Electronic Journal of Probability

Electron. J. Probab. **25** (2020), no. 56, 1–33.
ISSN: 1083-6489 https://doi.org/10.1214/20-EJP450

Kingman's coalescent with erosion

Félix Foutel-Rodier*† Amaury Lambert*† Emmanuel Schertzer*†

Abstract

Consider the Markov process taking values in the partitions of $\mathbb{N}$ such that each pair of blocks merges at rate one, and each integer is eroded, i.e., becomes a singleton block, at rate d. This is a special case of exchangeable fragmentation-coalescence process, called Kingman's coalescent with erosion. We provide a new construction of the stationary distribution of this process as a sample from a standard flow of bridges. This allows us to give a representation of the asymptotic frequencies of this stationary distribution in terms of a sequence of independent diffusions. Moreover, we introduce a new process valued in the partitions of $\mathbb{Z}$ called Kingman's coalescent with immigration, where pairs of blocks coalesce at rate one, and new blocks of size one immigrate according to a Poisson process of intensity d. By coupling Kingman's coalescents with erosion and with immigration, we are able to show that the size of a block chosen uniformly at random from the stationary distribution of the restriction of Kingman's coalescent with erosion to $\{1, \dots, n\}$ converges as $n \to \infty$ to the total progeny of a critical binary branching process.

Keywords: Kingman's coalescent with erosion; fragmentation-coalescence processes; Kingman's coalescent with immigration; mass-partition; species tree with introgresssion.
AMS MSC 2010: Primary 60G09, Secondary 60G10; 60J25; 60J60; 60J80; 92D15.
Submitted to EJP on July 12, 2019, final version accepted on April 6, 2020.
Supersedes arXiv:1907.05845.
Supersedes HAL:hal-02183351.

1 Introduction

1.1 Motivation

In evolutionary biology, speciation refers to the event when two populations from the same species lose the ability to exchange genetic material, e.g. due to the formation of a

*Laboratoire de Probabilités, Statistique et Modélisation (LPSM), CNRS UMR 8001, Sorbonne Université, Paris, France.
†Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS UMR 7241, INSERM U1050, PSL Research University, Paris, France
E-mail: felix.foutel-rodier@college-de-france.fr
E-mail: amaury.lambert@sorbonne-universite.fr
E-mail: emmanuel.schertzer@sorbonne-universite.fr

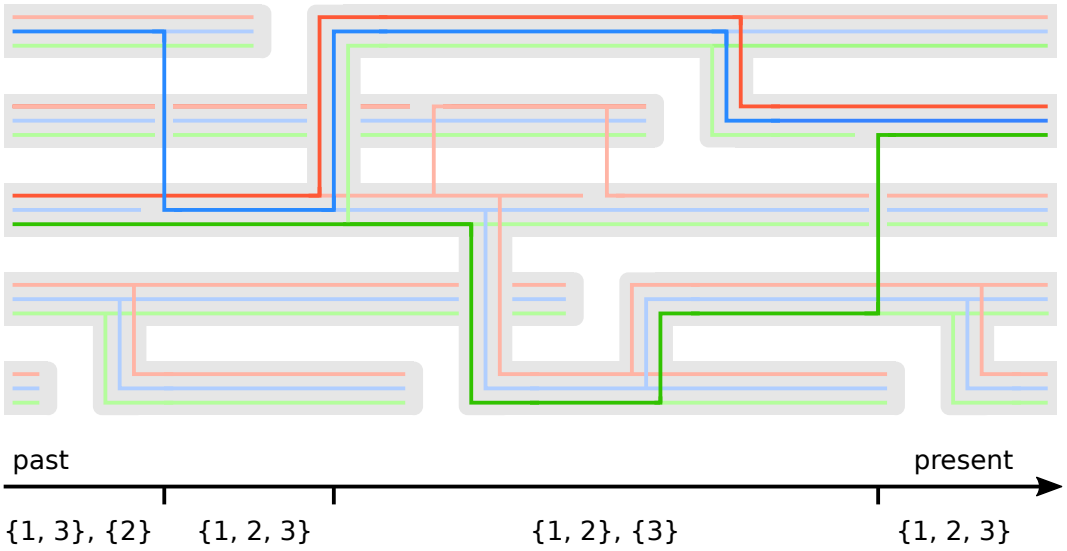

Figure 1: Illustration of the model with $N = 5$ species, represented by grey tubes, and $n = 3$ genes, represented by the colored lines inside the tubes. A species can split into two, simultaneously replicating its genome (speciation). A gene can replicate and move from one species to another and then replace its homologous copy in the recipient species (introgression). At present time a randomly chosen species is sampled: the ancestral lineages of its genes are represented with bolder colors. The green lineage is first subject to an introgression event and jumps to a new species. It is then brought back to the same species as the other genes by a coalescence event. The corresponding partition-valued process obtained by assigning the labels 1, 2 and 3 to the red, blue and green gene respectively is given.

new geographic barrier or accumulation of genetic incompatibilities. Even if speciation is usually thought of as irreversible, related species can often still exchange genetic material through exceptional hybridization, migration events or sudden collapse of a geographic barrier [19]. This can lead to the transmission of chunks of DNA between different species, a phenomenon known as introgression, which is currently considered as a major evolutionary force shaping the genomes of groups of related species [17]. Our study of Kingman's coalescent with erosion was first motivated by the following model of speciation incorporating rare migration events, depicted in Figure 1.

Consider a set of N species, each harboring a genome of n genes indexed by $\{1, \dots, n\}$. We suppose that the species are monomorphic, i.e., that all individuals in the same species carry the same alleles at all genes, and that their dynamics is given by a Moran model: at rate one for each pair of species (s_1, s_2), species s_2 dies, s_1 gives birth to a new species, replicates its genome and sends it into the daughter species. Moreover, we assume that the species are closely related and that they retain the ability to exchange genetic material at exceptional migration events. This effect is incorporated into the model by stating that at rate d for each gene $g \in \{1, \dots, n\}$ and each pair of species s_1 and s_2, the gene g is replicated, the new copy of g is sent from s_1 to s_2 and replaces its homolog in s_2.

The assumption that each migrant transmits at most one gene to the recipient species is strong. A more realistic model should allow any subset of the n genes to be transmitted, at a rate that depends in a complex way on the geometry of the genome due to the biological nature of the recombination process. However, if recombination is sufficiently

strong and if the number of individuals in each species is large, each time a migrant goes from species s_1 to s_2, its genome is rapidly broken into small segments due to frequent back-crosses with the resident. Each of these segments behaves almost independently from the other segments, and has a small probability to reach fixation. Thus, to the first order there should be at most one segment that can reach fixation at a time, as we have assumed.

Now consider a fixed large time T, and sample uniformly one species at that time. We follow backwards in time the ancestral lineages of its genes and the ancestral species to which those genes belong. This induces a process valued in the partitions of $\{1, \dots, n\}$ by declaring that i and j are in the same block at time t if the ancestral lineages of genes i and j sampled at T lie in the same ancestral species at time $T - t$.

At first ($t = 0$), all genes belong to the same ancestral species. Eventually this species receives a successful migrant from another species. Backwards in time, the gene that has been transmitted during this event is removed from its original species and placed in the migrant's original species. Such events occur at rate $(N-1)d$ for each gene, and the migrant species is then chosen uniformly in the population. Once genes belong to separate species, they can be brought back to the same species by coalescence events, corresponding to genome replication in foward time. Any two species find their common ancestor at rate one, and at such an event the genes from the two merging species are placed back into the same species.

This informal description shows that the partition-valued process has two kinds of transitions: each pair of blocks merges at rate one; each gene is placed in a new uniformly chosen species at rate $(N-1)d$. Setting the introgression rate to $d_N = d/N$ and letting $N \to \infty$, introgression events occur at rate d for each gene. At each such event the gene is sent to a new species that does not contain any of the other $n-1$ ancestral gene lineages, i.e., it is placed in a singleton block. This is the description of Kingman's coalescent with erosion, that we now more formally introduce.

1.2 Kingman's coalescent with erosion

Let $n \geq 1$, we define the n-Kingman coalescent with erosion as a Markov process $(\Pi_t^n)_{t\geq 0}$ taking values in the partitions of $[n] := \{1, \dots, n\}$. Its transition rates are the following. Started from a partition π of $[n]$, the process jumps to any partition π' obtained by merging two blocks of π at rate one. Moreover, at rate d for each $i \leq n$, the integer i is "eroded". This means that if C is the block of π containing i, then the process jumps to the partition π' obtained by replacing the block C with the blocks $C \setminus \{i\}$ and $\{i\}$. (Obviously if $C = \{i\}$, i.e., if i is in a singleton block, no such transition can occur.)

Kingman's coalescent with erosion is a special case of the more general class of partition-valued processes called *exchangeable fragmentation-coalescence processes*, introduced and studied in [1]. These processes are a combination of the well-studied fragmentation processes, where blocks can only split, and coalescence processes, where blocks are only allowed to merge. The main new feature of combining fragmentation and coalescence is that they can balance each other so that fragmentation-coalescence processes display non-trivial stationary distributions. In this work we will be interested into describing the stationary distribution associated to Kingman's coalescent with erosion. The following proposition, which is a direct consequence of Theorem 8 of [1], provides the existence and uniqueness of this distribution.

Proposition 1.1 (Berestycki, 2004 [1])**.** *There exists a unique process $(\Pi_t)_{t\geq 0}$ valued in the partitions of $\mathbb{N}$ such that for all $n \geq 1$, the restriction of $(\Pi_t)_{t\geq 0}$ to $[n]$ is distributed as the n-Kingman coalescent with erosion. Moreover, the process $(\Pi_t)_{t\geq 0}$ has a unique stationary distribution Π.*

Kingman's coalescent with erosion is an exchangeable process in the sense that for any finite permutation σ of $\mathbb{N}$,

$$(\sigma(\Pi_t))_{t\geq 0} \overset{(d)}{=} (\Pi_t)_{t\geq 0}.$$

It is then clear that the stationary distribution Π is also an exchangeable partition of $\mathbb{N}$. Exchangeable partitions of $\mathbb{N}$ are often studied through what is known as their asymptotic frequencies. Let $\Pi = (C_1, C_2, \dots)$ be the blocks of the partition Π. Then, Kingman's representation theorem [12] shows that for any i, the following limit exists a.s.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{k\in C_i\}} = f_i.$$

Let $(\beta_i)_{i\geq 1}$ be the non-increasing reordering of the sequence $(f_i)_{i\geq 1}$. We call $(\beta_i)_{i\geq 1}$ the *asymptotic frequencies* of Π. The sequence $(\beta_i)_{i\geq 1}$ is such that

$$\beta_1 \geq \beta_2 \geq \cdots \geq 0, \quad \sum_{i\geq 1} \beta_i \leq 1.$$

Such sequences are called *mass-partitions*. Mass-partitions are interesting because exchangeable partitions are entirely characterized by their asymptotic frequencies. The partition Π can be recovered from its asymptotic frequencies $(\beta_i)_{i\geq 1}$ through what is known as a *paintbox procedure*. Conditional on $(\beta_i)_{i\geq 1}$, let $(X_i)_{i\geq 1}$ be an independent sequence such that for $k \geq 1$, $\mathbb{P}(X_i = k) = \beta_k$, and $\mathbb{P}(X_i = -i) = 1 - \sum_{k\geq 1} \beta_k$. Then the partition Π' of $\mathbb{N}$ defined as

$$i \sim_{\Pi'} j \iff X_i = X_j$$

is distributed as Π [12]. We see that i is in a singleton block iff $X_i = -i$. The set of all singleton blocks is referred to as the *dust* of Π, and the partition Π has dust iff $\sum_{i\geq 1} \beta_i < 1$.

The main characteristics of the asymptotic frequencies of the stationary distribution of fragmentation-coalescence processes have already been derived in [1], see Theorem 8. In the case of Kingman's coalescent with erosion, these results specialize to the following theorem.

Theorem 1.2 (Berestycki, 2004 [1])**.** *Let $(\beta_i)_{i\geq 1}$ be the asymptotic frequencies of Π, the stationary distribution of Kingman's coalescent with erosion. Then*

$$\sum_{i\geq 1} \beta_i = 1, \quad \textit{and} \quad \forall i \geq 1,\ \beta_i > 0, \quad \textit{a.s.}$$

In other words, the partition Π has infinitely many blocks, and no dust.

Before stating our main two results, let us motivate them. Consider a partition $\hat{\Pi}$ obtained from a paintbox procedure on a random mass-partition $(\hat{\beta}_i)_{i\geq 1}$, and denote $\hat{\Pi}^n$ its restriction to $[n]$. There are two sources of randomness in $\hat{\Pi}^n$. One originates from the fact that $(\hat{\beta}_i)_{i\geq 1}$ is random. Moreover, conditional on $(\hat{\beta}_i)_{i\geq 1}$, $\hat{\Pi}^n$ is obtained by sampling a finite number of variables with distribution $(\hat{\beta}_i)_{i\geq 1}$. Thus, in addition to the randomness of $(\hat{\beta}_i)_{i\geq 1}$, $\hat{\Pi}^n$ is subject to a finite sampling randomness.

Suppose that $\hat{\Pi}$ has a finite number of blocks, say N, with asymptotic frequencies $(\hat{\beta}_1, \dots, \hat{\beta}_N)$. When n gets large, the finite sampling effects vanish and the sizes of the blocks of $\hat{\Pi}^n$ resemble $(n\hat{\beta}_1, \dots, n\hat{\beta}_N)$. However, when $\hat{\Pi}$ has infinitely many non-singleton blocks, there always exists a large enough i such that the size of the block with frequency $\hat{\beta}_i$ remains subject to finite sampling effects in $\hat{\Pi}^n$. In this case it is not entirely straightforward to go from the asymptotic frequencies $(\hat{\beta}_i)_{i\geq 1}$ to the size of the blocks of $\hat{\Pi}^n$, as this involves a non-trivial sampling procedure.

In this work our task will be twofold. First, we will investigate the size of the "large blocks" of Π^n by describing the distribution of the asymptotic frequencies $(\beta_i)_{i\geq 1}$. In order to get an insight into the distribution of the "small blocks" of Π^n, we will then also study the empirical distribution of the size of the blocks of Π^n, for large n. Let us now state the corresponding results.

1.3 Main results

We show two main results in this work. One is concerned with the size of the large blocks of the stationary distribution of Kingman's coalescent with erosion, and gives a representation of its asymptotic frequencies in terms of an infinite sequence of independent diffusions. The other is concerned with the size of the small blocks and provides the limit of the distribution of the size of a block chosen uniformly from the stationary partition when n is large. Let us start with the former result.

Size of the large blocks. Let $(Y_i)_{i\geq 1}$ be an i.i.d. sequence of diffusions verifying

$$\forall i \geq 1,\ \mathrm{d}Y_i = (1-Y_i)\mathrm{d}t + \sqrt{Y_i(1-Y_i)}\mathrm{d}W_i,$$

started from 0, and where $(W_i)_{i\geq 1}$ are independent Brownian motions. Each Y_i is distributed as a one-dimensional Wright-Fisher diffusion with mutation, see for example [7], Lemma 4.1. It represents the dynamics of the frequency of a focal allele in a population with constant size, where the mutation rate from any other allele to that focal allele is one, and there are no back mutations, i.e., the mutation rate from the focal allele to any other allele is 0. Moreover, it is known that each Y_i is also distributed as a Wright-Fisher diffusion (without mutation) conditioned on hitting 1, see for instance Proposition 2.3.4 in [15]. Thus we have

$$\forall i \geq 1,\ \lim_{t\to\infty} Y_i(t) = 1 \quad \text{a.s.}$$

Accordingly, we set $Y_i(\infty) = 1$. We build inductively a sequence of processes $(Z_i)_{i\geq 1}$ and time-changes $(\tau_i)_{i\geq 1}$ as follows. Set

$$\forall t \geq 0,\ Z_1(t) = Y_1(t), \quad \tau_1(t) = \int_0^t \frac{1}{1-Z_1(s)}\mathrm{d}s.$$

Then, suppose that $(Z_1,\dots,Z_i)$ and $(\tau_1,\dots,\tau_i)$ have been defined, and set

$$\forall t \geq 0,\ Z_{i+1}(t) = (1-Z_1(t)-\cdots-Z_i(t))Y_{i+1}(\tau_i(t)),$$
$$\forall t \geq 0,\ \tau_{i+1}(t) = \int_0^t \frac{1}{1-Z_1(s)-\cdots-Z_{i+1}(s)}\mathrm{d}s.$$

Then we have the following representation of the asymptotic frequencies of the stationary distribution of Kingman's coalescent with erosion.

Theorem 1.3. *Let $(Z_i)_{i\geq 1}$ be the sequence of diffusions defined previously. Then the non-increasing reordering of the sequence $(z_i)_{i\geq 1}$ defined as*

$$\forall i \geq 1,\ z_i = \int_0^\infty de^{-dt} Z_i(t)\mathrm{d}t,$$

is distributed as the frequencies of the blocks of the stationary distribution of Kingman's coalescent with erosion rate d.

Remark 1.4. Note that the previous result provides a coupling between the stationary distributions of Kingman's coalescent with erosion for various values of the erosion rate d.

Let us explain the intuition behind Theorem 1.3. Kingman's coalescent is dual to a measure-valued process called the Fleming-Viot process [6]. The Fleming-Viot process describes the family size distribution of a population with constant size, while Kingman's coalescent gives the genealogy of that population. By a classical duality argument, Kingman's coalescent at time t can be obtained by sampling individuals at time t from a Fleming-Viot process and placing in the same block those that have the same ancestor [2]. The link with Theorem 1.3 is that the diffusions $(Z_i)_{i\geq 1}$ correspond to the family sizes of the initial individuals of a Fleming-Viot process, ordered by extinction time of their descendance, see Section 5. The integral transformation is roughly due to the fact that in Kingman's coalescent with erosion, one needs to place in the same block the individuals that have the same ancestor at their last erosion event, which is an exponential variable with parameter d. This heuristic argument is made rigorous in Section 5, where Theorem 1.3 is proved.

Size of the small blocks. In order to capture the characteristics of the small blocks of Π^n, we study the empirical measure of the size of the blocks of Π^n. Let M^n be the total number of blocks of Π^n, and let $(|C_1^n|, \dots, |C_{M^n}^n|)$ be their sizes. For each $k \geq 1$, we denote by

$$\mu_k^n = \frac{1}{M^n} \operatorname{Card}(\{i : |C_i^n| = k\})$$

the frequency of blocks of size k. The probability vector $(\mu_k^n)_{k\geq 1}$ is the empirical measure of the size of the blocks of Π^n. We give the following characterization of the asymptotic law of $(\mu_k^n)_{k\geq 1}$ and M^n.

Theorem 1.5. (i) *The following convergence holds in probability*

$$\lim_{n\to\infty} \frac{M^n}{\sqrt{n}} = \sqrt{2d}.$$

(ii) *Moreover, for each $k \geq 1$, the following convergence holds in probability*

$$\lim_{n\to\infty} \mu_k^n = \frac{1}{2^{2k-1}} \frac{1}{k} \binom{2(k-1)}{k-1} = \mathbb{P}(J = k),$$

where J is half the return time to 0 of a simple symmetric random walk.

There is a natural interpretation of the random variable J involved in the previous proposition. Consider a Markov process on $\mathbb{N}$ starting from one that jumps from k to $k+1$ and from k to $k-1$ at rate k. It represents the size of a population where each individual gives birth and dies independently at rate one, and is called a critical binary branching process. Then the total progeny of this process, that is the total number of particles that have lived before the population goes extinct, is distributed as J. Actually, we will show the slightly stronger result that the genealogy of a block sampled uniformly from Kingman's coalescent with erosion is a critical binary branching process, see Remark 4.2.

Remark 1.6. It is interesting to notice that the limiting distribution of the vector $(\mu_k^n)_{k\geq 1}$ is deterministic and does not depend on the erosion coefficient d.

Remark 1.7. The convergence of the vector $(\mu_k^n)_{k\geq 1}$ is equivalent to the convergence in probability of the empirical measure of the size of the blocks of Π^n to the distribution of J in the weak topology.

Kingman's coalescent with immigration. The proof of Theorem 1.5 is based on the following heuristic. Erosion occurs at a rate proportional to the size of the blocks, i.e., a block of size k is eroded at rate dk, while coalescence events do not take the sizes of the blocks into account. As there are only few blocks with large size in Π^n, and many small

blocks, most coalescence events occur between small blocks, while most erosion events occur within these few large blocks. When restricting our attention to small blocks, we can neglect erosion, and consider that pairs of blocks coalesce at rate one, and that new blocks of size one appear at constant rate due to the erosion of the large blocks.

This heuristic led us to consider a process analogous to Kingman's coalescent with erosion, where pairs of blocks coalesce at rate one, but new singleton blocks immigrate according to a Poisson process with rate d. We call this process Kingman's coalescent with immigration, see Section 2.1 for a rigorous definition. We will first prove that the genealogy of a block sampled uniformly from Kingman's coalescent with immigration converges, as the immigration rate goes to infinity, to a critical binary birth-death process, see forthcoming Proposition 3.6. Then, we will use this result and a coupling between Kingman's coalescents with immigration and erosion, described in Section 2.4, to prove Theorem 1.5.

The main focus of the present work is the stationary distribution of Kingman's coalescent with erosion. We only use Kingman's coalescent with immigration to obtain information about this distribution. However, we believe that Kingman's coalescent with immigration is an interesting object in its own right, which could describe the genealogy of entities sampled at distinct time points. In a population genetics interpretation, Kingman's coalescent models the genealogy of genes that are all sampled at the current time. In this case, a new particle that immigrates corresponds to a gene that has been sampled in the past. Such a multi-temporal sampling could occur for example in two situations: in viral phylodynamics [9, 22] and in macroevolution [21, 10]. Viral phylodynamics is a field of evolutionary biology that studies viral phylogenies and their interaction with various characteristics of the underlying epidemics. Viral sequences are often sampled at several timepoints, corresponding for example to different viral outbreaks. Macroevolution studies the evolutionary history of speciation, extinction. In this context, fossil data correspond to remainders of individuals that have lived and been sampled at some time point in the far past.

Outline. The remainder of the paper is organized as follows. In Section 2 we provide two constructions of Kingman's coalescent with immigration, as well as a coupling between Kingman's coalescents with erosion and immigration. Section 3 is then devoted to giving the genealogy of the blocks of Kingman's coalescent with immigration. The main result of this section is Proposition 3.1, which is the reformulation of Theorem 1.5 in the immigration case. In Section 4, we use Proposition 3.1 and the coupling between Kingman's coalescents with erosion and immigration to prove Theorem 1.5. Finally, we prove Theorem 1.3 in Section 5.

Possible extensions. As we have mentioned, Kingman's coalescent is part of the more general class of fragmentation-coalescence processes. We now briefly discuss potential extensions of our results to such processes.

The main ingredient of our study of the size of small blocks is that fragmentation is faster for larger blocks, while coalescence occurs at the same speed regardless of the size of the blocks. This allows us to neglect fragmentation and consider a purely coalescing system where new blocks immigrate due to the fragmentation of the large blocks. This picture remains valid for Λ-coalescents with erosion, but the proofs would be more involved because computations could no longer be made explictly. Morever, we believe that this picture also remains valid for a broad class of binary fragmentation measures. The particles that are removed from the large blocks would no longer be of size one, but should not have time to split on the time-scale when small blocks are formed, yielding a situation similar to the erosion case.

Theorem 1.3 relies on a construction of the stationary distribution of Kingman's coalescent with erosion from a Fleming-Viot process that can be directly generalized to Λ-coalescents with erosion (and even to Ξ-coalescents with erosion) by using the corresponding Λ-Fleming-Viot process. However, the explicit expression of the size of the blocks in terms of independent diffusions cannot be achieved in general. Nevertheless see the end of Section 5 for a discussion of a possible extension of Theorem 1.3 to Beta-coalescents with erosion.

Overall, the techniques and ideas we use in this work are not entirely specific to Kingman's coalescent with erosion. Nevertheless, in this case, the proofs are greatly simplified because all calculations can be made explicitly. This reason led us to restrict our attention to Kingman's coalescent with erosion in this work, and to leave possible extensions for future work.

2 Kingman's coalescent with immigration

In this section we construct Kingman's coalescent with immigration as a partition-valued process such that pairs of blocks coalesce at rate one and new blocks immigrate at rate d. Then, we give an alternative construction of Kingman's coalescent with erosion from the flow of bridges of [2]. Finally, the coupling between Kingman's coalescents with erosion and with immigration is carried out in Section 2.4.

2.1 Definition

Consider a Poisson point process on $\mathbb{R}$ with intensity $d\mathrm{d}t$, and let $(T_i)_{i\in\mathbb{Z}}$ be its atoms labeled in increasing order such that $T_0 < 0 < T_1$. The sequence $(T_i)_{i\in\mathbb{Z}}$ corresponds to the immigration times of new particles in the system.

Fix $N \in \mathbb{Z}$, we will first define Kingman's coalescent with immigration for the particles that have a label larger that N, and then extend it to all particles by consistency. We do that in the following way. Initially, set

$$\forall t < T_N,\ \bar{\Pi}_t^N = \emptyset.$$

We now extend $\bar{\Pi}_t^N$ to all real times by induction. Suppose that $\bar{\Pi}_t^N$ has been defined on $(-\infty, T_k)$, for $k \geq N$. We first set

$$\bar{\Pi}_{T_k}^N = \bar{\Pi}_{T_k-}^N \cup \{k\}$$

to represent the immigration of the new particle with label k. We now let each pair of blocks of $\bar{\Pi}_t^N$ coalesce at rate one for $T_k \leq t < T_{k+1}$. One can achieve this by considering, conditional on $\left\{\bar{\Pi}_{T_k}^N = \bar{\pi}_k\right\}$, an independent version $(\Pi_t^k)_{t\geq 0}$ of Kingman's coalescent started from $\bar{\pi}_k$, and setting

$$\forall t < T_{k+1} - T_k,\ \bar{\Pi}_{T_k+t}^N = \Pi_t^k.$$

We say that the process $(\bar{\Pi}_t^N)_{t\in\mathbb{R}}$ is the *N-Kingman coalescent with immigration rate d*. The following proposition shows that we can extend consistently the N-Kingman's coalescent with immigration to a process taking its values in the partitions of $\mathbb{Z}$.

Proposition 2.1. (i) *There exists a unique process $(\bar{\Pi}_t)_{t\in\mathbb{R}}$, called Kingman's coalescent with immigration rate d, such that for all $N \in \mathbb{Z}$, its restriction to $\{i \in \mathbb{Z} : i \geq N\}$ is distributed as the N-Kingman coalescent with immigration.*

(ii) *With probability one, $\bar{\Pi}_t$ has finitely many blocks for all $t \in \mathbb{R}$.*

Proof. (i) Let $(\bar{\Pi}_t^N)_{t\in\mathbb{R}}$ be a N-Kingman's coalescent with immigration. It is sufficient to show that the restriction $(\bar{\Pi}_t^{N+1})_{t\in\mathbb{R}}$ of $(\bar{\Pi}_t^N)_{t\in\mathbb{R}}$ to $\{i \in \mathbb{Z} : i \geq N+1\}$ is distributed

as a $N+1$-Kingman's coalescent with immigration, and the result will follow from Kolmogorov's extension theorem. Obviously, the immigration times of $(\bar{\Pi}_t^{N+1})_{t\in\mathbb{R}}$ have the desired distribution. The result is now a simple consequence of the sampling consistency of Kingman's coalescent.

(ii) Let us now prove the second point. Kingman's coalescent has the property of coming down from infinity [13]. This means that even if Kingman's coalescent is started from a partition with an infinite number of blocks, then for all positive times it will have only finitely many blocks. Thus, as the number of immigrated particles is locally finite, Kingman's coalescent with immigration only has a finite number of blocks for all times a.s. □

In the remainder of this work we will make use of the process counting the number of blocks of Kingman's coalescent with immigration. More formally, for $t \in \mathbb{R}$, we define M_t as the (finite) number of blocks of $\bar{\Pi}_t$.

2.2 Preliminaries on flows of bridges

The previous construction of the Kingman coalescent with immigration is based on Kolmogorov's extension theorem. The aim of the next two sections is to give an alternative construction of Kingman's coalescent with immigration based on the flow of bridges of [2]. This construction will only be needed in Section 4 for the proof of Theorem 1.3. In this section we recall the material on flows of bridges that will be needed.

Bridges. We call a bridge [2] any random function of the form

$$\forall u \in [0,1],\ B(u) = (1-\sum_{i\geq 1}\beta_i)u + \sum_{i\geq 1}\beta_i \mathbb{1}_{\{u\geq V_i\}},$$

for some random mass-partition $(\beta_i)_{i\geq 1}$ and an independent i.i.d. sequence of uniform $[0,1]$ variables $(V_i)_{i\geq 1}$. For a bridge B, we define its inverse B^{-1} as

$$\forall u \in [0,1),\ B^{-1}(u) = \inf\{t\in[0,1] : B(t) > u\}, \quad B^{-1}(1) = 1.$$

Let $(U_i)_{i\geq 1}$ be a sequence of i.i.d. uniform variables. An exchangeable partition $\hat{\Pi}$ of $\mathbb{N}$ can be obtained from B and $(U_i)_{i\geq 1}$ by setting

$$i \sim_{\hat{\Pi}} j \iff B^{-1}(U_i) = B^{-1}(U_j).$$

Let $(C_1, C_2, \dots)$ be the blocks of $\hat{\Pi}$ labeled in decreasing order of their least elements, i.e., such that

$$i \leq j \iff \min(C_i) \leq \min(C_j).$$

To each block C_i is associated a unique random variable V_i' defined as

$$\forall j \in C_i,\ V_i' = B^{-1}(U_j).$$

If $\hat{\Pi}$ has finitely many blocks, say M, for $i > M$ we set $V_i' = \tilde{V}_i'$ where $(\tilde{V}_i')_{i\geq 1}$ is an independent sequence of i.i.d. uniform random variables. The sequence $(V_i')_{i\geq 1}$ will be referred to as the sequence of ancestors of the blocks of $\hat{\Pi}$. The key results on bridges from [2] is their Lemma 2 that we state here for later use.

Lemma 2.2 (Bertoin and Le Gall, 2003 [2])**.** *Consider a bridge B, and let $\hat{\Pi}$ and (V_i') be respectively the partition and sequence of ancestors obtained from B as above. Then $(V_i')_{i\geq 1}$ is independent of $\hat{\Pi}$, and $(V_i')_{i\geq 1}$ is a sequence of i.i.d. uniform variables.*

The standard flow of bridges. A flow of bridges is defined as follows.

Definition 2.3. *A flow of bridges is a family of bridges $(B_{s,t})_{s\leq t}$ such that:*

(i) *For any $s \leq u \leq t$, we have $B_{s,u} \circ B_{u,t} = B_{s,t}$.*

(ii) *For $t_1 \leq \cdots \leq t_p$, the bridges $B_{t_1,t_2}, \ldots, B_{t_{p-1},t_p}$ are independent, and B_{t_1,t_2} is distributed as B_{0,t_2-t_1}.*

(iii) *The limit $B_{0,t} \to \mathrm{Id}$ as $t \downarrow 0$ holds in probability in the Skorohod space.*

A flow of bridges encodes the dynamics of a population represented by the interval $[0,1]$. Let $t \in \mathbb{R}$ and $x < y$. If the interval $[x,y]$ is interpreted as a subfamily of the population at time t, then its progeny at time $s \leq t$ is represented by the interval $[B_{s,t}(x-), B_{s,t}(y)]$. (Notice that time is going backward: if t is the present, then $s \leq t$ represents the future of the population.)

By the independence and stationarity of the increments of the flow, the distribution of a flow of bridges is entirely characterized by the distribution of $B_{0,t}$, for $t \geq 0$. We will be particularly interested in the so-called *standard flow of bridges*, that can be described as follows. Let $t \geq 0$ and consider the bridge

$$\forall u \in [0,1],\ B_{0,t}(u) = \sum_{i=1}^{N_t} \beta_i \mathbb{1}_{\{V_i \leq u\}},$$

where

(i) The process $(N_t)_{t\geq 0}$ is distributed as a pure-death process started at ∞, and going from k to $k-1$ at rate $k(k-1)/2$.

(ii) Conditional on N_t, $(\beta_1, \ldots, \beta_{N_t})$ has a Dirichlet distribution with parameter $(1, \ldots, 1)$.

(iii) The variables $(V_i)_{i\geq 1}$ is an independent i.i.d. sequence of uniform variables.

Then we know [2] that there exists a flow of bridges $(B_{s,t})_{s\leq t}$ such that $B_{0,t}$ is distributed as above. It is called the standard flow of bridges.

Our interest in the standard flow of bridges is that is represents the dynamics of a population whose genealogy is given by Kingman's coalescent. Let $(U_i)_{i\geq 1}$ be a sequence of i.i.d. uniform variables, and let $\hat{\Pi}_t$ be the partition obtained from the bridge $B_{0,t}$ and the sequence $(U_i)_{i\geq 1}$. We stress that the *same* sequence is used for all t. Then the process $(\hat{\Pi}_t)_{t\geq 0}$ is distributed as Kingman's coalescent started from the partition of $\mathbb{N}$ into singletons [2].

The Fleming-Viot process. One of the main advantages of flows of bridges is that they couple a backward process, giving the genealogy of the population, and a forward process, giving the size of the progeny of the individuals in the population. This forward process is often encoded as a measure-valued process known as a Fleming-Viot process.

Let $(B_{s,t})_{s\leq t}$ be a standard flow of bridges. For each $t \geq 0$, $B_{-t,0}$ is the distribution function of some random measure ρ_t on $[0,1]$. The measure-valued process $(\rho_t)_{t\geq 0}$ is called a Fleming-Viot process [6]. A well-known fact that we will use is that the dynamics of the mass of n fixed disjoint intervals is distributed as the n first coordinates of a $(n+1)$-dimensional Wright-Fisher diffusion. More precisely, let $(I_1, \ldots, I_n)$ be n disjoint intervals, and define

$$\forall i \in \{1, \ldots, n\},\ \forall t \geq 0,\ X_i(t) = \rho_t(I_i)$$

and

$$\forall t \geq 0,\ X_{n+1}(t) = 1 - (X_1(t) + \cdots + X_n(t)).$$

Then, if we denote by $(|I_1|, \dots, |I_n|)$ the lengths of the intervals $(I_1, \dots, I_n)$, the process $(X_1, \dots, X_{n+1})$ is distributed as the unique solution to

$$\forall i \in \{1, \dots, n+1\},\ \mathrm{d}X'_i = \sum_{\substack{j=1 \\ j \neq i}}^{n+1} \sqrt{X'_i X'_j} \mathrm{d}W'_{i,j},$$

started from $(|I_1|, \dots, |I_n|, 1 - |I_1| - \dots - |I_n|)$, where $(W_{i,j})_{i<j}$ are independent Brownian motions and $W_{i,j} = -W_{j,i}$.

2.3 A flow of bridges construction of Kingman's coalescent with immigration

Let $(B_{s,t})_{s \leq t}$ be a standard flow of bridges. We now construct a version of Kingman's coalescent with immigration from $(B_{s,t})_{s \leq t}$. Consider a Poisson point process on $\mathbb{R} \times [0,1]$ with intensity $d\mathrm{d}t \otimes \mathrm{d}x$, and let $(T_i, U_i)_{i \in \mathbb{Z}}$ be its atoms, labeled in increasing order of their first coordinate such that $T_0 < 0 < T_1$. Similarly to Section 2.1, the times $(T_i)_{i \in \mathbb{Z}}$ correspond to immigration times of new particles. Here the sequence $(U_i)_{i \in \mathbb{Z}}$ represents the location in the population of these immigrated particles.

For each $t \in \mathbb{R}$, we define a partition $\bar{\Pi}_t$ of $\{i \in \mathbb{Z} : T_i \leq t\}$ by setting

$$i \sim_{\bar{\Pi}_t} j \iff B^{-1}_{T_i,t}(U_i) = B^{-1}_{T_j,t}(U_j).$$

The following proposition shows that $(\bar{\Pi}_t)_{t \in \mathbb{R}}$ is distributed as Kingman's coalescent with immigration.

Proposition 2.4. *The process $(\bar{\Pi}_t)_{t \in \mathbb{R}}$ defined from the flow of bridges is a version of Kingman's coalescent with immigration rate d.*

Proof. The proof is almost identical to the proof of Corollary 1 of [2]. The main difference is that here the flow of bridges is sampled at various times $(T_i)_{i \in \mathbb{Z}}$ while for the classical Kingman coalescent, the flow of bridges is only sampled at an initial time.

We work conditional on $(T_i)_{i \in \mathbb{Z}}$ and consider these times as fixed. Let $(\bar{\Pi}^N_t)_{t \in \mathbb{R}}$ be the restriction of $(\bar{\Pi}_t)_{t \in \mathbb{R}}$ to $\{i \in \mathbb{Z} : i \geq N\}$. It is sufficient to show that for all $N \in \mathbb{Z}$ the blocks of $(\bar{\Pi}^N_t)_{t \in \mathbb{R}}$ coalesce according to independent versions of Kingman's coalescent between immigration times.

Let $t \in \mathbb{R}$, and let $(C_1, \dots, C_{M_t})$ be the blocks of $\bar{\Pi}^N_t$, where M_t is the number of blocks, and where the blocks are labeled such that

$$i \leq j \iff \min(C_i) \leq \min(C_j).$$

Similarly to Section 2.2, we can define the sequence of ancestors of $\bar{\Pi}^N_t$ by setting

$$\forall j \in C_i,\ V'_i = B^{-1}_{T_j,t}(U_j),$$

and supplementing it with an independent sequence of i.i.d. uniform variables $(\tilde{V}'_i)_{i \geq 1}$, i.e., defining $\forall i > M_t$, $V'_i = \tilde{V}'_i$.

Let us show by induction that for all $k \geq N$,

(i) The ancestors $(V^{(k)}_i)_{i \geq 1}$ of $\bar{\Pi}^N_{T_k}$ are i.i.d. with uniform distribution.

(ii) The sequence $(V^{(k)}_i)_{i \geq 1}$ is independent of $(\bar{\Pi}^N_t)_{t \leq T_k}$.

(iii) $(\bar{\Pi}^N_t)_{t \leq T_k}$ is a version of the N-Kingman coalescent with immigration.

Fix $T_k \le t_1 < \cdots < t_{p+1} \le T_{k+1}$. By induction on p we can suppose that the sequence of ancestors of $\bar{\Pi}^N_{t_p}$, denoted by $(V_i^{(t_p)})_{i\ge1}$, is independent of $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_p}\big)$. Then (i) and (ii) are proved if we can show that the sequence of ancestors of $\bar{\Pi}^N_{t_{p+1}}$ is independent of $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_{p+1}}\big)$, and is a sequence of i.i.d. uniform variables.

Let us now call Π^* the partition obtained from the bridge $B_{t_p,t_{p+1}}$ and the sequence $(V_i^{(t_p)})_{i\ge1}$, i.e.,

$$i \sim_{\Pi^*} j \iff B^{-1}_{t_p,t_{p+1}}(V_i^{(t_p)}) = B^{-1}_{t_p,t_{p+1}}(V_j^{(t_p)}),$$

and let $(V_i^*)_{i\ge1}$ be the sequence of ancestors of Π^*, i.e.,

$$\forall j \in C_i^*,\ V_i^* = B^{-1}_{t_p,t_{p+1}}(V_j^{(t_p)}),$$

where $(C_1^*, C_2^*, \dots)$ denote the blocks of Π^* labeled in increasing order of their minimal elements as above. Using the fact that for $u \le s \le t$, $B^{-1}_{u,t} = B^{-1}_{s,t} \circ B^{-1}_{u,s}$, we get that for all $N \le i, j \le k$,

$$\begin{aligned} i \sim_{\bar{\Pi}_{t_{p+1}}} j &\iff B^{-1}_{t_p,t_{p+1}}(B^{-1}_{T_i,t_p}(U_i)) = B^{-1}_{t_p,t_{p+1}}(B^{-1}_{T_j,t_p}(U_j)) \\ &\iff B^{-1}_{t_p,t_{p+1}}(V^{(t_p)}_{b(i)}) = B^{-1}_{t_p,t_{p+1}}(V^{(t_p)}_{b(j)}) \\ &\iff b(i) \sim_{\Pi^*} b(j) \end{aligned} \tag{2.1}$$

where $b(i)$ denotes the label of the block of $\bar{\Pi}^N_{t_p}$ to which i belongs.

By independence of the increments of the flow of bridges, the bridge $B_{t_p,t_{p+1}}$ is independent of the collection of variables $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_p}, (V_i^{(t_p)})_{i\ge1}\big)$. Thus, $(B_{t_p,t_{p+1}}, (V_i^{(t_p)})_{i\ge1})$ are independent of $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_p}\big)$, and hence $(\Pi^*, (V_i^*)_{i\ge1})$ are independent of $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_p}\big)$. Using Lemma 2.2, we get that Π^* is independent of $(V_i^*)_{i\ge1}$. This shows that $(V_i^*)_{i\ge1}$ is independent of $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_p}, \Pi^*\big)$. Using (2.1), we see that $\bar{\Pi}^N_{t_{p+1}}$ can be recovered from $\bar{\Pi}^N_{t_p}$ and Π^*. Thus, the variables $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_{p+1}}\big)$ are independent of $(V_i^*)_{i\ge1}$.

In order to end the proof of the claim we need to distinguish two cases. First, suppose that $t_{p+1} < T_{k+1}$. Then, due to our labeling convention, we have that $(V_i^*)_{i\ge1} = (V_i^{(t_{p+1})})_{i\ge1}$ (up to the auxiliary variables $(\tilde{V}_i)_{i\ge1}$ that play no role). Conversely, if $t_{p+1} = T_{k+1}$, then one of the variables $(V_i^*)_{i\ge1}$ has to be replaced by the ancestor U_{k+1} of the block $\{k+1\}$. More precisely, if $\bar{\Pi}^N_{T_{k+1}}$ has M_{k+1} blocks, again by labeling convention, the block $\{k+1\}$ has label M_{k+1}. Thus, $(V_i^{(t_{p+1})})_{i\ge1}$ is recovered by setting $V_i^{(t_{p+1})} = V_i^*$ for $i \ne M_{k+1}$, and $V_i^{(t_{p+1})} = U_{k+1}$ for $i = M_{k+1}$. It is straightforward to see that, as U_{k+1} is independent of all other variables, $(V_i^{(t_{p+1})})_{i\ge1}$ remains a sequence of i.i.d. of uniform variables, independent of $\big((\bar{\Pi}^N_t)_{t\le T_k}, \bar{\Pi}^N_{t_1}, \dots, \bar{\Pi}^N_{t_{p+1}}\big)$ and thus the fact that points (i) and (ii) of the claim hold.

For $k \ge N$ and $t < T_{k+1} - T_k$ consider the partition Π^k_t of $\mathbb{N}$ defined as

$$i \sim_{\Pi^k_t} j \iff B^{-1}_{T_k,T_k+t}(V_i^{(k)}) = B^{-1}_{T_k,T_k+t}(V_j^{(k)})$$

As the sequence $(V_i^{(k)})_{i\ge1}$ is i.i.d. uniform and independent of $(\bar{\Pi}^N_t)_{t\le T_k}$, the process $(\Pi^k_t)_{t\ge0}$ is a version of Kingman's coalescent started from the partition into singletons, independent of $(\Pi^k_t)_{t\ge0}$. Using equation (2.1), we have that

$$i \sim_{\bar{\Pi}^N_{T_k+t}} j \iff b(i) \sim_{\Pi^k_t} b(j),$$

where $b(i)$ denotes the label of the block of $\bar{\Pi}^N_{T_k}$ to which i belongs. In other words, $(\bar{\Pi}^N_{T_k+t})_{0\le t<T_{k+1}-T_k}$ is obtained by letting the blocks of $\bar{\Pi}^N_{T_k}$ coalesce according to an independent version of Kingman's coalescent. This proves that $(\bar{\Pi}^N_t)_{t\le T_{k+1}}$ is distributed as a N-Kingman coalescent with immigration, and ends the proof of the result. □

2.4 Coupling erosion and immigration

We now explain the coupling between Kingman's coalescents with erosion and with immigration. Let $n \geq 1$, consider a Poisson point process P^n on $\mathbb{R}$ with intensity $nd\mathrm{d}t$ and let $(T_i)_{i\in\mathbb{Z}}$ be its atoms ordered increasingly such that $T_0 < 0 < T_1$. To each atom of the process we attach a uniform mark in $[n]$. We denote by ℓ_i the mark attached to T_i, so that $(\ell_i)_{i\in\mathbb{Z}}$ is a sequence of i.i.d. uniform variables on $[n]$.

Consider $t \in \mathbb{R}$. For each $k \in [n]$, let $\varphi_t(k)$ be the label of the last atom of P^n with mark k before time t, i.e., $\varphi_t(k) \in \mathbb{Z}$ is the unique i such that $\ell_i = k$ and there is no atom T of P^n with $T_i < T \leq t$ carrying mark k. Let $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ be Kingman's coalescent with immigration rate nd built from the Poisson process $(T_i)_{i\in\mathbb{Z}}$ as in Section 2.1. We define a partition Π_t^n of $[n]$ by setting

$$i \sim_{\Pi_t^n} j \iff \varphi_t(i) \sim_{\bar{\Pi}_t} \varphi_t(j).$$

In words, i and j belong to the same block of Π_t^n iff the most recently immigrated particles of $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ with marks i and j have coalesced before time t. The key point of this construction is that $(\Pi_t^n)_{t\in\mathbb{R}}$ is distributed as Kingman's coalescent with erosion.

Proposition 2.5. *The process $(\Pi_t^n)_{t\in\mathbb{R}}$ is a stationary version of the n-Kingman coalescent with erosion rate d.*

Proof. Let $k \in [n]$. By thinning, the set of atoms of P^n with mark k is a Poisson process on $\mathbb{R}$ with intensity $d\mathrm{d}t$, and these processes are independent. Thus new atoms of P^n with mark k arrive at rate d. Let us consider what happens at such an arrival time. Suppose that $\ell_i = k$. Then, by definition, we have $\varphi_{T_i}(k) = i$, as the atom T_i has mark k. Moreover, the particle i is a singleton of the partition $\bar{\Pi}_{T_i}$ (it is the particle that has newly immigrated). Thus at time T_i, the integer k is removed from its block and placed in a singleton block. This is the description of an erosion event, which occur at rate d.

Let us now describe the dynamics between two immigration times, say T_i and T_{i+1}. Conditional on $\bar{\Pi}_{T_i}$, the blocks of $(\bar{\Pi}_t)_{T_i\leq t\leq T_{i+1}}$ coalesce according to an independent version of Kingman's coalescent started from $\bar{\Pi}_{T_i}$. The labels of the atoms of P^n that are the last atoms with their marks form a subset of $\{j \in \mathbb{Z} : j \leq i\}$, say L. By sampling consistency of Kingman's coalescent, the restriction of $(\bar{\Pi}_t)_{T_i\leq t\leq T_{i+1}}$ to L is also distributed as Kingman's coalescent, starting from the restriction of $\bar{\Pi}_{T_i}$ to L. Thus, as the blocks of $(\Pi_t^n)_{T_i\leq t\leq T_{i+1}}$ are, up to an independent relabeling, the blocks of the restriction of $(\bar{\Pi}_t)_{T_i\leq t\leq T_{i+1}}$ to L, any two pairs of blocks of $(\Pi_t)_{T_i\leq t\leq T_{i+1}}$ coalesce at rate one.

The fact that $(\Pi_t)_{t\in\mathbb{R}}$ is stationary follows from the stationarity of the Poisson point process. □

Combined with the construction of Kingman's coalescent with immigration from the standard flow of bridges, this coupling gives an interesting construction of the stationary distribution of Kingman's coalescent with erosion.

Corollary 2.6. *Let $(B_{s,t})_{s\leq t}$ be a standard flow of bridges, $(T_i)_{i\geq 1}$ be an independent sequence of i.i.d. exponential variables with parameter d, and $(U_i)_{i\geq 1}$ be an independent sequence of i.i.d. uniform variables. Then the partition Π defined by*

$$i \sim_{\Pi} j \iff B^{-1}_{-T_i,0}(U_i) = B^{-1}_{-T_j,0}(U_j)$$

has the stationary distribution of Kingman's coalescent with erosion rate d.

Proof. Consider a Poisson process P^n on $\mathbb{R} \times [0,1]$ with intensity $nd\mathrm{d}t \otimes \mathrm{d}x$, and attach to each atom of P^n a uniform mark on $[n]$. If (T_i, U_i) denotes the last atom of P^n with mark i before $t = 0$, then T_i is exponentially distributed with parameter d, U_i is uniform

on $[0,1]$, and all these variables are independent. A combination of Proposition 2.5 and Proposition 2.4 now proves the result. □

Remark 2.7. The construction of Kingman's coalescent with immigration from Section 2.1 and the construction with the flow of bridges of Section 2.3 only rely on the sampling consistency of Kingman's coalescent. These constructions could be extended directly to a case where the coalescence events occur according to a Λ-coalescent [18, 20]. In particular, the construction of the stationary distribution of Kingman's coalescent with erosion of Corollary 2.6 extends directly to Λ-coalescents with erosion if one replaces the standard flow of bridges by the corresponding Λ-flow of bridges.

3 Size of the blocks of Kingman's coalescent with immigration

In this section we study Kingman's coalescent with immigration. The main result we will show is the following.

Proposition 3.1. *Let $n \geq 1$ and consider $(\bar{\Pi}_t^n)_{t\in\mathbb{R}}$ a version of Kingman's coalescent with immigration rate nd. Let $(|\bar{C}_1^n|, \dots, |\bar{C}_p^n|)$ be the size of p blocks chosen uniformly from $\bar{\Pi}_0^n$, then*

$$(|\bar{C}_1^n|, \dots, |\bar{C}_p^n|) \Longrightarrow (J_1, \dots, J_p)$$

where $(J_1, \dots, J_p)$ are i.i.d. variables distributed as the total progeny of a critical binary branching process.

We prove this result by choosing k blocks uniformly from $\bar{\Pi}_0^n$, and counting backwards in time the number of blocks that are ancestors of these blocks, i.e., that will further coalesce to form these blocks. We show that this process converges, under appropriate scaling, to k independent critical binary branching processes, yielding the result. In this section we work in both directions of time. We will index time by t when it is flowing forward, and by s when it is flowing backwards.

We first give a precise definition of the ancestral process counting the number of blocks in Section 3.1, along with its basic properties. The convergence is then carried out in Section 3.2.

3.1 The ancestral process

Let $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ be a version of Kingman's coalescent with immigration rate d. The process $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ is naturally endowed with a notion of ancestry between its blocks. For $t \in \mathbb{R}$, let M_t be the number of blocks of $\bar{\Pi}_t$. Let $(\bar{C}_1, \dots, \bar{C}_{M_t})$ be an enumeration of the blocks of $\bar{\Pi}_t$. We say that this enumeration is exchangeable if conditional on $\{M_t = k\}$, for any permutation σ of $[k]$,

$$(\bar{C}_1, \dots, \bar{C}_k) \overset{(d)}{=} (\bar{C}_{\sigma(1)}, \dots, \bar{C}_{\sigma(k)}).$$

We can always consider an exchangeable enumeration of the blocks of $\bar{\Pi}_t$ by changing the labels of any enumeration according to an independent uniform permutation.

For $u \leq t$, consider $\bar{\Pi}_t = (\bar{C}_1, \dots, \bar{C}_{M_t})$ and $\bar{\Pi}_u = (\bar{C}'_1, \dots, \bar{C}'_{M_s})$ an enumeration of the blocks of $\bar{\Pi}_t$ and $\bar{\Pi}_u$ respectively. In Kingman's coalescent with immigration, a block present at time u can only coalesce with other blocks. Thus, for any block $\bar{C}'_i$, there is a unique block $\bar{C}_j$ of $\bar{\Pi}_t$ such that $\bar{C}'_i \subseteq \bar{C}_j$. We say that $\bar{C}'_i$ is an ancestor of $\bar{C}_j$. We define the ancestral process of Kingman's coalescent with immigration as the vector counting the number of ancestors of the blocks of $\bar{\Pi}_0$, enumerated in an exchangeable way. This definition is illustrated in Figure 2.

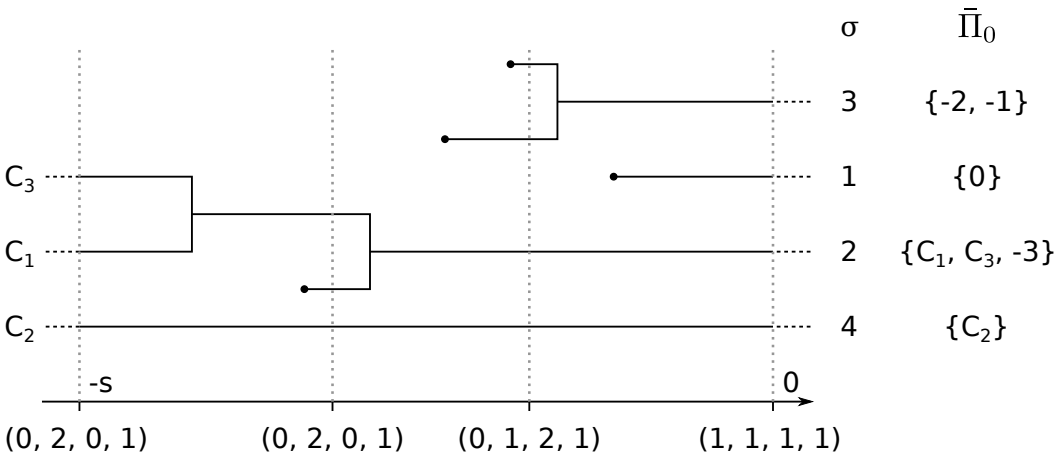

Figure 2: In this example, we have $\bar{\Pi}_{-s} = (C_1, C_2, C_3)$. Each black circle represents an immigration event, and the lines merge at the coalescence time of the blocks to which they correspond. At $s = 0$ the blocks of $\bar{\Pi}_0$ are labeled according to the permutation σ, and the value of $(\mathcal{A}_s)_{s\geq 0}$ is given below for some times.

Definition 3.2. *Let $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ be Kingman's coalescent with immigration, and denote by $(\bar{C}_1, \dots, \bar{C}_{M_0})$ the blocks of $\bar{\Pi}_0$ enumerated in an exchangeable order. For $s \geq 0$, let $(\bar{C}'_1, \dots, \bar{C}'_{M_{-s}})$ be the blocks of $\bar{\Pi}_{-s}$. We define the number of ancestors of the i-th block as*

$$\mathcal{A}_s(i) = \begin{cases} \mathrm{Card}\{j \leq M_{-s} : \bar{C}'_j \subseteq \bar{C}_i\} & \text{if } i \in \{1, \dots, M_0\} \\ 0 & \text{if } i > M_0. \end{cases}$$

The process $(\mathcal{A}_s)_{s\geq 0}$ defined as $\mathcal{A}_s := (\mathcal{A}_s(1), \mathcal{A}_s(2), \dots)$ is called the ancestral process associated to $(\bar{\Pi}_t)_{t\in\mathbb{R}}$.

The process $(\mathcal{A}_s)_{s\geq 0}$ can be seen as a particle system where at time 0, there are M_0 particles with distinct types, and $(\mathcal{A}_s(i))_{s\geq 0}$ records the number of particles with type i. As we have reversed time, each coalescence event now corresponds to the birth of a new particle, and each immigration event to the death of a particle.

Note that relative to the original population model described in the introduction, we have now reversed the time twice. As Kingman's coalescents with erosion and immigration represent genealogies, the future of these processes corresponds to the past of the population. Therefore, the "ancestors" of the blocks of Kingman's coalescent with immigration actually correspond to the descendants of these individuals in the population point of view.

Recall that $(M_t)_{t\in\mathbb{R}}$ stands for the number of blocks of $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ forward in time. For each $s \in \mathbb{R}$, we define $N_s := M_{-s}$, the number of blocks of $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ backwards in time. The process $(N_s)_{s\geq 0}$ also gives the number of particles of the ancestral process $(\mathcal{A}_s)_{s\geq 0}$, that is we have

$$\forall s \geq 0,\ N_s = \sum_{i\geq 1} \mathcal{A}_s(i).$$

The following proposition shows that the ancestral process is Markovian. This is a key feature that makes Kingman's coalescent with immigration easier to study than Kingman's coalescent with erosion.

Proposition 3.3. *Let $(\mathcal{A}_s)_{s\geq 0}$ be the ancestral process associated to Kingman's coalescent with immigration rate d, and let $(N_s)_{s\geq 0}$ be the number of particles of $(\mathcal{A}_s)_{s\geq 0}$. Then $(\mathcal{A}_s)_{s\geq 0}$ is a Markov process with initial condition*

$$\forall i \leq N_0,\ \mathcal{A}_0(i) = 1, \quad \forall i > N_0,\ \mathcal{A}_0(i) = 0.$$

Moreover, conditional on $\mathcal{A}_s$:

- *each particle gives birth to a new particle of its type at rate d/N_s.*
- *each particle dies at rate $(N_s - 1)/2$.*

The proof of Proposition 3.3 can be found in Appendix A, we only sketch it here. We will first show that the process $(M_t)_{t\in\mathbb{R}}$ is a stationary birth-death process, such that conditional on $M_t = k$, a birth occurs at rate d, and a death at rate $k(k-1)/2$. A simple calculation shows that it is actually a reversible process, i.e., with our notation, that $(N_s)_{s\geq 0}$ is distributed as $(M_t)_{t\geq 0}$. When $(N_s)_{s\geq 0}$ jumps from k to $k+1$, a particle has given birth to two particles. By exchangeability of our system, the particle that gives birth is chosen uniformly, i.e., each particle gives birth at the same rate d/k. Similarly, when $(N_s)_{s\geq 0}$ jumps from k to $k-1$ a particle chosen uniformly from the population dies. Thus each particle dies at rate $k(k-1)/(2k) = (k-1)/2$.

Making the above argument rigorous involves counting the number of trajectories of $(\bar{\Pi}_t)_{t\in\mathbb{R}}$ yielding a given trajectory of $(\mathcal{A}_s)_{s\geq 0}$. We postpone it until Appendix A.

In order to prove Proposition 3.1, we need to keep track of the number of ancestors of k blocks chosen uniformly from $\bar{\Pi}_0$. As we have chosen a uniform labeling of the blocks of $\bar{\Pi}_0$, this amounts to considering the process $(\mathcal{A}_s(1), \ldots, \mathcal{A}_s(k);\ s \geq 0)$. Proposition 3.3 directly gives us the distribution of this process.

Corollary 3.4. *The process $(\mathcal{A}_s(1), \ldots, \mathcal{A}_s(p), N_s;\ s \geq 0)$ is a Markov process such that conditional on $\{\mathcal{A}_s(1) = a_1, \ldots, \mathcal{A}_s(p) = a_p, N_s = k\}$, the process jumps to:*

- $(a_1, \ldots, a_i + 1, \ldots, a_p, k+1)$ *at rate* $\frac{d}{k} a_i$.
- $(a_1, \ldots, a_i - 1, \ldots, a_p, k-1)$ *at rate* $\frac{k-1}{2} a_i$.
- $(a_1, \ldots, a_p, k+1)$ *at rate* $\frac{d}{k}(k - a_1 - \cdots - a_p)$.
- $(a_1, \ldots, a_p, k-1)$ *at rate* $\frac{k-1}{2}(k - a_1 - \cdots - a_p)$.

Proof. We see from the expression of the transition rates of $(\mathcal{A}_s)_{s\geq 0}$ that the rate at which each particle splits or dies only depends on the rest of the population through the total population size N_s. This is enough to prove the result. □

3.2 Convergence

We now prove that the process $(\mathcal{A}_s(1), \ldots, \mathcal{A}_s(p);\ s \geq 0)$ converges to independent critical binary birth-death processes when time is rescaled by a factor $1/\sqrt{n}$. We start with the following lemma.

Lemma 3.5. *Let M^n have the stationary distribution of $(M^n_t)_{t\geq 0}$, the number of blocks of Kingman's coalescent with immigration rate dn. The sequence $(M^n/\sqrt{n};\ n \geq 1)$ is tight.*

Proof. Let $n \geq 1$ and consider a birth-death process $(X^n_t)_{t\geq 0}$ such that conditional on $\{X^n_t = k\}$, the process jumps to

- $k+1$ at rate dn;

- $k-1$ at rate μ_k,

where the death rate μ_k is defined as

$$\mu_k = \begin{cases} 0 & \text{if } k < \sqrt{2dn}+1, \\ \frac{(\sqrt{2dn}+1)\sqrt{2dn}}{2} & \text{else.} \end{cases}$$

The process $(X_t^n - \lfloor\sqrt{2dn}+1\rfloor;\, t \geq 0)$ is distributed as a simple random walk, reflected at 0. Thus it admits a geometric stationary distribution with parameter γ_n given by

$$\gamma_n = \frac{2dn}{(\sqrt{2dn}+1)\sqrt{2dn}} = \frac{1}{1+\sqrt{\frac{1}{2dn}}}.$$

This shows that the process $(X_t^n)_{t\geq 0}$ also admits a stationary distribution. If X^n has the stationary distribution of $(X_t^n)_{t\geq 0}$, then X^n is distributed as $\lfloor\sqrt{2dn}\rfloor + 1 + Y^n$, where Y^n has a geometric distribution with parameter γ_n.

Hence, for K and n large enough, we have

$$\begin{aligned} \mathbb{P}\left(X^n \leq K\sqrt{n}\right) &\leq \mathbb{P}\left(Y^n \leq K\sqrt{n} - \sqrt{2dn}\right) \\ &= 1 - \gamma_n^{(K-\sqrt{2d})\sqrt{n}} \\ &= 1 - \exp(-\frac{K-\sqrt{2d}}{\sqrt{2d}}) + o_n(1). \end{aligned}$$

Thus the sequence $(X^n/\sqrt{n};\, n \geq 1)$ is tight.

Recall that $(M_t^n)_{t\geq 0}$ is a birth-death process jumping from k to $k+1$ at rate dn, and from k to $k-1$ at rate $k(k-1)/2 \geq \mu_k$. Its stationary distribution is thus dominated by that of X^n, and this proves the result. □

We now prove our main convergence result. The proof will use a result from Chapter 11 of [8] on the a.s. convergence of rescaled Markov processes. In order to stick to their notation, we introduce

$$\forall s \geq 0,\ \hat{N}_s^n = N_{s/\sqrt{n}}^n, \quad \hat{\mathcal{A}}_s^n = \mathcal{A}_{s/\sqrt{n}}^n,$$

and

$$\forall x \geq 0,\ \beta_+(x) = d,\ \beta_-(x) = \frac{x^2}{2},\ F(x) = d - \frac{x^2}{2}.$$

Proposition 3.6. *Let $(\mathcal{A}_s^n)_{s\geq 0}$ be the ancestral process of Kingman's coalescent with immigration rate dn. Then*

$$\left(\hat{\mathcal{A}}_s^n(1), \ldots, \hat{\mathcal{A}}_s^n(p), \frac{\hat{N}_s^n}{\sqrt{n}};\, s \geq 0\right) \Longrightarrow \left(X_1(s), \ldots, X_p(s), \sqrt{2d};\, s \geq 0\right),$$

in the sense of convergence in distribution in the Skorohod space, and where the processes $(X_1, \ldots, X_p)$ are i.i.d. critical binary birth-death processes, with per-capita birth and death rate $\sqrt{d/2}$.

Proof. We start by showing that the process $(\hat{N}_s^n/\sqrt{n};\, s \geq 0)$ converges to the constant process with value $\sqrt{2d}$. By applying Proposition 3.3 (bearing in mind that in Proposition 3.3 the immigration rate is d, and not dn) the process $(\hat{N}_s^n)_{s\geq 0}$ is a Markov process jumping from

- k to $k+1$ at rate $d\sqrt{n} = \sqrt{n}\beta_+(\frac{k}{\sqrt{n}})$.

- k to $k-1$ at rate $\frac{k(k-1)}{2\sqrt{n}} = \sqrt{n}\beta_-(\frac{k}{\sqrt{n}}) - \frac{1}{2\sqrt{n}}$.

Thus, the process $(\hat{N}_s^n)_{s\geq 0}$ is of the same form as the processes considered in Theorem 2.1 of Chapter 11 of [8], except that the scaling is $\sqrt{n}$ and not n.

Let us consider a stationary version of the process $(\hat{N}_s^n)_{s\geq 0}$. Lemma 3.5 shows that the sequence $(\hat{N}_0^n/\sqrt{n};\, n \geq 1)$ is tight. We can thus find an increasing sequence of indices $(n_k)_{k\geq 1}$ such that the subsequence $(\hat{N}_0^{n_k}/\sqrt{n_k};\, k \geq 1)$ converges in distribution to a limiting variable N. Using Skorohod's representation theorem, see for example Theorem 6.7 in [3], we can assume that the convergence holds a.s.

Applying Theorem 2.1 of Chapter 11 of [8] shows that the sequence of processes $(\hat{N}_s^{n_k}/\sqrt{n_k};\, s \geq 0, k \geq 1)$ converges a.s. uniformly on compact sets to the solution of

$$\dot{x} = F(x) = d - \frac{x^2}{2}, \tag{3.1}$$

started from the random variable N. (The original theorem is given for a different scaling, but the proof is easily adapted to ours.) As each process $(\hat{N}_s^{n_k})_{s\geq 0}$ is stationary, the limiting process is a stationary solution to (3.1), i.e., it is the constant process with value $\sqrt{2d}$. This shows that each converging subsequence of $(\hat{N}_s^n/\sqrt{n};\, s \geq 0, n \geq 1)$ converges to the same constant process, and thus that the entire sequence converges.

Let us now prove the convergence of the ancestral processes. Consider independent Poisson processes $(P_i^-(s))_{s\geq 0}$, $(P_i^+(s))_{s\geq 0}$ for $i \leq p$, and $(P_N^-(s))_{s\geq 0}$, $(P_N^+(s))_{s\geq 0}$. Using for instance Theorem 4.1 from Chapter 6 of [8], there exists a unique strong solution to the following equations

$$\forall s \geq 0, \forall i \leq p,\ X_i^n(s) = P_i^+\Big(\int_0^s \frac{d\sqrt{n}X_i^n(u)}{Y^n(u)}\mathrm{d}u\Big) - P_i^-\Big(\int_0^s \frac{X_i^n(u)(Y^n(u)-1)}{2\sqrt{n}}\mathrm{d}u\Big),$$

and

$$\begin{aligned}\forall s \geq 0,\ Y^n(s) = P_N^+\Big(\int_0^s d\sqrt{n}(1 - \tfrac{\sum_i X_i^n(u)}{Y^n(u)})\mathrm{d}u\Big)\\ - P_N^-\Big(\int_0^s \tfrac{Y^n(u)(Y^n(u)-1)}{2\sqrt{n}}(1 - \tfrac{\sum_i X_i^n(u)}{Y^n(u)})\mathrm{d}u\Big) + \sum_{i=1}^p X_i^n(s).\end{aligned}$$

Moreover, this solution $(X_1^n, \ldots, X_p^n, Y^n)$ is distributed as $(\hat{\mathcal{A}}_s^n(1), \ldots, \hat{\mathcal{A}}_s^n(p), \hat{N}_s^n;\, s \geq 0)$.

As $Y^n/\sqrt{n}$ converges in probability to the constant process with value $\sqrt{2d}$, we can find a subsequence such that

$$\lim_{n\to\infty} \frac{d\sqrt{n}}{Y^n(s)} = \sqrt{\frac{d}{2}}, \quad \lim_{n\to\infty} \frac{(Y^n(s)-1)}{2\sqrt{n}} = \sqrt{\frac{d}{2}} \quad \text{a.s.}$$

holds uniformly in s on compact sets. This is sufficient to show that for each $i \leq p$, the subsequence of processes $(X_i^n(s))_{s\geq 0}$ converges a.s. in the Skorohod space to the solution $(X_i(s))_{s\geq 0}$ of

$$\forall s \geq 0, \forall i \leq p,\ X_i(s) = P_i^+\Big(\int_0^s \sqrt{\frac{d}{2}}X_i(u)\mathrm{d}u\Big) - P_i^-\Big(\int_0^s \sqrt{\frac{d}{2}}X_i(u)\mathrm{d}u\Big).$$

This proves that the entire sequence $(X_1^n, \ldots, X_p^n)$ converges in probability in the Skorohod topology to the solution of the previous equation. Finally, noting that the solutions of these equations are independent and distributed as critical binary branching processes with branching rate $\sqrt{d/2}$ ends the proof. □

We are now ready to prove Proposition 3.1.

Proof of Proposition 3.1. By construction, the size of p blocks of $\bar{\Pi}^n$ chosen uniformly is given by the total number of particles of the processes $(\hat{\mathcal{A}}^n_s(1), \dots, \hat{\mathcal{A}}^n_s(p);\, t \geq 0)$. Thus, in the limit, the size of these blocks converges to the total size of p independent critical binary branching processes. □

4 Proof of Theorem 1.5

In the previous section we have derived the limiting distribution of the sizes of blocks uniformly sampled from Kingman's coalescent with immigration. In this section we make use of the coupling between Kingman's coalescent with immigration and Kingman's coalescent with erosion from Section 2.4 to get the analogous result in the erosion case.

We first show the following result.

Corollary 4.1. *Let Π^n have the stationary distribution of the n-Kingman coalescent with erosion. Let $(|C^n_1|, \dots, |C^n_p|)$ be the size of p blocks chosen uniformly from Π^n. Then*

$$(|C^n_1|, \dots, |C^n_p|) \Longrightarrow (J_1, \dots, J_p),$$

where $(J_1, \dots, J_p)$ are i.i.d. variables distributed as the total progeny of a critical binary branching process.

Proof. Recall the coupling between Kingman's coalescent with erosion and Kingman's coalescent with immigration. Let $(T_i)_{i\in\mathbb{Z}}$ be the atoms of a Poisson point process P^n with intensity dn, labeled in increasing order such that $T_0 < 0 < T_1$. Consider an independent i.i.d. sequence of marks $(\ell_i)_{i\in\mathbb{Z}}$ that are uniformly distributed on $[n]$.

Let $\bar{\Pi}^n_0$ be the value at time 0 of the version of Kingman's coalescent with immigration rate nd built from $(T_i)_{i\in\mathbb{Z}}$ as in Section 2.1. We know from Proposition 2.5 that we can obtain a version Π^n of the stationary distribution of the n-Kingman coalescent with erosion rate d by placing i and j in the same block of Π^n if the most recent atoms of P^n in $(-\infty, 0]$ with mark i and j both belong to the same block of $\bar{\Pi}^n_0$.

Now let $(\bar{C}^n_1, \dots, \bar{C}^n_p)$ be p blocks chosen uniformly from $\bar{\Pi}_0$, and let $(|\bar{C}^n_1|, \dots, |\bar{C}^n_p|)$ be their respective sizes. For $k \leq p$, let

$$|C^n_k| = \mathrm{Card}\left\{i \in \bar{C}^n_k : (T_i, \ell_i) \text{ is the most recent atom in } (-\infty, 0] \text{ with mark } \ell_i\right\}.$$

Then conditional on $\left\{|C^n_1| \geq 1, \dots, |C^n_p| \geq 1\right\}$, $(|C^n_1|, \dots, |C^n_p|)$ are the sizes of p blocks chosen uniformly from Π^n. The result is thus proved if we can show that

$$\lim_{n\to\infty} \mathbb{P}\big(|C^n_1| = |\bar{C}^n_1|, \dots, |C^n_p| = |\bar{C}^n_p|\big) = 1.$$

Let us first explain intuitively why the previous claim holds. The ancestors of $\bar{C}^n_1$ have all immigrated on a time-scale of order $1/\sqrt{n}$. On this time-scale, there are of order $\sqrt{n}$ particles that have also immigrated. All these particles receive a uniform label in $[n]$. Thus the probability that an ancestor of $\bar{C}^n_1$ has received the same label as one of the other $\sqrt{n}$ particles, i.e., that it is not the most recent atom with its mark, is of order $1/\sqrt{n}$. Let us make this argument rigorous.

Set

$$\tau^n_1 := \min\left\{T_i : i \in \bar{C}^n_1\right\}$$

to be the total life-time of the ancestors of the block $\bar{C}^n_1$. (The variable τ^n_1 gives the immigration time of the first particle that forms the block $\bar{C}^n_1$.) The total number

of particles that have immigrated during the time interval $[\tau_1^n, 0]$ is then $P^n([\tau_1^n, 0])$. Consider the event

$$E_k = \Big\{ |\bar{C}_1^n| = k,\ \tau_1^n \in [-\tfrac{s}{\sqrt{n}}, 0],\ P^n([-\tfrac{s}{\sqrt{n}}, 0]) \le (1+\varepsilon) ds\sqrt{n} \Big\}.$$

On this event, if $|C_1^n| \neq |\bar{C}_1^n|$, then one of the k ancestors of $\bar{C}_1^n$ has received the same label as one of the particles that has immigrated in the time interval $[\tau_1^n, 0]$, that is, the same label as one of the $(1+\varepsilon)dt\sqrt{n}$ most recent atoms of P^n. As the labels are chosen uniformly, the probability that each of the k ancestors has a label distinct from the labels of the other $(1+\varepsilon)ds\sqrt{n} - 1$ most recent particles is

$$\Big(1 - \frac{1}{n}\Big) \dots \Big(1 - \frac{k-1}{n}\Big)\Big(1 - \frac{k}{n}\Big)^{(1+\varepsilon)ds\sqrt{n}-k}$$

which goes to 1 as n goes to infinity for all fixed k. Thus

$$\mathbb{P}\big(|C_1^n| \neq |\bar{C}_1^n|, E_k\big) \le \Big(1 - \frac{1}{n}\Big) \dots \Big(1 - \frac{k-1}{n}\Big)\Big(1 - \frac{k}{n}\Big)^{(1+\varepsilon)ds\sqrt{n}-k},$$

and

$$\begin{aligned} \mathbb{P}\big(|C_1^n| \neq |\bar{C}_1^n|\big) \le{}& \mathbb{P}\Big(\tau_1^n \notin [-\tfrac{s}{\sqrt{n}}, 0]\Big) + \mathbb{P}\big(|\bar{C}_1^n| \ge K\big) \\ &+ \mathbb{P}\Big(P^n([-\tfrac{s}{\sqrt{n}}, 0]) > (1+\varepsilon)ds\sqrt{n}\Big) + o_n(1). \end{aligned} \tag{4.1}$$

Now, by Proposition 3.6, the sequence $(-\sqrt{n}\tau_1^n)_{n\ge1}$ converges in distribution to the total life-time of a binary critical branching process and $(|\bar{C}_1^n|)_{n\ge1}$ converges to the total progeny of this process. Thus, the first two terms in the above equation can be made as small as desired uniformly in n by taking t and K large enough. For a fixed $\varepsilon > 0$, Chebishev's inequality shows that the last term goes to 0 as n goes to infinity. This proves the result for $p = 1$ and a simple union bound proves the result for any p. □

Remark 4.2. In the previous proof, on the event $\{|\bar{C}_1^n| = |C_1^n|\}$, not only the size of the blocks of Kingman's coalescents with erosion and immigration coincide, but also the genealogy of the blocks. Thus we have shown the slightly stronger result that, in the n-Kingman coalescent with erosion, the genealogy of a block chosen uniformly from the stationary distribution converges to that of a critical binary branching process.

We can now prove Theorem 1.5. Recall that μ_k^n denotes the frequency of blocks of size k of Π^n, i.e., if the blocks of Π^n are $(C_1^n, \dots, C_{M^n}^n)$, then

$$\mu_k^n = \frac{1}{M^n}\,\mathrm{Card}(\{i : |C_i^n| = k\}).$$

Proof of Theorem 1.5. (i) We start by proving that $M^n/\sqrt{n}$ converges to $\sqrt{2d}$ in probability. Let us consider a version $\bar{\Pi}^n$ of the stationary distribution of Kingman's coalescent with immigration rate nd, coupled with a version Π^n of the stationary distribution of Kingman's coalescent with erosion rate d on $[n]$. Let $\bar{M}^n$, resp. M^n, denote the number of blocks of $\bar{\Pi}^n$, resp. Π^n. Recall that the blocks of Π^n are subsets of the blocks of $\bar{\Pi}^n$, where a particle is retained if there are no other particles with the same label that have immigrated after it. Let $|\bar{C}^n|$ be the size of a block of $\bar{\Pi}^n$ chosen uniformly, and let $|C^n|$ be the size of the corresponding block of Π^n. Some blocks of $\bar{\Pi}^n$ are only composed of particles that are not retained to form Π^n. Such blocks have no corresponding blocks in Π^n, and $\bar{M}^n - M^n$ is exactly the number of such blocks. Thus

$$\mathbb{E}\Big[\frac{\bar{M}^n - M^n}{\bar{M}^n}\Big] = \mathbb{P}(|C^n| = 0) \le \mathbb{P}\big(|C^n| \neq |\bar{C}^n|\big) \longrightarrow 0,$$

where the convergence holds by (4.1). This shows that $M^n/\bar{M}^n$ goes to 1 in probability. Proposition 3.6 further shows that $\bar{M}^n/\sqrt{n}$ goes to $\sqrt{2d}$ in probability, and thus that $M^n/\sqrt{n}$ also goes to $\sqrt{2d}$ in probability.

(ii) We prove the second point using the method of moments. Let $(|C_1^n|, \dots, |C_p^n|)$ be the sizes of k uniformly sampled blocks of Π^n. Then, as the number of blocks M^n goes to infinity, Corollary 4.1 shows that

$$\lim_{n\to\infty} \mathbb{E}[(\mu_k^n)^p] = \lim_{n\to\infty} \mathbb{P}\big(|C_1^n| = \dots = |C_p^n| = k\big) = \mathbb{P}(J = k)^p,$$

where J is the total progeny of a binary critical branching process. The convergence of the moments readily implies convergence in distribution as the limit is a Dirac mass. □

5 Asymptotic frequencies of Kingman's coalescent with erosion

In this section we prove Theorem 1.3, which gives a representation of the asymptotic frequencies in terms of independent diffusions. First, we use the flow of bridges construction of Kingman's coalescent with erosion from Corollary 2.6 to give a correspondence between the frequencies of the blocks and the size of the families of a Fleming-Viot process.

5.1 Eves of a Fleming-Viot process

Let $(\rho_t)_{t\geq 0}$ be a Fleming-Viot process built from a standard flow of bridges as in Section 2.2. For each individual $x \in [0,1]$, denote by

$$\zeta(x) = \inf\{t \geq 0 : \rho_t(\{x\}) = 0\}$$

the extinction time of the offspring of x. It is clear that the set

$$\{x \in [0,1] : \zeta(x) > 0\} = \{x \in [0,1] : \rho_t(\{x\}) > 0 \text{ for some } t \geq 0\}$$

is countable. The elements of this set can actually be enumerated in decreasing order of their extinction time, that is, they can be written $(\mathbf{e}_i)_{i\geq 0}$ with

$$\zeta(\mathbf{e}_1) > \zeta(\mathbf{e}_2) > \dots$$

This fact can be found e.g. in [14], Theorem 1.6. The sequence $(\mathbf{e}_i)_{i\geq 0}$ is called the sequence of *Eves* of $(\rho_t)_{t\geq 0}$, and was introduced in [2] and [14], see also [5] for a similar notion for Continuous-State Branching Processes. The following result shows that the frequencies of the blocks of the stationary distribution of Kingman's coalescent with erosion can be recovered from the size of the offspring of the Eves.

Lemma 5.1. *Let $(\mathbf{e}_i)_{i\geq 1}$ be the Eves of a Fleming-Viot process $(\rho_t)_{t\geq 0}$. Then the non-increasing reordering of the sequence $(z_i)_{i\geq 1}$ defined as*

$$\forall i \geq 1,\ z_i = \int_0^\infty de^{-dt} \rho_t(\{\mathbf{e}_i\}) \mathrm{d}t$$

is distributed as the frequencies of the blocks of the stationary distribution of Kingman's coalescent with erosion rate d.

Proof. Consider a flow of bridges $(B_{s,t})_{s\leq t}$, and let $(T_i)_{i\geq 1}$, $(U_i)_{i\geq 1}$ be two independent i.i.d. sequences of exponential variables with parameter d, and uniform variables respectively. Again, as in Corollary 2.6, let Π be the partition of $\mathbb{N}$ defined as

$$i \sim_\Pi j \iff B^{-1}_{-T_i,0}(U_i) = B^{-1}_{-T_j,0}(U_j),$$

which has the stationary distribution of Kingman's coalescent with erosion. We denote by $\Pi = (C_1, C_2, \dots)$ the blocks of Π, ordered in increasing order of their least elements, i.e., such that

$$i \leq j \iff \min(C_i) \leq \min(C_j).$$

Then let us call

$$A_i = B^{-1}_{-T_j,0}(U_j),\ \forall j \in C_i,$$

the ancestor of the block C_i.

As the flow of bridges $(B_{s,t})_{s\leq t}$ is independent of the sequences $(U_i)_{i\geq 1}$ and $(T_i)_{i\geq 1}$, the sequence $(B^{-1}_{-T_i,0}(U_i))_{i\geq 1}$ is exchangeable. Thus, the law of large numbers shows that for any $i \geq 1$,

$$\frac{1}{n}\operatorname{Card}(C_i \cap [n]) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\left\{B^{-1}_{-T_j,0}(U_j)=A_i\right\}} \underset{n\to\infty}{\longrightarrow} \int_0^\infty de^{-dt}\rho_t(\{A_i\})\mathrm{d}t \quad \text{a.s.}$$

Thus the result is proved if we can show that a.s.

$$\{\mathbf{e}_i : i \geq 1\} = \{A_i : i \geq 1\}.$$

Clearly we have $\zeta(A_i) > 0$, as otherwise the frequency of the block C_i would be zero. Moreover, conditional on the flow of bridges, there exists a.s. some $j \geq 1$ such that

$$(U_j, T_j) \in \{(x,t) : B^{-1}_{-t,0}(x) = \mathbf{e}_i\}$$

as by definition of $\mathbf{e}_i$ this set has positive Lebesgue measure. Thus, a.s. $\mathbf{e}_i$ is the ancestor of some block of Π, and the result is proved. □

In order to prove Theorem 1.3, it remains to show that the sequence of processes $\big(\rho_t(\{\mathbf{e}_1\}), \rho_t(\{\mathbf{e}_2\}), \dots ; t \geq 0\big)$ has the same distribution as the sequence of diffusions introduced in Section 1.3. In the following section we characterize this distribution, and complete the proof in the last section.

5.2 Wright-Fisher diffusion conditioned on its extinction order

Consider a n-dimensional Wright-Fisher diffusion $(X_1, \dots, X_n)$. That is, $(X_1, \dots, X_n)$ is distributed as the unique solution to

$$\forall i \geq 1,\ \mathrm{d}X_i = \sum_{\substack{j=1\\ j\neq i}}^{n} \sqrt{X_i X_j}\mathrm{d}W_{i,j},$$

where $(W_{i,j})_{i<j}$ are independent Brownian motions, and $W_{j,i} = -W_{i,j}$, and started from an initial condition $(x_1, \dots, x_n) \in (0,1)^n$ verifying $x_1 + \dots + x_n = 1$. The Wright-Fisher diffusion describes the dynamics of a population with constant size, where individuals can be of n different types; X_i denotes the frequency of type i individuals in the population. Each process X_i is eventually absorbed at 0 or 1. We say that the family X_i reaches fixation if it gets absorbed at 1, and that it becomes extinct otherwise. Let

$$\zeta_i = \inf\{t \geq 0 : X_i = 0\}$$

denote its absorption time at 0.

In this section, we study the distribution of $(X_1, \dots, X_n)$ conditional on the event $\{\zeta_n < \dots < \zeta_1\}$. First, notice that as $X_1 + \dots + X_n = 1$, there is exactly one family that reaches fixation. Thus, on the event $\{\zeta_n < \dots < \zeta_1\}$, we have $\zeta_1 = \infty$ and X_1 reaches fixation; X_2 is the last family to go extinct, and X_n is the first family to go extinct. We

now express the distribution of the conditioned Wright-Fisher diffusion in terms of the diffusions introduced in Section 1.3.

We will work inductively, by first conditioning the process $(X_1, \dots, X_n)$ on ζ_1 being the largest extinction time, then on ζ_2 being the second largest and so on and so forth. The key point is that after conditioning on the fixation of X_1, the remainder of the population, $(X_2, \dots, X_n)$, is distributed as a rescaled, time-changed, unconditioned $(n-1)$-dimensional Wright-Fisher diffusion, independent of X_1.

Let us be more specific and let Y_1 be the solution of

$$\mathrm{d}Y_1 = (1 - Y_1)\mathrm{d}t + \sqrt{Y_1(1 - Y_1)}\mathrm{d}W_1, \tag{5.1}$$

for some Brownian motion W_1. Notice that Y_1 is distributed as a usual one-dimensional Wright-Fisher diffusion, conditioned on fixation. Consider the fixation time of Y_1 which is defined as

$$S_1 = \inf\{t \geq 0 : Y_1(t) = 1\}.$$

We further define a random time-change τ_1 as

$$\forall t < S_1,\ \tau_1(t) = \int_0^t \frac{1}{1 - Y_1(s)}\mathrm{d}s, \quad \forall t \geq S_1,\ \tau_1(t) = \infty.$$

We start by proving the following result.

Lemma 5.2. *Let Y_1 and τ_1 be as above and consider an independent $(n-1)$-dimensional Wright-Fisher diffusion $(X_2, \dots, X_n)$. Then, the process $(Z_1, \dots, Z_n)$ defined as*

$$Z_1 = Y_1, \quad \forall i > 1, \forall t \geq 0,\ Z_i(t) = (1 - Z_1(t))X_i(\tau_1(t)),$$

is distributed as a n-dimensional Wright-Fisher diffusion conditioned on $\{\zeta_1 = \infty\}$.

Remark 5.3. The time $\tau_1(t)$ is infinite with positive probability. However, each of the processes $(X_2, \dots, X_n)$ has an a.s. limit as t goes to infinity. On the event $\{\tau_1(t) = \infty\}$, we take $X_i(\tau_1(t))$ to be this limit, so that the process $(Z_1, \dots, Z_n)$ is now well-defined.

Before proving Lemma 5.2, we need the following fact that we prove for the sake of completeness.

Lemma 5.4. *Let $(W_t)_{t\geq 0}$ be a Brownian motion on $\mathbb{R}$ started at 1, and let T_0 be the first time it hits 0. Then for $\alpha \in \mathbb{R}$, a.s.*

$$\int_0^{T_0} W_s^\alpha \mathrm{d}s = \begin{cases} \infty & \text{if } \alpha \leq -2 \\ y_\alpha < \infty & \text{if } \alpha > -2. \end{cases}$$

Proof. Let us define

$$\forall t \geq 0,\ \xi_t = \tilde{W}_t - \frac{t}{2}, \quad \tau(t) = \inf\Big\{s \geq 0 : \int_0^s \exp(2\xi_u)\mathrm{d}u > t\Big\},$$

for a Brownian motion $(\tilde{W}_t)_{t\geq 0}$ with the convention that $\inf \emptyset = \infty$ and $\xi_\infty = -\infty$. The Lamperti representation of positive self-similar processes [16] shows that W_t stopped at T_0 satisfies the equality in distribution

$$(W_{t\wedge T_0})_{t\geq 0} \overset{(\mathrm{d})}{=} (\exp(\xi_{\tau(t)}))_{t\geq 0}.$$

Thus

$$\int_0^{t\wedge T_0} W_s^\alpha \mathrm{d}s \overset{(\mathrm{d})}{=} \int_0^t \exp(\alpha\xi_{\tau(s)})\mathrm{d}s = \int_0^{\tau(t)} \exp((2+\alpha)\xi_s)\mathrm{d}s,$$

and

$$\int_0^{T_0} W_s^\alpha \mathrm{d}s \overset{(\mathrm{d})}{=} \int_0^\infty \exp((2+\alpha)\xi_s)\mathrm{d}s,$$

which yields the result. □

Proof of Lemma 5.2. Consider a n-dimensional Wright-Fisher diffusion $(X_1, \dots, X_n)$. A calculation of Doob's h-transform using the harmonic function

$$h(x_1, \dots, x_n) = \mathbb{P}\Big(\lim_{t\to\infty} X_1(t) = 1 \,\Big|\, X_1(0) = x_1, \dots, X_n(0) = x_n\Big) = x_1$$

shows that the process $(X_1, \dots, X_n)$ conditioned on $\{\lim_{t\to\infty} X_1(t) = 1\} = \{\zeta_1 = \infty\}$ is distributed as the unique solution to the equation

$$\mathrm{d}X_1 = (1 - X_1)\mathrm{d}t + \sum_{j=2}^{n} \sqrt{X_1 X_j}\mathrm{d}W_{1,j},$$
$$\forall i \geq 2,\ \mathrm{d}X_i = -X_i\mathrm{d}t + \sum_{\substack{j=1\\ j\neq i}}^{n} \sqrt{X_i X_j}\mathrm{d}W_{i,j},$$

where $(W_{i,j})_{i<j}$ are independent Brownian motions, and $W_{i,j} = -W_{j,i}$. We will prove that the process $(Z_1, \dots, Z_n)$ solves this equation.

Now consider a $(n-1)$-dimensional Wright-Fisher diffusion $(X'_2, \dots, X'_n)$ independent of Y_1 which solves

$$\forall i \geq 2,\ \mathrm{d}X'_i = \sum_{\substack{j=2\\ j\neq i}}^{n} \sqrt{X'_i X'_j}\mathrm{d}W'_{i,j},$$

where $(W'_{i,j})_{i<j}$ are independent Brownian motions and $W'_{i,j} = -W'_{j,i}$. We start by giving the equation solved by the process $(Y_1, X'_2 \circ \tau_1, \dots, X'_n \circ \tau_1)$. Notice that here, only a subset of the processes are time-changed, and that τ_1 explodes in finite time. For these two reasons, let us realize the time-change carefully.

We transform τ_1 into a family of finite stopping times. Our first task is to prove that τ_1 goes continuously to infinity, we do this using the speed function and scale measures of the diffusion Y_1, see for instance [7]. If we define $D = 1/Y_1$, then by Itô's formula,

$$\mathrm{d}D = -\sqrt{D-1}D\mathrm{d}W_1, \quad \forall t \geq 0,\ [D, D]_t = \int_0^t (D(s) - 1)D(s)^2\mathrm{d}s.$$

Recall that S_1 stands for the first time when Y_1 hits one. Using Dubins-Schwarz theorem, see for instance Theorem 18.4 of [11], we obtain that

$$\int_0^{S_1} \frac{1}{1 - Y_1(s)}\mathrm{d}s = \int_0^{S_1} \frac{D(s)}{D(s) - 1}\mathrm{d}s = \int_0^{S_1} \frac{\tilde{W}_1([D, D]_s)}{\tilde{W}_1([D, D]_s) - 1}\mathrm{d}s = \int_0^{T_1} \frac{1}{(\tilde{W}_1(s) - 1)^2\tilde{W}_1(s)}\mathrm{d}s$$

where $\tilde{W}_1$ is a Brownian motion (on a possibly larger probability space) started at $1/Y_1(0)$, and T_1 is the first time when $\tilde{W}_1$ hits 1. We now know from Lemma 5.4 that this integral is a.s. infinite, and thus that τ_1 goes continuously to infinity, and does not "jump to infinity".

Further consider the times

$$\forall i \geq 2,\ S_i = \inf\{t \geq 0 : X'_i(t) = 1\}, \quad S = \min(S_2, \dots, S_n).$$

At time S, one of the families has reached fixation, and thus for $t \geq S$ we have $X'_i(t) = X'_i(S)$. Therefore, for all $t \geq 0$, we have $X'_i(\tau_1(t)) = X'_i(\tau_1(t) \wedge S)$, where the stopping time $\tau_1(t) \wedge S$ is now a.s. finite, and $t \mapsto \tau_1(t) \wedge S$ is continuous. (The continuity requires that τ_1 does not jump to infinity.) Thus, by making a time-change in the following integrals,

see for instance [11], Theorem 17.24, we obtain

$$\begin{aligned}\forall t \geq 0,\ X_i'(\tau_1(t)) &= X_i'(\tau_1(t) \wedge S) \\ &= \sum_{\substack{j=2 \\ j\neq i}}^{n} \int_0^{\tau_1(t)\wedge S} \sqrt{X_i'(s)X_j'(s)}\mathrm{d}W_{i,j}' \\ &= \sum_{\substack{j=2 \\ j\neq i}}^{n} \int_0^{t} \sqrt{X_i'(\tau_1(s)\wedge S)X_j'(\tau_1(s)\wedge S)}\mathrm{d}W_{i,j}'(\tau_1(s)\wedge S) \\ &= \sum_{\substack{j=2 \\ j\neq i}}^{n} \int_0^{t} \sqrt{\frac{X_i'(\tau_1(s))X_j'(\tau_1(s))}{1-Y_1(s)}}\mathrm{d}\tilde{W}_{i,j}\end{aligned}$$

where

$$\forall t \geq 0,\ \tilde{W}_{i,j}(t) = \int_0^t \sqrt{1-Y_1(s)}\mathrm{d}W_{i,j}'(\tau_1(s)\wedge S).$$

A direct computation of the quadratic variations gives

$$\forall i,j,t \geq 0,\ [\tilde{W}_{i,j}, \tilde{W}_{i,j}]_t = t \wedge S,$$

and the crossed variations are null. Thus a multidimensional version of Dubins-Schwarz theorem, see for instance Theorem 18.4 in [11], shows that we can find independent Brownian motions $(\hat{W}_{i,j})_{i<j}$ such that $\tilde{W}_{i,j}(t) = \hat{W}_{i,j}(t \wedge S)$. This proves that the time-changed processes solve

$$\forall t \geq 0,\ X_i'(\tau_1(t)) = \sum_{\substack{j=2 \\ j\neq i}}^{n} \int_0^{t} \sqrt{\frac{X_i'(\tau_1(s))X_j'(\tau_1(s))}{1-Y_1(s)}}\mathrm{d}\hat{W}_{i,j}.$$

Finally, setting $\hat{X}_i := X_i' \circ \tau_1$ and applying Itô's formula to the process $(Y_1, \hat{X}_2, \ldots, \hat{X}_n)$ with the function

$$(x_1, \ldots, x_n) \mapsto (x_1, (1-x_1)x_2, \ldots, (1-x_1)x_n)$$

we obtain that for all $i \geq 2$,

$$\begin{aligned}\mathrm{d}Z_i &= -\hat{X}_i\mathrm{d}Y_1 + (1-Y_1)\mathrm{d}\hat{X}_i \\ &= -\hat{X}_i(1-Y_1)\mathrm{d}t - \hat{X}_i\sqrt{Y_1(1-Y_1)}\mathrm{d}W_1 + \sum_{\substack{j=2 \\ j\neq i}}^{n} \sqrt{(1-Y_1)\hat{X}_i\hat{X}_j}\mathrm{d}\hat{W}_{i,j} \\ &= -Z_i\mathrm{d}t - Z_i\sqrt{\frac{Z_1}{1-Z_1}}\mathrm{d}W_1 + \sum_{\substack{j=2 \\ j\neq i}}^{n} \sqrt{\frac{Z_iZ_j}{1-Z_1}}\mathrm{d}\hat{W}_{i,j},\end{aligned}$$

where $(Z_1, \ldots, Z_n)$ is defined as in the statement of the result. A straightforward computation of the quadratic variations shows that $(Z_1, \ldots, Z_n)$ is distributed as $(X_1, \ldots, X_n)$ conditioned on $\{\zeta_1 = \infty\}$ and proves the result. □

We can now proceed inductively. Let us set up the notation for the proof. Consider i.i.d. processes $(Y_1, \ldots, Y_{n-1})$ such that

$$\forall i \geq 1,\ \mathrm{d}Y_i = (1-Y_i)\mathrm{d}t + \sqrt{Y_i(1-Y_i)}\mathrm{d}W_i$$

where $(W_1, \dots, W_{n-1})$ are independent Brownian motions. We set $\tilde{Z}_1 = Y_1$, and

$$\forall t \geq 0,\ \tilde{\tau}_1(t) = \int_0^t \frac{1}{1 - \tilde{Z}_1(s)} \mathrm{d}s.$$

We then define recursively, for $i < n-1$,

$$\forall t \geq 0,\ \tilde{Z}_{i+1}(t) = (1 - \tilde{Z}_1(t) - \cdots - \tilde{Z}_i(t)) Y_{i+1}(\tilde{\tau}_i(t))$$
$$\forall t \geq 0,\ \tilde{\tau}_{i+1}(t) = \int_0^t \frac{1}{1 - \tilde{Z}_1(s) - \cdots - \tilde{Z}_{i+1}(s)} \mathrm{d}s.$$

We finally set $\tilde{Z}_n = 1 - \tilde{Z}_1 - \cdots - \tilde{Z}_{n-1}$.

Proposition 5.5. *The process $(\tilde{Z}_1, \dots, \tilde{Z}_n)$ defined above is distributed as a n-dimensional Wright-Fisher diffusion conditioned on $\{\zeta_n < \cdots < \zeta_1\}$.*

Proof. We prove the result inductively. For $n = 2$, conditioning (X_1, X_2) on its extinction order amounts to conditioning it on the fixation of X_1, and Lemma 5.2 shows that the result holds.

Let $(Y_1, \dots, Y_{n-1})$ be the i.i.d. diffusions defined above. We first define

$$\forall t \geq 0,\ \tilde{Z}'_2(t) = Y_2(t), \quad \forall t \geq 0,\ \tilde{\tau}'_2(t) = \int_0^t \frac{1}{1 - \tilde{Z}'_2(s)} \mathrm{d}s$$

and then define inductively, for $i < n-1$,

$$\forall t \geq 0,\ \tilde{Z}'_{i+1}(t) = (1 - \tilde{Z}'_2(t) - \cdots - \tilde{Z}'_i(t)) Y_{i+1}(\tilde{\tau}'_i(t)),$$
$$\forall t \geq 0,\ \tilde{\tau}'_{i+1}(t) = \int_0^t \frac{1}{1 - \tilde{Z}'_2(s) - \cdots - \tilde{Z}'_{i+1}(s)} \mathrm{d}s,$$

and $\tilde{Z}'_n = 1 - \tilde{Z}'_2 - \cdots - \tilde{Z}'_{n-1}$. By induction, we can suppose that $(\tilde{Z}'_2, \dots, \tilde{Z}'_n)$ is distributed as a $(n-1)$-dimensional Wright-Fisher diffusion conditioned on its extinction order. We first claim that the process defined as

$$\forall t \geq 0,\ \tilde{Z}_1(t) = Y_1(t),$$
$$\forall i > 1, \forall t \geq 0,\ \tilde{Z}_i(t) = (1 - \tilde{Z}_1(t)) \tilde{Z}'_i(\tilde{\tau}_1(t))$$

is distributed as a n-dimensional Wright-Fisher diffusion conditioned on its extinction order.

To see this, let $(X_2, \dots, X_n)$ be a $(n-1)$-dimensional unconditioned Wright-Fisher diffusion, independent of Y_1, and recall the definition of $(Z_1, \dots, Z_n)$ from Lemma 5.2. Consider

$$\zeta'_i = \inf\{t \geq 0 : Z_i(t) = 0\}, \quad \zeta_i = \inf\{t \geq 0 : X_i(t) = 0\}$$

the extinction times of Z_i and X_i. Lemma 5.2 ensures that $(Z_1, \dots, Z_n)$ is distributed as a Wright-Fisher diffusion conditioned on the fixation of Z_1. Thus, the process $(Z_1, \dots, Z_n)$ further conditioned on $\{\zeta'_n < \cdots < \zeta'_2\}$ has the distribution of a Wright-Fisher diffusion conditioned on its extinction order. Now notice that

$$\{\zeta'_n < \cdots < \zeta'_2\} = \{\zeta_n < \cdots < \zeta_2\}.$$

Thus conditioning $(Z_1, \dots, Z_n)$ on $\{\zeta'_n < \dots \zeta'_2\}$ amounts to conditioning $(X_2, \dots, X_n)$ on $\{\zeta_n < \cdots < \zeta_2\}$, that is, conditioning it on its fixation order. As $\{\zeta_n < \cdots < \zeta_2\}$ is independent of Z_1, conditioning the process $(Z_1, \dots, Z_n)$ on this event is equivalent to

replacing $(X_2, \dots, X_n)$ by $(\tilde{Z}'_2, \dots, \tilde{Z}'_n)$ in the construction of $(Z_1, \dots, Z_n)$, and this proves the claim.

It only remains to show that $\tilde{Z}_{i+1}$ as defined in the proof can be written

$$\forall i > 1,\ \tilde{Z}_{i+1}(t) = (1 - \tilde{Z}_1(t) - \cdots - \tilde{Z}_i(t))Y_{i+1}(\tilde{\tau}_i(t)).$$

A direct calculation first shows that, for $i > 1$ and $t \geq 0$,

$$\begin{aligned}
\tilde{\tau}_i(t) &= \int_0^t \frac{1}{1 - \tilde{Z}_1(s) - \cdots - \tilde{Z}_i(s)} \mathrm{d}s \\
&= \int_0^t \frac{1}{1 - \tilde{Z}_1(s) - (1 - \tilde{Z}_1(s))\tilde{Z}'_2(\tilde{\tau}_1(s)) - \cdots - (1 - \tilde{Z}_1(s))\tilde{Z}'_i(\tilde{\tau}_1(s))} \mathrm{d}s \\
&= \int_0^t \frac{1}{1 - \tilde{Z}'_2(\tilde{\tau}_1(s)) - \cdots - \tilde{Z}'_i(\tilde{\tau}_1(s))} \ \frac{1}{1 - \tilde{Z}_1(s)} \mathrm{d}s \\
&= \tilde{\tau}'_i(\tilde{\tau}_1(t)),
\end{aligned}$$

and the result follows. □

We end this section by pointing out the following fact that will be required in the next section. We have only defined the Wright-Fisher diffusion conditioned on its extinction order for an initial condition $(x_1, \dots, x_n)$ such that for all $1 \leq i \leq n$, $x_i > 0$. Nevertheless, the processes Y_i have an entrance boundary at 0. Thus there exists a unique extension of the process $(Y_1, \dots, Y_{n-1})$ started from $(0, \dots, 0)$ that remains Feller, see [11], Theorem 23.3. This shows that a Wright-Fisher diffusion conditioned on its fixation order $(\tilde{Z}_1, \dots, \tilde{Z}_n)$ admits a Feller extension for the initial condition $(0, \dots, 0, 1)$.

5.3 Proof of Theorem 1.3

Let $(\rho_t)_{t \geq 0}$ be a Fleming-Viot process, and let $(\mathbf{e}_i)_{i \geq 1}$ be its Eves. In this section we end the proof of Theorem 1.3 by showing that the distribution of the sequence of processes $(\rho_t(\{\mathbf{e}_1\}), \rho_t(\{\mathbf{e}_2\}), \dots;\ t \geq 0)$ is that of a Wright-Fisher diffusion conditioned on its fixation order.

The result we want to prove is the direct extension of Theorem 4 of [2]. Reformulated in our setting, this theorem proves that $(\rho_t(\{\mathbf{e}_1\});\ t \geq 0)$ is distributed as the solution to eq. (5.1) started from 0. We now give a similar representation for the process $(\rho_t(\{\mathbf{e}_1\}), \dots, \rho_t(\{\mathbf{e}_n\});\ t \geq 0)$ giving the size of the progeny of the first n Eves.

Proposition 5.6. *Let $(\rho_t)_{t \geq 0}$ be a Fleming-Viot process, and $(\mathbf{e}_i)_{i \geq 1}$ be its Eves. Then for any $n \geq 1$, the process $(\rho_t(\{\mathbf{e}_1\}), \dots, \rho_t(\{\mathbf{e}_n\});\ t \geq 0)$ is distributed as $(\tilde{Z}_1, \dots, \tilde{Z}_n)$ where $(\tilde{Z}_1, \dots, \tilde{Z}_{n+1})$ is a $(n+1)$-dimensional Wright-Fisher diffusion conditioned on its extinction order, started from $(0, \dots, 0, 1)$.*

Proof. We realize a similar computation as in the proof of Theorem 4 of [2]. The proof requires three facts. First notice that

$$\lim_{m \to \infty} \rho_t\big(\big(\tfrac{\lfloor m\mathbf{e}_i \rfloor}{m}, \tfrac{\lfloor m\mathbf{e}_i + 1 \rfloor}{m}\big]\big) = \rho_t(\{\mathbf{e}_i\}).$$

Then, if $I_1, \dots, I_n$ are n disjoint intervals of length $1/m$, due to exchangeability of the increments of bridges, the process $(\rho_t(I_1), \dots, \rho_t(I_n);\ t \geq 0)$ is distributed as the process

$$\Big(\rho_t\big((0, \tfrac{1}{m}]\big), \dots, \rho_t\big(\big(\tfrac{n-1}{m}, \tfrac{n}{m}\big]\big);\ t \geq 0\Big)$$

which is distributed as the n first coordinates of a $(n+1)$-dimensional Wright-Fisher diffusion started from $(\frac{1}{m}, \dots, \frac{1}{m}, 1 - \frac{n}{m})$.

Finally, notice that on the event $\{\forall i \neq j \in \{1,\dots,n\},\ \lfloor m\mathbf{e}_i\rfloor \neq \lfloor m\mathbf{e}_j\rfloor\}$, conditioning the process

$$\Big(\rho_t\big((0,\tfrac{1}{m}]\big),\dots,\rho_t\big((\tfrac{n-1}{m},\tfrac{n}{m}]\big);\, t\geq 0\Big)$$

on its extinction order as in Section 5.2 is equivalent to conditioning it on the location of the Eves, i.e., on the event $\big\{\forall k\in\{1,\dots,n\},\ \mathbf{e}_k\in\big(\frac{k-1}{m},\frac{k}{m}\big]\big\}$.

We can now proceed to the calculation. Let $0\leq t_1<\dots<t_p$ and let $\varphi_1,\dots,\varphi_p$ be continuous bounded functions. Consider $(\tilde{Z}_1,\dots,\tilde{Z}_{n+1})$ a $(n+1)$-dimensional Wright-Fisher diffusion conditioned on its extinction order. Then

$$\begin{aligned}
&\mathbb{E}\Big[\varphi_1\big(\rho_{t_1}(\{\mathbf{e}_1\}),\dots,\rho_{t_1}(\{\mathbf{e}_n\})\big)\dots\varphi_p\big(\rho_{t_p}(\{\mathbf{e}_1\}),\dots,\rho_{t_p}(\{\mathbf{e}_n\})\big)\Big]\\
&=\lim_{m\to\infty}\sum_{i_1=0}^{m-1}\cdots\sum_{i_n=0}^{m-1}\mathbb{E}\Big[\varphi_1\big(\rho_{t_1}\big((\tfrac{i_1}{m},\tfrac{i_1+1}{m}]\big),\dots,\rho_{t_1}\big((\tfrac{i_n}{m},\tfrac{i_n+1}{m}]\big)\big)\dots\\
&\qquad\qquad\varphi_p\big(\rho_{t_p}\big((\tfrac{i_1}{m},\tfrac{i_1+1}{m}]\big),\dots,\rho_{t_p}\big((\tfrac{i_n}{m},\tfrac{i_n+1}{m}]\big)\big)\mathbb{1}_{\left\{\forall k\in\{1,\dots,n\},\ \mathbf{e}_k\in\left(\frac{i_k}{m},\frac{i_k+1}{m}\right]\right\}}\Big]\\
&=\lim_{m\to\infty}m^n\mathbb{E}\Big[\varphi_1\big(\rho_{t_1}\big((0,\tfrac{1}{m}]\big),\dots,\rho_{t_1}\big((\tfrac{n-1}{m},\tfrac{n}{m}]\big)\big)\dots\\
&\qquad\qquad\varphi_p\big(\rho_{t_p}\big((0,\tfrac{1}{m}]\big),\dots,\rho_{t_p}\big((\tfrac{n-1}{m},\tfrac{n}{m}]\big)\big)\mathbb{1}_{\left\{\forall k\in\{1,\dots,n\},\ \mathbf{e}_k\in\left(\frac{k-1}{m},\frac{k}{m}\right]\right\}}\Big]\\
&=\lim_{m\to\infty}\mathbb{E}\Big[\varphi_1\big(\tilde{Z}_1(t_1),\dots,\tilde{Z}_n(t_1)\big)\dots\varphi_p\big(\tilde{Z}_1(t_p),\dots,\tilde{Z}_n(t_p)\big)\mid\tilde{Z}_1(0)=\dots=\tilde{Z}_n(0)=\tfrac{1}{m}\Big]\\
&=\mathbb{E}\Big[\varphi_1\big(\tilde{Z}_1(t_1),\dots,\tilde{Z}_n(t_1)\big)\dots\varphi_p\big(\tilde{Z}_1(t_p),\dots,\tilde{Z}_n(t_p)\big)\mid\tilde{Z}_1(0)=\dots=\tilde{Z}_n(0)=0\Big],
\end{aligned}$$

where, the last line comes from the Feller property of the process $(\tilde{Z}_1,\dots,\tilde{Z}_{n+1})$. □

Our current proof of Theorem 1.3 relies on calculations specific to the Wright-Fisher diffusion. We end this section by discussing a potential alternative proof of this result that would more easily generalize to Beta-coalescents.

The Feller branching diffusion describes the size of a population where different individuals die and reproduce independently. Similarly to the Fleming-Viot process, it is possible to define a measure-valued process, called the Dawson-Watanabe process, that encodes the size of the offspring of each individual in the initial population, see for example [6]. (Note that there are no mutations here, i.e., no spatial motion of the particles.) Its total mass is then distributed as a Feller diffusion. Starting from a Dawson-Watanabe process, one can renormalize it by its total mass to obtain a process valued in the space of probability measures. Then the resulting renormalized process is distributed as a time-changed Fleming-Viot process, see [4].

Let us now discuss the results of Section 5.2 in the light of this new construction. The key point of Section 5.2 is that after removing one family from a Fleming-Viot process and renormalizing the remainder of the population to have mass one, the resulting process remains distributed as an independent time-changed Fleming-Viot process. Suppose that the Fleming-Viot process has been obtained by renormalizing a Dawson-Watanabe process. Then removing a family from the Fleming-Viot process amounts to removing a family from the original Dawson-Watanabe process. By the branching property, removing this family does not change the distribution of the remainder of the population, which remains distributed as an independent Dawson-Watanabe process. Thus when renormalizing the remainder of the population to have size one, we obtain a new time-changed Fleming-Viot process, independent of the removed family. In other words, the results of Section 5.2 essentially originate from the fact that the Fleming-Viot process can be seen as a renormalized branching measure-valued process.

A similar link has been obtained in [4] between the Λ-Fleming-Viot processes associated to Beta-coalescents and a family of α-stable measure-valued branching processes.

Thus we believe that one could derive a similar, but less explicit, representation of the asymptotic frequencies of the stationary distribution of the Beta-coalescents with erosion than the one obtained in Theorem 1.3.

Acknowledgement

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

References

[1] Julien Berestycki. Exchangeable fragmentation-coalescence processes and their equilibrium measures. *Electronic Journal of Probability*, 9:770–824, 2004. MR-2110018

[2] Jean Bertoin and Jean-François Le Gall. Stochastic flows associated to coalescent processes. *Probability Theory and Related Fields*, 126(2):261–288, 2003. MR-1990057

[3] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., second edition, 1999. MR-1700749

[4] Matthias Birkner, Jochen Blath, Marcella Capaldo, Alison Etheridge, Martin Möhle, Jason Schweinsberg, and Anton Wakolbinger. Alpha-stable branching and Beta-coalescents. *Electronic Journal of Probability*, 10:303–325, 2005. MR-2120246

[5] Thomas Duquesne and Cyril Labbé. On the Eve property for CSBP. *Electronic Journal of Probability*, 19:31 pp., 2014. MR-3164759

[6] Alison Etheridge. *An Introduction to Superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, 2000. MR-1779100

[7] Alison Etheridge. *Some Mathematical Models from Population Genetics*, volume 2012 of *École d'Été de Probabilités de Saint-Flour*. Springer Science & Business Media, 2011. MR-2759587

[8] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1986. MR-0838085

[9] Bryan T. Grenfell, Oliver G. Pybus, Julia R. Gog, James L.N. Wood, Janet M. Daly, Jenny A. Mumford, and Edward C. Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303:327–332, 2004.

[10] Tracy A. Heath, John P. Huelsenbeck, and Tanja Stadler. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111:E2957–E2966, 2014.

[11] Olav Kallenberg. *Foundations of Modern Probability*. Probability and its Applications. Springer-Verlag New York, second edition, 2002. MR-1876169

[12] John F.C. Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, s2-18:374–380, 1978. MR-0509954

[13] John F.C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13:235–248, 1982. MR-0671034

[14] Cyril Labbé. From flows of Λ-Fleming-Viot processes to lookdown processes via flows of partitions. *Electronic Journal of Probability*, 19:49 pp., 2014. MR-3227064

[15] Amaury Lambert. Population dynamics and random genealogies. *Stochastic Models*, 24:45–163, 2008. MR-2466449

[16] John Lamperti. Semi-stable markov processes. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 22:205–225, 1972. MR-0307358

[17] James Mallet, Nora Besansky, and Matthew W. Hahn. How reticulated are species? *BioEssays*, 38:140–149, 2016.

[18] Jim Pitman. Coalescents with multiple collisions. *Annals of Probability*, 27:1870–1902, 1999. MR-1742892

[19] Camille Roux, Christelle Fraïsse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLOS Biology*, 14:1–22, 2016.

[20] Serik Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *Journal of Applied Probability*, 36:1116–1125, 1999. MR-1742154

[21] Graham J. Slater, Luke J. Harmon, and Michael E. Alfaro. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution: International Journal of Organic Evolution*, 66:3931–3944, 2012.

[22] Erik M. Volz, Katia Koelle, and Trevor Bedford. Viral phylodynamics. *PLoS Computational Biology*, 9, 2013. MR-3048921

A Proof of Proposition 3.3

In this section, we prove that the ancestral process of Kingman's coalescent with immigration is Markovian. To do this, consider a version of Kingman's coalescent with immigration $(\bar{\Pi}_t)_{t\in\mathbb{R}}$, and let $(\bar{\Pi}_i)_{i\in\mathbb{Z}}$ be its embedded chain, i.e., the sequence of states visited by $(\bar{\Pi}_t)_{t\in\mathbb{R}}$, where $\bar{\Pi}_0$ is the state at time $t=0$. We count the number of trajectories of $(\bar{\Pi}_i)_{i\in\mathbb{Z}}$ that produce a given trajectory of $(\mathcal{A}_i)_{i\geq 0}$, the embedded chain of $(\mathcal{A}_t)_{t\geq 0}$.

First, note that given the values of $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_0)$ and a uniform permutation σ of the blocks of $\bar{\Pi}_0$, one can uniquely reconstruct the values of $(\mathcal{A}_0,\dots,\mathcal{A}_n)$. We now fix a sequence $(a_0,\dots,a_n)$ of possible values of $(\mathcal{A}_0,\dots,\mathcal{A}_n)$, and a partition $\bar{\pi}_{-n}$ with $|a_n|$ blocks, where $|a_n|$ is the total number of particles of a_n. Our first task is to count the number of trajectories of $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_0)$ starting from $\bar{\pi}_{-n}$, and of labelings σ of the blocks of $\bar{\Pi}_0$ such that $(\mathcal{A}_0,\dots,\mathcal{A}_n)=(a_0,\dots,a_n)$. Before stating the result we need to introduce one notation. The variable $\mathcal{A}_{k+1}$ is obtained from $\mathcal{A}_k$ by splitting or killing one particle. Let us denote ℓ_k the label of this particle. That is, ℓ_k is the unique integer such that

$$|\mathcal{A}_{k+1}(\ell_k)-\mathcal{A}_k(\ell_k)|=1,\quad \forall i\neq \ell_k,\ |\mathcal{A}_{k+1}(i)-\mathcal{A}_k(i)|=0.$$

Lemma A.1. *Fix a sequence of states $(a_0,\dots,a_n)$ of $(\mathcal{A}_0,\dots,\mathcal{A}_n)$, and a partition $\bar{\pi}_{-n}$ of $\{i\in\mathbb{Z}: i\leq -n\}$ with $|a_n|$ blocks. Then the number of trajectories of $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_0)$ and labelings of the blocks of $\bar{\Pi}_0$ such that $(\mathcal{A}_0,\dots,\mathcal{A}_n)=(a_0,\dots,a_n)$ and $\bar{\Pi}_{-n}=\bar{\pi}_{-n}$ is*

$$\frac{|a_n|!}{2^b}a_0(\ell_0)\dots a_{n-1}(\ell_{n-1}),$$

where b is the number of birth events along the sequence $(a_0,\dots,a_n)$.

Proof. Each trajectory of $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_0)$ naturally encodes a forest that can be built through the following procedure, which is illustrated in Figure 2. Choose any labeling of the blocks of $\bar{\Pi}_{-n}$, and for each block add an initial leaf with the corresponding label. Suppose that the forest corresponding to $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_{-k})$ has been built. If $\bar{\Pi}_{-k+1}$ is obtained from $\bar{\Pi}_{-k}$ by immigrating a new particle, then add a new isolated vertex. Otherwise, a coalescence event has occurred between two blocks of $\bar{\Pi}_{-k}$. Then add a new internal node and connect it to the nodes corresponding to the two blocks that have coalesced. Once the forest representing $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_0)$ is built, by construction the nodes corresponding to $\bar{\Pi}_0$ all belong to different trees. We set them to be the roots of their respective trees, and label them according to the partition σ. (Notice that the resulting forest is endowed with some additional structure: the nodes added along the procedure are totally ordered by the induction step at which they have been added.)

Counting trajectories of $(\bar{\Pi}_{-n},\dots,\bar{\Pi}_0)$ now amounts to counting forests. Instead of building the forests by starting from the leaves as above, we build a forest with ancestral sequence $(a_0,\dots,a_n)$ by starting from the roots. Initially, consider a set of $|a_0|$ roots, labeled by $\{1,\dots,|a_0|\}$, that represent the particles of a_0. Nodes can be in two states: active or inactive. Active nodes represent the particles that are still alive in the population while inactive nodes represent the dead particles. Initially all roots are active.

We build the forest recursively. Suppose that at step k we have built a forest such that for all i there are $a_k(i)$ nodes that are active in the tree with root i. If a particle with label ℓ_k has died from a_k to a_{k+1}, we inactivate one of the nodes belonging to the tree with root ℓ_k. There are $a_k(\ell_k)$ such nodes. Similarly, if a particle has split from a_k to a_{k+1}, we inactivate one node in the tree ℓ_k, and connect it to two new active nodes. There are again $a_k(\ell_k)$ active nodes in the tree ℓ_k. After step n, we have built a forest with ancestral sequence $(a_0, \dots, a_n)$. We assign the blocks of $\bar{\Pi}_{-n}$ to the remaining active nodes of the forest by choosing one of the $|a_n|!$ permutations of the blocks.

There are

$$|a_n|!\, a_0(\ell_0) \dots a_{n-1}(\ell_{n-1})$$

outputs of the previous construction, and all forests with ancestral sequence $(a_0, \dots, a_n)$ can be obtained that way. However, due to symmetries, some forests can be obtained multiple times through this construction. More precisely, at each birth event, the two daughter nodes are indistinguishable. Interchanging the trees corresponding to the offspring of these two nodes yields the same forest. Thus, the actual number of forests with ancestral sequence $(a_0, \dots, a_n)$ is

$$\frac{|a_n|!}{2^b} a_0(\ell_0) \dots a_{n-1}(\ell_{n-1})$$

where b is the number of birth events, and the result is proved. □

Lemma A.2. *Let $(M_t)_{t\in\mathbb{R}}$ be the process counting the number of blocks of Kingman's coalescent with immigration. Then $(M_t)_{t\in\mathbb{R}}$ is a stationary Markov process such that conditional on $\{M_t = k\}$, it jumps to*

- *$k+1$ at rate d;*
- *$k-1$ at rate $k(k-1)/2$.*

Moreover $(M_t)_{t\in\mathbb{R}}$ is a reversible process.

Proof. Let us consider a version of Kingman's coalescent with immigration built from a Poisson point process P. Let us first show that $(M_t)_{t\in\mathbb{R}}$ is a Markov process. Conditional on $M_t = k$, each of the $k(k-1)/2$ pairs of blocks coalesce a rate one, and new atoms of P immigrate at rate d. Thus, $(M_t)_{t\in\mathbb{R}}$ goes to $k-1$ at rate $k(k-1)/2$ and to $k+1$ at rate d.

Let us now argue that the family of variables $(M_t)_{t\in\mathbb{R}}$ is tight. Fix $t \in \mathbb{R}$ and let T be the location of the most recent atom of P before time t. Then $t - T$ is exponentially distributed with parameter d, and M_t is distributed as the number of blocks of Π_{t-T}, where $(\Pi_s)_{s\geq 0}$ is a version of Kingman's coalescent started with M_T blocks. Thus M_t is stochastically dominated by the number of blocks Π'_{t-T}, where $(\Pi'_s)_{s\geq 0}$ is a version of Kingman's coalescent started from an infinite number of blocks. As each variable M_t is stochastically dominated by the same variable, the family is tight.

It is not hard to see that a Markov process jumping from k to $k+1$ at rate d, and from k to $k-1$ at rate $k(k-1)/2$ admits a unique stationary distribution. As it is irreducible we have

$$\forall k \geq 1,\ \mathbb{P}(M_t = i \mid M_s = j) \underset{t\to\infty}{\longrightarrow} \mathbb{P}\big(M_\infty = i\big).$$

Thus, using the tightness of $(M_t)_{t\geq 0}$, we have

$$\mathbb{P}\big(M_t = i\big) = \sum_{j\geq 1} \mathbb{P}\big(M_s = j\big)\mathbb{P}(M_{t-s} = i \mid M_s = j) \underset{s\to-\infty}{\longrightarrow} \mathbb{P}\big(M_\infty = i\big).$$

Let us compute the stationary distribution of $(M_t)_{t\in\mathbb{R}}$. As $(M_t)_{t\in\mathbb{R}}$ jumps from k to $k+1$ at rate d and from k to $k-1$ at rate $k(k-1)/2$, a usual calculation shows that its stationary distribution $(\nu_k)_{k\geq 1}$ is

$$\forall k \geq 1,\ \nu_k \propto \frac{(2d)^k}{k!\,(k-1)!}$$

where the renormalization constant is obtained by summing over all the terms. Thus a direct calculation now proves that $(\nu_k)_{k\geq 1}$ fulfills the detailed balance equation

$$\forall k \geq 1,\ d\nu_k = \frac{k(k+1)}{2}\nu_{k+1}$$

and thus that $(M_t)_{t\in\mathbb{R}}$ is reversible. □

We are now ready to prove Proposition 3.3

Proof of Proposition 3.3. Recall the notations from Section 3.1. As proved in Lemma A.2, the process $(M_t)_{t\in\mathbb{R}}$ that counts the number of the blocks of Kingman's coalescent with immigration is a reversible Markov process. Thus, the process $(N_t)_{t\geq 0}$ that gives the number of particles of $(\mathcal{A}_t)_{t\geq 0}$ is a stationary process jumping from k to $k+1$ at rate d, and from k to $k-1$ at rate $k(k-1)/2$. Hence, the result is proved if we show that conditional on the sequence of states $(N_0,\dots,N_n)$ visited by $(N_t)_{t\geq 0}$, the type of the particle that dies or splits from $\mathcal{A}_k$ to $\mathcal{A}_{k+1}$ is chosen with a probability proportional to the vector $\mathcal{A}_k$.

Let b denote the number of birth events along the sequence $(a_0,\dots,a_n)$. (Hence, forward in time, there are $n-b$ immigration events.) We have

$$\mathbb{P}\big(\mathcal{A}_0 = a_0,\dots,\mathcal{A}_n = a_n\big) = \sum_{(\bar\pi_{-n},\dots,\bar\pi_0)}\sum_{s}\mathbb{P}\big(\forall i<n,\ \bar\Pi_{-i}=\bar\pi_{-i},\ \sigma = s \,\big|\, \bar\Pi_{-n}=\bar\pi_{-n}\big)\,\mathbb{P}\big(\bar\Pi_{-n}=\bar\pi_{-n}\big)$$

where the sum is taken over all partitions $\bar\pi_{-n}$ of $\{i\in\mathbb{Z} : i \leq -(n-b)\}$ with $|a_n|$ blocks, all trajectories $(\bar\pi_{-n+1},\dots,\bar\pi_0)$ and labelings s of the blocks of $\bar\pi_0$ such that $(\mathcal{A}_0,\dots,\mathcal{A}_n) = (a_0,\dots,a_n)$. Now notice that the probability of seeing such a trajectory and labeling does only depend on the sequence of number of blocks $(|a_0|,\dots,|a_n|)$. Indeed, conditional on $(|a_0|,\dots,|a_n|)$, two trajectories $(\bar\Pi_{-n},\dots,\bar\Pi_0)$ are identical up to the choice of the pairs of blocks that merge at each coalescence event, and these pairs are chosen uniformly.

Thus the probability of the event $\{\mathcal{A}_0 = a_0,\dots,\mathcal{A}_n = a_n\}$ is proportional to the number of terms in the sum, and thus to the number of trajectories of $(\bar\Pi_{-n},\dots,\bar\Pi_0)$ that correspond to this ancestral sequence. Hence, Lemma A.1 shows that

$$\mathbb{P}\big(\mathcal{A}_0 = a_0,\dots,\mathcal{A}_n = a_n\big) = C(|a_0|,\dots,|a_n|)a_0(\ell_0)\dots a_{n-1}(\ell_{n-1}),$$

where the coefficient $C(|a_0|,\dots,|a_n|)$ only depends on $(|a_0|,\dots,|a_n|)$. This proves the result. □

Let us end this section by discussing a possible extension to Λ-coalescents. The key point here is that conditional on the block counting process, the particles that die or split are chosen uniformly in the population. This is a consequence of 1) Lemma A.1 and 2) the fact that all trajectories with a given sequence of number of blocks have the same probability. The second point is a consequence of exchangeability so remains valid for Λ-coalescents. As for Lemma A.1, the proof could be easily adapted to Λ-coalescents

with immigration. (The factor 2^b should be replaced by the product of the number of blocks involved in coalescence events.)

Thus, the only difference between Kingman's coalescent with immigration and more general Λ-coalescents with immigration is that the block counting process is no longer reversible. Hence we cannot obtain a closed form for the transition rates of the corresponding ancestral processes. Nevertheless, we believe that in some cases it should be possible to obtain a result similar to Theorem 1.5 by using the same techniques as in this paper, if one can derive a good enough approximation for the stationary distribution of the number of blocks.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System `http://www.vtex.lt/en/ejms.html`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `http://www.imstat.org/publications/open.htm`