fisher model. *Theoretical Population Biology*, 69(4):385–394, 2006.

[16] A. Greven, P. Pfaffelhuber, and A. Winter. Tree-valued resampling dynamics: Martingale problems and applications. ArXiv 2009, `http://arxiv.org/pdf/0806.2224v3`. MR2520129

[17] R. C. Griffiths. Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoret. Population Biol.*, 17(1):37–50, 1980. MR0568666

[18] T. Huillet. On Wright-Fisher diffusion and its relatives. *J. Stat. Mech.: Theory and Experiment*, 11, ID P11006, 2007.

[19] N. L. Johnson and S. Kotz. *Urn models and their application*. John Wiley & Sons, New York-London-Sydney, 1977. An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics. MR0488211

[20] M. Kimura and T. Ohta. The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics*, 63(3):701–709, 1969.

[21] M. Kimura and T. Ohta. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61:763–771, 1969.

[22] J. F. C. Kingman. Exchangeability and the evolution of large populations. In *Exchangeability in probability and statistics (Rome, 1981)*, pages 97–112. North-Holland, Amsterdam, 1982. MR0675968

[23] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag, London, 1993. MR1287609

[24] M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, 29(4):1547–1562, 2001. MR1880231

[25] P. A. P. Moran. Random processes in genetics. *Proc. Cambridge Philos. Soc.*, 54:60–71, 1958. MR0127989

[26] P. Pfaffelhuber and A. Wakolbinger. The process of most recent common ancestors in an evolving coalescent. *Stochastic Process. Appl.*, 116(12):1836–1859, 2006. MR2307061

[27] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 1999. MR1742892

[28] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):1116–1125, 1999. MR1742154

[29] J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, 5:Paper no. 12, 50 pp. (electronic), 2000. MR1781024

[30] D. Simon and B. Derrida. Evolution of the most recent common ancestor of a population with no selection. *J. Stat. Mech.: Theory and Experiment*, 5, ID P05002, 27 pp., 2006.

[31] D. Steinsaltz and S. N. Evans. Quasistationary distributions for one-dimensional diffusions with killing. *Trans. Amer. Math. Soc.*, 359(3):1285–1324 (electronic), 2007. MR2262851

[32] F. Tajima. Relationship between DNA polymorphism and fixation time. *Genetics*, 125(2):447–454, 1990.

[33] C. Wiuf and P. Donnelly. Conditional genealogies and the age of a neutral mutant. *Theor. Pop. Biol.*, 56:183–201, 1999.

[34] S. Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.