ON THE TWO OLDEST FAMILIES FOR THE WRIGHT-FISHER PROCESS

JEAN-FRANÇOIS DELMAS, JEAN-STÉPHANE DHERSIN, AND ARNO SIRI-JEGOUSSE

ABSTRACT. We extend some of the results of Pfaffelhuber and Walkobinger on the process of the most recent common ancestors in evolving coalescent (MRCA) by taking into account the size of the oldest family which contains the immortal individual. 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 quasy-stationary distribution of the Wrigh-Fisher process using the process of the relative size of one of the oldest family, which can be seen as a resurrected Wright-Fisher process.

1. INTRODUCTION

Many models have been introduced to describe population dynamics in population genetics. Fisher [14], Wright [33] and Moran [24] 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 [16] 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 [21] (see also [26, 27, 23, 28] for general coalescent processes associated with Cannings' model). One of the associated object of interest is the most recent common ancestor (MRCA) of the population currently alive, which is also the depth of their genealogical tree (see [12, 9]). In the case of Kingman's coalescent, each couple of particle merges at rate one, which gives an MRCA of expectation 2, or an expectation equivalent to $2N$ generations in the discrete case (see [12] for more results on this approximation and [fu2006ecw for exact coalescent for the Wright-Fisher Model). MRCAs have been studied for many other models (see e.g. [4] for a more relevant model for human population, where the position MRCA of a population of N individuals is almost $\log_2 N$ generations ago). In the special case of Fleming-Viot processes, which genealogies at a fixed time are given by Kingman's coalescent, [15] have introduced a tree-valued process as solution of a martingale problem that represents the evolving genealogies.

In Moran model (finite population size) and in Wright-Fisher model with infinite population size, only two lineages can merge at a time and the genealogy at a given time is given by Kingman's coalescent. At time t 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 this two oldest families will disappear (in the future). Let Y_t be the relative size of the oldest family which will fixate: either $Y_t = X_t$ or $Y_t = 1 - X_t$. In a sense Y_t is the size of the oldest family to which belongs the immortal line of descent (or the immortal individual). Notice that X_t can be estimated (for example using

DNA analysis of neutral mutation) at time t whereas Y_t is not observable at time t. When X_t hits 0 or 1, that is when Y_t hits 1, one of the two oldest family disappear and there is a change of MRCA. At this time two new oldest families appear. This corresponds to a jump of the processes $\mathbf{X} = (X_t, t \in \mathbb{R})$ and $\mathbf{Y} = (Y_t, t \in \mathbb{R})$.

For the Wright-Fisher model with infinite population size, 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. The two absorbing states 0 and 1 are reached in finite time. In between two jumps the process $\mathbf{Y}$ is a WF diffusion on $[0, 1]$ conditioned not to reach 0: $dY_t = \sqrt{Y_t(1 - Y_t)}dB_t + (1 - Y_t)dt$. The WF diffusion and its conditioned version have been largely used to model allelic frequencies in a neutral two-types population, see [12, 17, 9]. Using the look-down representation for the genealogy introduced by Donnelly and Kurtz [7, 8], we prove rigorously, see Corollary 2.2, that at a jump time the law of $\mathbf{X}$, μ_0, is the uniform distribution on $[0, 1]$, and that the law of $\mathbf{Y}$, μ_1, is the size-biased distribution of μ_0, that is the beta $(2, 1)$ distribution. The process $\mathbf{X}$ can be seen as a resurrected WF diffusion with resurrection distribution μ_0. It is then easy to check that μ_0 is also the invariant measure of $\mathbf{X}$. Indeed, according to Lemma 2.1 of [5] (see also the pioneer work of [13] in a discrete setting), μ is a quasi-stationary measure of a process killed when it reaches a set Δ if and only if μ is the stationary measure of the corresponding resurrected process which jumps with resurrection measure μ when it reaches the set Δ. See Section 3.1 for a precise statement. Then the conclusion follows as μ_0 (resp. μ_1) is the Yaglom limit of the (resp. conditioned) WF diffusion, see [12, 17] and also [3]. In Section 3, we check that the measure μ_0 (resp. μ_1) is the only quasi-stationary measure for the (resp. conditioned) WF diffusion. A similar result is also true for the Moran model. In this case also, the quasi-stationary measure can be seen as the distribution of the size of one of the two oldest family. There is no such interpretation for the WF model for finite population, see Remark 3.2.

To establish Corollary 2.2, we use the look-down process, which gives a representation of the genealogy for the WF model of a population with infinite size. Following Pfaffelhuber and Walkobinger [25], we are interested in the distribution of the following quantities:

- A: the birth time of MRCA for the current population
- $\tau \geq 0$: the time to wait before a change of MRCA happens (the hitting time of $\{0, 1\}$ for $\mathbf{X}$).
- $L' \in \mathbb{N}^*$: the number of living individuals which will have descendants at time τ.
- $Z \in \{0, \ldots, L'\}$: the number of living individuals which will become MRCA in the future.

Recent papers give an exhaustive study of birth dates and death times of MRCA, see [25] and also [29] (see also [11] for genealogies of continuous state branching processes). In particular the birth dates of MRCA, as well as the death times of MRCA, are distributed according to a Poisson process, see [25].

The distribution of (τ, L', Z) is given in [25]. In particular, τ is an exponential random variable with mean 1. We give, see Corollary 1.1 below, the joint distribution of (A, τ, L', Z) at time t conditionally on Y_t or X_t, where t is either fixed or an MRCA death time. The study of this conditional distribution is also motivated by the fact that the relative size of the current two oldest families, X_t, can be inferred from available DNA data at time t. By stationarity, for fixed t, this distribution does not depend on t. It is also the same, but for A, at the death time of an MRCA (the argument is the same as in the proof of Theorem 2 in [25]). This property is the analogue of the so-called PASTA (Poisson Arrivals See Time Average) property in queuing theory, see [1] for a review on this subject.

Let $(E_k, k \in \mathbb{N}^*)$ be independent exponential random variables with mean 1. We denote by $T_K = \sum_{k \geq 1} \frac{2}{k(k+1)} E_k$ the lifetime of a Kingman coalescent process, and $T_T = \sum_{k \geq 2} \frac{2}{k(k+1)} E_k$.

Corollary 1.1. *At a fixed time t or at the death time of an MRCA, we have:*

(1) *A is independent of (Y, X, τ, L', Z), and is distributed as T_K at a fixed time and as T_T at the death time of an MRCA.*

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

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

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

(5) *Conditionally on Y, L' is geometric with parameter $1 - Y$.*

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

(7) *For $u \in [0, 1]$, and $a \geq 1$,*

$$\mathbb{E}[u^Z | Y, L' = a] = \begin{cases} 1 & \text{if} \quad a = 1 \\ \dfrac{u}{3} \dfrac{a+1}{a-1} \prod_{k=2}^{a-1} \left(1 + \dfrac{2u}{(k-1)(k+2)}\right) & \text{if} \quad a \geq 2 \end{cases}$$

We also give the first two moments of Z and τ in Section 2. The proof of 1) is a direct consequence of Kingman's coalescent (for fixed t) or of [31] (for the death time of MRCA) and the fact that the coalescent times (and thus the birth time of the MRCA A) does not depend on the coalescent tree shape. This last property can be deduced from [32], Section 3. In particular, A does not depend on (X, Y, L', Z) neither on τ which conditionally on the past depends only on the coalescent tree shape (see Section 2.4).

We deduce 2-7 from Section 2, where $L' = L - 1$. Our results also give a detailed proof of the heuristic arguments of Remarks 3.2 and 7.3 in [25]. From the conditional distribution of τ, we recover formula from Kimura and Otha [19, 20] of its conditional expectation (11) and conditional variance (12). Notice the Laplace transform of τ conditionally on $X_t = x$: $f(x) = \mathbb{E}[e^{-\lambda \tau} | X_t = x]$ solves the ODE: $Lf = \lambda f$, $f(0) = f(1) = 1$, where L is the generator of the WF diffusion: $Lh(x) = x(1-x)h''(x)$ in $(0, 1)$. Remark that 6) implies that τ is an exponential random variable with mean 1.

We shall end by a proposition on the connection between Z and τ.

Proposition 1.2. *We have for all $\lambda \geq 0$:*
$$\mathbb{E}[e^{-\lambda T_K}] \mathbb{E}[(1 + \lambda)^Z | Y, L'] = \mathbb{E}[e^{-\lambda \tau} | Y, L'].$$

In particular, we also have

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

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

(2) $$\mathbb{E}[Z | Y, L'] = 2 - \mathbb{E}[\tau | Y, L'],$$

(3) $$\mathbb{E}[Z^2 | Y, L'] = \mathbb{E}[\tau^2 | Y, L'] - 5 \mathbb{E}[\tau | Y, L'] + 18 - \frac{4\pi^2}{3}.$$

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

Remark 1.3. At the MRCA death time, we have a new MRCA which is born at A in the past and will die at τ in the future. On one hand, by looking at the death time of this new MRCA,

Kingman's coalescent theory implies that $A + \tau$ is distributed as T_K. As the coalescent times are independent of the structure of the coalescent tree, we get that A and τ are independent. On the other hand, there are Z living future MRCA and one new MRCA. We deduce that the new MRCA is the $Z+1$-th point in the past of the birth dates of MRCA process. This latter is a Poisson point process with intensity 1, which is a direct consequence of the look-down representation of the genealogy. Intuitively, we could think that the new MRCA date of birth, A, is distributed as the sum of $Z + 1$ independent exponentials of parameter 1. This result is false, as one can easily check by computing Laplace transform; this is because the Poisson point process of the MRCA births is not independent of Z. However, this result is partially true at least for the conditional expectation thanks to (2). The link between the distribution of T_K and the joint distribution of τ and Z (which are independent conditionally on (Y, L)) is given by equation (1).

.

The results are presented in Sections 2 and 3, and 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 [7] and the modified look-down process [8] have been introduced by Donnelly an Kurtz 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 for the 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 particle 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 particle having level 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).
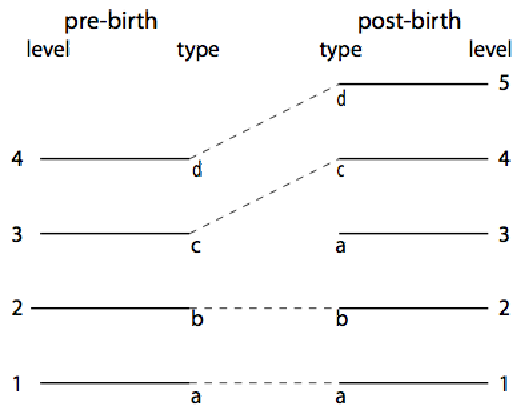

FIGURE 1. A look-down event between levels 1 and 3. Each individual living at level at least 3 before birth is pushed one level up after the birth.

For any fixed time t_0, we can introduce the following family of equivalence relations $\mathcal{R}^{(t_0)} = (\mathcal{R}_s^{(t_0)}, s \geq 0)$: $i\mathcal{R}_s^{(t_0)}j$ if the two individuals i and j living at time t_0 have a common ancestor

at time $t_0 - s$. It is then easy to show that the coalescent process on $\mathbb{N}^*$ defined by $\mathcal{R}^{(t_0)}$ is the Kingman's coalescent. See Figure 2 for a graphical representation.
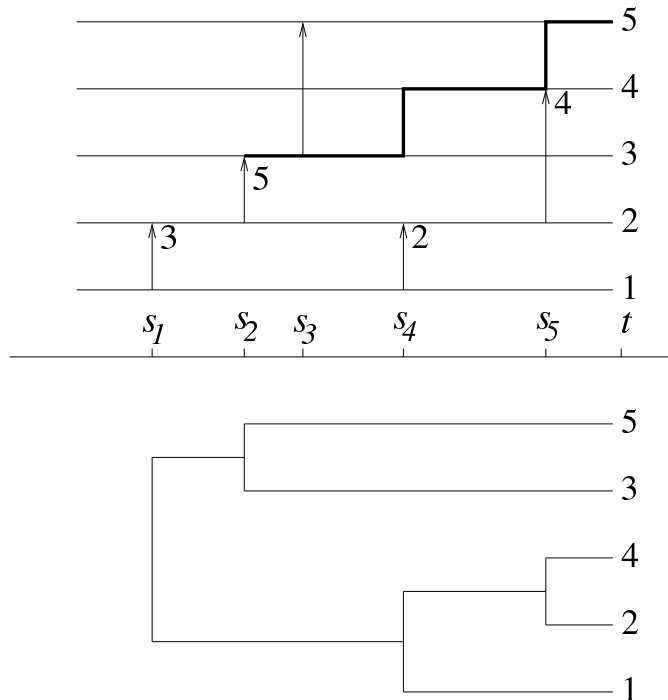

FIGURE 2. The look down process and its associated coalescent tree, started at time t for the 5 first levels. At each look-down event, a new curve is born. We indicate at which level this curve is at time t. In bold the curve of the individual who is at level 5 at time t.

We can describe the path of an individual born at level $j \geq 2$ at time s_0 as a curve in E

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

where for $k \in \mathbb{N}^*$, s_k is the first birth time after s_{k-1} of an individual with level less than $j + k + 1$. In particular G describes the different levels occupied by the individual born at time s_0. In fact we shall identify an individual with its curve. We shall write $b_G = s_0$ for the birth time of individual G. We say that $d_G = \lim_{k \to \infty} s_k$ is the death time of this individual. We say an individual or a curve is alive at time t if $b_G \leq t < d_G$, and k is the level of G at time t if $(t, k) \in G$. The set of all the curves, $\mathcal{G}$, is a partition of $E^* = \mathbb{R} \times \{2, \ldots\}$. We write $\mathcal{G}_t$ for the set of all curves alive at time t. Notice that the individual at level 1 is immortal and that by definition, its curve $\mathbb{R} \times \{1\}$ is not in $\mathcal{G}$. An individual at level j is pushed at rate $\binom{j}{2}$ at level $j+1$ (since there are $\binom{j}{2}$ possible independent look-down events which arrive at rate 1 and which push an individual living at level j). Since $\sum_{j \geq 2} 1/\binom{j}{2} < \infty$, we get that any individual but the one at level 1 dies in finite time.

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

For a fixed time t, let G_t be the living MRCA of the whole population living at time t. Notice the birth of the MRCA is $A_t = \inf\{b_G; G \in \mathcal{G}_t\} = b_{G_t}$. It corresponds to the birth time of the highest fixation curve living at time t. Let $Z_t + 1$ denote the number of fixation curves living at time t: $Z_t \geq 0$ is the number of future MRCA living at time t. 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. The joint distribution of $(Z_t, L_0(t), L_1(t), \ldots, L_{Z_t}(t))$ is given in Theorem 2 of [25], and the distribution of Z_t, the number of future MRCA alive at fixed time t, is given in Theorem 3 of [25]. We consider the partition of the population into the two oldest family given by the equivalence relation $\mathcal{R}^{(t)}_{t-A_t}$. This corresponds to the partition of individuals alive at time t whose ancestor is either G_t or the immortal individual. We shall denote by Y_t the relative proportion of the sub-population whose ancestor at time A_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.

By stationarity, we have that the distribution of $H_t = (X_t, Y_t, Z_t, L_0(t), L_1(t), \ldots, L_{Z_t}(t))$ does not depend on t. In between two MRCA deaths, the process $(X_t, t \in \mathbb{R})$ is a Wright-Fisher diffusion with generator $\frac{1}{2} x(1-x)\partial_x^2$ and the process $(Y_t, t \in \mathbb{R})$ is a Wright-Fisher diffusion conditioned to die at 1 with generator $\frac{1}{2} x(1-x)\partial_x^2 + (1-x)\partial_x^1$, see [9, 17]. Notice the distribution of Z_t conditionally on X_t is of interest, as the relative proportion of the two oldest families at time t, X_t, can be well estimated by DNA analysis if the (neutral) mutation rate is strong enough.

We are interested in the law of H_t at (random) times where the MRCA changes, as well as the distribution of the labels of the individuals of the same oldest family. The distributions of H_t is the same if we consider a fixed time t or this random time (the argument is the same as in the proof of [25, Theorem 2]). 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 where the MRCA changes. For this reason, we shall omit the subscript and write H, and carry the proofs at the death time of an MRCA. Our results will give a detailed proof of the heuristic arguments of Remarks 3.2 and 7.3 in [25].

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 of death of an MRCA. More precisely, let G_* be a fixation curve and G be the next fixation curve: the individual G is the next MRCA after the MRCA G_*. 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 of the MRCA G. Let $N \geq 2$. Notice that at time s_{N-1}, only the individuals with level 1 to N will survive up to 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 $\mathcal{R}^{(s_{N-1})}_{s_{N-1}-s_0}$ which corresponds to the partition of individuals alive at time s_{N-1} with labels 1 to N whose ancestor is either G or the immortal individual. Consider the ancestor at time s_0 of the individual at level $k \in \{1, \ldots, N\}$ and time s_{N-1}, and let $\sigma_N(k) = 1$ if it is the immortal individual and $\sigma_N(k) = 0$ if it is G. 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_0 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 at the death time of the MRCA G_*. By construction the process $(\sigma_N, N \in \mathbb{N}^*)$ is Markov.
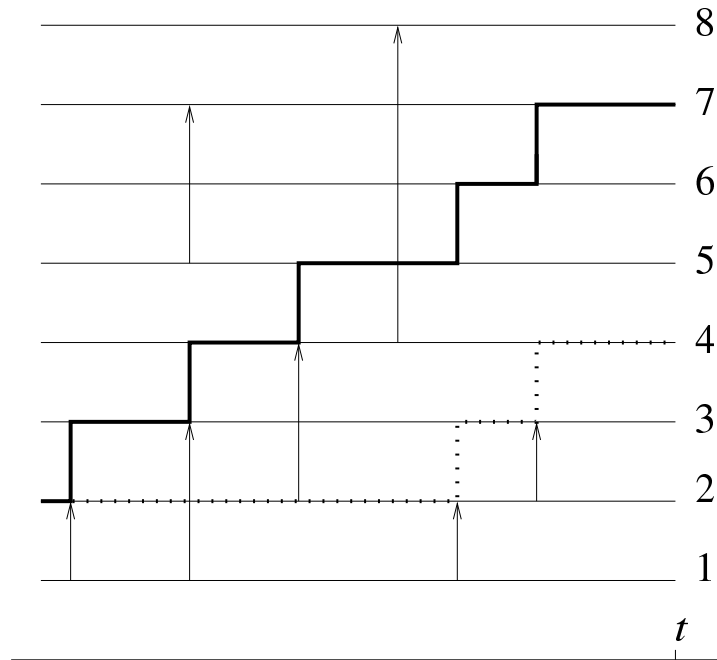

FIGURE 3. In this example, at time s_5, we have $\sigma_6 = (1, 1, 1, 0, 1, 0)$ and $V_6 = 4$.

In order to give the law of (V_N, σ_N) we first recall some facts on Pólya's urns, see [18]. 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 $\ell \in \{0, \ldots, N\}$

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

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

$$(4) \qquad \qquad \mathbb{P}(S_N^{(2,1)} = k + 1) = \frac{2k}{(N + 2)(N + 1)}.$$

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

(1) *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 $\{1, \ldots, N - 1\}$, i.e.*

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

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

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 [18, Section 6.3.3], 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 2.2. *When the MRCA changes, the relative proportion Y of the new oldest family which contains the immortal individual 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 [25]. 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 $L^{(N)}$ 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 $\{2, \ldots, V_N + 1\}$. 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 (6) which was already given by [25, Lemma 7.1]. Note that the process is an inhomogeneous Markov chain [25, Lemma 6.1].

Proposition 2.3. *Let $N \geq 2$.*

(1) *For $k \in \{1, \ldots, N-1\}$ and $i \in \{2, \ldots, k+1\}$,*

(5)
$$\mathbb{P}(L^{(N)} = i, V_N = k) = 2\frac{(N-i)!}{N!}\frac{k!}{(k-i+1)!}\frac{N-k}{N-1},$$

and for all $i \in \{2, \ldots, N\}$,

(6)
$$\mathbb{P}(L^{(N)} = i) = \frac{N+1}{N-1}\frac{2}{(i+1)i}.$$

(2) *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, $L-1$ is geometric with parameter $1 - Y$.*

The level L corresponds to the level of the MRCA, just after a change of MRCA. Recall $L_1(t)$ is the level at time t of the second fixation curve. We use the convention that $L_1(t) = 1$ if there is only one fixation curve i.e. $Z(t) = 0$. At the random time d_{G_*} of the death of the fixation curve G_*, we have $L_1(d_{G_*}) = L$. And a straightforward computation gives that for $i \in \mathbb{N}^*$

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

At a fixed time t, by stationarity, the distribution of $L_1(t)$ does not depend on t, and equation (3.4) from [25] gives that $L_1(t)$ is distributed as $L-1$ (and that L is distributed as $L_1(t)$ conditionally on $\{L_1(t) \geq 2\}$). This is already pointed out in [25]. In view of Remark 4.1 in [25], notice the result is also similar for $M/M/k$ queue where the invariant measure for the queue process and the queue process at arrivals time are the same (up to translation of 1 because of our convention). This is known as the PASTA property.

2.4. **Next fixation time.** We consider the time d_{G_*} of death of the MRCA. 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. τ 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

$$(7) \qquad \tau = \sum_{k=L}^{\infty} \frac{2}{k(k-1)} E_k$$

where E_k are independent exponential random variables with parameter 1.

Proposition 2.4. *Let $a \geq 1$. The distribution of the waiting time for the next fixation time is given by: For $\lambda \in \mathbb{R}_+$,*

$$(8) \qquad \mathbb{E}[e^{-\lambda \tau} | Y, L = a + 1] = \prod_{k=a}^{\infty} \left(\frac{k(k+1)}{k(k+1) + 2\lambda} \right).$$

Its first two moments are given by:

$$(9) \qquad \mathbb{E}[\tau | Y, L = a + 1] = \frac{2}{a} \quad and \quad \mathbb{E}[\tau^2 | Y, L = a + 1] = 8 \left[\sum_{k \geq a} \frac{1}{k^2} - \frac{1}{a} \right].$$

We also have: For $y \in (0, 1)$,

$$(10) \qquad \mathbb{E}[e^{-\lambda \tau} | Y = y] = (1 - y) \sum_{l=2}^{\infty} y^{l-2} \prod_{k=l-1}^{\infty} \left(\frac{k(k+1)}{k(k+1) + 2\lambda} \right).$$

Notice that using (10) we immediately get that

$$\mathbb{E}[e^{-\lambda \tau}] = \frac{1}{1 + \lambda},$$

and recover that τ is exponential with mean 1.

Using the second part of the proposition, we recover the well known results from Kimura and Ohta [19, 20] (see also [12]):

$$(11) \qquad \mathbb{E}[\tau | Y = y] = -2 \frac{(1 - y) \log(1 - y)}{y}, \quad and$$

$$(12) \qquad \mathbb{E}[\tau^2 | Y = y] = 8 \left(\frac{(1 - y) \log(1 - y)}{y} - \int_y^1 \frac{\log(1 - z)}{z} dz \right),$$

and

$$(13) \qquad \mathbb{E}[\tau | X = x] = -2 \left(x \log(x) + (1 - x) \log(1 - x) \right), \quad and$$

$$(14) \qquad \mathbb{E}[\tau^2 | X = x] = 8 \left(x \log(x) + (1 - x) \log(1 - x) - x \int_x^1 \frac{\log(1 - z)}{z} dz \right.$$
$$\left. - (1 - x) \int_{1-x}^1 \frac{\log(1 - z)}{z} dz \right).$$

We deduce from (9) that $\mathbb{E}[\tau | L = a + 1] = \frac{2}{a}$, which was already in Theorem 1 in [25].

2.5. **Number of future MRCA already living.** We keep notations from Sections 2.1 and 2.3. We set $Z = Z_{d_{G_*}}$ the number of future MRCA living at time d_{G_*} of death of the MRCA G_*. Let $(L, L_1, ..., L_Z) = (L_0(d_{G_*}), ..., L_Z(d_{G_*}))$ be the levels of the fixation curves at the death time of G_*. Recall notations from Section 2.2. The following Lemma and Proposition 2.3 characterize the joint distribution of $(Y, Z, L, L_1, ..., L_Z)$.

Lemma 2.5. *Conditionally on (L, Y) the distribution of $(Z, L_1, ..., L_Z)$ does not depend on Y. Conditionally on $\{L = N + 1\}$, $(Z, L_1, ..., L_Z)$ is distributed as follows:*

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

We are now able to give the distribution of Z conditionally on Y or X.

Proposition 2.6. *Let $a \geq 1$. We have $\mathbb{P}(Z = 0|L = 2) = 1$ and for $k \geq 1$,*

$$(15) \qquad \mathbb{P}(Z = k|Y, L = a + 1) = \frac{2^{k-1}}{3} \frac{a+1}{a-1} \sum_{1 < a_k < \cdots < a_2 < a} \prod_{i=2}^{k} \frac{1}{(a_i - 1)(a_i + 2)};$$

for all $u \in [0, 1]$,

$$(16) \qquad \mathbb{E}[u^Z|Y, L = a + 1] = \begin{cases} 1 & if \quad a = 1 \\ \dfrac{u}{3} \dfrac{a+1}{a-1} \prod_{k=2}^{a-1} \left(1 + \dfrac{2u}{(k-1)(k+2)}\right) & if \quad a \geq 2 \end{cases}$$

with the convention that $\prod_\emptyset = 1$. We also have

$$(17) \quad \mathbb{E}[Z|Y, L = a + 1] = 2 - \frac{2}{a} \quad and \quad \mathbb{E}[Z^2|Y, L = a + 1] = 18 - \frac{\pi^2}{3} - \frac{18}{a} + 8 \sum_{k \geq a} \frac{1}{k^2}.$$

Corollary 2.7. *Let $y \in [0, 1]$. We have $\mathbb{P}(Z = 0|Y = y) = 1 - y$, and: for all $k \in \mathbb{N}^*$,*

$$(18) \quad \mathbb{P}(Z = k|Y = y) = \frac{2^{k-1}}{3}(1 - y) \sum_{1 < a_k < \cdots < a_1 < \infty} (a_1 + 1)(a_1 + 2)y^{a_1 - 1} \prod_{i=1}^{k} \frac{1}{(a_i - 1)(a_i + 2)};$$

for all $u \in [0, 1]$

$$(19) \qquad \mathbb{E}[u^Z|Y = y] = (1 - y) + u\frac{1 - y}{3} \sum_{a=2}^{\infty} \frac{a+1}{a-1} y^{a-1} \prod_{\ell=2}^{a-1} \left(1 + \frac{2u}{(\ell - 1)(\ell + 2)}\right),$$

with the convention that $\prod_\emptyset = 1$. We also have

$$(20) \qquad \qquad \mathbb{E}[Z|Y = y] = 2\left(1 + \frac{1 - y}{y} \log(1 - y)\right).$$

We then deduce the following result.

Corollary 2.8. *Let $x \in [0, 1]$. We have $\mathbb{P}(Z = 0|X = x) = 2x(1 - x)$, and: for all $k \in \mathbb{N}^*$,*

$$(21) \quad \mathbb{P}(Z = k|X = x)$$
$$= \frac{2^{k-1}}{3}x(1 - x) \sum_{1 < a_k < \cdots < a_1 < \infty} (a_1 + 1)(a_1 + 2)(x^{a_1 - 2} + (1 - x)^{a_1 - 2}) \prod_{i=1}^{k} \frac{1}{(a_i - 1)(a_i + 2)};$$

for all $u \in [0, 1]$,

(22)

$$\mathbb{E}[u^Z | X = x] = 2x(1-x) + u\frac{x(1-x)}{3} \sum_{a=2}^{\infty} \frac{a+1}{a-1}(x^{a-1} + (1-x)^{a-1}) \prod_{\ell=2}^{a-1} \left(1 + \frac{2u}{(\ell-1)(\ell+2)}\right),$$

with the convention that $\prod_{\emptyset} = 1$. We also have

(23)
$$\mathbb{E}[Z | X = x] = 2\left(1 + x\log(x) + (1-x)\log(1-x)\right).$$

The second moment of Z conditionally on Y (resp. X) can be deduced from (19) (resp. (22)) or from (3) and (12) (resp. (14)).

Some elementary computations give:

$$\mathbb{P}(Z = 0 | X = x) = 2x(1-x),$$

$$\mathbb{P}(Z = 1 | X = x) = \frac{1}{3}\left[x^2 + (1-x)^2 - 2x(1-x)\ln(x(1-x))\right],$$

$$\mathbb{P}(Z = 2 | X = x) = \frac{2}{3}\left[\frac{11}{6}(x^2 + (1-x)^2) - (1-x)\ln(1-x) - x\ln(x)\right]$$

$$+ \frac{2}{3}x(1-x)\left[2 - \frac{\pi^2}{3} + 2\ln(x)\ln(1-x) - \frac{1}{3}\ln(x(1-x))\right].$$

We recover by integration of the previous equations the following results from [25], that

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

3. STATIONARY DISTRIBUTION OF THE RELATIVE SIZE FOR THE TWO OLDEST FAMILIES

3.1. **Resurrected process and quasi-stationary measure.** 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 of U if

$$\mathbb{P}_\nu(U_t \in A | U_t \notin \Delta) = \nu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}),$$

where we write $\mathbb{P}_\nu$ to say that the distribution of U_0 is ν. See also [30] for quasi-stationary distributions for diffusions with killing.

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

(1) U_0 is distributed according to ν and $U_t^\mu = U_t$ for $t \in [0, T_0[$, where $T_0 = \inf\{s \geq 0; U_s \in \Delta\}$.
(2) Conditionally on $(T_0, \{T_0 < \infty\}, (U_t^\mu, t \in [0, T_0)))$, $(U_{t+T_0}^\mu, t \geq 0)$ is distributed as U^μ under $\mathbb{P}_\mu$.

According to Lemma 2.1 of [5], the distribution μ is a quasi-stationary measure of U if and only if μ is a stationary measure of U^μ. See also the pioneer work of [13] in a discrete setting.

The uniqueness of quasy-stationary distributions is an open question in general. We will give a genealogical representation of the quasi-stationary distribution 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

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

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

3.2. **The resurrected Wright-Fisher diffusion.** From Corollary 2.2 and comments below it, we get that the relative proportion of one of the two oldest families at a change of MRCA is distributed according to the uniform distribution over $[0, 1]$. Then the relative proportion evolves according to a Wright-Fisher (WF) diffusion with generator $\frac{1}{2}x(1-x)\partial_x^2$. In particular it hits the absorbing state of the WF diffusion, $\{0, 1\}$, in finite time. At this time one of the two oldest family dies out and there is (again) a change of MRCA.

The Yaglom distribution of the WF diffusion exists and is the uniform distribution, see [12, p. 161], or [17] for an explicit computation. From Section 3.1, we get that in stationary regime, for fixed t (and of course at time when the MRCA changes) the relative size of one of the two oldest families taken at random, X_t, 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 quasy-stationary distribution of the WF diffusion. Lemma 2.1 in [5] implies there is no other resurrection measure which is also the stationary measure of the resurrected process.

3.3. **The oldest family with the immortal individual.** Let $Y = (Y_t, t \in \mathbb{R})$ be the process of relative size for the oldest family containing the immortal individual. From Corollary 2.2, we get that Y at a change of MRCA 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 $\frac{1}{2}x(1-x)\partial_x^2 + (1-x)\partial_x^1$, see [9, 17]. Therefore Y is a resurrected Wright-Fisher diffusion conditioned not to hit 0, with beta $(2, 1)$ resurrection measure.

The Yaglom distribution of the Wright-Fisher diffusion conditioned not to hit 0 exists and is the beta $(2, 1)$ distribution, see [17] for an explicit computation. In fact the Yaglom distribution is the only quasy-stationary distribution according to the next proposition.

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

Lemma 2.1 in [5] 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 measure μ has stationary measure 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 [24]. 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 quasy-stationary distribution.

Let μ be a distribution on $\{1, \ldots, N-1\}$. We consider the resurrected process $(\xi_t^\mu, t \geq 0)$ with resurrection measure 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:

$$\Lambda^\mu(1,k) = \left(\mu(k) + \mathbf{1}_{\{k=2\}}\right)\frac{N-1}{N} \quad \text{for } k \in \{2,\ldots,N-1\},$$

$$\Lambda^\mu(k,k+\varepsilon) = \frac{k(N-k)}{N} \quad \text{for } \varepsilon \in \{-1,1\} \text{ and } k \in \{2,\ldots,N-2\},$$

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

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 quasi-stationary distribution 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 measure 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 $\{1,\ldots,N-1\}$ when there is a change of MRCA. The PASTA property also implies that this is the stationary distribution. If, when there is a change of MRCA, 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 measure, 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 measure, μ_N (which is not the uniform distribution), see [6]. We deduce that the resurrected process ζ^μ has stationary measure μ if and only if $\mu = \mu_N$. Notice, that in this example there is no interpretation of μ_N as the size of one of the oldest family at a change of MRCA.

4. Proofs

4.1. **Proof of Theorem 2.1.** We denote by

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

so that $\mathbb{P}(V_1 = k_1,\ldots,V_N = k_N) > 0$ if and only if $(k_1,\ldots,k_N) \in A_N$. Remark that, using (4), it is enough to show that $(V_N, N \geq 2)$ is a Markov process, and that its transition probabilities are given by
(24)

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

for $N \geq 2$ and $(k_1,\ldots,k_{N+1}) \in A_{N+1}$ to prove the first part of Theorem 2.1.

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

$$\Delta_{p,q} = \{a = (a_1,\ldots,a_p) \in \{0,1\}^p, a_1 = 1, \#\{i, a_i = 1\} = q\}.$$

Notice that $\mathrm{card}(\Delta_{p,q}) = \binom{p-1}{q-1}$. Hence the second part of Theorem 2.1 can be rewritten as follows: for all $(k_1, \ldots, k_N) \in A_N$, and all $a \in \Delta_{N,k_N}$,

$$\mathbb{P}(\sigma_N = a | V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1) = \frac{1}{\binom{N-1}{k_N-1}}.$$

We proceed by induction on N for the proof of Theorem 2.1. The result is obvious for $N = 2$. We suppose it is true for a fixed N. We denote by I_N and J_N, $I_N < J_N$, the two levels involved for the jump at time s_N. Notice that (I_N, J_N) and σ_N are independent. This pair is chosen uniformly so that

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

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

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

For $a = (a(1), \ldots, a(N+1)) \in \{0,1\}^{N+1}$ and $j \in \{1, \ldots, N+1\}$, we set $a_X^j = (a(1), \ldots, a(j-1), a(j+1), \ldots, a(N+1)) \in \{0,1\}^N$.

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

$$\mathbb{P}(\sigma_{N+1} = a | V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1).$$

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

$$\mathbb{P}(\sigma_{N+1} = a | V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{1 \le i < j \le N+1} \mathbb{P}(I_N = i, J_N = j, \sigma_{N+1} = a | V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{1 \le i < j \le N+1, a(i)=a(j)=1} \mathbb{P}(I_N = i, J_N = j, \sigma_N = a_X^j | V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{1 \le i < j \le N+1, a(i)=a(j)=1} \mathbb{P}(I_N = i, J_N = j) \mathbb{P}(\sigma_N = a_X^j | V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{1 \le i < j \le N+1, a(i)=a(j)=1} \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N-1}}$$

$$= \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N-1}} \frac{k_{N+1}(k_{N+1} - 1)}{2}$$

$$(25)$$

$$= \frac{k_{N+1}!(N - k_{N+1} + 1)!}{(N+1)!}.$$

We have 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,

$$\mathbb{P}(V_{N+1} = k_{N+1}|V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{a \in \Delta_{N+1,k_{N+1}}} \mathbb{P}(\sigma_{N+1} = a|V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \binom{N}{k_{N+1} - 1} \frac{k_{N+1}!(N - k_{N+1} + 1)!}{(N+1)!}$$

(26)
$$= \frac{k_N + 1}{N + 1}.$$

Moreover,

$$\mathbb{P}(\sigma_{N+1} = a|V_{N+1} = k_{N+1}, V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

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

$$= \frac{1}{\binom{N}{k_{N+1}-1}},$$

which proves that, conditionally on $\{V_{N+1} = k_{N+1}, V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1\}$, σ_{N+1} is uniformly distributed on $\Delta_{N+1,k_{N+1}}$.
2nd case: $k_{N+1} = k_N$**.** Similarly, we get that

$$\mathbb{P}(\sigma_{N+1} = a|V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{1 \le i < j \le N+1} \mathbb{P}(I_N = i, J_N = j, \sigma_{N+1} = a|V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{1 \le i < j \le N+1, a(i)=a(j)=1} \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N-1}}$$

$$= \frac{2}{(N+1)N} \frac{1}{\binom{N-1}{k_N-1}} \frac{(N+1-k_N)(N-k_N)}{2}$$

$$= \frac{(N-k_N)(k_N-1)!(N-k_N+1)!}{(N+1)!}.$$

Hence,

$$\mathbb{P}(V_{N+1} = k_{N+1}|V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \sum_{a \in \Delta_{N+1,k_{N+1}}} \mathbb{P}(\sigma_{N+1} = a|V_N = k_N, V_{N-1} = k_{N-1}, \ldots, V_1 = k_1)$$

$$= \binom{N}{N - k_{N+1} + 1} \frac{(N-k_N)(k_N-1)!(N-k_N+1)!}{(N+1)!}$$

$$= 1 - \frac{k_N + 1}{N + 1}.$$

Together with (26), this implies (24). Moreover,

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

which proves that, conditionally on $\{V_{N+1} = k_{N+1}, V_N = k_N, V_{N-1} = k_{N-1}, \dots, V_1 = k_1\}$, σ_{N+1} is uniformly distributed on $\Delta_{N+1, k_{N+1}}$.

4.2. **Proof of Proposition 2.3.** Theorem 2.1 shows that the distribution of σ_N conditionally on V_N is uniform. Then, if $V_N = k$, we can see $L^{(N)} - 1$ 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 $i \in \{2, \dots, k+1\}$,

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

This gives (5).

To get $\mathbb{P}(L^{(N)} = i)$, we sum over k in (5):

$$\begin{aligned}\mathbb{P}(L^{(N)} = i) &= \frac{2(N-i)!}{N!(N-1)} \sum_{k=i-1}^{N-1} \frac{k!}{(k-i+1)!} (N-k) \\ &= \frac{2(N-i)!}{N!(N-1)} \sum_{k=0}^{N-i} \frac{(k+i-1)!}{k!} (N-i-k+1) \\ &= \frac{2(N-i)!}{N!(N-1)} \sum_{k=0}^{N-i} \frac{(k+i-1)!}{k!} \sum_{l=k}^{N-i} 1 \\ &= \frac{2(N-i)!}{N!(N-1)} \sum_{l=0}^{N-i} \sum_{k=0}^{l} \frac{(k+i-1)!}{k!}.\end{aligned}$$

Let us prove by induction that

$$(27) \qquad\qquad \forall j \geq 0, \quad \sum_{k=0}^{j} \frac{(k+i-1)!}{k!} = \frac{(i+j)!}{j! i}.$$

It is clear for $j = 0$. We suppose it is true for a fixed j.

$$\begin{aligned}\sum_{k=0}^{j+1} \frac{(k+i-1)!}{k!} &= \frac{(i+j)!}{j! i} + \frac{(i+j)!}{(j+1)!} \\ &= \frac{(i+j)!(j+1) + (i+j)! i}{(j+1)! i} \\ &= \frac{(i+j+1)!}{(j+1)! i},\end{aligned}$$

which achieves to prove (27). Hence,

$$
\begin{aligned}
\mathbb{P}(L^{(N)} = i) &= \frac{2(N-i)!}{N!(N-1)} \sum_{l=0}^{N-i} \sum_{k=0}^{l} \frac{(k+i-1)!}{k!} \\
&= \frac{2(N-i)!}{N!(N-1)} \sum_{l=0}^{N-i} \frac{(i+l)!}{l!i} \\
&= \frac{2(N-i)!}{N!(N-1)} \frac{(N+1)!}{(N-i)!(i+1)i} \\
&= \frac{N+1}{N-1} \frac{2}{(i+1)i},
\end{aligned}
$$

where we have used (27) for the first and the second equality.

In order to obtain the asymptotic behavior of $(L^{(N)}, V_N/N)$, we fix $i \geq 2$ and $v \in [0,1]$.

$$
\begin{aligned}
\mathbb{P}\left(L^{(N)} = i, \frac{V_N}{N} \leq v\right) &= \sum_{k=i-1}^{\lfloor Nv \rfloor} \mathbb{P}\left(L^{(N)} = i, \frac{V_N}{N} = k\right) \\
&= \sum_{k=i-1}^{\lfloor Nv \rfloor} 2\frac{(N-i)!}{N!} \frac{k!}{(k-i+1)!} \frac{N-k}{N-1} \\
&= \frac{2}{N} \sum_{k=i-1}^{\lfloor Nv \rfloor} \frac{k}{N-1} \frac{k-1}{N-2} \cdots \frac{k-i+2}{N-i+1} \frac{N-k}{N-1} \\
&\xrightarrow{N\to+\infty} 2 \int_0^v y^{i-1}(1-y)dy.
\end{aligned}
$$

4.3. **Proof of Proposition 2.4.** The Laplace transform (8) comes from (7). To get the moments, we set $g(\lambda) = \mathbb{E}[e^{-\lambda\tau} \mid Y, L = a+1]$ and $c_k = k(k+1)$. We have

$$
g'(\lambda) = -g(\lambda) \sum_{k \geq a} \frac{1}{c_k + 2\lambda}.
$$

Hence

$$
\mathbb{E}[\tau \mid Y, L = a+1] = -g'(0) = \sum_{k \geq a} \frac{2}{k(k+1)} = \frac{2}{a}.
$$

We also have

$$
g''(\lambda) = g(\lambda) \sum_{k \geq a} \frac{1}{(c_k + 2\lambda)^2} + g(\lambda) \sum_{\ell,k \geq a} \frac{1}{c_k + 2\lambda} \frac{1}{c_\ell + 2\lambda}.
$$

Thus we get

$$\mathbb{E}[\tau^2|Y, L = a + 1] = g''(0)$$

$$= 4 \sum_{k \geq a} \frac{1}{k^2(k+1)^2} + 4 \sum_{\ell, k \geq a} \frac{1}{k(k+1)} \frac{1}{\ell(\ell+1)}$$

$$= 8 \sum_{k \geq a} \frac{1}{k(k+1)} \sum_{\ell \geq k} \frac{1}{\ell(\ell+1)}$$

$$= 8 \sum_{k \geq a} \frac{1}{k^2(k+1)}$$

$$= 8 \sum_{k \geq a} \frac{1}{k^2} - 4 \sum_{k \geq a} \frac{2}{k(k+1)}$$

$$= 8 \sum_{k \geq a} \frac{1}{k^2} - 4\mathbb{E}[\tau|Y, L = a+1].$$

We get (10) from (8) and Proposition 2.3.

4.4. **Proofs of Lemma 2.5, Proposition 2.6 and Corollaries 2.7 and 2.8.** We first notice that Corollary 2.7 is a direct consequence of Proposition 2.6 and the second part of Proposition 2.3. Corollary 2.8 is a direct consequence of Corollary 2.7. Indeed, the law of X conditionally on Y is given by:

$$Y = \begin{cases} X & \text{with probability} \quad X \\ 1 - X & \text{with probability} \quad 1 - X \end{cases}$$

Hence we get that

$$\mathbb{P}(Z = k|X = x) = \mathbb{P}(Z = k|Y = x)\mathbb{P}(Y = x|X = x) + \mathbb{P}(Z = k|Y = 1 - x)\mathbb{P}(Y = 1 - x|X = x)$$
$$= x\mathbb{P}(Z = k|Y = x) + (1 - x)\mathbb{P}(Z = k|Y = 1 - x).$$

Then (18) implies (21). Same arguments give (22) from (19) and (23) from (20).

Proof of Lemma 2.5 Let us fix $N \geq 2$. We have introduced $L^{(N)}$ 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 2.5 is straightforward for $(V_N/N, Z_N, L^{(N)}, L_1^{(N)}, L_2^{(N)}, ..., L_{Z_N}^{(N)})$ instead of $(Y, Z, L, L_1, ..., L_Z)$. Now, using same arguments as for the proof of the second part of Proposition 2.3, we get that $\left((V_N/N, Z_N, L^{(N)}, L_1^{(N)}, L_2^{(N)}, ..., L_{Z_N}^{(N)}), N \geq 2\right)$ converges a.s. to $(Y, Z, L, L_1, ..., L_Z)$ which ends the proof.

Proof of Propositions 2.6 First notice that using Lemma 2.5, it is enough to compute the quantities $\mathbb{P}(Z = k|L = a + 1)$. By definition of L and Z, $\mathbb{P}(Z = 0|L = 2) = 1$ and $\mathbb{P}(Z = 0|L = a + 1) = 0$ for $a \geq 2$. We suppose that $a \geq 2$. We get $\mathbb{P}(Z = k|L = a + 1)$ by induction on k:

For $k \geq 1$,

$$
\begin{aligned}
\mathbb{P}(Z = k | L = a + 1) &= \sum_{1 < a_2 < a} \mathbb{P}(Z = k, L_1 = a_2 + 1 | L = a + 1) \\
&= \sum_{1 < a_2 < a} \mathbb{P}(Z = k | L_1 = a_2 + 1, L = a + 1) \mathbb{P}(L_1 = a_2 + 1 | L = a + 1) \\
&= \sum_{1 < a_2 < a} \mathbb{P}(Z = k - 1 | L_1 = a_2 + 1) \mathbb{P}(L^{(a)} = a_2) \\
&= \sum_{1 < a_k < \cdots a_2 < a} \mathbb{P}(L^{(a_k)} = 2) \mathbb{P}(L^{(a_{k-1})} = a_k) \cdots \mathbb{P}(L^{(a)} = a_2),
\end{aligned}
$$

where we have used Lemma 2.5 for the third and last equalities. Using (6), equation (15) follows.

An expansion of $\prod_{k=2}^{a-1} \left(1 + \frac{2u}{(k-1)(k+2)} \right)$ and (15) immediately give (16). The result of the first two moments (17) follow from (9) and Proposition 1.2 proved in the next section.

4.5. **Proof of Proposition 1.2.** We set $c_k = k(k + 1)$, $b_k = c_k - 2 = (k - 1)(k + 2)$ and $w = u - 1$. Using (16), we have for $a \geq 3$

$$
\begin{aligned}
\mathbb{E}[u^Z | Y, L = a + 1] &= \frac{u}{3} \frac{a + 1}{a - 1} \prod_{k=2}^{a-1} \frac{b_k + 2u}{b_k} \\
&= \frac{u}{3} \frac{a + 1}{a - 1} \prod_{k=2}^{a-1} \frac{c_k + 2w}{b_k} \\
&= \prod_{k=1}^{a-1} \frac{c_k + 2w}{c_k}.
\end{aligned}
$$

This equality is also true for $a = 2$. And for $a = 1$, we have $\mathbb{E}[u^Z | Y, L = a + 1] = 1$. The conclusion is then clear from (8) as

$$
\mathbb{E}[\mathrm{e}^{-\lambda T_K}] = \prod_{k=1}^{\infty} \frac{c_k}{c_k + 2\lambda}.
$$

4.6. **Proof of Proposition 3.1.** Let μ_0 be the beta $(2, 1)$ distribution. Using [5], it is enough to prove that μ_0 is the only probability measure μ on $[0, 1)$ such that μ is invariant for Y^μ. Since $x \mapsto \mathbb{E}_x[\tau]$ is bounded (see (13)), we get that $\mathbb{E}_\mu[\tau] < \infty$. As $\mathbb{E}_\mu[\tau] < \infty$, it is straight forward to deduce from standard results on Markov chain having one atom with finite mean return time (see e.g. [22] for discrete time Markov chains) that Y^μ has a unique invariant probability 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

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

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

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

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

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

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 (28) that

(29) $$\langle \mu, g_\lambda \rangle = \lambda \mathbb{E}_\mu[\tau]\langle \mu, g_\lambda \rangle.$$

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

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

Notice that the function $\sum_{n=0}^\infty a_n^\lambda x^n$ solves $Lf = -\lambda f$ on $[0,1)$. For $\lambda = \dfrac{N(N+1)}{2}$ and $N \in \mathbb{N}^*$, 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 L with eigenvalue $-N(N+1)/2$, and such that $P_N(1) = 0$ (as $LP_N(1) = 0$). Notice that $P_1(x) = 1 - x$, which implies that $\langle \mu, P_1 \rangle > 0$. We deduce from (29) 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 get $P_N(x) = (1-x)Q_{N-1}(x)$, where Q_{N-1} is a polynomial function of degree $N-1$. For the probability measure $\bar\mu(dx) = \dfrac{1-x}{\langle \mu, P_1 \rangle}\mu(dx)$, we get that for all $N \geq 1$

(30) $$\langle \bar\mu, Q_N \rangle = \frac{\langle \mu, P_{N+1} \rangle}{\langle \mu, P_1 \rangle}.$$

Since $\bar\mu$ is a probability measure, it is characterized by (30). To conclude, we then just have to check that $\bar\mu_0$ satisfies (30). In fact, we shall check the stronger condition that $\langle \mu_0, g_\lambda \rangle = 0$ for any C^2 function g_λ eigenvector of L with eigenvalue $-\lambda$ such that $g_\lambda(1) = 0$ and $\lambda \neq 1$. We have

$$\begin{aligned}
-\lambda \langle \mu_0, g_\lambda \rangle &= -\lambda \int_0^1 2x g_\lambda(x)dx \\
&= \int_0^1 x^2(1-x)g_\lambda''(x)dx + \int_0^1 2x(1-x)g_\lambda'(x)dx \\
&= \left[x^2(1-x)g_\lambda'(x)\right]_0^1 - \int_0^1 (2x(1-x) - x^2)g_\lambda'(x)dx + \int_0^1 2x(1-x)g_\lambda'(x)dx \\
&= \int_0^1 x^2 g_\lambda'(x)dx \\
&= \left[x^2 g_\lambda(x)\right]_0^1 - \int_0^1 2x g_\lambda(x)dx,
\end{aligned}$$

which implies $\langle \mu_0, g_\lambda \rangle = 0$ unless $\mu = 1$.

References

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

[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.

[3] P. Cattiaux, P. Collet, A. Lambert, S. Martinez, S. Méléard, and J. S. Martin. Quasi-stationarity distributions and diffusion models in population dynamics, 2006.

[4] J. T. Chang. Recent common ancestors of all present-day individuals. *Adv. in Appl. Probab.*, 31(4):1002–1038, 1999. With discussion and reply by the author.

[5] 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.

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

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

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

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

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

[11] S. N. Evans and P. L. Ralph. Dynamics of the time to the most recent common ancestor in a large branching population, 2008.

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

[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.

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

[15] A. Greven, P. Pfaffelhuber, and A. Winter. Tree-valued resampling dynamics: Martingale problems and applications, 2008.

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

[17] T. Huillet. On wright-fisher diffusion and its relatives. *J. Stat. Mech.*, 2007.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

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

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

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

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

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

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

[29] D. Simon and B. Derrida. Evolution of the most recent common ancestor of a population with no selection, 2006.

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

[31] F. Tajima. Relationship between DNA polymorphism and fixation time. *Genetics*, 1990.

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

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

CERMICS, École des Ponts, ParisTech, 6-8 av. Blaise Pascal, Champs-sur-Marne, 77455 Marne La Vallée, France.
E-mail address: delmas@cermics.enpc.fr

MAP5, Université Paris Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France
E-mail address: jean-stephane.dhersin@parisdescartes.fr

MAP5, Université René Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France
E-mail address: Arno.Jegousse@parisdescartes.fr