

THE SEQUENTIAL LOSS OF ALLELIC DIVERSITY

BY GUILLAUME ACHAZ, AMAURY LAMBERT AND EMMANUEL SCHERTZER

Abstract

In this paper we give a new flavour to what Peter Jagers and his co-authors call ‘the path to extinction’. In a neutral population of constant size N , assume that each individual at time 0 carries a distinct type, or allele. Consider the joint dynamics of these N alleles, for example the dynamics of their respective frequencies and more plainly the nonincreasing process counting the number of alleles remaining by time t . Call this process the extinction process. We show that in the Moran model, the extinction process is distributed as the process counting (in backward time) the number of common ancestors to the whole population, also known as the block counting process of the N -Kingman coalescent. Stimulated by this result, we investigate whether it extends (i) to an identity between the frequencies of blocks in the Kingman coalescent and the frequencies of alleles in the extinction process, both evaluated at jump times, and (ii) to the general case of Λ -Fleming–Viot processes.

Keywords: Coalescent; Λ -Fleming–Viot; look-down; extinction; population genetics; urn model; intertwining

2010 Mathematics Subject Classification: Primary 92D10

Secondary 60-06; 60G09; 60J20

1. Introduction

In this paper we are interested in so-called neutral models of population genetics, that is, where all individuals are exchangeable in the face of death and reproduction [5], [6]. In such a population, *genetic drift* refers to the randomness of births and deaths and their effect on the composition of the population. If individuals are given a type, also called *allele*, which is transmitted faithfully to their offspring, one can follow the fluctuations of the numbers of carriers of each allele through time under the sole action of genetic drift [5], [6]. In the absence of mutation, the number of alleles present in the population can only decrease, and does so exactly at times when the last carrier of a given allele dies. In a population where births and deaths compensate so as to keep its total size constant equal to N , the number of alleles decreases sequentially until one single allele remains present, an event called *fixation*. By exchangeability, under the assumption that all individuals initially carry a distinct allele, each allele has the same probability $1/N$ to be the one that fixes.

Here, we study this process of sequential loss of allelic diversity and its limit as $N \rightarrow \infty$. The genealogical model we consider is a classical model of population genetics called the Λ -Fleming–Viot process [2], [4], whose law is characterized by a finite measure Λ on $[0, 1]$. When $\Lambda = \Lambda(\{0\})\delta_0$, this process is also known as the Moran process [5], [6]. In this case, all offspring sizes at birth times are a.s. equal to 2. In all other cases, roughly speaking, each offspring size ξ is binomially distributed with parameters N and p , where, conditional on $\xi \geq 2$, p is drawn from the measure $x^{-2}\Lambda(dx)$.

Let us introduce the coalescent (i.e. backwards in time) view of the Λ -Fleming–Viot process. The Λ -Fleming–Viot process is stationary, so one can fix an arbitrary time called present time, and count the number of *ancestors* A_t^N common to the whole population alive at present time, t units of time before the present. We can define more generally the vector $\rho^N(t)$ of frequencies of the descendants at present time of these A_t^N ancestors (uniformly ordered at time 0, with uniform reinsertion of the new frequency at each coalescence event). The counter A_t^N decreases as t increases and we can study the sequence of its jump times $(T^{N,k})_{k=2}^N$, where $T^{N,k}$ is the time at which A_t^N decreases by crossing or leaving k , as time runs backwards (i.e. as t increases). We are also interested in the sequence $(\pi_k^N)_{k=2}^N$, called the *ancestral block process*, where $\pi_k^N = \rho^N(T^{N,k})$ is the embedded chain of ρ^N (with possible repeats).

As $N \rightarrow \infty$, the processes $(A_t^N; t \geq 0)$ converge in the sense of finite-dimensional distributions (FDDs) to some process $(A_t^\infty; t \geq 0)$. It is known that, under a condition on the measure Λ usually known as CDI (‘coming down from infinity’) [14], $A_t^\infty < \infty$ for all $t > 0$ a.s. We assume that this condition is met throughout the paper. Then the two sequences $(T^{N,k})_{k=2}^N$ and $(\pi_k^N)_{k=2}^N$ converge as $N \rightarrow \infty$ in the sense of FDDs. In addition, the limiting sequences $(T^k)_{k \geq 2}$ and $(\pi_k^\infty)_{k \geq 2}$ are respectively the jump times and the embedded chain of the Λ -coalescent [11], [13].

Now let us consider the Λ -Fleming–Viot process as time runs forward. Fix again an initial time, give a distinct allele to each individual present in the population at this time and define \widehat{A}_t^N to be the number of alleles present in the population t units of time later. The topic that gave its title to this note is the study of the sequence of extinction times $(\widehat{T}^{N,k})_{k=2}^N$, where $\widehat{T}^{N,k}$ is the time at which \widehat{A}_t^N decreases by crossing or leaving k . Similarly as before, we can also define $\widehat{\rho}^N(t)$ the vector of frequencies of these \widehat{A}_t^N alleles at time t (where the N initial alleles are uniformly ordered at time 0) and the sequence $(\widehat{\pi}_k^N)_{k=2}^N$, called the *haplotype block process*, where $\widehat{\pi}_k^N = \widehat{\rho}^N(\widehat{T}^{N,k})$ is the embedded chain of $\widehat{\rho}^N$ (with possible repeats). Here again, under the CDI condition, both sequences converge in the sense of FDDs as $N \rightarrow \infty$ (see Proposition 1).

We show or recall in Theorem 4 two distributional identities: for each fixed t , $\rho^N(t)$ and $\widehat{\rho}^N(t)$ have the same law (and thus also A_t^N and \widehat{A}_t^N) and, for each fixed $k \in \{2, \dots, N\}$, $T^{N,k}$ and $\widehat{T}^{N,k}$ have the same law. We are interested in generalizing these identities from one-dimensional marginals to processes. Note though that the processes ρ^N and $\widehat{\rho}^N$ cannot be equally distributed because ρ^N remains constant between jump times while $\widehat{\rho}^N$ does not. We thus focus our study on jump times and embedded chains.

Our first, striking result (Theorem 1) is that in the binary case for which $\Lambda = \Lambda(\{0\})\delta_0$, the process $(A_t^N; t \geq 0)$ that counts the number of ancestors backwards in time and the process $(\widehat{A}_t^N; t \geq 0)$ that counts the number of remaining alleles forwards in time, have the same law (where $N \leq \infty$). In particular, the sequences $(T^{N,k})$ and $(\widehat{T}^{N,k})$ have the same law.

This result triggers two questions.

- (i) Does this result extend to an identity in law between π_k^∞ and $\widehat{\pi}_k^\infty$ in the binary case?
- (ii) Does this result extend to an identity in law between $(A_t^\infty; t \geq 0)$ and $(\widehat{A}_t^\infty; t \geq 0)$ in the general case?

The answer to question (i) is ‘yes’: for each fixed k , π_k^∞ and $\widehat{\pi}_k^\infty$ are equally distributed in the binary case. Their common distribution is the uniform distribution on the simplex of dimension $k - 2$. Therefore, the limiting processes ρ^∞ and $\widehat{\rho}^\infty$ jump at the same rate and have the same one-dimensional distribution at each jump time. We have seen that they cannot be equally distributed and actually, neither can their embedded chains $(\pi_k^\infty)_{k \geq 2}$ and $(\widehat{\pi}_k^\infty)_{k \geq 2}$. This can

be seen thanks to Theorem 6 where we prove that $(\widehat{\pi}_k^\infty)_{k \geq 2}$ is a Markov chain (as k decreases, that is, in the natural underlying direction of time) whose transitions are elegantly characterized thanks to a random coupon collection procedure which is obviously distinct from the merging procedure of the Kingman coalescent.

The answer to question (ii) is ‘no’ in general, simply because the jumps of $(A_t^\infty; t \geq 0)$ can take arbitrary values in \mathbb{Z}^- , while those of $(\widehat{A}_t^\infty; t \geq 0)$ are in most known cases of measures Λ , all equal to -1 a.s. Labbé [9] conjectured that this last property always holds. Our Proposition 3 gives a new criterion equivalent to this property based on urn models, and this might be helpful in the future to prove the conjecture.

In the next two sections, we state our main results, first in the binary case and then in the general case. The remainder of the paper is devoted to proving these results.

From now on, we denote the set $\{1, \dots, N\}$ by $[N]$, and the set $\mathbb{N} \cup \{0\}$ by \mathbb{N}_0 .

2. The binary case

In this section, we confine attention to the binary case, i.e. the case that $\Lambda = \Lambda(\{0\})\delta_0$, corresponding to the Moran process in forward time, and to the Kingman coalescent in backward time. We introduce our main tools and results in this section, and generalize them in the next.

2.1. The Moran model and the N -Kingman coalescent

Consider a population of constant size N and assume that at time $t = 0$, every individual is assigned a distinct type, with values in $[N]$ say. Assume that the population evolves according to classical Moran dynamics. Every ordered pair (a, b) of individuals is equipped with an independent Poisson clock of rate 1. When a clock associated to a pair (a, b) rings, a gives birth to a new individual inheriting her type and b dies simultaneously. Starting from the present, one can trace backwards in time the genealogy of the population. It is well known [8] that this genealogy is described in terms of the N -Kingman coalescent.

2.2. Block counting and extinction processes

In the framework introduced previously, we define the *extinction process* as

$$\widehat{A}_t^N = \#\{\text{alleles present at time } t\}, \quad t \geq 0, \tag{1}$$

and

$$\widehat{T}^{N,k} = \sup\{t > 0 : \widehat{A}_t^N \geq k\}, \quad k \in \{2, \dots, N\}, \tag{2}$$

so that $\widehat{T}^{N,2}$ is the fixation time of the population, i.e. the first time when one of the initial alleles has invaded the whole population. In the Kingman coalescent, define the *block counting process* as

$$A_t^N = \#\{\text{ancestors at time } t\}, \quad t \geq 0, \tag{3}$$

and

$$T^{N,k} = \sup\{t > 0 : A_t^N \geq k\}, \quad k \in \{2, \dots, N\}, \tag{4}$$

so that $T^{N,k}$ is the first time that the number of blocks goes from k to $k - 1$. It is well known that the block counting process A_t^N is a pure-death process starting from N with transition rate $k(k - 1)/2$ from level k to $k - 1$ for $k \in \{2, \dots, N\}$.

The next result states that the sequential decrease of ancestral lineages in the Kingman coalescent has the same law as the sequential loss of allelic diversity forwards in time. This theorem is proved in Section 5.

Theorem 1. *For every $N \in \mathbb{N}$, the extinction process \widehat{A}^N and the block counting process A^N are identical in law.*

2.3. The ancestral and haplotype block processes

For each $l \in \mathbb{N}$, define the set of mass partitions of $[0, 1]$ with cardinality l by

$$E_l = \left\{ u = (u_1, \dots, u_l) \in [0, 1]^l : \sum_{i=1}^l u_i = 1 \right\},$$

and the set of finite mass partitions of $[0, 1]$ as

$$E = \bigcup_{l \in \mathbb{N}} E_l.$$

For each $u \in E$, we let $\lambda(u)$ denote its cardinality.

Define $\pi_k^N \in E$ to be the vector of uniformly ordered frequencies of the blocks in the N -Kingman coalescent at time $T^{N,k}$. Note that π_k^N has exactly $k - 1$ coordinates. Call the process $(\pi_k^N; k \in \{2, \dots, N\})$ the *ancestral block process*. In the next theorem, \mathcal{D}_k refers to the random mass partition induced by a Dirichlet random variable with parameter

$$\underbrace{(1, \dots, 1)}_{k \text{ times}},$$

that is, the uniform distribution on the simplex of dimension $k - 1$. We recall the next result.

Theorem 2. (The ancestral block process [8].) *As $N \rightarrow \infty$,*

$$(\pi_k^N)_{k=2}^N \Rightarrow (\pi_k^\infty)_{k=2}^\infty \text{ in the sense of FDD,}$$

where the limiting vector is uniquely determined by the property that, for every $K \in \mathbb{N}$, the chain $(\pi_{K-k}^\infty; k \in \{0, \dots, K - 2\})$ is a Markov chain such that

- (i) the initial distribution π_K^∞ is distributed as \mathcal{D}_{K-1} ; and
- (ii) the mass partition at time k is obtained from the mass partition at time $k - 1$ by merging two entries chosen uniformly at random among all possible pairs of entries and reinserting the resulting mass uniformly in the vector deprived of these two entries.

Remark 1. Bertoin and Goldschmidt [1] considered the case $k = 1$ of this sequence. In particular, the reversed chain $(\pi_k^\infty; k \in \{2, \dots, K\})$ is also Markovian. The mass partition at time k is obtained from the mass partition at time $k - 1$ by uniformly fragmenting one block chosen as a size-biased pick from the mass partition at time $k - 1$.

We now draw a connection with the forward dynamics. In the Moran model, we can partition the population into blocks of individuals carrying the same allele. Analogously to the ancestral block process, define $\hat{\pi}_k^N \in E$ to be the mass partition induced by the alleles at time $\hat{T}^{N,k}$, where the N initial alleles are uniformly labelled from 1 to N at time 0. Note that $\hat{\pi}_k^N$ has at most $k - 1$ coordinates. Call the process $(\hat{\pi}_k^N; k \in \{2, \dots, N\})$ the *haplotype block process*.

In order to describe our next result, we need some further notation. First, say that a mass partition is *nondegenerate* if all the coordinates are strictly positive. For a given mass partition $m \in E$, we can think of m as a discrete probability distribution on $[\lambda(m)]$. Then let $(Y_n; n \geq 1)$ be an infinite sequence of i.i.d. random variables with distribution m and define

$$T = \inf\{n \geq 1 : \text{for all } i \in [\lambda(m)], \text{ there exists } k \in [n] \text{ for which } Y_k = i\};$$

such n is a.s. finite if m is nondegenerate and otherwise taken equal to $+\infty$ by convention. In the context of the coupon collector problem, T is the time needed to collect all the coupons $1, \dots, \lambda(m)$, where the probability to draw coupon i is equal to m_i . In view of this observation,

for any nondegenerate m , we define

$$B_i^{(m)} = \sum_{k=1}^{T-1} 1_{Y_k=i}, \quad i \in [\lambda(m)],$$

as well as $J^{(m)} = Y_T$, so that $J^{(m)}$ is the unique i such that $B_i^{(m)} = 0$. Then call the random vector

$$B^{(m)} = (B_i^{(m)}; i \in [\lambda(m)], i \neq J^{(m)})$$

the random coupon collection associated with m .

Theorem 3. (The haplotype block process.) As $N \rightarrow \infty$,

$$(\widehat{\pi}_k^N)_{k=2}^n \Rightarrow (\widehat{\pi}_k^\infty)_{k=2}^\infty \quad \text{in the sense of FDD,}$$

where the limiting vector is uniquely determined by the property that, for every $K \in \mathbb{N}$, the chain $(\widehat{\pi}_{K-k}^\infty; k \in \{0, \dots, K-2\})$ is a Markov chain such that

- (i) the initial distribution $\widehat{\pi}_K^\infty$ is distributed as \mathcal{D}_{K-1} ; and
- (ii) the mass partition at time i is obtained from the mass partition at time $i-1$ according to
 - (a) conditional on $\widehat{\pi}_{K-i}^\infty$, generate (X_i) distributed as the random coupon collection associated to the mass partition $\widehat{\pi}_{K-i}^\infty$, and
 - (b) conditional on X_i , the block process at step $i+1$ (i.e. $\widehat{\pi}_{K-(i+1)}^\infty$) is distributed as a Dirichlet random variable with parameters (X_i) .

This last result is a special case of Theorem 6, and it is proved in Section 7.

Remark 2. The vector $\widehat{\pi}_k^N$ gives the frequencies of all the alleles present at time $\widehat{T}^{N,k}$. For each of these alleles, the subpopulation carrying this allele at time $\widehat{T}^{N,k}$ has a certain number of ancestors living at time $\widehat{T}^{N,k+1}$. This number is precisely the number of coupons corresponding to this allele in the random coupon collection procedure mentioned in Theorem 3.

Remark 3. From Theorems 2 and 3, observe that the two mass partitions π^∞ and $\widehat{\pi}^\infty$ have the same one-dimensional marginal distributions, but certainly have different probability transitions. In relation to Remark 1 and Bertoin and Goldschmidt’s work [1], we know that (π_k^∞) is Markovian both as k decreases and as k increases. The previous statement ensures that $(\widehat{\pi}_k^\infty)$ is Markovian as k decreases. It remains an open question as to whether $(\widehat{\pi}_k^\infty)$ is also Markovian as k increases.

3. The general case

3.1. The Λ -coalescent and Λ -Fleming–Viot process

We now turn to more general exchangeable population models. Let Λ be a finite measure on $[0, 1]$. Consider a population of constant size N evolving according to the following dynamics. As before, at time $t = 0$, we assign a distinct allele to each individual in the population. Each k -tuple ($2 \leq k \leq N$) of individuals is equipped with a Poisson clock with rate

$$\lambda^N(k) := \int_0^1 x^{k-2}(1-x)^{N-k} \Lambda(dx).$$

When the clock associated with a group of k individuals rings, one of them is picked uniformly at random, gives birth to $k-1$ new individuals inheriting her type, while the $k-1$ others

die simultaneously. We call such a process an (N, Λ) -Fleming–Viot (FV) process. The case $\Lambda = \Lambda(\{0\})\delta_0$ corresponds to the classical Moran model presented earlier.

It is well known that the genealogy of the population can be described in terms of the (N, Λ) -coalescent, i.e. the projection of the Λ -coalescent on $[N]$ [2], [4].

3.2. The block counting and extinction processes (continued)

For the (N, Λ) -FV process and (N, Λ) -coalescent, define the extinction process \widehat{A}^N and the block counting process A^N as in (1) and (3), respectively, along with their jump times $\widehat{T}^{N,k}$ and $T^{N,k}$ as in (2) and (4), respectively. The next result gives some identities between one-dimensional marginals of the block counting and the extinction processes in the general case.

Theorem 4. *Let $N \in \mathbb{N}$.*

- (i) *For every fixed $t > 0$, A_t^N and \widehat{A}_t^N have the same law.*
- (ii) *(Hénard [7]) For every fixed $k \in \{2, \dots, N\}$, $T^{N,k}$ and $\widehat{T}^{N,k}$ have the same law.*

Proof. We prove only (i). Consider the population at time $t > 0$. By stationarity of the FV process, the number of ancestors a_t^N at time 0 of this population is equal in distribution to A_t^N . Since each of these a_t^N ancestors carries a distinct allele, the number of alleles present in the population at time t is a_t^N , that is, $\widehat{A}_t^N = a_t^N$. This shows that \widehat{A}_t^N has the same law as A_t^N . The same argument actually shows that the identity also holds for the frequencies of descendants (see the introduction). □

In view of Theorem 4, it is natural to conjecture that the block counting process and the extinction process are identical in law for any measure Λ . We show now that this is not the case in general. To this end, we let the size of the population go to ∞ .

In the following, we are interested in the particular case that, for every $k \geq 2$, $(T^{N,k}, \widehat{T}^{N,k}; N \geq 0)$ is a tight sequence of random variables as $N \rightarrow \infty$. This corresponds to the case that the Λ -coalescent *comes down from infinity* (CDI), which is equivalent to the condition [14]

$$\int_1^\infty \frac{du}{\psi(u)} < \infty, \quad \text{where} \quad \psi(u) = \Lambda(\{0\})u^2 + \int_{(0,1)} \frac{e^{-xu} - 1 + xu}{x^2} \Lambda(dx). \quad (5)$$

Proposition 1. *Under condition (5), there exist two processes A^∞ and \widehat{A}^∞ coming down from infinity such that, as $N \rightarrow \infty$, A^N and \widehat{A}^N converge in the sense of FDDs to A^∞ and \widehat{A}^∞ , respectively. In addition, the sequences $(T^{N,k})_k$ and $(\widehat{T}^{N,k})_k$ converge in the sense of FDDs to the corresponding jump times of A^∞ and \widehat{A}^∞ , respectively.*

This proposition is proved in Section 4 where the previous statement is reformulated in terms of the look-down process.

Theorem 5. ([9, Theorem 1.6].) *If Λ is such that ψ is regularly varying at ∞ with index in $(1, 2]$ then there are no simultaneous extinction events in the Λ -Fleming–Viot process, i.e.*

$$\mathbb{P}(\text{there exists } t > 0: |\widehat{A}_t^\infty - \widehat{A}_{t-}^\infty| \geq 2) = 0. \quad (6)$$

Conjecture 1. ([9].) *Equation (6) holds for any Λ -FV in the CDI class (i.e. satisfying (5)).*

Now note that, when $\Lambda \neq \Lambda(\{0\})\delta_0$, it is obvious that

$$\mathbb{P}(\text{there exists } t > 0: |A_t^\infty - A_{t-}^\infty| \geq 2) = 1,$$

because $(A_t^\infty; t > 0)$ is the number of blocks in the Λ -coalescent which by assumption has multiple mergers. Thus, under the assumptions of the previous theorem and whenever $\Lambda \neq \Lambda(\{0\})\delta_0$, the block counting process and the extinction process cannot have the same law. Thus, even if Theorem 4 suggests a strong relation between the block counting and the extinction processes, those two processes are in general very different when considering coalescent processes with multiple mergers.

3.3. Λ -urn. Haplotype block process (continued)

In this section, we present an extension of Theorem 3 for Λ -coalescent processes in the CDI class. Define Y^n to be the $[n - 1]$ -valued random variable (RV) with the following distribution:

$$\mathbb{P}(Y^n = k - 1) = \frac{\int_{[0,1]} \binom{n}{k} x^{k-2} (1-x)^{n-k} \Lambda(dx)}{\int_{[0,1]} (1-n(1-x)^{n-1} - (1-x)^n) \Lambda(dx)}, \quad k \in \{2, \dots, n\}. \tag{7}$$

With this RV on hand, we now define an urn model as follows. At step n , the urn contains U_n colored balls; let $B(n)$ denote the vector containing the numbers of balls of each color. Let $B(0) = (B_1(0), \dots, B_c(0)) \in \mathbb{N}^c$ be the initial configuration of the urn, where c is the initial number of colors in the urn. Update the configuration of the urn as follows: conditional on $B(n)$, the configuration of the urn at time $n + 1$ is obtained by drawing a ball uniformly at random from the urn and by replacing the ball together with y^n balls of the same color, where y^n is an independent random variable distributed as Y^{U_n} . We call this urn model a Λ -urn with initial condition $B(0)$.

Proposition 2. *Let $(B(n); n \geq 0)$ be a Λ -urn with initial condition \mathcal{B} , and let $c = \lambda(\mathcal{B})$ be the number of colors. Let $m_n^{\mathcal{B}} \in E_c$ be the mass partition induced by the color frequencies*

$$\left(\frac{B_i(n)}{\sum_{j=1}^c B_j(n)} \right)_{i \in [c]}.$$

Then there exists a random mass partition $m_\infty^{\mathcal{B}} \in E_c$ such that, as $n \rightarrow \infty$,

$$m_n^{\mathcal{B}} \rightarrow m_\infty^{\mathcal{B}} \text{ a.s.}$$

Proof. As in a standard Pólya urn, the proof follows by noting that each color frequency is a nonnegative martingale. □

Remark 4. Note that the limit $m_\infty^{\mathcal{B}}$ is potentially *degenerate*, i.e. some of the coordinates might be equal to 0.

In the following, we define \mathcal{D}_k^Λ to be the limit RV $m_\infty^{\mathcal{B}}$ when we start from the initial condition

$$\mathcal{B} = \underbrace{(1, \dots, 1)}_{k \text{ times}}.$$

In the case $\Lambda = \Lambda(\{0\})\delta_0$, the Λ -urn coincides with the classical Pólya urn, so that \mathcal{D}_k^Λ is the Dirichlet distribution \mathcal{D}_k . Consequently, Theorem 3 is a special case of the next result.

Theorem 6. *Assume that Λ belongs to the CDI class and that there are no simultaneous extinction events in \hat{A}^∞ a.s. Then*

- (i) \mathcal{D}_K^Λ is nondegenerate a.s.;
- (ii) as $N \rightarrow \infty$,

$$(\hat{\pi}_k^N)_{k=2}^n \Rightarrow (\hat{\pi}_k^\infty)_{k=2}^\infty \text{ in the sense of FDDs,}$$

where the limiting vector is uniquely determined by the property that, for every $K \in \mathbb{N}$,

the discrete process $(\widehat{\pi}_{K-k}^\infty; k \in \{0, \dots, K - 2\})$ is a Markov process such that

- (a) the initial distribution $\widehat{\pi}_K^\infty$ is distributed as $\mathcal{D}_{K-1}^\Lambda$, and
- (b) the configuration at time i is obtained from the configuration at time $i - 1$ as follows:
 - conditional on $\widehat{\pi}_{K-i}^\infty$, let X_i be the coupon collection associated to $\widehat{\pi}_{K-i}^\infty$, and
 - conditional on X_i , the block process at step $i + 1$ (i.e. $\widehat{\pi}_{K-(i+1)}^\infty$) is distributed as $m_\infty^{X_i}$.

We prove this result in Section 7.

Remark 5. In the binary case, we have characterized the limits of the vectors of frequencies of descendances at jump times of both the common ancestors (π^∞ , Theorem 2) and the surviving alleles ($\widehat{\pi}^\infty$, Theorem 3). Note that in the general case, we do so only for the latter (Theorem 6).

Finally, we end with a result that may shed some light on Conjecture 1.

Proposition 3. *The following two statements are equivalent:*

- (i) *there are no simultaneous extinctions in \widehat{A}^∞ ;*
- (ii) *for every $k \in \mathbb{N}$, the limiting mass partition \mathcal{D}_k^Λ is nondegenerate.*

This proposition is proved in Section 7 together with Theorem 6.

In the next section we recall the construction of the (N, Λ) -FV process from the look-down (LD) process of Donnelly and Kurtz [4]. This construction is the cornerstone of our comparison between extinction and coalescence. In Section 5 we specialize to the case that $\Lambda = \Lambda(\{0\})\delta_0$ and give a proof of Theorem 1. In Section 6 we present a general urn model (encompassing the Λ -urn models) and give a particle representation *à la* Donnelly and Kurtz [4]. We then use this representation to prove Theorem 6 (and so Theorem 3 as a special case) in Section 7.

4. The look-down construction

Our results are based on the look-down (LD) process as defined by Donnelly and Kurtz [4]. In the sequel, we work directly with the LD process, without referring to Λ -FV processes anymore. We abuse notation, and use the same symbols for the coalescence, extinction times, etc. as those used in the introduction. This abuse of notation is based on the fact that all those quantities in the LD process and their analog for the (N, Λ) -FV process are known to be identical in distribution [4], [7]. At time $t = 0$, each level is assigned a distinct allele, where here alleles are i.i.d. uniform in $[0, 1]$. Now let ν denote a Poisson point measure in $\mathbb{R} \times \mathcal{P}(\mathbb{N})$ with intensity measure $dt \otimes dL$, where L is defined by

$$L = \int_{(0,1]} x^{-2} \Lambda(dx) P_x + \Lambda(\{0\}) \sum_{i < j} Q_{ij}, \tag{8}$$

where P_x is the law of the set of 1s in a sequence of i.i.d. Bernoulli RVs with parameter x and Q_{ij} is the Dirac mass at $\{i, j\}$. Each atom (t, A) of ν is called an *LD event*. At each LD event (t, A) , all levels $i \in A$ inherit the allele carried by level $\min A$. In terms of the genealogy, the lineage present at level $\min(A)$ splits at time t and gives birth simultaneously to new lineages whose locations are the levels labelled by A . All other lineages are untouched, but are shifted

upwards so that lineages do not cross (see Figure 1 in Section 5). More rigorously, if (t, A) is an LD event, the lineage present at level $i \notin A$ at $t -$ jumps to level $i + \Delta$ at time t , where

$$\Delta = \#((A \setminus \min A) \cap [i]).$$

Note that $\Delta \in \{0, 1\}$ a.s. when Λ charges only 0.

Remark 6. Given the LD construction for the infinite population model, the LD process for the model with population size N is given by the restriction of the LD process to the first N levels. This coupling between all LD processes with different population sizes and with infinite population is used repeatedly in the remainder of the paper without further notice. To couple also the vectors of allele frequencies $(\hat{\pi}^{N,k})_k$ for $N \leq \infty$, we assume that each allele at time 0 is labelled with an independent uniform variable. The order in $(0, 1)$ of these uniform labels induces a (uniform) order of the first N initial alleles which is consistent in N .

Fixation lines. Let $k \in \mathbb{N}$ and let U_k denote the type carried by level k at time 0. Define $(L_t^k; t \geq 0)$, the k th fixation line [3], [10], as the right-continuous jump process equal at time t to the minimum of all levels carrying type U_k at time t . In particular, $L_0^k = k$. The fixation line can be constructed forwards in time as follows. Conditional on $L_t^k = j$, set

$$v := \inf\{s > t: (s, A) \text{ is an LD event with } \#(A \cap [j]) \geq 2\}.$$

Then $L_s^k = j$ on $[t, v)$ and $L_v^k = j + \Delta$, where $\Delta = \#((A \setminus \min A) \cap [j])$. Define

$$\hat{T}^{N,k} = \inf\{t > 0: L_t^k \geq N + 1\}, \quad \hat{T}^k = \sup\{t > 0: L_t^k < \infty\}. \tag{9}$$

Since L_t^k is the minimum of all levels carrying type U_k , $\hat{T}^{N,k}$ is exactly the time when U_k disappears from the population occupying the first N levels. Recall that \hat{A}_t^N denotes the number of alleles present in this population at time t . Observe that, by construction, $L_t^{k-1} < L_t^k$ for all t a.s., so that

- for all $t < \hat{T}^{N,k}$, $\hat{A}_t^N \geq k$, and
- for all $t \geq \hat{T}^{N,k}$, $\hat{A}_t^N < k$,

so $\hat{T}^{N,k} = \sup\{t > 0: \hat{A}_t^N \geq k\}$ exactly as defined in (2). Also, $\hat{\pi}^{N,k}$ is the vector of frequencies of each of the N initial alleles at time $\hat{T}^{N,k}$, where the N initial alleles are uniformly labelled from 1 to N at time 0. Note that the length of $\hat{\pi}^{N,k}$ is equal to or smaller than $k - 1$ and that its components take values in $\{0, 1/N, 2/N, \dots, 1\}$.

Proof of Proposition 1. We prove the result only for the forward process \hat{A}^N ; the backward process A^N can be handled in a similar way.

Assume that $(\hat{T}^{N,k}; k \geq 1)_N$ and \hat{T}^k are coupled by the LD process (recall Remark 6). It is clear that, for every k , the sequence $\{\hat{T}^{N,k}; N \geq k\}$ is nondecreasing and converges to \hat{T}^k a.s. Furthermore, since $\hat{T}^{N,k} = T^{N,k}$ in law (by Theorem 4(ii)), and since $T^{N,k}$ converges in distribution to the k th coalescence time T^k of a Λ -coalescent coming down from infinity (under Proposition 1), it follows that \hat{T}^k is identical in law with T^k and so is finite a.s. Let us now consider the process

$$\hat{A}_t^\infty = \sup\{k \geq 1: \hat{T}^k > t\},$$

so that the jump times of \hat{A}^N converge a.s. to those of \hat{A}^∞ . For every deterministic t , it is not hard to see that $\mathbb{P}(\hat{A}_t^\infty - \hat{A}_t^N > 0) = 0$ (i.e. \hat{A}^∞ does not have any fixed point of discontinuity), and, thus, for every $t \geq 0$, the sequence $\{\hat{A}_t^N; N \geq 1\}$ converges to \hat{A}_t^∞ a.s. It follows that \hat{A}^N converges to \hat{A}^∞ in the sense of FDDs. □

From now on, assume that the Λ -FV dynamics are in the CDI class, that is, condition (5) is enforced. Assume also that there are no simultaneous extinctions, i.e.

$$\widehat{T}^k < \infty \quad \text{and} \quad \widehat{T}^{k+1} < \widehat{T}^k \quad \text{a.s. for every } k \geq 2.$$

According to Theorem 5, these two assumptions are met as soon as ψ is regularly varying at ∞ with index in $(1, 2]$. In particular, this encompasses the binary case $\Lambda = \Lambda(\{0\})\delta_0$.

5. Proof of Theorem 1

In this section we consider the extinction process defined from the LD process when $\Lambda = \Lambda(\{0\})\delta_0$. Recall that in this case there are no simultaneous extinction events, so that all jumps of \widehat{A}_t^N are equal to -1 a.s., just as for the block-counting process A_t^N of the Kingman coalescent. In order to prove Theorem 1, we therefore only need to prove the equality in law

$$(T^{N,k}; 2 \leq k \leq N) \stackrel{\mathcal{L}}{=} (\widehat{T}^{N,k}; 2 \leq k \leq N), \tag{10}$$

where here $T^{N,k}$ is the time at which the block counting process goes from k to $k - 1$. First, note that definition (9) immediately implies that

$$\widehat{T}^{N,k} = e_k + \dots + e_N,$$

where e_j is the time needed for the k th fixation line to make a transition from level j to $j + 1$. By definition of the LD process, the RVs e_j are independent and e_j follows the exponential distribution with parameter $\binom{j}{2}$. On the other hand, by considering the successive coalescence times in the N -Kingman coalescent, $T^{N,k}$ can be decomposed analogously, i.e.

$$T^{N,k} = e'_N + \dots + e'_k,$$

where e'_j is the time needed for the coalescent to go from j blocks to $j - 1$ block(s), that is, $(e'_j; k \leq j \leq N)$ is identical in law to $(e_j; k \leq j \leq N)$.

The previous argument shows that the one-dimensional marginals of the two vectors in (10) coincide. In order to prove (10), it remains to show that $(\widehat{T}^{N,k-1} - \widehat{T}^{N,k}; 3 \leq k \leq N)$ is a sequence of independent RVs. By a simple induction, this boils down to proving that

$$\widehat{T}^{N,k-1} - \widehat{T}^{N,k} \quad \text{is independent of} \quad (\widehat{T}^{N,i}; k \leq i \leq N) \quad \text{for all } k \leq j \leq N. \tag{11}$$

Define

$$R^{N,k} := L_{\widehat{T}^{N,k}}^{k-1},$$

i.e. $R^{N,k}$ is the position of the $(k - 1)$ th fixation line at extinction time $\widehat{T}^{N,k}$ (see Figure 1). The crucial observation is that $\widehat{T}^{N,k-1} - \widehat{T}^{N,k}$ depends only on

- (a) $R^{N,k}$, and
- (b) the LD events after time $\widehat{T}^{N,k}$.

We now argue that both (a) and (b) are *jointly* independent of $(\widehat{T}^{N,i}; k \leq i \leq N)$, thus showing (11). We start by showing that $R^{N,k}$ is independent of $(\widehat{T}^{N,i}; k \leq i \leq N)$. Since $L_t^{k-1} < L_t^k$, we note that L_t^{k-1} can jump only if L_t^k jumps (again see Figure 1). More precisely, if t is a jump time for the fixation line L_t^k and if

$$L_{t-}^k = k + j \quad \text{and} \quad L_{t-}^{k-1} = i \quad \text{for } i < k + j,$$

then $L_t^{k-1} = i + 1$ if and only if the LD event at time t involves a pair (i_1, i_2) with $i_1, i_2 \leq i$. This happens with probability $\binom{i}{2} / \binom{k+j}{2}$.

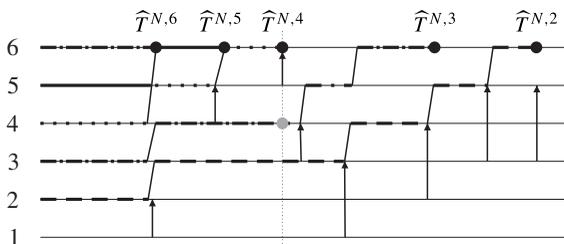


FIGURE 1: In this example $N = 6$ and $k = 4$, so that $R^{N,k} = 4$. The value of the fixation L^{k-1} (dotted lines) at time $\widehat{T}^{N,k}$ depends only on the type of arrow at the jumps of L^k (dot-dash lines). L^{k-1} jumps only at the first jump of L^k .

Now push the previous observation a bit further. For every $j \in [N - k]$, consider

$$Z(j) = L_{\inf\{t>0: L_t^k=k+j\}}^{k-1}.$$

In words, when L_t^k reaches level $k + j$, record the position of the $(k - 1)$ th fixation curve. In particular, $Z(0) = k - 1$ and $Z(N + 1 - k) = R^{N,k}$. Arguing as before, observe that Z is a discrete-time Markov chain with initial condition $Z(0) = k - 1$ and, for every $j \in [N - k]$,

$$\mathbb{P}(Z(j + 1) = i + 1 \mid Z(j) = i) = 1 - \mathbb{P}(Z(j + 1) = i \mid Z(j) = i) = \binom{i}{2} / \binom{j + k}{2},$$

and by a standard property of a Poisson point process, all these events are independent of the jump times of $(L_t^k, t \in [0, \widehat{T}^{N,k}))$. Note that this type of argument is already present in Pfaffelhuber and Wakolbinger [10]. Consequently, Z is independent of the σ -field \mathcal{F}_k generated by $(L_t^k, t \in [0, \widehat{T}^{N,k}))$ and the LD events strictly above L_t^k , that is, the LD events $(t, (u, v))$ such that $L_t^k < v$. Since $Z(N + 1 - k) = R^{N,k}$, $R^{N,k}$ is independent of \mathcal{F}_k .

We now claim that $(\widehat{T}^{N,i}; k \leq i \leq N)$ is \mathcal{F}_k -measurable. Indeed, for $i \in \{k, \dots, N\}$, the jumps of L_t^i are the jumps of L_t^k in addition to jumps caused by LD events strictly above L_t^k . This shows that the path $(L_t^i, t \in [0, \widehat{T}^{N,i}))$ is \mathcal{F}_k -measurable and so the extinction time $\widehat{T}^{N,i}$ is \mathcal{F}_k -measurable. Since $R^{N,k}$ is independent of \mathcal{F}_k , $R^{N,k}$ is independent of $(\widehat{T}^{N,i}; k \leq i \leq N)$.

Now let \mathcal{B}_k denote the σ -field generated by the LD events after time $\widehat{T}^{N,k}$. Note that both $(\widehat{T}^{N,i}; k \leq i \leq N)$ and $R^{N,k}$ depend only on the LD events before time $\widehat{T}^{N,k}$. By the strong Markov property, this implies that \mathcal{B}_k is independent of $(R^{N,k}, (\widehat{T}^{N,i}; k \leq i \leq N))$. Using the fact that if A, B, C are three RVs with $A \perp\!\!\!\perp C$ and $B \perp\!\!\!\perp (A, C)$ then $(A, B) \perp\!\!\!\perp C$, we conclude that $R^{N,k}$ and \mathcal{B}_k are jointly independent of $(\widehat{T}^{N,i}; k \leq i \leq N)$. This completes the proof.

6. A general urn model

6.1. Some definitions

In what follows, $c \in \mathbb{N}$ and corresponds to the number of colors in the urn. Furthermore, $(p^l)_{l \in \mathbb{N}}$ is a family of probability distributions on \mathbb{N} indexed by \mathbb{N} . We consider a discrete time increasing Markov process U_n such that, conditional on $U_n = l$, the jump probability distribution at time n is given by p^l . We think of U_n as the size of the urn at time n .

For every $V \in \bigcup_{l \in \mathbb{N}} [c]^l$, $\lambda(V)$ denotes the length of the vector V . For every finite subset $S \subset \mathbb{N}$, and every $u \in [c]$, define $\sigma_S^u(V)$ to be the unique vector W such that

- (i) for every $j \in S, W_j = u$, and
- (ii) if we remove the lines of W with indices in S , the resulting vector is equal to V .

Now, for every $B \in \mathbb{N}_0^c$, letting $l = \sum_{u=1}^c B_u$, define \mathbb{P}_B to be the uniform probability measure on

$$\left\{ V \in [c]^l : \text{for all } u \in [c], \sum_{i=1}^l \mathbf{1}_{V_i=u} = B_u \right\}.$$

6.2. A general urn model

We now define an urn model, whose state at step n is denoted by $B(n) \in \mathbb{N}_0^c$, where in particular $U_n = \sum_{u=1}^c B_u(n)$ is the number of balls in the urn. At step $n + 1$, conditional on $B(n)$, $B(n + 1)$ is generated by the following procedure.

- (i) Sample an independent RV b_n according to p^{U_n} .
- (ii) Conditional on b_n , sample a ball uniformly at random from the urn and return it with b_n additional balls of the same color, so that in particular $U_{n+1} = U_n + b_n$.

When $p^l = \delta_1$ for every $l \in \mathbb{N}_0$, this model coincides with the classical Pólya urn.

6.3. A particle system

In the spirit of Donnelly and Kurtz [4], we represent our urn as a particle system. This particle system is embodied by a $\bigcup_{l \in \mathbb{N}} [c]^l$ -valued process $(V(n); n \geq 0)$, where the length $U_n = \lambda(V(n))$ of the vector $V(n)$ is the number of particles at time n .

At step $n + 1$, conditional on $V(n)$, $V(n + 1)$ is generated by the following procedure.

- (R1) Sample an independent RV b_n according to p^{U_n} .
- (R2) Conditional on b_n , draw a random finite set S of cardinality $b_n + 1$ uniformly in $[U_n + b_n]$.
- (R3) Conditional on S , letting $S' = S \setminus \{\min S\}$ and $u = V_{\min(S)}(n)$, define

$$V(n + 1) = \sigma_{S'}^u(V(n)).$$

Our next result provides a particle representation of the urn model defined above.

Proposition 4. *Define the \mathbb{N}_0^c -valued process $(\bar{B}(n); n \geq 0)$ such that, for every color $u \in [c]$, $\bar{B}_u(n) = \sum_{i=1}^{\lambda(V(n))} \mathbf{1}_{V_i(n)=u}$, so that \bar{B} records the number of types in V . If the law of $V(0)$ is given by $\mathbb{P}_{B(0)}$ then*

- (i) $(\bar{B}(n); n \geq 0)$ and $(B(n); n \geq 0)$ are identical in law; and
- (ii) for every $n \in \mathbb{N}_0$ and $\mathcal{B} \in \mathbb{N}_0^c$, conditional on the event that $\{\bar{B}(n) = \mathcal{B}\}$, the law of $V(n)$ is given by $\mathbb{P}_{\mathcal{B}}$.

Proof. The first part of the result can be seen as a direct consequence of Theorem 1.1 of Donnelly and Kurtz [4]. The second part is apparent in the proof of the same result. For the sake of completeness, we provide an alternative proof of Proposition 4 based on the intertwining relation of Pitman and Rogers [12].

For any space E , define $F(E)$ to be the set of bounded functions from E to \mathbb{R} . Define the operator \mathcal{V} from $F(\bigcup_{l \in \mathbb{N}_0} [c]^l)$ to $F(\mathbb{N}_0^c)$ through the relation

$$\mathcal{V}f(B) = \mathbb{E}_B(f(V)) \quad \text{for all } B \in \mathbb{N}_0^c,$$

where the right-hand side means that the average is taken with respect to the random variable V distributed according to \mathbb{P}_B .

According to Pitman and Rogers [12], Proposition 4 follows if we can show that, for every $B \in \mathbb{N}_0^c$, and every bounded function $h \in F(\bigcup_{l \in \mathbb{N}} [c]^l)$,

$$\nu \widehat{G}h(B) = G \mathcal{V}h(B),$$

where G is the generator associated to the Pólya urn and \widehat{G} is the generator of the particle process defined above.

Let $l \equiv \sum_{u=1}^c B_u$. Note first that $\widehat{G}h(V)$ equals

$$\left[\sum_{u=1}^c \sum_{b=1}^l \frac{p^l(b)}{\binom{l+b}{b+1}} \sum_{S \subseteq [l+b]: |S|=b} |\{1 \leq x < \min(S) : V(x) = u\}| h(\sigma_S^u(V)) \right] - h(V).$$

Let I denote the term between brackets. In the following, for every $\mathcal{B} \in \mathbb{N}_0^c$ and every $(u, b) \in [c] \times \mathbb{N}_0$, $\theta_{u,b}(B)$ refers to the vector obtained from B by incrementing the u coordinate of B by b units. Then

$$\mathbb{E}_B(I) = \sum_{u=1}^c \sum_{b=1}^l \frac{p^l(b)}{\binom{l+b}{b+1}} \sum_{S \subseteq [l+b]: |S|=b} \mathbb{E}_B(|\{1 \leq x < \min(S) : V_x = u\}| h(\sigma_S^u(V))).$$

The innermost sum here equals

$$\begin{aligned} & \sum_{S \subseteq [l+b]: |S|=b} \mathbb{E}_B(|\{1 \leq x < \min(S) : \sigma_S^u(V)_x = u\}| h(\sigma_S^u(V))) \\ &= \sum_{S \subseteq [l+b]: |S|=b} \mathbb{E}_{\theta_{u,b}(B)}(|\{1 \leq x < \min(S) : V_x = u\}| h(V) \mid \text{for all } j \in S, V_j = u) \\ &= \sum_{S \subseteq [l+b]: |S|=b} \frac{\mathbb{E}_{\theta_{u,b}(B)}(|\{1 \leq x < \min(S) : V_x = u\}| \mathbf{1}_{\text{for all } j \in S, V_j = u} h(V))}{\mathbb{P}_{\theta_{u,b}(B)}(\text{for all } j \in S, V_j = u)}, \end{aligned}$$

where in the first expression above we used the fact that $\sigma_S^u(V)_x = V_x$ for $x < \min(S)$, and the next relation follows from the fact that, for every bounded function m ,

$$\mathbb{E}_B(m(\sigma_S^u(V))) = \mathbb{E}_{\theta_{u,b}(B)}(m(V) \mid \text{for all } j \in S, V_j = u).$$

On the other hand, if B is such that $B_u > b$ then

$$\sum_{S \subseteq [l+b]: |S|=b} \mathbf{1}_{\text{for all } j \in S, V_j = u} |\{1 \leq x < \min(S) : V_x = u\}| = \binom{B_u}{b+1}, \quad \mathbb{P}_B\text{-a.s.},$$

whereas

$$\mathbb{P}_B(\text{for all } j \in S, V_j = u) = \frac{B_u \cdots (B_u - (|S| - 1))}{l \cdots (l - (|S| - 1))}, \quad \text{where } l = \sum_{u=1}^c B_u,$$

and, thus, applying the two previous formulae after replacing B by $\theta_{u,b}(B)$,

$$\begin{aligned} \mathbb{E}_B(I) &= \sum_{u=1}^c \sum_{b=1}^l \frac{p^l(b)}{\binom{l+b}{b+1}} \frac{(l+b) \cdots (l+1)}{(B_u+b) \cdots (B_u+1)} \binom{B_u+b}{b+1} \mathbb{E}_{\theta_{u,b}(B)}(h(V)) \\ &= \sum_{u=1}^c \frac{B_u}{l} \sum_{b=1}^l p^l(b) \mathbb{E}_{\theta_{u,b}(B)}(h(V)). \end{aligned}$$

Furthermore, since

$$\mathbb{E}_B(h(V)) = \sum_{u=1}^c \frac{B_u}{l} \sum_{b=1}^l p^l(b) \mathbb{E}_B(h(V)),$$

we obtain

$$\mathcal{V}\widehat{G}h(B) = \sum_{u=1}^c \frac{B_u}{l} \sum_{b=1}^l p^l(b) (\mathbb{E}_{\theta_{u,b}(B)}(h(V)) - \mathbb{E}_B(h(V))) = G\mathcal{V}h(B).$$

As already mentioned, this completes the proof of Proposition 4 by an application of the intertwining theorem of Rogers and Pitman [12]. □

7. Proofs of Theorem 6 and Proposition 3

7.1. Definition of $\widehat{\pi}^k$

We start by defining a quantity $\widehat{\pi}^k$ in the LD process; this is the analog of the quantity $\widehat{\pi}^{\infty,k}$ introduced in Theorem 6. Consider the embedded Markov chain associated to the Markov process

$$(\xi_1(t), \dots, \xi_{L_t^k(t)-1}(t)),$$

where $\xi(t)$ is the LD process at time t . More specifically, $\xi_i(t)$ is the allele carried by level i at time t , assuming that at time $t = 0$, types are i.i.d. uniform in $[0, 1]$. From the definition of L^k , there are exactly $k - 1$ different values (alleles) along the coordinates of the chain at any time n . Consequently, this chain can be mapped to a $[k - 1]$ -valued chain $(V^k(n); n \geq 0)$ by replacing each type by its rank in the ordered statistics of alleles (i.e. the smallest $\xi_i(t)$ present in the sample is assigned type 1, etc.). It is not hard to check that $(V^k(n); n \geq 0)$ is identical in law to the particle process described in Section 6 with p^{U_n} , the jump law of the vector size at level U_n , being given by the law of Y^{U_n} as defined in (7). Indeed, conditional on $U_n = l$, where $U_n = \lambda(V^k(n))$ is the length of $V^k(n)$, the next LD-event with an effect on the LD process up to level l is the first LD event (t, A) such that $\#A \cap [l] \geq 2$ and then the level that gives birth is $\min A$ and the number of newborn levels is $\#A - 1$. Consequently,

$$\mathbb{P}(U_{n+1} - U_n = b - 1) = \frac{L(\{A \subset [l]: \#A = b\})}{L(\{A \subset [l]: \#A \geq 2\})} \quad \text{for } b \in \{2, \dots, l\},$$

and a quick calculation based on the definition of L given by (8) shows that this is exactly the distribution of Y^l given in (7). Therefore, we have proved that rule (R1) as specified in the definition of the particle system in Section 6 is satisfied. Now, conditional on $U_{n+1} - U_n = b - 1$, the Poissonian construction of the LD event (t, A) implies that A is uniformly chosen among the subsets of $[l]$ with cardinality b , so that rule (R2) is also satisfied. Finally, rule (R3) describing the sharing of new particles amongst old particles was exactly designed to fit the reordering prescribed by the LD process at a LD event.

Furthermore, the initial distribution of the initial configuration $V^k(0)$ is given by

$$\mathbb{P}(\underbrace{1, \dots, 1}_{k-1}).$$

Define

$$B_u^k(n) = \sum_{i=1}^{\lambda(V^k(n))} \mathbf{1}_{V_i^k(n)=u}, \quad u \in [k - 1].$$

Then, according to Proposition 4, $(B^k(n); n \geq 0)$ defines a Λ -urn with initial condition

$$\underbrace{(1, \dots, 1)}_{k-1}.$$

By Proposition 2, there exists a mass partition which is obtained as the limit of the sequence of mass partitions $(B_u^k(n) / \sum_{u=1}^{k-1} B_u^k(n); u \in \{1, \dots, k-1\})$. Denote this mass partition by $\hat{\pi}^k$. Note that, by definition, $\hat{\pi}^k$ is distributed as $\mathcal{D}_{k-1}^\Lambda$. (When $\Lambda = \Lambda(\{0\})\delta_0$, this is the standard Dirichlet mass partition.) We now show that $\hat{\pi}^k$ corresponds to the quantity $\hat{\pi}^{\infty,k}$ introduced in Theorem 6.

Lemma 1. (Large population limit.) *For every $k \geq 2$, $\hat{\pi}^{N,k} \rightarrow \hat{\pi}^k$ a.s. as $N \rightarrow \infty$.*

Proof. Recall that the $\hat{\pi}^{N,k}$ encapsulate the frequencies of the $k-1$ alleles strictly below the fixation line starting at time 0 at level k (uniformly ordered once for all at time 0, regardless of the value of $N \geq k$) at $\hat{T}^{N,k}$, i.e. when the fixation line goes strictly above level N . This can be reformulated by saying that $\hat{\pi}^{N,k}$ is the vector of allele frequencies in the Λ -urn (as defined in the previous paragraph) when the number of balls exceeds N for the first time. Analogously, $\hat{\pi}^k$ was defined as the vector of asymptotic frequencies in the urn. The result then follows by a direct application of Proposition 2. □

Remark 7. Note that $\hat{\pi}^k$ is potentially degenerate.

7.2. LD Configuration at time \hat{T}^k : proof of Theorem 6(i)

As a simple extension of Proposition 4(ii) (by going to the limit), it is not hard to see that, conditional on $\hat{\pi}^k$, the LD process at time \hat{T}^k is obtained by assigning alleles independently for every level according to the distribution $\hat{\pi}^k$. Next, we note that the definition of the fixation line L^{k-1} readily implies that

$$L_t^{k-1} = \inf\{l \in \mathbb{N} : \#\{\text{types carried by levels in } [l] \text{ at time } t\} = k-1\} \quad \text{for all } t \geq 0.$$

Assume now that $\hat{\pi}^k$ is degenerate. The previous two observations imply that $L_{\hat{T}^k}^{k-1} = +\infty$. Since, by assumption, there are no simultaneous extinctions, this yields a contradiction. This completes the proof of the first part of Theorem 6.

Before proceeding with the second part of the proof, we mention the following lemma which is a direct consequence of the previous observations.

Lemma 2. *Under the assumptions of Theorem 6, $\hat{\pi}^k$ is nondegenerate, and, conditional on $\hat{\pi}^k$, the mass partition*

$$B = \left(\sum_{i=1}^{L_{\hat{T}^k}^{k-1}-1} \mathbf{1}_{V_i=u}; u \in [k-2] \right)$$

is identical in law to the random coupon collection associated to $\hat{\pi}^k$.

Also, conditional on B , the particle configuration $(\xi_1, \dots, \xi_{L_{\hat{T}^k}^{k-1}-1})$ is distributed as \mathbb{P}_B .

7.3. Markov property and transition probabilities

Corollary 1. *For every $K \in \mathbb{N}$, $(\hat{\pi}^{K-k}; k \in \{0, \dots, K-2\})$ is a Markov process.*

Proof. The result is obviously true for a finite population, at least if we look at the original (N, Λ) -FV process (not the one defined from the LD process). The proof is completed by going to the limit. □

In order to describe the transition probabilities of the Markov process given in Corollary 1, we fix $k \geq 3$ and describe the distribution of $\hat{\pi}^{k-1}$ given $\hat{\pi}^k$. Define \mathcal{N}^k to be the number of jumps of the fixation line L^{k-1} right before the explosion time \hat{T}^k (see Figure 1), and set

$$\tilde{V}^{k-1}(n) = V^{k-1}(n + \mathcal{N}^k), \quad n \in \mathbb{N}_0.$$

Intuitively, \tilde{V}^{k-1} corresponds to the description of the urn with index $k - 1$ just after the explosion of the urn with index k . By definition, $\hat{\pi}^{k-1}$ is obtained by computing the asymptotic frequencies of the $k - 2$ types present in the shifted process \tilde{V}^{k-1} . Analogously to the proof of Theorem 1 (again see Figure 1), we first note that \tilde{V}^{k-1} is a function of

- (a) the types carried by the levels strictly below level $L_{\hat{T}^k}^{k-1}$ at time \hat{T}^k (in the original LD process), and
- (b) the LD events after time \hat{T}^k .

From Lemma 2, it follows that, conditional on $\hat{\pi}^k$, the process $\tilde{B}^{k-1}(0) := B^{k-1}(\mathcal{N}^k)$ is identical in law to the random coupon collection associated to the nondegenerate $\hat{\pi}^k$, and, furthermore, conditional on $\tilde{B}^{k-1}(0)$, the particle configuration $\tilde{V}^{k-1}(0)$ is distributed as $\mathbb{P}_{\tilde{B}^{k-1}(0)}$.

Using again Proposition 4 and the fact that (b) above is independent of $\hat{\pi}^k$, it follows that, conditional on $\hat{\pi}^k$, $\hat{\pi}^{k-1}$ is identical in law to a Λ -urn whose initial condition can be described in terms of a coupon collection generated from $\hat{\pi}^k$. This is exactly the transition mechanism described in Theorem 6.

Proof of Theorem 6. The result follows on combining Lemma 1, Corollary 1, and the correlation between $\hat{\pi}^k$ and $\hat{\pi}^{k-1}$ exposed above. □

Proof of Proposition 3. (ii) \Rightarrow (i). If \mathcal{D}_k^Λ has exactly k coordinates a.s., Theorem 6 then implies that at extinction time \hat{T}^{k+1} , exactly k types remain (and, furthermore, those types have a strictly positive frequency). Thus we must have $\hat{T}^k > \hat{T}^{k+1}$.

(i) \Rightarrow (ii). This is shown in Section 7.2. □

Acknowledgement

The authors thank the *Center for Interdisciplinary Research in Biology* (Collège de France) for funding.

References

- [1] BERTOIN, J. AND GOLDSCHMIDT, C. (2004). Dual random fragmentation and coagulation and an application to the genealogy of Yule processes. In *Mathematics and Computer Science. III*, Birkhäuser, Basel, pp. 295–308.
- [2] BERTOIN, J. AND LE GALL, J.-F. (2003). Stochastic flows associated to coalescent processes. *Prob. Theory Relat. Fields* **126**, 261–288.
- [3] DELMAS, J.-F., DHERSIN, J.-S. AND SIRI-JEGOUSSE, A. (2010). On the two oldest families for the Wright-Fisher process. *Electron. J. Prob.* **15**, 776–800.
- [4] DONNELLY, P. AND KURTZ, T. G. (1999). Particle representations for measure-valued population models. *Ann. Prob.* **27**, 166–205.
- [5] ETHERIDGE, A. (2011). *Some Mathematical Models from Population Genetics* (Lectures Notes Math. **2012**). Springer, Heidelberg.
- [6] EWENS, W. J. (2012). *Mathematical Population Genetics 1: Theoretical Introduction* (Interdisciplinary Appl. Math. **27**). Springer, New York.
- [7] HÉNARD, O. (2015). The fixation line in the Lambda-coalescent. *Ann. Appl. Prob.* **25**, 3007–3032.
- [8] KINGMAN, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13**, 235–248.
- [9] LABBÉ, C. (2014). From flows of Λ -Fleming-Viot processes to lockdown processes via flows of partitions. *Electron. J. Prob.* **19**, 49pp.

- [10] PFAFFELHUBER, P. AND WAKOLBINGER, A. (2006). The process of most recent common ancestors in an evolving coalescent. *Stoch. Process. Appl.* **116**, 1836–1859.
- [11] PITMAN, J. (1999). Coalescents with multiple collisions. *Ann. Prob.* **27**, 1870–1902.
- [12] ROGERS, L. C. G. AND PITMAN, J. W. (1981). Markov functions. *Ann. Prob.* **9**, 573–582.
- [13] SAGITOV, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Prob.* **36**, 1116–1125.
- [14] SCHWEINSBERG, J. (2000). A necessary and sufficient condition for the Λ -coalescent to come down from infinity. *Electron. Commun. Prob.* **5**, 1–11.

GUILLAUME ACHAZ, *Sorbonne Université and Collège de France*

Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle (MNHN), Sorbonne Université, 75005 Paris, France.

AMAURY LAMBERT, *Sorbonne Université and Collège de France*

Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université, Case courrier 158, 4 Place Jussieu, 75252 Paris, France. Email address: amaury.lambert@upmc.fr

EMMANUEL SCHERTZER, *Sorbonne Université and Collège de France*

Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université, Case courrier 158, 4 Place Jussieu, 75252 Paris, France.