EXCHANGEABLE COALESCENTS

Jean BERTOIN

Nachdiplom Lectures

ETH Zürich, Autumn 2010

The purpose of this series of lectures is to introduce and develop some of the main aspects of a class of random processes evolving by coalescence, which arise in the study of the genealogy of certain large populations. Let us first present the naive idea.

For many years, probability theory has provided models of population which evolve forward in time: Galton-Watson processes, branching processes, Moran and Fleming-Viot processes, birth and death processes, etc., so roughly speaking the genealogy is view from the ancestor(s). Kingman has been the first to investigate the dual point of view point, i.e. from the offspring going backward in time. Typically, consider a population at the present date. For the sake of simplicity, let us assume that this population is haploid with non-overlapping generations. We can decompose it into siblings (brothers and sisters), or families of grand-children, and so on. For each integer n, we have a natural partition into families of individuals having the same ancestor n generations backwards. Plainly, these partitions get coarser as n increases, and more precisely a merging of sub-families corresponds to coalescence of ancestral lineages.

Loosely speaking, we are interested in the study of such coalescent processes for large populations and after an appropriate rescaling of time. In this direction, we shall first develop some basic material on various natural spaces of partitions and their connexions, including the fundamental result due to Kingman that characterizes exchangeable random partitions as mixtures of paint-box processes. Then we will introduce the coalescent of Kingman. At the end of previous millennium, Pitman, Sagitov, Möhle and Schweinsberg considered natural generalizations of Kingman that rely crucially on the concept of exchangeability; we will provide a synthetic presentation of their contributions. We shall then study the dual population models whose genealogy is described by such exchangeable coalescent processes; this leads us to the so-called generalized Fleming-Viot processes which have been introduced first by Donnelly and Kurtz and then by Bertoin and Le Gall. The last chapters are devoted to two particular and important examples of exchangeable coalescents, namely the Bolthausen-Sznitman coalescent and the Beta coalescents.

At this point, I would like to stress in an informal way an important feature of this framework. On the one hand, because populations are large, we have to rescale families

and therefore, in some sense, to work with continuous limits. On the other hand, even though populations are large, they consist of individuals who thus can be sampled. In practice, one samples a few individuals (hundreds to thousands) from a very large population (billions of individuals) and investigate their genealogy by comparing their DNA. One then expects to learn something on the whole population from the analysis of a small sample. It turns out that the statistical nature of this point of view will have a prominent role in the study, even for its most theoretical and purely probabilistic aspects.

This text mainly consists of a re-arrangment of some sections of my book (*Random Fragmentation and Coagulation Processes.* Cambridge University Press, 2006) and some hacking of contributions of other authors on the subject. Nathanaël Berestycki has recently written down very pedagogical lecture notes [5] to which we refer as a parallel presentation of the subject.

Table of Contents

Chapter 1

Random partitions

We start by presenting different aspects of partitions in the deterministic setting, and then develop the fundamental concept of exchangeable random partition due to Kingman. We shall conclude this chapter by discussing the special case of Poisson-Dirichlet partitions, which arises in a variety of situations.

1.1 Various notions of partitions

1.1.1 Partitions of a unit mass

Roughly speaking, a mass partition may be thought of as the sequence, ranked in decreasing order, of the masses of clusters in a universe with unit total mass.

Definition 1.1 *A* **mass-partition** *is an infinite numerical sequence*

$$\mathbf{s} = (s_1, s_2, \ldots)$$

such that

$$s_1 \geq s_2 \geq \cdots \geq 0 \quad \textit{and} \quad \sum_{i=1}^{\infty} s_i \leq 1\,.$$

The terms s_i are referred to as fragments of the mass-partition $\mathbf{s}$*. The space of mass-partitions is denoted by $\mathcal{P}_{\mathtt{m}}$.*

We stress that the total mass of the clusters, $\sum_{i=1}^{\infty} s_i$, can be strictly less than 1. In this direction, it is convenient to define

$$s_0 := 1 - \sum_{i=1}^{\infty} s_i\,,$$

a quantity which can be thought of as the total mass of *dust* (i.e. infinitesimal particles) in the universe. A mass-partition $\mathbf{s}$ is called *proper* if there is no dust[1], that is if $s_0 = 0$,

[1]The terminology is perhaps better understood in French: *une partition de masse est dite propre si elle n'a pas de poussière.*

and *improper* otherwise. It is sometimes convenient to think of a mass-partition as the ranked sequence of the masses of the atoms of some probability measure; then the mass of dust corresponds to the mass of the continuous component of this probability measure.

It is easy to check that the space of mass-partitions enjoys nice topological properties.

Proposition 1.1 *The space $\mathcal{P}_{\mathtt{m}}$, endowed with the uniform distance*

$$d(\mathbf{s},\mathbf{s}') \;=\; \max\left\{|s_i - s'_i|, i\in\mathbb{N}\right\}\,, \qquad \mathbf{s},\mathbf{s}'\in\mathcal{P}_{\mathtt{m}},$$

is compact, and the induced topology coincides with that of pointwise convergence.

Proof Because terms of a mass-partition are ranked in decreasing order and their sum does not exceed 1, we have

$$|s_i - s'_i| \le 1/i \qquad \text{for all } i\in\mathbb{N} \text{ and } \mathbf{s},\mathbf{s}'\in\mathcal{P}_{\mathtt{m}}.$$

This implies that pointwise convergence in $\mathcal{P}_{\mathtt{m}}$ is equivalent to uniform convergence.

Consider now some sequence $(\mathbf{s}^{(n)}, n\in\mathbb{N})$ in $\mathcal{P}_{\mathtt{m}}$. By the diagonal procedure, we may extract some subsequence, which we still denote by $(\mathbf{s}^{(n)}, n\in\mathbb{N})$ for convenience, such that $\lim_{n\to\infty} s_i^{(n)} := s_i^{(\infty)}$ exists for each $i\in\mathbb{N}$. Plainly, one has $s_1^{(\infty)}\ge s_2^{(\infty)}\ge\cdots\ge 0$ and, by Fatou's lemma, $\sum_{i=1}^{\infty} s_i^{(\infty)}\le 1$. By the theorem of Bolzano-Weierstrass, this establishes the compactness. □

The fact that $\mathcal{P}_{\mathtt{m}}$ is a compact metric space will be especially useful later in the text when we will deal with the construction of $\mathcal{P}_{\mathtt{m}}$-valued random variables or processes. We now turn our attention to some alternative representations of mass-partitions, which we shall also use for different purposes.

1.1.2 Interval-partitions

It is well-known that every open subset of the unit interval $I=]0,1[$ can be decomposed into a unique (at most) countable collection of disjoint open intervals. In our framework, we may thus view such open sets as interval-partitions; points in the complementary closed set can be thought of as dust, that is isolated infinitesimal particles.

Definition 1.2 *The collection of the interval-components of an arbitrary open set $\vartheta\subseteq]0,1[$ is called an* **interval-partition**. *By a slight abuse of notation, we shall often identify this interval-partition with the open set ϑ. The space of interval-partitions is denoted by $\mathcal{P}_{\mathrm{I}}$.*

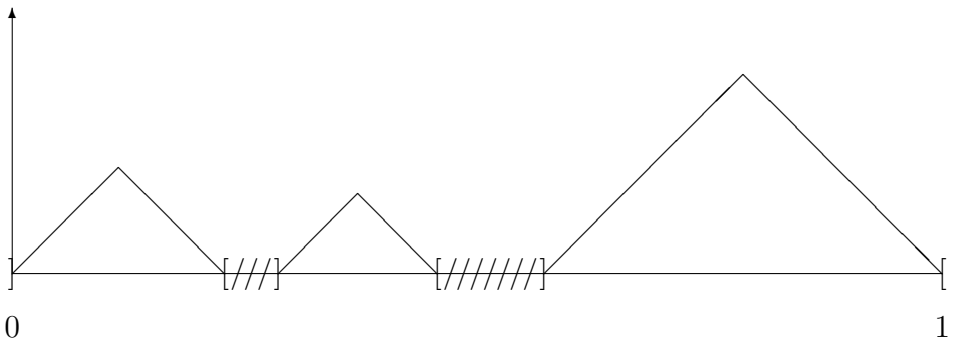
The lengths of the interval components of an interval-partition ϑ are called **spacings**. *We denote by $|\vartheta|^{\downarrow}$ the sequence of spacings, ranked in decreasing order and completed by an infinite sequence of 0 when ϑ only has finitely many interval components, so that $|\vartheta|^{\downarrow}$ is a mass-partition.*

Given an arbitrary mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$, it is not difficult to construct an interval-partition $\vartheta \in \mathcal{P}_{\mathrm{I}}$ with $|\vartheta|^{\downarrow} = \mathbf{s}$; such ϑ will then be called an *interval-representation* of $\mathbf{s}$. Clearly, there are in general many different interval-representations of the same mass-partition. Observe also that a mass-partition is proper if and only if some (and then all) of its interval-representations has full Lebesgue measure.

There is a natural distance on $\mathcal{P}_{\mathrm{I}}$ that we now introduce. Each interval-partition ϑ is determined by the function

$$\chi_{\vartheta}(x) \,=\, \min\{|y-x|, y \in \vartheta^c\}\,, \qquad x \in [0,1]\,,$$

where $\vartheta^c = [0,1]\backslash\vartheta$ stands for the complementary set of ϑ in $[0,1]$ (so ϑ^c is a closed set which always contains the boundary points 0 and 1).



Interval-partition ϑ with 3 interval components and graph of χ_{ϑ}

We then define the distance

$$d(\vartheta, \vartheta') \,=\, \max\{|\chi_{\vartheta}(x) - \chi_{\vartheta'}(x)|, x \in [0,1]\}\,.$$

We mention that $d(\vartheta, \vartheta')$ can also be viewed as the Hausdorff distance between the closed sets ϑ^c and ϑ'^c.

Proposition 1.2 *The space $(\mathcal{P}_{\mathrm{I}}, d)$ is compact, and the map $\vartheta \to |\vartheta|^{\downarrow}$ is continuous from $\mathcal{P}_{\mathrm{I}}$ to $\mathcal{P}_{\mathtt{m}}$.*

Proof We have to check sequential compactness, that is for every sequence $(\vartheta_n, n \in \mathbb{N})$ in $\mathcal{P}_{\mathrm{I}}$, we can extract a subsequence which converges in $\mathcal{P}_{\mathrm{I}}$. For each $n \in \mathbb{N}$, let $]a_{1,n}, b_{1,n}[,]a_{2,n}, b_{2,n}[, \ldots$ denote the sequence of the interval components of ϑ_n ordered by decreasing lengths (if ϑ_n has only k interval components, then we agree that $a_{j,n} = b_{j,n} = 0$ for $j > k$, and if two or more interval components have the same positive length, then we order them from the left to the right). By diagonal extraction procedure, we suppose that for each i, the sequences $(a_{i,n}, n \in \mathbb{N})$ and $(b_{i,n}, n \in \mathbb{N})$ converge, say to a_i and b_i,

where $0 \leq a_i \leq b_i \leq 1$. Plainly the intervals $]a_i, b_i[$, $i \in \mathbb{N}$, must be pairwise disjoint, so we may consider the interval-partition

$$\vartheta := \bigcup_{i \in \mathbb{N}}]a_i, b_i[\, .$$

Now, by the triangle inequality, we have that for every integers n, k

$$d(\vartheta_n, \vartheta) \leq \max_{i=1,\ldots,k} (|a_{i,n} - a_i| + |b_{i,n} - b_i|) + \max_{j>k} (|b_{j,n} - a_{j,n}| + |b_j - a_j|) \, .$$

As the j-th largest spacing of an interval-partition is at most $1/j$, we see that the second maximum in the right-hand side is bounded by $2/k$. Furthermore, for each fixed k, the first maximum converges to 0 as $n \to \infty$, so $\lim_{n\to\infty} d(\vartheta_n, \vartheta) = 0$, and the compactness is proven.

Next, observe that for every interval-partition $\vartheta \in \mathcal{P}_{\mathrm{I}}$ and every integer k, the sum of the lengths of the k largest interval components of ϑ can be expressed as

$$\sum_{i=1}^{k} |\vartheta|_i^{\downarrow} = \max \sum_{j=1}^{2k} |\chi_\vartheta(x_j) - \chi_\vartheta(x_{j-1})| \, ,$$

where in the right-hand side, the maximum is taken over the set of subdivisions $0 = x_0 < x_1 < \cdots < x_{2k} = 1$ of $[0, 1]$. We deduce that if $(\vartheta_n, n \in \mathbb{N})$ is a sequence of interval-partitions which converges to $\vartheta \in \mathcal{P}_{\mathrm{I}}$, then

$$\lim_{n\to\infty} \sum_{i=1}^{k} |\vartheta_n|_i^{\downarrow} = \sum_{i=1}^{k} |\vartheta|_i^{\downarrow} \, ,$$

so $|\vartheta_n|^{\downarrow}$ converges pointwise to $|\vartheta|^{\downarrow}$. □

The order relation $\leq$ on the unit interval induces a natural order (from the left to the right) for the components of an interval-partition, and thus we can think of an interval representation $\vartheta \in \mathcal{P}_{\mathrm{I}}$ of a mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ as reordering the terms of $\mathbf{s}$. We shall now present two useful procedures for reordering randomly the terms of mass-partitions: the uniform random order and the size-biased random order.

The first is doubtless the simplest and more intuitive. It is obvious that given a sample of n variables from some continuous distribution on $\mathbb{R}$, the random permutation that reorder the sample and produces the order statistics has the uniform distribution on the space of permutations of n elements. When we deal with an infinite sequence $\mathbf{s} = (s_1, \ldots)$ defined by a mass partition , we introduce an i.i.d. sequence of uniform variables, $U_1, \ldots$ and decide that to put the fragment s_i at the left of s_j whenever $U_i < U_j$. More precisely, we consider the random function

$$F(x) = x s_0 + \sum_{i=1}^{\infty} s_i \mathbf{1}_{\{x \leq U_i\}} \, ,$$

so $F : [0,1] \to [0,1]$ is right-continuous non-decreasing with $F(0) = 0$ and $F(1) = 1$, i.e. F is the distribution function of some random probability measure on $[0,1]$. More precisely, the U_i's are the locations of the atoms, and the masses $s_i = F(U_i) - F(U_i-)$ of the atoms are given by the fragments of the mass-partition $\mathbf{s}$; recall that they also coincide with the lengths of the gaps in the range of F. The spacings induced by the open set

$$(0, F(1))\backslash\{F(x) : x \in [0,1]\}^{\mathrm{cl}}$$

may thus be viewed as a random interval representation of the mass partition $\mathbf{s}$ in which fragments appear in the uniform random order.

We next turn our attention to size-biased reordering. A *proper* mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ can be viewed as a discrete probability measure; more precisely we may associate to $\mathbf{s}$ an integer-valued random variable 1^* with distribution

$$\mathbb{P}(1^* = i) \,=\, s_i\,, \qquad i \in \mathbb{N}\,.$$

One then calls any random variable s_1^* distributed as s_{1^*} a *size-biased sample* from $\mathbf{s}$. The law of a size-biased sample s_1^* is referred to as the *structural distribution* of $\mathbf{s}$. In other words, the structural distribution of $\mathbf{s}$ is given by $\sum_{i=1}^{\infty} s_i \delta_{s_i}$, where δ_s stands for the Dirac point mass at s, and plainly one can recover $\mathbf{s}$ from its structural distribution. When S is a *random* mass-partition which is proper a.s., we call a variable whose conditional distribution given $S = \mathbf{s}$ is that of a size-biased sample of $\mathbf{s}$ a size-biased sample of S. We stress that in general, one cannot recover the law of a random mass-partition from that of its size-biased sample.

It will also be convenient to define size-biased sampling for possibly *improper* mass-partitions. In this more general situation, we consider a random variable 1^* with values in $\overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$ with law given by

$$\mathbb{P}(1^* = i) \,=\, s_i \quad \text{ for } \quad i \in \mathbb{N}\,, \quad \text{ and } \quad \mathbb{P}(1^* = \infty) \,=\, s_0 \,=\, 1 - \sum_{i=1}^{\infty} s_i\,,$$

and we agree that $s_\infty = 0$. Just as above, we call a random variable s_1^* distributed as s_{1^*} a size-biased sample of $\mathbf{s}$.

By sampling recursively with a size-bias and without replacement the terms of some proper mass-partition $\mathbf{s}$, we obtain a so-called size-biased reordering of $\mathbf{s}$. Before giving a formal definition, it is convenient to introduce the space $\mathcal{S}_{[0,1]}$ of numerical sequences $\mathbf{x} = (x_1, \ldots)$ with values in $[0,1]$. This space is equipped with the distance

$$\delta(\mathbf{x}, \mathbf{x}') := \sum_{i=1}^{\infty} 2^{-i} |x_i - x_i'|\,, \qquad \mathbf{x} = (x_1, \ldots), \mathbf{x}' = (x_1', \ldots)\,.$$

This makes $\mathcal{S}_{[0,1]}$ compact, and the induced topology coincides with that of pointwise convergence.

For the sake of simplicity, we again focus on *proper* mass-partitions $\mathbf{s} = (s_1, \ldots)$ and give a formal definition. [2]

[2]There would be at least two natural definitions for size-biased reordering for possibly improper mass-partitions, which only coincide in the proper case. Rather than choosing one arbitrarily, we focus here on the proper case.

Definition 1.3 *Let* $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ *be a proper mass-partition. A* **size-biased permutation** *(based on* $\mathbf{s}$*) is a random map* $\sigma : \mathbb{N} \to \mathbb{N}$ *whose finite-dimensional distributions are given as follows:*

• *For every* $n \in \mathbb{N}$ *such that* $s_n > 0$ *and every* n*-tuple* $k_1, \ldots, k_n$ *of distinct integers,*

$$\mathbb{P}\left(\sigma(1) = k_1, \ldots, \sigma(n) = k_n\right) \;=\; \prod_{i=1}^{n} \frac{s_{k_i}}{1 - \left(s_{k_1} + \cdots + s_{k_{i-1}}\right)} \,,$$

where by convention, the term in the product above corresponding to $i = 1$ *is equal to* s_{k_1}*.*

• *When* $\ell := \inf\left\{n \in \mathbb{N} : s_n = 0\right\} < \infty$*, we agree that for every* $n \geq \ell$ *and every* n*-tuple* $k_1, \ldots, k_n$ *of distinct integers with* $k_i = i$ *for every* $i \geq \ell$*,*

$$\mathbb{P}\left(\sigma(1) = k_1, \ldots, \sigma(n) = k_n\right) \;=\; \prod_{i=1}^{\ell} \frac{s_{k_i}}{1 - \left(s_{k_1} + \cdots + s_{k_{i-1}}\right)} \,.$$

The random sequence $\mathbf{s}^* = (s_{\sigma(1)}, s_{\sigma(2)}, \ldots)$ *is then called a* **size-biased reordering** *of* $\mathbf{s}$*.*

More generally, when S is a random mass-partition which is proper a.s., we call a random variable S^* with values in $\mathcal{S}_{[0,1]}$ whose conditional distribution given $S = \mathbf{s}$ is that of a size-biased reordering of $\mathbf{s}$ a size-biased reordering of S. Note in particular that S_1^* is a size-biased sample from S (so our notation is coherent), and moreover that with probability one, the unordered families $\{S_i, i \in \mathbb{N}\}$ and $\{S_i^*, i \in \mathbb{N}\}$ coincide. In particular, we may recover S from S^* by ranking the terms of the latter in decreasing order.

Here is a simple procedure for constructing size-biased reordering. Fix a proper mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$, and let $\vartheta_{\mathbf{s}}$ be some interval representation of $\mathbf{s}$. So $\vartheta_{\mathbf{s}} \subseteq]0,1[$ is an open set with Lebesgue measure 1. Let $U_1, \ldots$ be a sequence of i.i.d. uniform variables on $[0,1]$. Next define I_1^* as the interval component of $\vartheta_{\mathbf{s}}$ which contains U_1, and then recursively I_n^* for $n \geq 2$ as the interval component of $\vartheta_{\mathbf{s}}$ which contains U_{n^*}, where

$$n^* \;:=\; \inf\left\{ k \in \mathbb{N} : U_k \in \vartheta_{\mathbf{s}} \setminus \bigcup_{j=1}^{n-1} I_j^* \right\} .$$

In words, I_n^* is the n-th interval component of $\vartheta_{\mathbf{s}}$ which is visited by the sequence $U_1, \ldots$ As usual, we write $|I_n^*|$ for its length. When $\ell := \inf\{j \in \mathbb{N} : s_j = 0\} < \infty$, $\vartheta_{\mathbf{s}}$ has exactly $\ell - 1$ non-empty interval-components, and we agree that $I_n^* = \varnothing$ and $|I_n^*| = 0$ for $n \geq \ell$.

Lemma 1.1 *The sequence* $(|I_1^*|, \ldots)$ *is a size-biased reordering of* $\mathbf{s}$*.*

Proof As the variable U_1 is uniformly distributed on $\vartheta_{\mathbf{s}}$, the length $|I_1^*|$ of the interval component of $\vartheta_{\mathbf{s}}$ that contains U_1 is a size-biased sample from $\mathbf{s}$. An easy induction completes the proof. □

1.1.3 Partitions of discrete sets

Recall that $\mathbb{N} = \{1, \ldots\}$ stands for the set of positive integers; a *block* is a subset $B \subseteq \mathbb{N}$. Hereafter, the block formed by the k first integers will play a special role, and it will be convenient to use the notation

$$[k] := \{1, \ldots, k\} .$$

In this direction, we also agree that $[\infty] := \mathbb{N}$ for $k = \infty$.

Definition 1.4 (i) *A partition of $B \subseteq \mathbb{N}$ is a countable collection $\pi = \{\pi_i, i \in \mathbb{N}\}$ of pairwise disjoint blocks such that $\bigcup_{i\in\mathbb{N}} \pi_i = B$, which are always enumerated in increasing order of their least element, that is*

$$\min \pi_i \leq \min \pi_j \qquad \text{for every } i \leq j ,$$

with the convention that $\min \varnothing = \infty$.

(ii) *We write $\mathcal{P}_B$ for the set of partitions of B. In the special case when $B = [k]$ for some $k \in \overline{\mathbb{N}}$, we simply write $\mathcal{P}_k := \mathcal{P}_{[k]}$; in particular $\mathcal{P}_\infty := \mathcal{P}_{\mathbb{N}}$.*

(iii) *We denote by*

$$\#\pi := \#\{i \in \mathbb{N} : \pi_i \neq \varnothing\} = \sup\{i \in \mathbb{N} : \pi_i \neq \varnothing\}$$

the cardinal of the set of non-empty blocks of a partition π.

If $B' \subseteq B$ is a subset of some block B and $\pi \in \mathcal{P}_B$ a partition of B, we write $\pi_{|B'}$ for the obvious restriction of π to B', that is the partition of B' induced by the sequence of blocks $(\pi_i \cap B', i \in \mathbb{N})$. Restricted partitions naturally yield the notion of compatibility.

Definition 1.5 *A sequence $\pi^{[1]}, \pi^{[2]}, \ldots$ of partitions of $[1], [2], \ldots$ is called* **compatible** *if for all integers $k' \leq k$, $\pi^{[k']}$ coincides with the restriction of $\pi^{[k]}$ to $[k']$.*

Plainly, if $\pi \in \mathcal{P}_\infty$ is a partition of $\mathbb{N}$, then the sequence of its restrictions $(\pi_{|[n]}, n \in \mathbb{N})$ is compatible. It is easy to see that the converse holds; here is a formal statement.

Lemma 1.2 *A sequence of partitions $(\pi^{[n]} : \pi^{[n]} \in \mathcal{P}_n$ and $n \in \mathbb{N})$ is compatible if and only if there exists $\pi \in \mathcal{P}_\infty$ such that $\pi_{|[n]} = \pi^{[n]}$ for every $n \in \mathbb{N}$. Moreover, π is then uniquely determined by the sequence $(\pi^{[n]}, n \in \mathbb{N})$.*

Proof The compatibility assumption and our rule for labelling blocks of partitions show that for each $i \in \mathbb{N}$, the sequence of blocks $(\pi_i^{[n]}, n \in \mathbb{N})$ increases. If we define

$$\pi_i := \bigcup_{n\in\mathbb{N}} \pi_i^{[n]} , \qquad i \in \mathbb{N} ,$$

then $(\pi_i, i \in \mathbb{N})$ is a partition of $\mathbb{N}$, say π, and plainly $\pi_{|[n]} = \pi^{[n]}$ for every n. □

This elementary observation points at the rooted tree structure of the set of partition $\mathcal{P}_\infty$. More precisely, the root is given by the unique partition of $\mathcal{P}_1$, $\mathcal{P}_n$ corresponds to the set of vertices at generation n, and there is an edge between $\pi \in \mathcal{P}_n$ and $\pi' \in \mathcal{P}_{n+1}$ if and only if π is the restriction of π' to $[n]$. In this framework, an infinite branch from the root consists in a compatible sequence of partitions of $[1], [2], \ldots$, and the associated leaf is the partition of $\mathbb{N}$ that corresponds to this sequence. In particular, this enables us to define a natural metric on $\mathcal{P}_\infty$.

Lemma 1.3 *The space $\mathcal{P}_\infty$ is endowed with the ultra-metric*

$$d(\pi, \pi') \;=\; 1/\max\left\{k \in \mathbb{N} : \pi_{|[k]} = \pi'_{|[k]}\right\}, \qquad \pi, \pi' \in \mathcal{P}_\infty,$$

with the convention that $1/\max \mathbb{N} = 0$. Then $(\mathcal{P}_\infty, d)$ is compact.

The terminology *ultra-metric* refers to the fact that the usual triangle inequality can be reinforced into

$$d(\pi, \pi'') \leq \max(d(\pi, \pi'), d(\pi', \pi'')), \qquad \text{for every } \pi, \pi', \pi'' \in \mathcal{P}_\infty.$$

Proof Given a sequence $\left(\pi^{(n)}, n \in \mathbb{N}\right)$ in $\mathcal{P}_\infty$, we may extract by the diagonal procedure a subsequence, which is still denoted by $\left(\pi^{(n)}, n \in \mathbb{N}\right)$ for the sake of convenience, such that for every $k \in \mathbb{N}$, the restriction of $\pi^{(n)}$ to $[k]$ is the same for all sufficiently large n. More precisely

$$\pi^{(n)}_{|[k]} \;=\; \pi^{(n')}_{|[k]} \qquad \text{whenever } n, n' \geq k.$$

This defines a compatible sequence of partitions of $[k]$ when k varies in $\mathbb{N}$, which can thus be expressed in the form $\pi_{|[k]}$ for some $\pi \in \mathcal{P}_\infty$, thanks to Lemma 1.2. It is now straightforward to see that, by construction, $d(\pi, \pi^{(n)}) \leq 1/n$, which establishes the compactness. □

Next, we turn our attention to the notion of asymptotic frequency of blocks, which provides a simple map from $\mathcal{P}_\infty$ to the space $\mathcal{P}_{\mathtt{m}}$ of mass-partitions.

Definition 1.6 (i) *We say that a block B possesses an* **asymptotic frequency** *if and only if the limit*

$$|B| \;:=\; \lim_{n\to\infty} \frac{1}{n} \#(B \cap [n]) \;=\; \lim_{n\to\infty} \frac{1}{n} \#\left\{k \in B : k \leq n\right\}$$

exists.

(ii) *If each block of some partition π has an asymptotic frequency, then we say that π possesses asymptotic frequencies. We then write $|\pi| = (|\pi_1|, \ldots)$, and then $|\pi|^{\downarrow} = (|\pi|^{\downarrow}_1, \ldots)$ for the mass-partition [3] given by the decreasing rearrangement of the sequence $|\pi|$.*

[3] Fatou's lemma implies that when a partition possesses asymptotic frequencies, then their sum is cannot exceed 1.

(iii) *We say that a partition π has* **proper** *asymptotic frequencies if π possesses asymptotic frequencies with*

$$\sum_{i=1}^{\infty} |\pi_i| \;=\; 1\,.$$

When some block of a partition π does not have an asymptotic frequency, we decide to write $|\pi| = |\pi|^{\downarrow} = \partial$, where ∂ stands for some extra point added to $\mathcal{P}_{\mathtt{m}}$. This allows us to define a natural map $\pi \to |\pi|^{\downarrow}$ from $\mathcal{P}_\infty$ to $\mathcal{P}_{\mathtt{m}} \cup \{\partial\}$ which is measurable but not continuous. Indeed, for any $\pi \in \mathcal{P}_\infty$ and any $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$, one can easily construct a sequence of partitions $(\pi^{(n)}, n \in \mathbb{N})$ that converges to π and such that $|\pi^{(n)}|^{\downarrow}$ tends to $\mathbf{s}$ as $n \to \infty$. This lack of continuity will be a source of some technical difficulties as we continue.

1.2 Exchangeable random partitions

In this section, we develop the fundamental observation made by Kingman [39] that mass-partitions can be conveniently encoded by certain random partitions of $\mathbb{N}$. This observation, which is very natural from the point of view of statistics, will have a fundamental importance later in the text; roughly it will provide us with a very powerful method of discretization. The coding can be understood as follows.

Imagine that we have an object with a unit mass, for instance the unit interval endowed with Lebesgue measure, which is split into fragments (i.e. we consider some interval-partition). The ranked sequence of the masses of these fragments is thus given by some mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$. One then introduces a sequence of i.i.d. random points $U_1, \ldots$ which are picked according to the mass distribution of the object (so the U_i are simply i.i.d. uniform variables in the case of the unit interval), and considers the random partition π of $\mathbb{N}$, specified by the rule that two indices, say i and j, belong to the same block of π if and only if the points U_i and U_j belong to the same fragment. An application of the law of large numbers shows that the masses of the fragments can be recovered as the asymptotic frequencies of the blocks of the partition. Conversely, although in general a partition of $\mathbb{N}$ does not necessarily correspond to a mass-partition (because the asymptotic frequencies of blocks may well not exist), there is a bijective correspondence between the laws of *exchangeable* random partitions of $\mathbb{N}$ and probability measures on $\mathcal{P}_{\mathtt{m}}$.

In this section, we shall consider a natural family of probability measures on $\mathcal{P}_\infty$. To that end, it will be sometimes convenient to identify a partition π with an equivalence relation on $\mathbb{N}$, in the sense that $i \overset{\pi}{\sim} j$ if and only if i and j belong to the same block of the partition π.

For every $n \in \mathbb{N}$, a permutation of $[n]$ is a bijection $\sigma : [n] \to [n]$. For $n = \infty$ (recall that $[\infty] = \mathbb{N}$), we call permutation of $\mathbb{N}$ any bijection $\sigma : \mathbb{N} \to \mathbb{N}$ such that $\sigma(k) = k$ when k is large enough, so that the space of permutations of $\mathbb{N}$ is the group generated by transpositions. The group of permutations acts on $\mathcal{P}_n$; more precisely for every permutation σ and every partition π, the relation

$$i \sim j \iff \sigma(i) \overset{\pi}{\sim} \sigma(j)\,, \qquad i, j = 1, \ldots, n$$

is an equivalence relation which can be identified as a partition denoted by $\sigma(\pi)$. In other words, the blocks of $\sigma(\pi)$ are the images of the blocks of π by σ^{-1}, the inverse permutation of σ.

Definition 1.7 *Let $n \in \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$. A random partition π of $[n]$ is called* **exchangeable** *if for every permutation σ of $[n]$, $\sigma(\pi)$ has the same law as π.*

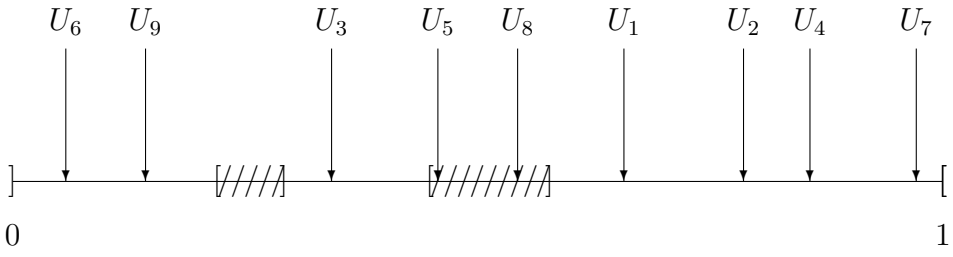
Plainly, a random permutation π of $\mathbb{N}$ is exchangeable if and only if the restrictions $\pi_{|[n]}$ are exchangeable for all $n \in \mathbb{N}$. Later in this text, we shall be mainly concerned with the case $n = \infty$; however, for the sake of simplicity, we shall then often omit to specify that we are working with partitions of $\mathbb{N}$. Of course, whenever we will be dealing with random partitions of a finite set, this will be mentioned explicitly.

We now dwell on the informal connection between mass-partitions and partitions of $\mathbb{N}$ which was sketched at the beginning of this section. Let us fix $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$, and consider an interval representation ϑ of $\mathbf{s}$. Recall that this means that ϑ is an open subset of $]0,1[$ such that the ranked sequence of the lengths of its interval components is given by $\mathbf{s}$. Let $U_1, \ldots$ be an i.i.d. sequence of uniform variables on $[0,1]$, and consider the random partition π of $\mathbb{N}$ induced by the following equivalence relation:

$$i \overset{\pi}{\sim} j \iff (i = j) \text{ or } (U_i \text{ and } U_j \text{ belong to the same interval component of } \vartheta).$$

The alternative in the right-hand side is needed because when the Lebesgue measure of ϑ is strictly less than 1, it may happen that U_i does not belong to ϑ and in that case $\{i\}$ is a singleton of π.

There is a simple pictorial interpretation of the equivalence relation π; see the figure below. Think of the unit interval as a paint-box, in which a different color is assigned to each interval component of ϑ. Every integer i then receives the color of the interval to which the variable U_i belongs, and i receives no color if $U_i \notin \vartheta$. The classes of equivalence are given by the families of indices with the same color, where we agree that indices with no color form singletons. We will refer to the random partition π defined above as the *paint-box* based on the mass-partition $\mathbf{s}$ (or on the interval-partition ϑ).



Paint-box construction: $\pi = (\{1,2,4,7\},\{3\},\{5\},\{6,9\},\{8\})$

Lemma 1.4 *The paint-box based on $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ is a random exchangeable partition. Its law does not depend on the choice of the interval representation ϑ of $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$; it will be denoted by $\varrho_{\mathbf{s}}$.*

Proof Consider any permutation σ; the partition $\sigma(\pi)$ is specified by equivalence relation:

$$i \overset{\sigma(\pi)}{\sim} j \iff (i = j) \text{ or } (U_{\sigma(i)} \text{ and } U_{\sigma(j)} \text{ belong to the same interval component of } \vartheta).$$

The variables $U_{\sigma(1)}, \ldots$ are i.i.d. and uniformly distributed on $[0,1]$; this shows that π is exchangeable.

Next, let ϑ' be another interval representation of $\mathbf{s}$. We can then find a measurable map $f : [0,1] \to [0,1]$ which preserves the Lebesgue measure and induces a bijection from ϑ to ϑ' such that the image by f of the interval components of ϑ are precisely the interval components of ϑ'. By construction

$$i \overset{\pi}{\sim} j \iff (i = j) \text{ or } (U'_i \text{ and } U'_j \text{ belong to the same interval component of } \vartheta');$$

which yields our claim, since the variables $U'_n := f(U_n)$ for $n \in \mathbb{N}$, are again i.i.d. and uniformly distributed on $[0,1]$. □

Lemma 1.4 yields another simple construction which is equivalent to the paint-box: Recall that for every measure m, a point $\mathtt{a}$ such that $m(\{\mathtt{a}\}) > 0$ is called an atom of m. Given a mass-partition $\mathbf{s}$, consider a probability distribution F on $\mathbb{R}$, such that the ranked sequence of the masses of the atoms of F, namely $(F(\mathtt{a}_i) - F(\mathtt{a}_i -), i \in I)$, coincides with the sequence of strictly positive terms in $\mathbf{s}$. Next, let $\xi_1, \ldots$ be a sequence of i.i.d. variables with distribution F, and π the random partition of $\mathbb{N}$ such that i and j belong to the same block if and only if $\xi_i = \xi_j$. The simple observation that we may choose $\xi_i = F^{-1}(U_i)$ easily yields that π is a paint-box based on $\mathbf{s}$.

Next, we state some elementary properties of paint-boxes.

Proposition 1.3 *Let π be a paint-box based on some mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$. Then the following assertions hold.*

(i) *The paint-box π possesses asymptotic frequencies. More precisely, $|\pi|^{\downarrow} = \mathbf{s}$ and $|\pi_1|$ is a size-biased sample of $\mathbf{s}$ a.s.*

(ii) *For every $i \in \mathbb{N}$, if $|\pi_i| = 0$, then π_i is either a singleton or empty a.s.*

(iii) *More precisely, some blocks of π are reduced to singletons if and only if $\mathbf{s}$ is improper, and in that case, the set of singletons $\pi_0 := \{i \in \mathbb{N} : i \text{ is a singleton of } \pi\}$ has an asymptotic frequency given by $|\pi_0| = s_0 = 1 - \sum_{i=1}^{\infty} s_i$ a.s.*

(iv) *If $\mathbf{s}$ is proper, then the sequence $|\pi|$ of the asymptotic frequencies of the blocks of π is a size-biased reordering of $\mathbf{s}$.*

Proof If $A \subseteq]0,1[$ is some measurable set and $U_1, \ldots$ a sequence of i.i.d. uniform variables, then $\left(\mathbb{1}_{\{U_i \in A\}}, i \in \mathbb{N}\right)$ is an i.i.d. sequence of Bernoulli variables with mean given by the Lebesgue measure $|A|$ of A. It thus follows from the law of large numbers that the random block $B := \{i \in \mathbb{N} : U_i \in A\}$ has an asymptotic frequency $|B| = |A|$.

Assertions (i–iii) now follow immediately from the paint-box construction and the preceding observation. Finally, (iv) is a consequence of Lemma 1.1 and our convention for labelling the blocks of a partition. □

The fundamental result about exchangeable random fragmentations states that the paint-box construction induces a bijection between probability measures on mass-partitions and exchangeable probability measures on $\mathcal{P}_\infty$.

Theorem 1.1 (Kingman) *Let π be an exchangeable random partition of $\mathbb{N}$. Then π possesses asymptotic frequencies a.s. More precisely, the law of π can be expressed as a mixture of paint-boxes:*

$$\mathbb{P}\left(\pi \in \cdot\right) \,=\, \int_{\mathcal{P}_\mathbb{m}} \mathbb{P}\left(|\pi|^{\downarrow} \in \mathrm{d}\mathbf{s}\right) \varrho_{\mathbf{s}}(\cdot)\,,$$

where $\varrho_{\mathbf{s}}$ stands for the law of the paint-box based on $\mathbf{s}$, which is defined in Lemma 1.4.

In words, the second part of the statement claims that one can construct a version of an arbitrary exchangeable random partition π as follows: One first considers some (random) interval representation ϑ_S of the ranked asymptotic frequencies $|\pi|^{\downarrow} := S$ of π and an independent sequence $U_1, \ldots$ of i.i.d. uniform variables on $[0,1]$. Then the mixture of paint-boxes constructed from ϑ_S and the U_i has the same law as π.

Kingman's original proof of Theorem 1.1 uses a martingale argument. We shall present here a simpler approach, due to Aldous [1]. It relies on de Finetti's theorem for sequences of exchangeable variables, which claims that exchangeable sequences of variables are mixtures of i.i.d. sequences. Here is the formal statement.

Theorem 1.2 (de Finetti) *Let $\xi_1, \ldots$ be an exchangeable sequence of real-valued variables, that is for every permutation σ of $\mathbb{N}$, $(\xi_1, \ldots)$ and $(\xi_{\sigma(1)}, \ldots)$ have the same finite-dimensional distributions. Then the sequence of empirical distributions*

$$\mu_n(\mathrm{d}x) \,:=\, \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i}(\mathrm{d}x)$$

converges a.s. when $n \to \infty$, in the sense of weak convergence of probability measures on $\mathbb{R}$, to some random probability measure $\mu(\mathrm{d}x)$. Moreover, conditionally on μ, the variables $\xi_1, \ldots$ are i.i.d. with joint distribution μ.

Informally, the set of laws of exchangeable sequences forms a closed convex space, and its extremal points are the distributions of i.i.d. sequences. A well-known theorem due to Choquet then ensures that the law of a general exchangeable sequence can be expressed as a mixture of extremal points. We refer to [1] for a proof of de Finetti's theorem, and now establish Theorem 1.1.

Proof For a given partition of $\mathbb{N}$, let us call any function $b : \mathbb{N} \to \mathbb{N}$ that maps all the points of each block of the partition to the same point of this block a *choice function*. For instance, $b(i)$ can be the smallest element of the block that contains i.

Now let b be some choice function for the exchangeable random partition π and $U_1, \ldots$ a sequence of i.i.d. uniform variables on $[0,1]$, which is independent of π and b, and set

$\xi_i = U_{b(i)}$. It should be plain that the law of the sequence $(\xi_i, i \in \mathbb{N})$ does not depend on how the choice function b has been chosen. More precisely, for every $n \in \mathbb{N}$ and every partition γ of $[n]$ with non-empty blocks $B_1, \ldots, B_k$, we have that

$$\mathbb{E}(f_1(U_{b(1)}) \cdots f_n(U_{b(n)}) \mid \pi_{|[n]} = \gamma) = \prod_{j=1}^{k} \left(\int_0^1 \prod_{i \in B_j} f_i(u) \mathrm{d}u \right),$$

where $f_1, \ldots, f_n : [0,1] \to \mathbb{R}_+$ denote generic measurable functions.

The key lies in the observation that the sequence $\xi_1, \ldots$ is exchangeable. Indeed, let σ be a permutation of $\mathbb{N}$; we then have

$$\xi_{\sigma(i)} \; = \; U_{b(\sigma(i))} \; = \; U'_{b'(i)} \, ,$$

where $U'_j = U_{\sigma(j)}$ and $b' = \sigma^{-1} \circ b \circ \sigma$. It is immediately seen that b' is a choice function for the partition $\sigma(\pi)$, and that $U'_1, \ldots$ are i.i.d. uniform variables on $[0,1]$ which are jointly independent of the partition $\sigma(\pi)$ and its choice function b'. By assumption, π is exchangeable and independent of the U_i, so $((U'_i)_{i \in \mathbb{N}}, \sigma(\pi))$ has the same law as $((U_i)_{i \in \mathbb{N}}, \pi)$, and the sequence $\xi_1, \ldots$ is exchangeable.

Next, we observe that we may recover a.s. the partition π from the sequence $(\xi_1, \ldots)$, since the blocks of π are precisely the sub-families of indices i for which the variables ξ_i take the same value. By de Finetti's theorem, there is some random probability measure μ on $[0,1]$ such that conditionally on μ, the variables $\xi_1, \ldots$ are i.i.d. with law μ.

We now work conditionally on μ and write q for the quantile function of μ, that is the inverse of its distribution function. Introduce the open set of flat points of q,

$$\vartheta \; := \; \{x \in]0,1[\colon \exists \varepsilon > 0 \text{ such that } q(x) = q(y) \text{ whenever } |y - x| < \varepsilon\} \, ,$$

so that the lengths of the intervals components of ϑ coincide with the masses of the atoms of μ. Introduce an independent sequence $V_1, \ldots$ of i.i.d. uniform variables on $[0,1]$, so that the sequence $(q(V_1), \ldots)$ has the same law as $(\xi_1, \ldots)$ conditionally on μ. Define a random partition π' by declaring that two distinct indices, say i and j, belong to the same block of π' if and only if $q(V_i) = q(V_j)$, that is if and only if V_i and V_j belong to the same interval component of ϑ. Then π' has the same law as π, which shows that conditionally on μ, π is distributed as a paint-box based on ϑ. □

We now conclude this section with an important result of continuity in distribution for mixtures of paint-boxes. Indeed, one difficulty related to this representation is that the map $\pi \to |\pi|^{\downarrow}$ that enables one to recover the mass-partition from a paint-box is not continuous. However, this difficulty vanishes when one considers convergence in law.

Proposition 1.4 *Consider for each $n \in \overline{\mathbb{N}}$, a random exchangeable partition $\pi^{(n)}$, and write $|\pi^{(n)}|^{\downarrow}$ for the mass-partition given by the ranked sequence of the asymptotic frequencies of its blocks. The following conditions are equivalent:*

(i) *When $n \to \infty$, $|\pi^{(n)}|^{\downarrow}$ converges in distribution on $\mathcal{P}_{\mathtt{m}}$ to $|\pi^{(\infty)}|^{\downarrow}$.*

(ii) *When $n \to \infty$, $\pi^{(n)}$ converges in distribution on $\mathcal{P}_\infty$ to $\pi^{(\infty)}$.*

Proof It is convenient to denote by $S^{(n)}$ a random mass-partition distributed as $|\pi^{(n)}|^{\downarrow}$.

Suppose (i) holds. Since the space of mass-partitions $\mathcal{P}_{\mathtt{m}}$ is metric and compact, we may apply Skorokhod representation theorem (see for example Billingsley [15]) and assume that $\lim_{n\to\infty} S^{(n)} = S^{(\infty)}$ a.s. In particular $\lim_{n\to\infty} S_k^{(n)} = S_k^{(\infty)}$ a.s. for each $k \in \mathbb{N}$. Set for $n \in \overline{\mathbb{N}}$ and $k \in \mathbb{N}$

$$\Sigma_0^{(n)} = 0 \quad \text{and} \quad \Sigma_k^{(n)} = S_1^{(n)} + \cdots + S_k^{(n)},$$

so $\lim_{n\to\infty} \Sigma_k^{(n)} = \Sigma_k^{(\infty)}$ a.s. for every $k \in \mathbb{N}$. Consider the random interval-partition $\vartheta^{(n)} \in \mathcal{P}_{\mathrm{I}}$

$$\vartheta^{(n)} = \bigcup_{k\in\mathbb{N}}]\Sigma_{k-1}^{(n)}, \Sigma_k^{(n)}[,$$

so each $\vartheta^{(n)}$ is an interval representation of $S^{(n)}$. If we introduce a sequence $U_1, \ldots$ of i.i.d. uniform variables on $[0,1]$ which is independent of the $S^{(n)}$, we may suppose that each $\pi^{(n)}$ is the mixture of paint-boxes based on $\vartheta^{(n)}$ and $(U_1, \ldots)$.

Now for each $i, k \in \mathbb{N}$, we have that

$$\lim_{n\to\infty} \mathbb{1}_{\{\Sigma_{k-1}^{(n)} < U_i < \Sigma_k^{(n)}\}} = \mathbb{1}_{\{\Sigma_{k-1}^{(\infty)} < U_i < \Sigma_k^{(\infty)}\}} \qquad \text{a.s.},$$

which implies that with probability one, for every $\ell \in \mathbb{N}$, the restrictions of $\pi^{(n)}$ and $\pi^{(\infty)}$ to $[\ell]$ coincide when n is sufficiently large. Thus $\lim_{n\to\infty} \pi^{(n)} = \pi^{(\infty)}$ a.s., which shows that (i) $\Rightarrow$ (ii).

Next, suppose (ii) holds. As $\mathcal{P}_{\mathtt{m}}$ is a compact metric space, the space of probability measures on $\mathcal{P}_{\mathtt{m}}$ is also metric and compact by Prohorov's Theorem (cf. Section 6 in Billingsley [15]), and thus from any subsequence of $(S^{(n)}, n \in \mathbb{N})$ we can extract a subsequence, say $(\tilde{S}^{(n)}, n \in \mathbb{N})$, which converges weakly to some random mass-partition $\tilde{S}^{(\infty)}$. We deduce from the first part of the proof that $\pi^{(\infty)}$ is distributed as a mixture of paint-boxes based on $\tilde{S}^{(\infty)}$, and in particular that the distribution of $\tilde{S}^{(\infty)}$ does not depend on the chosen subsequence (more precisely, $\tilde{S}^{(\infty)}$ has the same law as $|\pi^{(\infty)}|^{\downarrow}$). Thus $S^{(n)}$ converges in distribution to $\tilde{S}^{(\infty)}$, which proves that (ii) $\Rightarrow$ (i). □

1.3 Poisson-Dirichlet partitions

In short, the preceding sections have been devoted to the exposition of the necessary background on state spaces of partitions (from several distinct points of view) and partition-valued random variables. Now in the case of real-valued random variables, certain distributions such as Gaussian, Poisson, ..., play a prominent role. The Poisson-Dirichlet partitions form an important family of random partitions which appear in a number of circumstances. In this section, we shall introduce the one-parameter family, postponing the more general two-parameter family to Section 5.2. We start by presenting some elementary material on a well-known family of probability measures on the $(N-1)$-dimensional simplex related to beta and gamma variables.

1.3.1 Multidimensional Dirichlet distributions

To start with, recall that for every $\theta, c > 0$, the gamma law with parameter (θ, c) is the probability measure on $\mathbb{R}_+$ which has density

$$\frac{c^\theta}{\Gamma(\theta)} x^{\theta-1} \mathrm{e}^{-cx}\,, \qquad x > 0\,.$$

The parameter c has a very minor role in this section, due to the fact that if a variable γ has the gamma(θ, c) distribution, then $c\gamma$ has the gamma$(\theta, 1)$ distribution.

Next, for every $a, b > 0$, the beta law with parameter (a, b) is the probability measure on $]0, 1[$ which has density

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\,, \qquad 0 < x < 1\,.$$

It is well-known, and easy to check, that if γ and γ' are two independent gamma variables with respective parameters (a, c) and (b, c), then

$$\frac{\gamma}{\gamma+\gamma'} \quad \text{has the beta}(a,b) \text{ law and is independent of } \gamma+\gamma'\,. \tag{1.1}$$

Moreover, $\gamma + \gamma'$ has the gamma$(a + b, c)$ distribution.

Throughout this section, $N \geq 2$ denotes a fixed integer. Dirichlet distributions form an important family of probability laws on the simplex

$$\Delta_{N-1} := \left\{ x = (x_1, \ldots, x_N) : x_i \geq 0 \text{ for every } i = 1, \ldots, N \text{ and } \sum_{i=1}^{N} x_i = 1 \right\}.$$

Definition 1.8 *For every $\alpha_1, \ldots, \alpha_N > 0$, the probability measure on the simplex Δ_{N-1} with density*

$$\frac{\Gamma(\alpha_1 + \cdots + \alpha_N)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_N)} x_1^{\alpha_1 - 1} \cdots x_N^{\alpha_N - 1}\,, \qquad (x_1, \ldots, x_N) \in \Delta_{N-1}$$

*is called the $(N-1)$-**dimensional Dirichlet distribution** with parameter $(\alpha_1, \ldots, \alpha_N)$. The special case when $\alpha_1 = \cdots = \alpha_N = 1$ will be referred to as the* **uniform distribution** *on Δ_{N-1}.*

We now recall a useful representation of Dirichlet distribution based on independent gamma variables, which extends (1.1).

Lemma 1.5 (i) *Fix $\alpha_1, \ldots, \alpha_N, c > 0$, and let $\gamma_1, \ldots, \gamma_N$ be independent gamma variables with respective parameters $(\alpha_1, c), \ldots, (\alpha_N, c)$. Set $\gamma := \gamma_1 + \cdots + \gamma_N$, so γ has a gamma distribution with parameter $(\alpha_1 + \cdots + \alpha_N, c)$. Then the N-tuple*

$$(\gamma_1/\gamma, \ldots, \gamma_N/\gamma)$$

has the $(N-1)$*-dimensional Dirichlet distribution with parameter* $(\alpha_1, \ldots, \alpha_N)$ *and is independent of* γ*.*

(ii) *Let* $0 < V_1 < \cdots < V_{N-1} < 1$ *be the ordered statistics of a family of* $N-1$ *i.i.d. uniform* $[0,1]$ *variables. Then the* N*-tuple of the increments*

$$(V_1, V_2 - V_1, \ldots, V_{N-1} - V_{N-2}, 1 - V_{N-1})$$

has the uniform distribution on the simplex Δ_{N-1}*.*

Proof (i) Let $\mathtt{x} = (\mathtt{x}_1, \ldots, \mathtt{x}_N)$ be a Dirichlet variable with parameter $(\alpha_1, \ldots, \alpha_N)$ and γ an independent gamma variable with parameter (α, c) where $\alpha := \alpha_1 + \cdots + \alpha_N$. Set $\gamma_i := \gamma \mathtt{x}_i$ for $i = 1, \ldots, N$, so for every measurable function $f : \mathbb{R}^N \to \mathbb{R}_+$, we have

$$\begin{aligned}
&\mathbb{E}\left(f(\gamma_1, \ldots, \gamma_N)\right) \\
&= \frac{c^\alpha}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_N)} \int_{\Delta_{N-1}} \mathrm{d}x_1 \cdots \mathrm{d}x_{N-1} x_1^{\alpha_1 - 1} \cdots x_N^{\alpha_N - 1} \int_0^\infty \mathrm{d}y y^{\alpha - 1} \mathrm{e}^{-cy} f(yx_1, \ldots, yx_N) \\
&= \frac{c^{\alpha_1} \cdots c^{\alpha_N}}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_N)} \int_{\mathbb{R}_+^N} \mathrm{d}z_1 \cdots \mathrm{d}z_N z_1^{\alpha_1 - 1} \cdots z_N^{\alpha_N - 1} \mathrm{e}^{-c(z_1 + \cdots + z_N)} f(z_1, \ldots, z_N)\,,
\end{aligned}$$

where at the second equality, we used the change of variables

$$z_1 = yx_1\,,\ \ldots\ ,\ z_{N-1} = yx_{N-1}\,,\ z_N = yx_N = y - z_1 - \cdots - z_{N-1}\,.$$

Thus $\gamma_1, \ldots, \gamma_N$ are independent gamma variables with respective parameters (α_1, c), ..., (α_N, c).

(ii) This follows readily from the fact that $(V_1, \ldots, V_{N-1})$ is uniformly distributed on the set $\{(v_1, \ldots, v_{N-1}) : 0 < v_1 < \cdots < v_{N-1} < 1\}$. □

1.3.2 Gamma subordinators and Dirichlet processes

Throughout this section, $\theta, c > 0$ are two fixed real numbers. A gamma subordinator with parameter (θ, c) is an increasing process $(\gamma(t), t \geq 0)$ with independent and stationary increment with no drift and Lévy measure

$$\Lambda(\mathrm{d}x) \,=\, \theta x^{-1} \mathrm{e}^{-cx} \mathrm{d}x\,, \qquad x > 0.$$

Its Laplace exponent is thus given by

$$\Phi(q) \,=\, \theta \int_0^\infty (1 - \mathrm{e}^{-qx}) x^{-1} \mathrm{e}^{-cx} \mathrm{d}x \,=\, \theta \ln(1 + q/c)\,, \qquad q \geq 0;$$

note that $\gamma(t)$ has the gamma distribution with parameter $(\theta t, c)$. The parameter c will have a very minor role, due to the easy fact that $c\gamma(\cdot)$ is a gamma subordinator with parameter $(\theta, 1)$. In this direction, it might be also interesting to point out that for every $a \in]0,1[$, $(\gamma(at), 0 \leq t \leq 1)$ is a gamma subordinator with parameter $(a\theta, c)$.

We start by observing that, thanks to the independence and stationarity of the increments of the subordinator, we may rewrite Lemma 1.5 as follows: For every $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$, the n-tuple

$$\left(\frac{\gamma(t_1) - \gamma(t_0)}{\gamma(1)}, \ldots, \frac{\gamma(t_n) - \gamma(t_{n-1})}{\gamma(1)} \right) \tag{1.2}$$

has the $(n-1)$-dimensional Dirichlet distribution with parameter $(\theta(t_1 - t_0), \ldots, \theta(t_n - t_{n-1}))$. Intuitively, if we let n tend to ∞ taking finer and finer subdivisions $\{t_0, t_1, \ldots, t_{n-1}, t_n\}$ of $[0,1]$, and re-rank the n-tuple (1.2) in decreasing order, we should get as limit the ranked sequence of the jumps of $\gamma(\cdot)/\gamma(1)$. This leads to the following definition.

Definition 1.9 *Let $(\gamma(t), t \in [0,1])$ be a gamma subordinator with parameter (θ, c), and denote by $\mathtt{a}_1 > \mathtt{a}_2 > \cdots > 0$ the ranked sequence of its jumps.*

(i) *The process $(\gamma(t)/\gamma(1), t \in [0,1])$ is called a* **Dirichlet process** *with parameter θ.*

(ii) *The distribution of the random mass-partition $(\mathtt{a}_1/\gamma(1), \ldots)$ is called the* **Poisson-Dirichlet** *law with parameter $(0, \theta)$ and is denoted by $\mathrm{PD}(0, \theta)$.*

Remark. We recall that the Lévy-Itô decomposition shows that the jump process of a subordinator is described by a Poisson point process with intensity given by the Lévy measure. In the present setting, as gamma subordinators have no drift, this means that

$$\gamma(1) = \sum_{i=1}^{\infty} \mathtt{a}_i$$

and the random point measure

$$\sum_{i=1}^{\infty} \delta_{\mathtt{a}_i}$$

is Poisson with intensity measure Λ. See the Appendix for background.

The Poisson-Dirichlet distribution $\mathrm{PD}(0,1)$ arises in several remarkable limit theorems about the component frequency spectrum of large structures in combinatorics, including the decomposition of permutations in cycles and the decomposition of integers in prime factors. The one-parameter family denoted here by $\mathrm{PD}(0,\theta)$ for $\theta > 0$ has appeared in a variety of contexts, such as Bayesian statistics, population genetics (more precisely to describe the frequency of species in population models with neutral mutations), invariant distributions for certain split and merge transformations. We refer to the monograph [2] by Arratia, Barbour and Tavaré for further examples and references.

We now state some elementary properties of Dirichlet processes.

Proposition 1.5 (i) *The one-dimensional distributions of the Dirichlet process are beta laws. More precisely $\gamma(t)/\gamma(1)$ has a beta$(\theta t, \theta(1-t))$ law.*

(ii) *The Dirichlet process $(\gamma(t)/\gamma(1), t \in [0,1])$ is independent of $\gamma(1)$.*

(iii) *For every $x > 0$, the conditional distribution of $(x^{-1}\mathtt{a}_1, x^{-1}\mathtt{a}_2, \ldots)$ given $\gamma(1) = x$ is $\mathrm{PD}(0, \theta)$.*

Proof The first claim merely rephrases (1.1) and the second follows also from (1.1) by an immediate induction using the independence of the increments of the gamma process. The last part of the statement derives readily from the second. □

One can derive from Proposition 1.5 and the Palm formula for Poisson random measures (cf. the Appendix) the so-called *residual allocation model* (also sometimes called *stick breaking scheme*) for PD$(0,\theta)$ distributions: Take a sequence $\beta_1,\ldots$ of i.i.d. beta$(1,\theta)$ variables, we break the stick into two smaller sticks with size β_1 and $1-\beta_1$, keep the former, break the latter into two smaller sticks with size $(1-\beta_1)\beta_2$ and $(1-\beta_1)(1-\beta_2)$, and so on ... Then the random sequence of the lengths is distributed as the size-biased re-ordering of a PD$(0,\theta)$-variable. Here is a formal statement.

Corollary 1.1 *Let S be a random mass-partition with the* PD$(0,\theta)$ *distribution, $S^* = (S_1^*, S_2^*, \ldots)$ a size-biased reordering of S, and set*

$$\beta_1 := S_1^*\,,\ \beta_2 := S_2^*/(1-\beta_1)\,,\ \ldots\,,\ \beta_{n+1} := S_{n+1}^* \times \prod_{i=1}^{n}(1-\beta_i)^{-1}\,,\ \ldots$$

Then the variables $\beta_1,\ldots$ are i.i.d. with the beta$(1,\theta)$ distribution, that is $\mathbb{P}(\beta_i \in \mathrm{d}a) = \theta(1-a)^{\theta-1}\mathrm{d}a$ for $a \in [0,1]$.

Note that Corollary 1.1 enables us to construct a size-biased reordering of a PD$(0,\theta)$ variable from a sequence of i.i.d. beta variables, since, in the notation used there, we have

$$S_1^* = \beta_1\,,\ S_2^* = (1-\beta_1)\beta_2\,,\ \ldots\,,\ S_{n+1}^* = \beta_{n+1}\prod_{i=1}^{n}(1-\beta_i)\,,\ldots$$

The size-biased reordering S^* of a Poisson-Dirichlet PD$(0,\theta)$ variable S is often called a Griffiths-Engen-McCloskey variable with parameter θ and denoted by GEM(θ).

1.3.3 Ewens sampling formula

Let π be an exchangeable random partition. Fix an integer $n \in \mathbb{N}$ and consider a partition $\varphi = (B_1,\ldots,B_k,\varnothing,\ldots)$ of $[n]$ with $B_k \neq \varnothing$. The exchangeability implies that the probability that the restriction of π to $[n]$ coincides with φ can be expressed in the form

$$\mathbb{P}\left(\pi_{|[n]} = \varphi\right) \,=\, \mathtt{p}(\#B_1,\ldots,\#B_k)$$

where $\#B$ denotes the cardinal of the finite block B and $\mathtt{p}$ is called the *exchangeable partition probability function* (in short, EPPF) of π.

Note that $\mathtt{p}$ is a symmetric function of a finite (but not fixed) number of variables which are all positive integers. By symmetry, one usually writes the argument of EPPF in decreasing order. For every $\theta > 0$, write $\mathtt{p}_{0,\theta}$ for the EPPF of an $(0,\theta)$-partition, that is a random exchangeable partition π such that its ranked sequence of asymptotic frequencies $|\pi|^{\downarrow}$ has the PD$(0,\theta)$ distribution.

Theorem 1.3 (Ewens Sampling Formula) *Pick integers $k \leq n$ and $n_1, \ldots, n_k$ such that $n_1 + \cdots + n_k = n$. We have :*

$$\mathtt{p}_{0,\theta}(n_1, \ldots, n_k) \;=\; \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{k} (n_i - 1)!$$

We stress that in the literature, the Ewens Sampling formula is often expressed in a different (but of course equivalent) form. Namely, is gives the probability that a $(0, \theta)$-partition restricted to $[n]$ has k_1 blocks of size 1, k_2 blocks of size 2, ..., k_n blocks of size n, where $\sum_{i=1}^{n} ik_i = n$.

Proof The residual allocation model described in Corollary 1.1 provides a construction of the size-biased reordering of the PD$(0, \theta)$-variable using an i.i.d. sequence $\beta_1, \ldots$ of standard beta variables with parameter $(1, \theta)$. Elementary combinatorics then yield

$$\begin{aligned}
&\mathtt{p}_{0,\theta}(n_1, \ldots, n_k) \\
&\quad = \mathbb{E}\left(\left(\beta_1^{n_1 - 1}(1-\beta_1)^{n_2 + \cdots + n_k}\right) \cdots \left(\beta_{k-1}^{n_{k-1}-1}(1-\beta_{k-1})^{n_k}\right) \beta_k^{n_k - 1} \right) \\
&\quad = \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{k} (n_i - 1)!
\end{aligned}$$

where in the second equality we used the fact that the moments of a beta variable with parameter (a, b), say β, are given by

$$\mathbb{E}\left(\beta^k (1-\beta)^\ell\right) \;=\; \frac{\Gamma(a+b)\Gamma(a+k)\Gamma(b+\ell)}{\Gamma(a)\Gamma(b)\Gamma(a+b+k+\ell)}\,.$$

□

Let us now present a useful consequence of Theorem 1.3, which provides a remarkable recursive construction of PD$(0, \theta)$-partitions due to Dubins and Pitman. It can be described as follows.

Imagine a *Chinese Restaurant* having an infinite number of tables with infinite capacity, in the sense that each table can accommodate an infinite number of customers. Tables are denoted by $T_1, T_2, \ldots$; initially all the tables are empty. Customers arrive one after the other and pick a table according to a random process that we now explain. Fix $\theta > 0$. The first customer, denoted by 1, sits at the first table T_1. For every $n \geq 1$, if at the time when the $(n+1)$-th customer enters the restaurant there are k non-empty tables, $T_1, \ldots, T_k$, this new customer decides to sit alone at table T_{k+1} with probability $\theta/(n+\theta)$, and at table T_i for $1 \leq i \leq k$ with probability $\#B_i(n)/(n+\theta)$, where $B_i(n)$ is the block of $[n]$ formed by the customers already sat at table T_i. So for each n, the occupation of the tables yields a partition $\pi(n) = (B_1(n), \ldots)$ of $[n]$. Note that by construction, blocs are labeled according to the increasing order of their least element, in agreement with our convention. Clearly, these partitions are compatible as n varies, so by Lemma 1.2, there exists a unique (random) partition $\pi(\infty)$ of $\mathbb{N}$ such that $\pi(n) = \pi(\infty)_{|[n]}$ for each n.

Remark. We point out that in the special case $\theta = 1$, the Chinese restaurant is a simple algorithm that produces a uniform random permutation on $[n]$ together with its decomposition into cycles. More precisely, we may decide that when the customer $n+1$ arrives, (s)he sits at a new table with probability $1/(n+1)$, and at the right of some previously sat customer also with probability $1/(n+1)$. So the probability that (s)he sits at a table already occupied by j customers is $j/(n+1)$ as required. The blocks of the partition corresponding to customers sat at the same table are then naturally endowed with a cyclic order, which yields a description of the decomposition of a random permutation into cycles. It should be plain that at each step, this algorithm produces a uniform random permutation.

Corollary 1.2 *The random partition $\pi(\infty)$ constructed above is a* $\mathrm{PD}(0,\theta)$*-partition.*

Note that Corollary 1.2 implies in particular that the random partition $\pi(\infty)$ is exchangeable, a property which is not obvious from the construction.

Proof An immediate check by iteration can be made that for every partition of $[n]$ with k non-empty blocks, say $(B_1, \ldots, B_k, \varnothing, \ldots)$, the probability that $\pi(n) = (B_1, \ldots, B_k, \varnothing, \ldots)$ is given by

$$\mathbb{P}(\pi(\infty)_{|[n]} = (B_1, \ldots, B_k, \varnothing, \ldots)) \;=\; \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{k} (\#B_i - 1)!\,.$$

The comparison with Theorem 1.3 establishes the claim. □

In the special case when θ is an integer, the Chinese Restaurant can be interpreted as a variation of Pólya's urn model, see [34]: Let $c_0, c_1, \ldots$ denote a sequence of different colors, and consider an urn which contains initially one ball of color c_1 and θ balls with color c_0. At the first step, we pick a ball at random in the urn, note its color c and replace it in the urn together with a new colored ball. More precisely, if $c = c_1$, then the color of the new ball which is added to the urn is c_1; whereas if $c = c_0$, then the color of the new ball is c_2. We iterate the process in an obvious way. After n steps, there are $\theta + n + 1$ balls in the urn, with colors $c_0, c_1, \ldots, c_k$. We pick a ball at random, uniformly and independently of the preceding drawing. If this ball has color c_ℓ for some $\ell = 1, \ldots, k$, then we replace it in the urn together with an additional ball with color c_ℓ. If the ball has color c_0, we replace it in the urn together with an additional ball with the new color c_{k+1}. Plainly, the distribution of the numbers of balls with respective colors $c_1, \ldots$ in the urn after n steps is the same as that of the numbers of customers sat at table $T_1, \ldots$ in a Chinese Restaurant process when the total number of customers is $n+1$. We refer to Section 3.1 in [48] for much more on Chinese Restaurants and their applications.

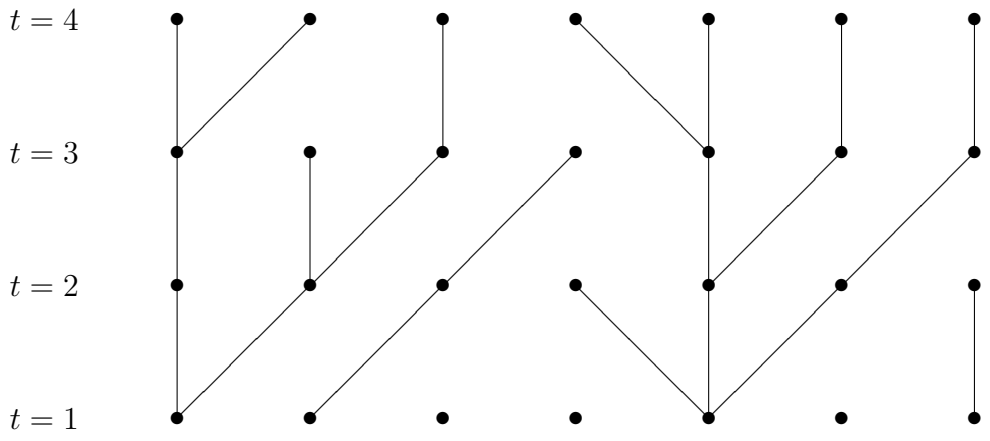
Chapter 2

Kingman's coalescent

Kingman's coalescent is a natural and most important Markov process with values in the space of partitions, which evolves by binary coagulation. Its introduction has been motivated by the study of the genealogy of large populations, and we shall briefly present some applications in this setting.

2.1 Genealogy of populations in the Wright-Fisher model

Coalescence naturally arises when one studies the genealogy of populations; we first briefly explain why. Following Kingman [39], this will lead us to introduce a natural Markov process with values in the space $\mathcal{P}_\infty$ of partitions of $\mathbb{N}$.



Coalescing ancestral lineages: population of fixed size 7 *with* 4 *generations*

Imagine at time $T > 0$ a population with size n which can be identified with the

set $[n] = \{1, \ldots, n\}$. Assume the population is haploid, meaning that each individual has exactly one parent at the previous generation, so we may follow its ancestral lineage backwards in time. Plainly, ancestral lineages coalesce, in the sense that when distinct individuals have the same ancestor at some generation t, they necessarily have the same ancestor at any generation $t' \leq t$. For every $t < T$, consider the partition, say $\pi(t) \in \mathcal{P}_n$, of the population into sub-families (i.e. blocks of $[n]$) having the same ancestor at the generation $T - t$. When t increases, the partitions $\pi(t)$ get coarser and, more precisely, the partition $\pi(t + s)$ can be obtained from $\pi(t)$ by the coagulation of the sub-families such that their respective ancestors at the generation $T - t$ have the same ancestor at the generation $T - t - s$.

To start the mathematical study of this phenomenon, we need a model for the evolution of populations. Here, we shall consider one of the simplest, which was introduced by Wright and Fisher around 1930. In the Wright-Fisher model, time is discrete, the size of the population is fixed, the generations do not overlap and, finally, each individual at the generation $k+1$ picks its parent from the individuals at the generation k according to the uniform probability, independently of the other individuals. In particular, the number of children ξ of a typical individual has a binomial$(n, 1/n)$ distribution:

$$\mathbb{P}(\xi = k) \,=\, \left(\begin{array}{c} n \\ k \end{array} \right) n^{-k} (1 - 1/n)^{n-k} \,, \qquad k = 0, 1, \ldots, n \,.$$

Plainly, the probability that two distinct individuals at the same generation have the same parent at the preceding generation equals $1/n$. Since the generations are supposed to be independent, we see that the probability that the ancestral lines of these two individuals remain distinct during at least k generations equals $(1 - 1/n)^k$. Hence, the time of coalescence of the ancestral lines, that is the age of the most recent common ancestor of two distinct individuals, has the geometric distribution with mean n. More generally, if we select $\ell \leq n$ distinct individuals at the same generation, then the probability that all have distinct parents at the preceding generation is the proportion of injections among the maps from $[n]$ to itself, that is $(1 - 1/n) \cdots (1 - (\ell - 1)/n)$. Thus the probability that the ancestral lines of these ℓ individuals do not coalesce before k generations is $(1 - 1/n)^k \cdots (1 - (\ell - 1)/n)^k$.

This suggests a diffusion-approximation (cf. Ethier and Kurtz [30]) when the size n of the population and the number of generations are large. Specifically, let us renormalize time in such a way that one time unit corresponds to n generations, and let $n \to \infty$. Thus the probability that the ancestral lines of ℓ distinct individuals at the same generation remain distinct at least up to time t (i.e. during at least $k = \lfloor nt \rfloor$ generations) converges when $n \to \infty$ to

$$\mathrm{e}^{-t(1 + \cdots + (\ell - 1))} \,=\, \exp(-t\ell(\ell - 1)/2) \,.$$

In other words, the time of the first coalescence for the ancestral lines of ℓ distinct individuals, converges in distribution to an exponential variable with parameter $\ell(\ell - 1)/2$, that is the minimum of $\ell(\ell - 1)/2$ independent standard exponential variables. Observe that there are precisely $\ell(\ell - 1)/2$ pairs of ancestral lines that can be built from ℓ distinct individuals. These elementary observations have led Kingman to introduce a remarkable

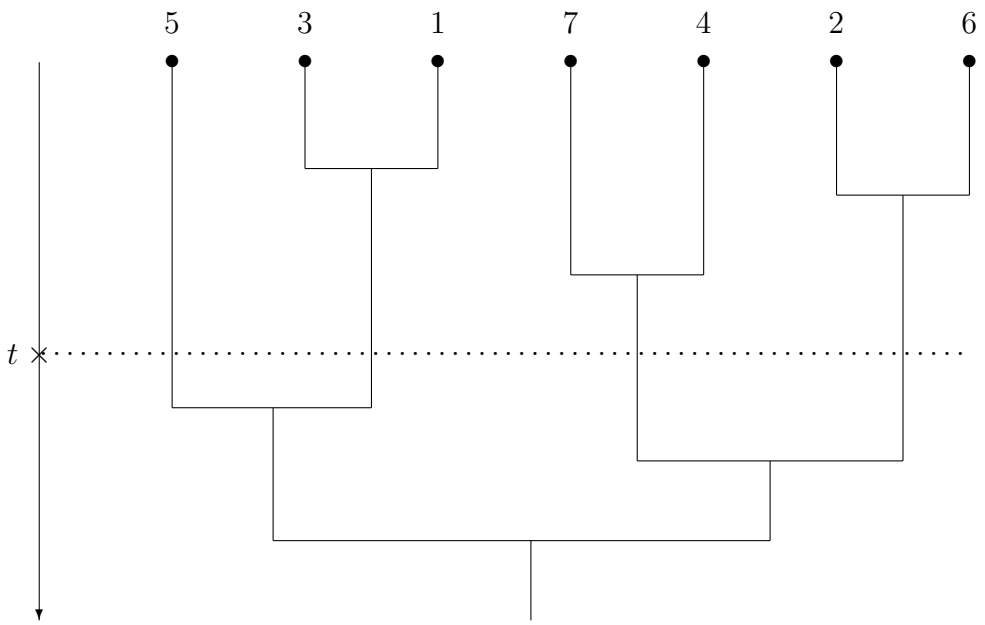
Markov process on the space of partitions of $\mathbb{N}$, which will be developed in the next section.

2.2 Construction of Kingman's coalescent

We start by introducing the simple notion of *coagulation* of pairs of blocks for partitions, referring to Section 1.1.3 for the basic notation. Let π and π' be two partitions of some set $E \subseteq \mathbb{N}$ (later in the text, we shall mostly be concerned with the cases $E = [n]$ and $E = \mathbb{N}$). We say that π' can be obtained from π by the coagulation of (exactly) two of its (non-empty) blocks if there exists $1 \leq i < j$ such that $\pi_i, \pi_j \neq \varnothing$, $\pi'_i = \pi_i \cup \pi_j$, and for all $n' \neq i$, there is some $n \in \mathbb{N} \backslash \{i, j\}$ such that $\pi'_{n'} = \pi_n$. Recall the notation

$$\#\pi \;=\; \sup\{i : \pi_i \neq \varnothing\}$$

for the number of non-empty blocks of a partition π.



k-coalescent for $k = 7$

$$\Pi^{[k]}(t) = (\{1,3\}, \{2,6\}, \{4,7\}, \{5\})$$

The discussion of the preceding section leads us to consider for each fixed $n > 0$ a Markov chain in continuous time with values in the space $\mathcal{P}_n$ of partitions of $[n]$, denoted by $\Pi^{[n]} = \big(\Pi^{[n]}(t), t \geq 0\big)$ and called the n-coalescent, which is governed by the following

dynamics. The trivial partition $\mathbf{1}_{[n]} = ([n], \varnothing, \ldots)$ is an absorbing state. If $\pi, \pi' \in \mathcal{P}_n$ are two partitions such that π' can be obtained from π by the coagulation of two of its blocks, then the jump rate from π to π' is one. All the other jump rates are equal to zero. This means that when the initial state of the chain is given by some partition π with $\#\pi = k \geq 2$, $\Pi^{[n]}$ stays at π for an exponential time with parameter $k(k-1)/2$ (i.e. the number of pairs of non-empty blocks), and then jumps at one of the $k(k-1)/2$ partitions which can be obtained from π by the coagulation of two of its blocks, according to the uniform probability.

For the sake of simplicity, we shall assume throughout this section (except in the proof of the forthcoming Lemma 2.1) that $\Pi^{[n]}$ starts from the partition of $[n]$ into singletons, as the case of different initial configurations can easily be reduced to that one. We can then rephrase the description of the dynamics of $\Pi^{[n]}$ by saying that the process of the number of non-empty blocks,

$$\#\Pi^{[n]} \;=\; \big(\#\Pi^{[n]}(t), t \geq 0\big)\;,$$

is a pure death process, that is a continuous-time Markov chain on $[n]$ in which the only admissible steps are from k to $k-1$; more precisely, the death rate at level $k \geq 1$ is $k(k-1)/2$. We call the sequence of the successive states visited by $\Pi^{[n]}$ the *state sequence*; specifically, for every $m = 1, \ldots, n$, we denote by $\mathtt{K}^{[n]}(m)$ the partition of $[n]$ with m non-empty blocks that is obtained after $n-m$ coagulations in the n-coalescent process. In particular $\mathtt{K}^{[n]}(n)$ is the partition into singletons, and $\mathtt{K}^{[n]}(1)$ the trivial partition $\mathbf{1}_{[n]}$. The state sequence $(\mathtt{K}^{[n]}(n), \mathtt{K}^{[n]}(n-1), \ldots, \mathtt{K}^{[n]}(1))$ is Markovian, that is for every $1 \leq k < n$

$$\mathbb{P}\left(\mathtt{K}^{[n]}(k) = \cdot \mid \mathtt{K}^{[n]}(n), \ldots, \mathtt{K}^{[n]}(k+1)\right) \;=\; \mathbb{P}\left(\mathtt{K}^{[n]}(k) = \cdot \mid \mathtt{K}^{[n]}(k+1)\right)\;,$$

and further this state sequence is independent of the death process $\#\Pi^{[n]}$. Its transition probabilities from a partition π with $\#\pi = k \geq 2$ is simply given by the uniform probability on the set of the $k(k-1)/2$ partitions which can be obtained from π by the coagulation of two of its blocks. By definition, we have

$$\Pi^{[n]}(t) \;=\; \mathtt{K}^{[n]}(\#\Pi^{[n]}(t))\,, \qquad t \geq 0\,.$$

The one-dimensional distributions of the state sequence $\mathtt{K}^{[n]}$ can be computed explicitly.

Proposition 2.1 *Fix $n \in \mathbb{N}$ and let $(B_1, \ldots, B_k, \varnothing, \ldots)$ be some partition of $[n]$ with $B_k \neq \varnothing$. Then*

$$\mathbb{P}\left(\mathtt{K}^{[n]}(k) = (B_1, \ldots, B_k, \varnothing, \ldots)\right) \;=\; \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \prod_{i=1}^{k} \#B_i!\,.$$

Observe that this formula shows that the random partition $\mathtt{K}^{[n]}(k)$ is exchangeable.

Proof The proof uses a backwards induction on the number of blocks k. The case when $k = n$ corresponds to the initial state of $\Pi^{[n]}$, that is the partition into singletons. So pick

$2 \leq k \leq n$ and assume that the stated formula holds for all the partitions of $[n]$ with k blocks or more.

For every $1 \leq \ell < b$ and every block B with size b, there are $\binom{b}{\ell}$ blocks $B' \subset B$ with size ℓ, and each such block yields a partition of B into two smaller blocks B' and $B \backslash B'$ with respective sizes ℓ and $b - \ell$. Observe that the same partition of B is obtained by choosing first the block $B \backslash B'$ amongst the $\binom{b}{b-\ell} = \binom{b}{\ell}$ sub-blocks of B with size $b - \ell$. In short, the application which maps a block B' with $\varnothing \neq B' \subsetneq B$ to a non-degenerate partition of B into two smaller blocks is surjective and two-on-one.

We now use the identity

$$\mathbb{P}\left(\mathtt{K}^{[n]}(k-1) = (B_1, \ldots, B_{k-1}, \varnothing, \ldots)\right) \;=\; \frac{2}{k(k-1)} \sum \mathbb{P}(\mathtt{K}^{[n]}(k) = \pi)\,,$$

where, in the right-hand side, the summation is taken over the family of partitions $\pi \in \mathcal{P}_n$ with $\#\pi = k$, such that the partition $(B_1, \ldots, B_{k-1}, \varnothing, \ldots)$ can result from π after the coagulation of two blocks of the latter. The preceding observation yields

$$\begin{aligned}
&\mathbb{P}\left(\mathtt{K}^{[n]}(k-1) = (B_1, \ldots, B_{k-1}, \varnothing, \ldots)\right) \\
&\quad= \frac{1}{k(k-1)} \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \sum_{i=1}^{k-1} \sum_{\ell=1}^{\#B_i - 1} \frac{\#B_1! \cdots \#B_{k-1}!}{\#B_i!}\, \ell!(\#B_i - \ell)! \binom{\#B_i}{\ell} \\
&\quad= \frac{(n-k)!(k-1)!(k-2)!}{n!(n-1)!} \#B_1! \cdots \#B_{k-1}! \sum_{i=1}^{k-1} \sum_{\ell=1}^{\#B_i - 1} 1\,.
\end{aligned}$$

Since

$$\sum_{i=1}^{k-1} \sum_{\ell=1}^{\#B_i - 1} 1 \;=\; \sum_{i=1}^{k-1} (\#B_i - 1) \;=\; n - (k-1)\,,$$

this establishes the formula for partitions of $[n]$ with $k-1$ blocks. □

Proposition 2.1 enables us to compute the transition probabilities of the reversed state sequence $\mathtt{K}^{[n]}(1), \ldots, \mathtt{K}^{[n]}(n)$. Specifically, let $\xi = (B_1, \ldots, B_k, \varnothing, \ldots)$ be some partition of $[n]$ with $k \geq 2$ non-empty blocks, and γ the partition of $[n]$ obtained from ξ by the coagulation of the blocks B_i and B_j with $1 \leq i < j \leq k$. By Bayes' rule, we have

$$\mathbb{P}(\mathtt{K}^{[n]}(k) = \xi \mid \mathtt{K}^{[n]}(k-1) = \gamma) \;=\; \frac{2}{k(k-1)} \times \frac{\mathbb{P}(\mathtt{K}^{[n]}(k) = \xi)}{\mathbb{P}(\mathtt{K}^{[n]}(k-1) = \gamma)}\,,$$

and after cancellations one gets

$$\mathbb{P}(\mathtt{K}^{[n]}(k) = \xi \mid \mathtt{K}^{[n]}(k-1) = \gamma) \;=\; \frac{2}{n-k+1} \times \frac{\#B_i!\,\#B_j!}{(\#B_i + \#B_j)!}\,.$$

It is interesting to observe that these transition probabilities can also be described as follows. First, recall that the application which maps a block B' with $\varnothing \neq B' \subsetneq B$ to a non-degenerate partition of B into two smaller blocks is surjective and two-on-one.

We deduce that conditionally on $\mathtt{K}^{[n]}(k-1)$, the probability that $\mathtt{K}^{[n]}(k)$ is obtained by splitting a given block with cardinal ℓ of $\mathtt{K}^{[n]}(k-1)$ equals

$$\frac{1}{n-k+1}\sum_{m=1}^{\ell-1}\frac{m!(\ell-m)!}{\ell!}\times\binom{\ell}{m}\;=\;\frac{\ell-1}{n-k+1}.$$

Second, conditionally on this event, the probability that the split of the latter block yields two blocks, say B' and B'', is

$$\frac{2}{\ell-1}\times\frac{\#B'!\,\#B''!}{\ell!}\,.$$

We now arrive at a fundamental property of n-coalescents, namely the following compatibility result.

Lemma 2.1 *For every $n\geq 2$, the restriction $\Pi^{[n]}_{|[n-1]}$ of an n-coalescent to $[n-1]$ is an $(n-1)$-coalescent.*

This result can be both rather surprising and intuitively obvious, depending of the point of point of view. On the one hand, the restriction of partitions to $[n-1]$, $\pi\to\pi_{|[n-1]}$ is a many-to-one map on $\mathcal{P}_n$, and in general, transforming a Markov process by a non-injective function destroys the Markov property. On the other hand, if we think of an n-coalescent as the process describing the genealogy of a sample of n individuals in the Wright-Fisher model with a large population, then discarding the ancestral lineage of the n-th individual obviously yields the genealogy of $(n-1)$ individuals, which is thus given by an $(n-1)$-coalescent.

Proof Fix $\gamma\in\mathcal{P}_{n-1}$ and pick an arbitrary $\pi\in\mathcal{P}_n$ such that $\gamma=\pi_{|[n-1]}$. We work with an n-coalescent $\Pi^{[n]}$ started from π and aim at studying the instant and the location of the first jump of the restricted chain $\Pi^{[n]}_{|[n-1]}$. We shall show that its distribution only depends on γ and more precisely coincides with that one would obtain for an $(n-1)$-coalescent $\Pi^{[n-1]}$ started from γ.

Suppose first that the block of π that contains n is not reduced to the singleton $\{n\}$ and let $i<n$ be the smallest integer in this block. Plainly, for every $t\geq 0$, we may recover $\Pi^{[n]}(t)$ from its restriction to $[n-1]$ simply by adding n to the block of $\Pi^{[n]}_{|[n-1]}(t)$ which contains i. This immediately implies that $\Pi^{[n]}_{|[n-1]}$ is a Markov chain with the same transitions as the $(n-1)$-coalescent.

Suppose now that $\{n\}$ is a block of π and set $k=\#\pi$. Let τ denote the first jump time of $\Pi^{[n]}$, so τ has an exponential law with parameter $k(k-1)/2$, and is independent of $\pi':=\Pi^{[n]}(\tau)$. Consider the event A that the block $\{n\}$ is not involved into the coagulation that occurs at time τ, and A' the complementary event. Clearly $\mathbb{P}(A)=1-2/k$ and, conditionally on A, the restriction of π' to $[n-1]$ is the uniform distribution on the set of $(k-1)(k-2)/2$ partitions obtained from $\pi_{|[n-1]}$ by the coagulation of two of its blocks.

On the event A', let us denote by τ' the waiting time for the second jump of $\Pi^{[n]}$, so τ' has an exponential distribution with parameter $(k-1)(k-2)/2$. Moreover, by the first part of the proof, the restriction of $\pi'':=\Pi^{[n]}(\tau+\tau')$ to $[n-1]$ is independent of τ and τ',

and is uniformly distributed on the set of the $(k-1)(k-2)/2$ partitions obtained from $\pi_{|[n-1]}$ by the coagulation of two of its blocks.

Now the process $\Pi^{[n]}_{|[n-1]}$ stays at $\pi_{|[n-1]}$ up to time $\tau + \mathbb{1}_{A'}\tau'$ and then jumps at $\pi'_{[n-1]}$ on the event A, and at $\pi''_{[n-1]}$ on the event A'. It is easily seen from the classical properties of independent exponential variables that $\tau + \mathbb{1}_{A'}\tau'$ has an exponential distribution with parameter $(k-1)(k-2)/2$ (this is essentially a consequence of the lack of memory). Moreover, this time is independent of the random partition of $[n-1]$ that equals $\pi'_{|[n-1]}$ on the event A and $\pi''_{|[n-1]}$ on A'.

Putting the pieces together, we get that $\Pi^{[n]}_{|[n-1]}$ is a continuous time Markov chain with the following characteristics. When this chain starts from $\gamma \in \mathcal{P}_{n-1}$ with $\#\gamma = k \geq 2$, the next step of the chain occurs after an exponential time with parameter $k(k-1)/2$, and then the chain jumps at one of the $k(k-1)/2$ partitions obtained from γ by the coagulation of two of its blocks, independently of the waiting time and according to the uniform distribution. In other words, $\Pi^{[n]}_{|[n-1]}$ is an $(n-1)$-coalescent. □

The compatibility stated in Lemma 2.1 can be combined with Kolmogorov's extension theorem. Specifically, this enables us to construct simultaneously for all $n \in \mathbb{N}$, a family of processes $\left(\Pi^{[n]}(t), t \geq 0\right)$ such that each $\Pi^{[n]}$ is an n-coalescent, and for every $t \geq 0$, $\Pi^{[n]}(t)$ coincides with the restriction of $\Pi^{[n+1]}(t)$ to $[n]$. In particular, the sequence $\Pi^{[1]}(t), \Pi^{[2]}(t), \ldots$ is compatible, and by Lemma 1.2, this yields the following definition.

Definition 2.1 *There exists a unique (in law) process denoted by* $\Pi^{\mathrm{K}} = (\Pi^{\mathrm{K}}(t), t \geq 0)$*, with values in* $\mathcal{P}_\infty$ *and such that for every* $n \in \mathbb{N}$*, the process induced by the restriction to* $[n]$*,* $\Pi^{\mathrm{K}}_{|[n]} = (\Pi^{\mathrm{K}}_{|[n]}(t), t \geq 0)$*, is an* n*-coalescent. The process* Π^{K} *is called* **Kingman's coalescent**.

We point out that each restriction $\Pi^{\mathrm{K}}_{|[n]}$ has càdlàg (i.e. right-continuous with left-limits) paths with values in $\mathcal{P}_n$, and by the definition of the metric on $\mathcal{P}_\infty$ (cf. Lemma 1.3), this implies that Π^{K} has càdlàg paths valued in $\mathcal{P}_\infty$, a.s. Note that it is implicitly assumed that $\Pi^{\mathrm{K}}(0)$ is the partition of $\mathbb{N}$ into singletons. We also stress that

$$\text{Kingman's coalescent is an exchangeable process,} \tag{2.1}$$

in the sense that for every permutation ς of $\mathbb{N}$, $\left(\varsigma\left(\Pi^{\mathrm{K}}(t)\right), t \geq 0\right)$ has the same distributions as Π^{K}. Indeed, it should be plain from the description of its jump rates that the n-coalescent is an exchangeable process. By compatibility, this implies that the processes Π^{K} and $\varsigma\left(\Pi^{\mathrm{K}}\right)$ have the same finite-dimensional distributions, and since both processes have càdlàg paths a.s., they thus have the same law (see for example Theorem 14.5 in [15] or Section VI.1 in [36]).

We now present a simple description of the evolution of Kingman's coalescent.

Theorem 2.1 (i) *Kingman's coalescent comes down from infinity, that is for every* $t > 0$*, the number* $\#\Pi^{\mathrm{K}}(t)$ *of the non-empty blocks of* $\Pi^{\mathrm{K}}(t)$ *is finite a.s. More precisely,* $(\#\Pi^{\mathrm{K}}(t), t > 0)$ *is a pure death process with death rate* $k(k-1)/2$ *at level* k*.*

(ii) *For every $n \in \mathbb{N}$, let $\mathtt{K}(n)$ denote the state of Π^{K} when $\#\Pi^{\mathrm{K}} = n$. Then the state sequence $(\ldots, \mathtt{K}(n+1), \mathtt{K}(n), \ldots, \mathtt{K}(1))$ is Markovian and independent of the death process $\#\Pi^{\mathrm{K}}$. Further*

$$\lim_{n\to\infty} \mathtt{K}(n) = \mathbf{0}_{[\infty]} \qquad a.s.$$

where $\mathbf{0}_{[\infty]} = \{\{1\}, \{2\}, \ldots\}$ stands for the partition into singletons.

(iii) *The conditional distribution of $\mathtt{K}(n)$ given $\mathtt{K}(n+1)$ is the uniform probability on the set of the $n(n+1)/2$ partitions of $\mathbb{N}$ which can be obtained from $\mathtt{K}(n+1)$ by the coagulation of two of its blocks.*

Proof Clearly, for an arbitrary partition $\pi \in \mathcal{P}_\infty$, the number of non-empty blocks of π can be obtained from its restrictions as the increasing limit

$$\#\pi \;=\; \lim_{n\to\infty} \#\pi_{|[n]}\,. \tag{2.2}$$

So let us fix $k \in \mathbb{N}$ and compute the probability that $\Pi^{\mathrm{K}}_{|[n]}(t)$ has at least k non-empty blocks. By the description of $\#\Pi^{\mathrm{K}}_{|[n]}$ as a death process, we have

$$\mathbb{P}\left(\#\Pi^{\mathrm{K}}_{|[n]}(t) \geq k\right) \;=\; \mathbb{P}\left(\sum_{j=k}^{n} \frac{2}{j(j-1)}\mathbf{e}_j > t\right) \;\leq\; \mathbb{P}\left(\sum_{j=k}^{\infty} \frac{2}{j(j-1)}\mathbf{e}_j > t\right)\,,$$

where $\mathbf{e}_1, \ldots$ is a sequence of i.i.d. standard exponential variables. Since

$$\lim_{k\to\infty} \sum_{j=k}^{\infty} \frac{2}{j(j-1)}\mathbf{e}_j \;=\; 0$$

with probability one, we conclude that $\sup_{n\in\mathbb{N}} \mathbb{P}\left(\#\Pi^{\mathrm{K}}_{|[n]}(t) \geq k\right)$ tends to 0 as $k \to \infty$, that is Kingman's coalescent comes down from infinity.

The description of $\#\Pi^{\mathrm{K}}(\cdot)$ as a death process follows now from that for $\#\Pi^{\mathrm{K}}_{|[n]}(\cdot)$ and (2.2). Moreover if we denote by $\mathtt{K}^{[n]}(\cdot)$ the state sequence of the n-coalescent $\Pi^{\mathrm{K}}_{|[n]}$, then for every fixed integer k, (2.2) yields the identity

$$\mathtt{K}^{[n]}(j) \;=\; \mathtt{K}_{|[n]}(j)\,, \qquad j = 1, \ldots, k\,,$$

provided that n is sufficiently large. The stated properties for the state sequence $\mathtt{K}$ are now readily derived from those of $\mathtt{K}^{[n]}$. □

We stress that Theorem 2.1(i) entirely characterizes the law of $\#\Pi^{\mathrm{K}}$, as there is a unique entrance law from ∞ for such a pure death process. Indeed, it is seen from the proof that its one-dimensional distributions are necessarily given by

$$\mathbb{P}\left(\#\Pi^{\mathrm{K}}(t) \leq k\right) \;=\; \mathbb{P}\left(\sum_{j=k+1}^{\infty} \frac{2}{j(j-1)}\mathbf{e}_j \leq t\right)\,, \qquad t > 0, k \in \mathbb{N}\,,$$

where $\mathbf{e}_1, \ldots$ is a sequence of i.i.d. standard exponential variables. At this point, it is not clear however that the Markovian dynamics of the reversed chain $\mathtt{K}(\cdot)$ in Theorem 2.1 (iii) and the fact that $\mathtt{K}(\infty)$ is the partition into singletons do characterize the law of $\mathtt{K}(\cdot)$. More precisely, one still needs to verify the uniqueness of the entrance law from the partition into singletons for these dynamics. In other words, we do not know yet that the description of Theorem 2.1 does characterize the one-dimensional distributions of $\mathtt{K}(\cdot)$. That this is indeed the case derives from the following statement.

Proposition 2.2 *Consider a Markovian sequence of random partitions*

$$(\ldots, \mathtt{K}'(n+1), \mathtt{K}'(n), \ldots, \mathtt{K}'(1))$$

such that for every integer n, $\#\mathtt{K}'(n) = n$ and the conditional distribution of $\mathtt{K}'(n)$ given $\mathtt{K}'(n+1)$ is the uniform probability measure on the set of the $n(n+1)/2$ partitions which can be obtained from $\mathtt{K}'(n+1)$ by coagulation of two of its blocks. Suppose also that $lim_{n\to\infty}\mathtt{K}'(n) = \mathbf{0}_{[\infty]}$ is the partition into singletons.

Introduce a death process $(D(t), t > 0)$ distributed as $(\#\Pi^{\mathtt{K}}(t), t > 0)$, that is with death rate $k(k-1)/2$ at level $k \in \mathbb{N}$. We assume that $D(\cdot)$ is independent of the chain $(\ldots, \mathtt{K}'(n+1), \mathtt{K}'(n), \ldots, \mathtt{K}'(1))$. Set $\Pi'(t) = \mathtt{K}'(D(t))$ for every $t > 0$ and let $\Pi'(0)$ be the partition of $\mathbb{N}$ into singletons. Then the process $(\Pi'(t), t \geq 0)$ constructed above is a version of Kingman's coalescent.

Proof By construction, the process $(\Pi'(t), t > 0)$ is Markovian and has the same transition probabilities as Kingman's coalescent; we have to check that it has the same one-dimensional distributions as Π^{K}.

Let us introduce for every $n \in \mathbb{N}$, the first passage time of the death process D at level n,

$$T(n) \,=\, \inf\{t > 0 : D(t) = n\}\ .$$

Plainly,

$$\lim_{n\to\infty} T(n) \,=\, 0 \qquad \text{a.s.} \tag{2.3}$$

Observe also that $(D(T(n)+t), t \geq 0)$ is a death process started from n, with death rate $j(j-1)/2$ at level j, and which is of course independent of the sequence $\mathtt{K}'(\cdot)$.

Then define for each $t \geq 0$ a $\mathcal{P}_n$-valued random variable $\Gamma^{[n]}(t)$, by declaring that $i, j = 1, \ldots, n$ belong to the same block of $\Gamma^{[n]}(t)$ if and only if the blocks $\mathtt{K}'_i(n) = \Pi'_i(T(n))$ and $\mathtt{K}'_j(n) = \Pi'_j(T(n))$ are parts of the same block of the partition $\Pi'(T(n)+t)$. By inspection of the jump rates of the n-coalescent $\Gamma^{[n]}(\cdot)$ that are seen from the dynamics of $\Pi'(\cdot)$, we get that $(\Gamma^{[n]}(t), t \geq 0)$ is an n-coalescent.

Next, fix $k \in \mathbb{N}$. On the one hand, it is immediate that whenever n is sufficiently large, for every $i \leq k$, the restriction of the i-th block of $\mathtt{K}'(n)$ to $[k]$, $\mathtt{K}'_i(n) \cap [k]$, is reduced to the singleton $\{i\}$. The very construction of $\Gamma^{[n]}$ implies that

$$\Gamma^{[n]}_{|[k]}(t) \,=\, \Pi'_{|[k]}(T(n)+t)\,, \qquad t \geq 0\,,$$

provided that n is large enough. On the other hand, by the compatibility property stated in Lemma 2.1, the restriction of $\Gamma^{[n]}$ to $[k]$ is a k-coalescent. Taking the limit as $n \to \infty$ and using (2.3), we conclude that for each $k \in \mathbb{N}$, $\Pi'_{|[k]}(\cdot)$ is a k-coalescent, which completes the proof. □

The property of coming down from infinity has a fundamental importance in the study of Kingman's coalescent. As an example of application and development, we point at the following asymptotic result for the number of blocks when time tends to 0.

Proposition 2.3 *We have*

$$\lim_{t\to 0+} t\#\Pi^{\mathrm{K}}(t) = 2 \qquad a.s.$$

Proof Denote for every integer n by $T(n)$ the first instant when the coalescent has n blocks, which can be expressed in the form

$$T(n) = \sum_{j=n+1}^{\infty} \frac{2}{j(j-1)}\mathbf{e}_j$$

where $\mathbf{e}_1, \ldots$ is a sequence of i.i.d. standard exponential variables. In particular

$$\mathbb{E}(T(n)) = 2/n \quad \text{and} \quad \mathrm{Var}(T(n)) = \sum_{j=n+1}^{\infty} \left(\frac{2}{j(j-1)}\right)^2 = O(n^{-3})\,.$$

We get from Chebychev's inequality that

$$\mathbb{P}(|T(n) - 2/n| > 1/n\ln n) = O(n^{-1}\ln^2 n)\,,$$

and then deduce from the Borel-Cantelli lemma that

$$\lim_{k\to\infty} k^2 T(k^2) = 2 \qquad \text{a.s.}$$

By a standard argument of monotonicity, the above can be reinforced into

$$\lim_{k\to\infty} kT(k) = 2 \qquad \text{a.s.},$$

which is equivalent to our claim. □

We now complete this section with the observation that the time of total coalescent, that is the first instant t such that $\Pi^{\mathrm{K}}(t)$ is the trivial partition $\mathbf{1}_{[n]}$ is distributed as

$$\sum_{j=2}^{\infty} \frac{2}{j(j-1)}\mathbf{e}_j\,.$$

If we think of Kingman's coalescent as a model for the genealogy of large populations, this can also be viewed as the age of the Most Recent Common Ancestor (MRCA); we refer to Pfaffelhuber and Wakolbinger [46] and to Simon and Derrida [56] for interesting results about the evolution of the MRCA.

2.3 Interval representation of Kingman's coalescent

Theorem 2.1 describes precisely the transitions of Kingman's coalescent; however, this is not entirely satisfactory as the one-dimensional distributions are not specified. The purpose of this section is to present an explicit construction of the state sequence of Kingman's coalescent based on a natural family of partitions of the unit interval, and which can be viewed as a multidimensional version of the paint-box construction (in this direction, recall from (2.1) that $\Pi^{\mathrm{K}}(\cdot)$ is an exchangeable process).

The starting point is a simple consequence of the representation in Lemma 1.5 of the uniform distribution on the simplex Δ_{N-1}, which provides a natural example of duality between fragmentation and coagulation. Specifically, fragmentation consists of picking a term in a mass-partition by size-biased sampling and splitting this term into two parts using a uniform variable, whereas coagulation consists of merging two distinct terms picked uniformly at random.

Corollary 2.1 *Let* $\mathtt{x} = (\mathtt{x}_1, \ldots, \mathtt{x}_N)$ *be uniformly distributed on the simplex* Δ_{N-1}.

(i) *Introduce a random variable* $\mathtt{k}$ *whose conditional distribution given* $\mathtt{x}$ *is*

$$\mathbb{P}(\mathtt{k} = k \mid \mathtt{x}) \;=\; \mathtt{x}_k\,, \qquad k = 1, \ldots, N,$$

so that $\mathtt{x}_\mathtt{k}$ *is a size-biased sample of* $\mathtt{x}$. *Let* U *be an independent uniform variable on* $[0,1]$. *Then the* $(N+1)$*-tuple*

$$(\mathtt{x}_1, \ldots, \mathtt{x}_{\mathtt{k}-1}, U\mathtt{x}_\mathtt{k}, (1-U)\mathtt{x}_\mathtt{k}, \mathtt{x}_{\mathtt{k}+1}, \ldots, \mathtt{x}_N)$$

has the uniform distribution on the simplex Δ_N.

(ii) *Suppose* $N \geq 3$, *and pick two indices* $\mathtt{j}$ *and* $\mathtt{k}$ *in* $\{1, \ldots, N\}$, *uniformly at random without replacement and independently of* $\mathtt{x}$. *Next, denote by* $\mathtt{x}'$ *the sequence with* $N-1$ *terms which is obtained from* $\mathtt{x}$ *after replacing its* $\mathtt{j}$*-th term* $\mathtt{x}_\mathtt{j}$ *by* $\mathtt{x}_\mathtt{j} + \mathtt{x}_\mathtt{k}$, *and removing its* $\mathtt{k}$*-th term* $\mathtt{x}_\mathtt{k}$. *Then* $\mathtt{x}'$ *is uniformly distributed on the simplex* Δ_{N-2}.

Proof (i) Thanks to Lemma 1.5(ii), we may suppose that $\mathtt{x}$ is the sequence of the lengths of the interval components (ranked from the left to the right) of a random interval-partition $\vartheta :=]0,1[\setminus\{U_1, \ldots, U_{N-1}\}$, where $U_1, \ldots, U_{N-1}$ are i.i.d. uniform variables. Let U_N be another uniform variable which is independent of the preceding. By Lemma 1.1, the length of the interval component I_* of ϑ which contains U_N is a size-biased picked sample from $\mathtt{x}$. Moreover, it should be plain that conditionally on ϑ and I_*, U_N is uniformly distributed on I_*. Hence, the sequence of the lengths of the interval components (ranked from the left to the right) of the interval-partition $]0,1[\setminus\{U_1, \ldots U_N\}$ is distributed as the $(N+1)$-tuple of the statement. Another application of Lemma 1.5(ii) shows that this variable is uniformly distributed on the simplex Δ_N.

(ii) It is convenient to work with the oriented unit circle $\mathcal{C}$. Let $U_1, \ldots, U_N$ be i.i.d. uniform variables on $\mathcal{C}$ which we use to split $\mathcal{C}$ into N arcs. Specifically, we write A_i for the arc with left-extremity U_i; it should be plain from Lemma 1.5(ii) that the sequence $\mathtt{x} := (|A_1|, \ldots, |A_n|)$ is uniformly distributed on Δ_{N-1}.

Next, pick an index $\mathtt{k}$ uniformly at random in $\{1,\dots,N\}$, independently of the U_i. Then $U_{\mathtt{k}}$ is the right extremity of some arc, say $A_{\mathtt{j}} =]U_{\mathtt{j}}, U_{\mathtt{k}}[$. It is easily checked that conditionally on $\mathtt{k}$, $\mathtt{j}$ is uniformly distributed in $\{1,\dots,N\}\backslash\{\mathtt{k}\}$ and is independent of $\mathtt{x}$. Merging the arcs $A_{\mathtt{j}}$ and $A_{\mathtt{k}}$ amounts to splitting the circle $\mathcal{C}$ using $N-1$ independent uniform variables, namely $\{U_1,\dots,U_{\mathtt{k}-1},U_{\mathtt{k}+1},\dots,U_N\}$, and the resulting sequence of the arc lengths has the uniform distribution on Δ_{N-2}. This establishes the claim. □

This invites us to introduce $U_1,\dots$ and $U'_1,\dots$, two independent sequences of i.i.d. uniform variables on $[0,1]$. Set $\vartheta(1) =]0,1[$ and for each $n \in \mathbb{N}$, consider the random interval-partition

$$\vartheta(n+1) =]0,1[\setminus \{U'_1,\dots,U'_n\}\,.$$

We denote by $\mathtt{K}'(n)$ the paint-box constructed from the random interval-partition $\vartheta(n)$ and the U_i, so that two indices i and j belong to the same block of $\mathtt{K}'(n)$ if and only if the variables U_i and U_j belong to the same interval component of $\vartheta(n)$.

Proposition 2.4 *The sequence* $(\dots,\mathtt{K}'(n+1),\mathtt{K}'(n),\dots,\mathtt{K}'(1))$ *has the same distribution as the state sequence* $(\dots,\mathtt{K}(n+1),\mathtt{K}(n),\dots,\mathtt{K}(1))$ *of Kingman's coalescent*

Proof It is then seen from the proof of Corollary 2.1(ii), that the sequence

$$(\dots,|\vartheta(n+1)|^{\downarrow},|\vartheta(n)|^{\downarrow},\dots,|\vartheta(1)|^{\downarrow})$$

is Markovian, and more precisely, the conditional distribution of the mass-partition $|\vartheta(n)|^{\downarrow}$ given $|\vartheta(n+1)|^{\downarrow}$ is the uniform distribution on the set of the $(n+1)n/2$ mass-partitions that can be obtained from $|\vartheta(n+1)|^{\downarrow}$ by merging two of its non-zero terms and then rearranging in decreasing order. This property can be shifted to paint-boxes and then $(\dots,\mathtt{K}'(n+1),\mathtt{K}'(n),\dots,\mathtt{K}'(1))$ is a Markov chain governed by the same transitions as the state sequence $\mathtt{K}(\cdot)$ of Kingman's coalescent. Further, as $n\to\infty$, $\mathtt{K}'(n)$ clearly converges to the partition into singletons. We can conclude from Proposition 2.2. □

Note that by the exchangeability property (2.1), the state sequence $\mathtt{K}(\cdot)$ is also formed by exchangeable random partitions, and hence each possesses asymptotic frequencies. As an immediate consequence of Proposition 2.4 and Lemma 1.5(ii), the distribution of the latter is given as follows.

Corollary 2.2 *For every integer* $n \geq 2$, $|\mathtt{K}(n)|^{\downarrow}$ *has the law of the decreasing rearrangement of a variable that is uniformly distributed on the simplex* Δ_{n-1}.

Corollary 2.2 can be used to provide a more intuitive proof of Proposition 2.1, which was established by guessing the formula beforehand. We refer to Corollary 2.3 in [5] for a nice argument. It is also interesting to observe from Corollary 2.1(i) that time-reversal transforms the process of the ranked asymptotic frequencies of the state sequence in Kingman's coalescent into a sequence $(|\mathtt{K}(1)|^{\downarrow},|\mathtt{K}(2)|^{\downarrow},\dots)$ which can be viewed as a fragmentation sequence.

Another interesting consequence of Proposition 2.4 is that it readily yields precise estimates for the asymptotic frequency of blocks in Kingman's coalescent for small times.

Corollary 2.3 *Let* $|\mathtt{K}(n)|^{\downarrow} = (\mathtt{K}_1^{(n)}, \ldots, \mathtt{K}_n^{(n)})$ *denote the sequence ranked in the decreasing order of the asymptotic frequencies of blocks of the chain sequence at the* n*-th step. Then*

$$n\mathtt{K}_1^{(n)} - \ln n \implies G$$

and

$$n^2\mathtt{K}_n^{(n)} \implies \mathbf{e}$$

where the notation $\implies$ *means convergence in distribution as* $n \to \infty$*,* G *stands for a Gumbel variable (i.e.* $\mathbb{P}(G \leq x) = \exp(-\mathrm{e}^{-x})$ *for every* $x \in \mathbb{R}$*) and* $\mathbf{e}$ *for a standard exponential variable.*

Further, the empirical measure of the rescaled fragments

$$\epsilon(n) = n^{-1} \sum_{i=1}^{n} \delta_{n\mathtt{K}_i^{(n)}}$$

converges with probability 1 *as* $n \to \infty$ *towards the standard exponential distribution, in the sense of weak convergence of probability measures.*

Proof The combination of Lemma 1.5 and Proposition 2.4 shows that if γ_n denotes an independent gamma variable with parameter $(n, 1)$, then $\gamma_n|\mathtt{K}(n)|^{\downarrow}$ has the distribution of a family of n independent standard exponential variables ranked in the decreasing order. As by the central limit theorem, $|\gamma_n - n| = o(n^{2/3})$ when $n \to 1$, the first claim follows immediately from elementary extreme values theory. The second is even simpler as the minimum of n independent standard exponential variables has the exponential distribution with parameter n, and the third derives similarly from the Glivenko-Cantelli theorem. □

We stress that Corollary 2.3 can also be restated directly in terms of the coalescent $\Pi^{\mathrm{K}}(t)$, thanks to the simple estimate for the number of blocks at time t stated in Proposition 2.3. More precisely, as $t \to 0+$

$$\frac{2}{t}\mathtt{K}_1^{(n)} - \ln \frac{2}{t} \implies G$$

and the random measure

$$\frac{t}{2} \sum_i \delta_{\frac{2}{t}|\Pi^{\mathrm{K}}(t)|_i^{\downarrow}}$$

converges weakly to the standard exponential distribution, with probability one.

2.4 Neutral mutations and allelic partition

ROUGH DRAFT

One can represent Kingman's coalescent as a random tree, where the set of leaves is given by $\mathbb{N}$, internal nodes correspond to blocks that appear in the coalescent process

(more precisely the block associated to an internal node is the set of leaves of the subtree which is rooted at that node), and the length of the branches between two adjacent nodes is the time elapsed between the formation of the smaller block and the time when it coalesces with some other block to form the larger one. If we reduce this random tree to the first n leaves, we simply get the genealogical tree of the first n individuals in the population (see the figure at the beginning of Section 2.2).

This tree has a root which corresponds to the most recent common ancestor (MRCA), and the distance of any leave to the root (i.e. the height of the tree) is just the age of the MRCA (which a finite. Recall that this height can be expressed as

$$\sum_{i=1}^{\infty} \frac{2}{i(i-1)} \mathbf{e}_i$$

where $(\mathbf{e}_i : i \in \mathbb{N})$ is a sequence of i.i.d. standard exponential variables. It is also easy to study the total length L_n of the tree reduced to the first n leaves. Elementary considerations about the n-coalescent show that

$$L_n \sim \sum_{i=2}^{n} \frac{2}{i-1} \mathbf{e}_i \,.$$

Using the standard properties of the sequences of independent exponential variables, we can also express the preceding identity in distribution as

$$L_n \sim \max\{2\mathbf{e}_i : 1 \leq i \leq n-1\} \,.$$

It then follows from a well-known limit theorem for extremes that

$$\frac{1}{2} L_n - \ln n \implies G$$

where G denotes a standard Gumbel variable, i.e. with cumulative distribution function $\exp(-\mathrm{e}^{-x})$.

Imagine now that neutral mutations affect lineages with rate $\theta/2$; i.e. superpose on Kingman's tree a Poisson process of marks with intensity $\theta/2$ per unit of length on the skeleton. Use the marks to cut the tree; this yields a partition of the set of leaves which is referred to as the allelic partition. In other words, two individuals (i.e. leaves) belong to the same block of the allelic partition if and only if there is no mark on the unique path which connects these two individuals, that is if and only if their two ancestral lineages coalesce before being marked, which means, from a biological point of view, that this two individuals have the same genetic type.

It should be clear that the allelic partition is exchangeable, and thus we can measure the frequencies of the sub-families (blocks). An important question in biology is to determine the statistics of this allelic partition, as it provides an important test for non-Darwinian evolution (to decide whether or not selection has a role in the diversity of gene frequencies in a population).

Theorem 2.2 *If we rank the frequencies of the blocks of the allelic partition according to the decreasing age of the alleles, then we observe a* GEM(θ)*-partition.*

Sketch of the proof It is convenient to re-label the individuals using the allelic forest, by first assigning to each individual a length as follows. If the individual k is the smallest of its allelic block, then we call k the prime of its allelic block and we define $\ell(k)$ as the age of its allele, that is the distance from the leaf k to the first mark on its ancestral lineage. Otherwise, the ancestral lineage of k coalesces with the lineage of some individual $j < k$ before any of the two lineages receive a mark, and we decide that $\ell(k)$ is the time when this coalescence occurs. The new labels of individuals are then obtained by reordering in the decreasing order of the lengths (i.e. the individual with label 1 is the one who has the largest length, label 2 is assigned to the individual with the second largest length, and so on).

The motivation for this new labeling stems from the fact that it enables us to study simultaneously the genealogy and the mutation process in a simpler way. It can be checked that when k is large, the label assigned to k is $\sim k$, and therefore this reordering does not affect the frequencies of blocks. Observe also that the increasing order of the labels of the ancestors of allelic groups is precisely the decreasing order of the ages of alleles. Our goal is to check that if we consider individuals one after the other in the order of their labels, and gather them according to their allele, then we observe precisely the dynamics of a Chinese restaurant.

For every time $t \geq 0$, consider the restriction of $\Pi^{\mathrm{K}}(t)$ to individuals which have no mutation marks on their ancestral lineages up-to time t. The choice for labeling implies that if at time t there are k blocks, then the individuals with labels $1, \ldots, k$ belong to distinct blocks; and the first block which will disappear either due to a coalescence or a mutation is the block containing the individual with label k. More precisely the dynamics of Kingman's coalescence and the fact that mutations occur at a fixed rate $\theta/2$ on each lineage (i.e. branch) shows that, when there are k blocks left, the first event occurs after an exponential time with parameter $k(k-1)/2 + k\theta/2 = (k-1+\theta)k/2$. It corresponds to a mutation with probability $\theta/(k-1+\theta)$ and to a coalescence with probability $(k-1)/(k-1+\theta)$.

Now time-reverse the process from the time when the oldest allele appears (so now time runs as the biological time) The first event occurs after an exponential time with parameter $\theta/2$, it is a mutation with probability $\theta/(1+\theta)$ and a coalescent with probability $1/(1+\theta)$. By iteration, when there are $k-1$ blocks, the k-th block appears due to a mutation event with probability $\theta/(k-1+\theta)$ and due to a coalescent with probability $(k-1)/(k-1+\theta)$. More precisely, in the latter case, it joins one of the $k-1$ previous blocks uniformly at random. Thus we have precisely the dynamics of the Chinese Restaurant for the parameter θ, which produces the GEM(θ) partition. □

In particular, Theorem 3.1 implies that if we ignore the age of alleles and simply rank the allelic frequencies in the decreasing order, we observe a Poisson-Dirichlet partition PD$(0, \theta)$, which rephrases the celebrated result of Ewens. Recalling that the GEM(θ) partition derives from PD$(0, \theta)$ by size-biased reordering, we also deduce that the conditional probability given the allelic frequencies that an individual sampled at random bears the oldest allele is simply given by the frequency of the allele of that individual. As

a consequence, the unconditional probability of that event is

$$\mathbb{E}(\text{beta}(1,\theta)) = \frac{1}{1+\theta}\,.$$

In the same vein, one sees that the probability that the oldest allele is not found in a sample of k individuals is

$$\mathbb{E}(\text{beta}(\theta,1)^k) = \frac{\theta}{k+\theta}\,.$$

Chapter 3

General exchangeable coalescents

The fundamental point in the construction of Kingman's coalescent is the use of the space $\mathcal{P}_\infty$ of partitions of $\mathbb{N}$ as the natural state space. Restricting partitions to a finite set is a powerful discretization technique which enables the application of the elementary theory of continuous time Markov chains. In particular, this is the key to circumventing the major difficulty that infinitely many coagulations occur immediately after the initial time. In this section, we shall develop this idea further by introducing a general class of coalescent processes where coagulations of infinitely many blocks may occur and different merging may take place simultaneously. These coalescents have been considered first by Möhle and Sagitov [44] and Schweinsberg [53], extending earlier works by Pitman [47] and Sagitov [52].

3.1 Coagulation of partitions

We start by introducing the basic notion of coagulation of partitions which has a central role in this chapter. Recall that blocks in a partition are enumerated in the increasing order of their least element, that for every block $B \subseteq \mathbb{N}$, $\mathcal{P}_B$ denotes the set of partitions of B, and that for $n \in \mathbb{N}$, $\mathcal{P}_n = \mathcal{P}_{[n]}$ with $[n] = \{1, \ldots, n\}$.

Definition 3.1 *Let $B \subseteq \mathbb{N}$ and $k' \in \overline{\mathbb{N}}$. A pair of partitions $(\pi, \pi') \in \mathcal{P}_B \times \mathcal{P}_{k'}$ is called* **admissible** *if the number of non-empty blocks of π is $\#\pi \leq k'$. For every admissible pair of partitions (π, π'), we call* **coagulation of π by π'** *and write $\mathrm{Coag}(\pi, \pi')$ for the partition $\pi'' = (\pi''_j, j \in \mathbb{N})$ of B given by*

$$\pi''_j := \bigcup_{i \in \pi'_j} \pi_i \,, \qquad j \in \mathbb{N} \,,$$

where $(\pi_i, i \in \mathbb{N})$ and $(\pi'_j, j \in \mathbb{N})$ denote the sequence of the blocks of π and π', respectively.

For instance, for $B = [10]$, if

$$\pi = (\{1, 6, 7\}, \{2, 4, 5\}, \{3, 8\}, \{9, 10\}) \quad \text{and} \quad \pi' = (\{1, 3\}, \{2, 4\}) \,,$$

then $\mathrm{Coag}(\pi, \pi')$ results from the merging of the first and third and, respectively, the second and fourth blocks of π. We get

$$\mathrm{Coag}(\pi, \pi') = (\{1, 3, 6, 7, 8\}, \{2, 4, 5, 9, 10\})\,.$$

Plainly, the partition $\pi'' = \mathrm{Coag}(\pi, \pi')$ is *coarser* than π, in the sense that each block of the latter is contained in some block of the former. We also stress that the blocks π''_j above are labeled according to the increasing order of their least elements, in agreement with our convention.

The partition into singletons will have a special role in this section. It is convenient to introduce the notation

$$\mathbf{0}_{[\infty]} \,:=\, (\{1\}, \{2\}, \ldots)\,.$$

Indeed, $\mathbf{0}_{[\infty]}$ serves as a neutral element for the coagulation operator, in the sense that

$$\mathrm{Coag}(\pi', \mathbf{0}_{[\infty]}) \,=\, \mathrm{Coag}(\mathbf{0}_{[\infty]}, \pi') \,=\, \pi'\,.$$

More generally, for every block $B \subseteq \mathbb{N}$, we denote by $\mathbf{0}_B$ the partition of B into singletons, that is $\mathbf{0}_B$ is the restriction of $\mathbf{0}_{[\infty]}$ to B. Henceforth, we shall also use the notation $\mathbf{0}$ for the mass-partition in $\mathcal{P}_{\mathtt{m}}$ which is identical to 0; we hope this slight abuse of notation should be helpful rather than confusing.

Recall that the notion of restriction enables us to endow $\mathcal{P}_\infty = \mathcal{P}_{\mathbb{N}}$ with a tree-structure and an induced ultra-metric. It is important to observe the following easy fact.

Lemma 3.1 *Coagulation is a Lipschitz-continuous function from $\mathcal{P}_\infty \times \mathcal{P}_\infty$ to $\mathcal{P}_\infty$, and associative in the sense that*

$$\mathrm{Coag}\left(\pi, \mathrm{Coag}\left(\pi', \pi''\right)\right) \,=\, \mathrm{Coag}\left(\mathrm{Coag}\left(\pi, \pi'\right), \pi''\right)$$

whenever (π, π') and (π', π'') are admissible pairs.

Proof Indeed, the fact that the labels of the blocks of a partition are assigned according to the order of their least element yields that for every $n \in \mathbb{N}$ and $\pi, \pi' \in \mathcal{P}_\infty$

$$\mathrm{Coag}(\pi, \pi')_{|[n]} \,=\, \mathrm{Coag}(\pi_{|[n]}, \pi') \,=\, \mathrm{Coag}(\pi_{|[n]}, \pi'_{|[n]})\,. \tag{3.1}$$

These identities and the very definition of the topology on $\mathcal{P}_\infty$ imply our first claim. The second is immediately checked. □

Another useful observation is that the coagulation operator preserves exchangeability in the following sense.

Lemma 3.2 *Let π and π' be two independent exchangeable random partitions. Then the random partition $\mathrm{Coag}(\pi, \pi')$ is also exchangeable.*

Proof Let σ be some permutation of $\mathbb{N}$, and $\sigma(\pi'')$ the image of $\pi'' = \mathrm{Coag}(\pi, \pi')$ by the action of σ. So the blocks of $\sigma(\pi'')$ are the images of those of π'' by the inverse permutation σ^{-1}, that is they are given by

$$\sigma^{-1}(\pi''_j) \;=\; \bigcup_{i \in \pi'_j} \sigma^{-1}(\pi_i)\,, \qquad j \in \mathbb{N}\,,$$

where $(\pi_i, i \in \mathbb{N})$ and $(\pi'_j, j \in \mathbb{N})$ denote the sequence of the blocks of π and π', respectively. The $\sigma^{-1}(\pi_i)$ form the family of the blocks of $\sigma(\pi) := \tilde{\pi}$, which is distributed as π. We stress that in general, these blocks are not labeled according to the increasing order of their least elements, so one should not believe that $\sigma(\pi'') = \mathrm{Coag}(\sigma(\pi), \pi')$.

Nonetheless, let σ' be the permutation of $\mathbb{N}$ such that the i-th block of $\sigma(\pi)$ is $\sigma^{-1}(\pi_{\sigma'(i)})$ (if the number of non-empty blocks of π is finite, we decide that $\sigma'(n) = n$ for every $n > \#\pi$, which then specifies σ' uniquely). By construction, we now have $\sigma(\pi'') = \mathrm{Coag}(\sigma(\pi), \sigma'(\pi'))$.

Finally, observe that σ' is independent of π', and it follows from the exchangeability of π' that the pair $(\sigma(\pi), \sigma'(\pi'))$ has the same law as (π, π'). This shows that $\sigma(\pi'')$ has the same law as π''. □

We also point out that more generally, the same argument (with heavier notation) shows that if $\pi^{(1)}, \ldots, \pi^{(n)}$ are independent exchangeable random partitions of $\mathbb{N}$, and we define by induction $\gamma^{(1)} := \pi^{(1)}$ and $\gamma^{(i+1)} := \mathrm{Coag}(\gamma^{(i)}, \pi^{(i+1)})$ for $i = 1, \ldots, n-1$, then the n-tuple $(\gamma^{(1)}, \ldots, \gamma^{(n)})$ is jointly exchangeable, in the sense that its distribution is invariant by the action of permutations of $\mathbb{N}$ on the space $(\mathcal{P}_\infty)^n$.

3.2 Exchangeable coalescents and coagulation rates

We now turn our attention to Markov processes with values in the space $\mathcal{P}_\infty$ of partitions of $\mathbb{N}$. In this direction, the notion of coagulation leads naturally to the following definition.

Definition 3.2 *Fix $n \in \overline{\mathbb{N}}$, and let $\Pi = (\Pi(t), t \geq 0)$ be a Markov process with values in $\mathcal{P}_n$ which is continuous in probability.*

(i) Π *is called an* **exchangeable coalescent** *if its semigroup can be described as follows. For every $t, t' \geq 0$, the conditional distribution of $\Pi(t+t')$ given $\Pi(t) = \pi$ is the law of $\mathrm{Coag}(\pi, \pi')$, where π' is some exchangeable random partition (whose law only depends on t').*

(ii) *We call Π a* **standard** *exchangeable coalescent if Π is an exchangeable coalescent started from $\mathbf{0}_{[n]}$, the partition of $[n]$ into singletons.*

Remark. Plainly, Kingman's coalescent Π^{K} is a remarkable example of a standard exchangeable coalescent. At this point, it may be interesting to draw a parallel with Brownian motion and Lévy processes, anticipating some results which will be developed later on. In short, Brownian motion is essentially the unique Markov process in the Euclidean

space which has continuous sample paths and a semigroup given by convolution operators (i.e. the increments are independent and stationary). If we remove the hypothesis of continuity of the paths, we arrive at the more general notion of Lévy process. The fine structure of Lévy processes is revealed by the Lévy-Itô decomposition which underlines the Poissonnian nature of the jumps. In the framework of coalescent, the notion of coagulation operator based on exchangeable random partitions should be viewed as the counterpart of convolution semigroup. Loosely speaking, continuity of the paths then corresponds to the hypothesis that coagulation events are binary (i.e. involve only two blocks), while a multiple coagulation should be viewed as a jump. Hence Kingman's coalescent plays the same role as Brownian motion, and one may expect that in general, exchangeable coalescent should have a Poissonnian structure. We shall see in this section that this informal comparison holds rigorously.

Later in this text, we shall be mostly interested in $\mathcal{P}_\infty$-valued processes. We make a few simple observations related to this definition which will be important in this setting.

First, the elementary property (3.1) of the coagulation operator implies that if Π is an exchangeable coalescent with values in $\mathcal{P}_\infty$, then for every $n \in \mathbb{N}$, the restriction $\Pi_{|[n]}$ of Π to $[n]$ is also an exchangeable coalescent [1]. More precisely, for every $t, t' \geq 0$, the conditional distribution of $\Pi_{|[n]}(t+t')$ given $\Pi_{|[n]}(t) = \pi$ is the law of $\mathrm{Coag}(\pi, \Pi(t')) = \mathrm{Coag}(\pi, \Pi_{|[n]}(t'))$. Conversely, it follows from the Compatibility Lemma 1.2 that if Π is a process with values in $\mathcal{P}_\infty$ such that its restriction $\Pi_{|[n]}$ is an exchangeable coalescent for each $n \in \mathbb{N}$, then Π itself is an exchangeable coalescent.

Second, we may work with a version of Π with regular paths. Indeed, since $\mathcal{P}_n$ is a finite space, we may consider for each n a version of the Markov chain $\Pi_{|[n]}$ with càdlàg paths, so that, by the very definition of the distance on $\mathcal{P}_\infty$, Π then also has càdlàg paths a.s. From now on, we shall always consider such regular version.

Third, we point out that the exchangeable random partition π' in (i) which is used to specify the semigroup, has obviously the law of $\Pi(t')$ when Π is standard (simply take $t = 0$ in the definition).

Finally, there is no loss of generality in focussing on standard exchangeable coalescents. Indeed, thanks to the associativity of the coagulation operator, if $\pi \in \mathcal{P}_\infty$ and Π is a standard exchangeable coalescent, then the process $(\mathrm{Coag}(\pi, \Pi(t)), t \geq 0)$ is an exchangeable coalescent started from π, with the same semigroup as Π. In particular, the semigroup of an exchangeable coalescent is given in terms of its standard version Π by

$$\pi \ \to \ \mathbb{E}\left(\varphi(\mathrm{Coag}(\pi, \Pi(t)))\right), \qquad \pi \in \mathcal{P}_\infty, \tag{3.2}$$

where $\varphi : \mathcal{P}_\infty \to \mathbb{R}$ stands for a generic measurable bounded function. Recall that the space of partitions $\mathcal{P}_\infty$ is metrizable and compact.

Proposition 3.1 *The semigroup of an exchangeable coalescent enjoys the Feller property. This means that for every continuous function $\varphi : \mathcal{P}_\infty \to \mathbb{R}$, the map (3.2) is continuous*

[1] The compatibility property (3.1) is crucial. Indeed, the restriction map $\pi \to \pi_{[n]}$ is not injective, and therefore, the restriction to $[n]$ of a Markov process with values in $\mathcal{P}_\infty$ may well not be Markovian.

for each $t \geq 0$, *and*

$$\lim_{t\to 0} \mathbb{E}\left(\varphi(\mathrm{Coag}(\pi, \Pi(t)))\right) \,=\, \varphi(\pi)\,, \qquad \pi \in \mathcal{P}_\infty\,.$$

Proof This follows immediately from Lemma 3.1. □

The Feller property provides another argument for the existence of a càdlàg version of the process. It also ensures that the (completed) natural filtration of Π is right-continuous and enables us to extend the Markov property to stopping times. See for example Section III.2 in Revuz and Yor [50] for details.

Let us now investigate the transitions of an exchangeable coalescent, using the simple structure of their restrictions. Indeed, exchangeable coalescents with values in a finite space like $\mathcal{P}_n$ are Markov chains. Thus, the distribution of the restricted chain $\Pi_{|[n]}$ can be characterized by its jump rates. In this direction, denote the jump rate of the standard chain $\Pi_{|[n]}$ from $\mathbf{0}_{[n]}$ to π by

$$q_\pi \,:=\, \lim_{t\to 0+} \frac{1}{t}\mathbb{P}\left(\Pi_{|[n]}(t) = \pi\right)\,, \qquad \pi \in \mathcal{P}_n\backslash\{\mathbf{0}_{[n]}\}\,.$$

For instance, in the case of Kingman's coalescent, we have $q_\pi = 1$ if the partition π consists in a unique pair and else singletons, and $q_\pi = 0$ otherwise. It is immediately seen from this definition that the jump rates inherit an exchangeability property from Π, in the sense that for every permutation σ of $[n]$, there is the identity

$$q_\pi \,=\, q_{\sigma(\pi)}\,. \tag{3.3}$$

Note that for every $n' \geq n$, q_π also gives the jump rate of $\Pi_{|[n']}$ from π' to $\mathrm{Coag}(\pi', \pi)$, where $\pi' \in \mathcal{P}_{n'}$ with $\#\pi' = n$, and that any jump which is not of this type must have a rate equal to zero. Hence, the family $(q_\pi, \pi \in \mathcal{P}_n\backslash\{\mathbf{0}_{[n]}\}$ and $n \in \mathbb{N})$ characterizes the transitions of all the restricted chains $\Pi_{|[n]}$, and thus of the exchangeable coalescent Π.

A first basic result about these jump rates is that they can be represented by a single measure on $\mathcal{P}_\infty$. In this direction, introduce the notation

$$\mathcal{P}_{n',\pi} \,=\, \left\{\pi' \in \mathcal{P}_{n'} : \pi'_{|[n]} = \pi\right\}\,,$$

where $\pi \in \mathcal{P}_n$ and $n' \in \{n, n+1, \ldots, \infty\}$. Recall that $\mathcal{P}_\infty$ has a natural tree-structure; in particular $\mathcal{P}_{\infty,\pi}$ corresponds to the subtree of $\mathcal{P}_\infty$ spanned by vertex π.

Proposition 3.2 *Let* $(q_\pi : \pi \in \mathcal{P}_n\backslash\{\mathbf{0}_{[n]}\}$ *and* $n \in \mathbb{N})$ *be the family of jump rates of some exchangeable coalescent* Π. *There exists a unique measure* μ *on* $\mathcal{P}_\infty$, *called the* **coagulation rate** *of* Π, *such that* $\mu(\{\mathbf{0}_{[\infty]}\}) = 0$ *and*

$$\mu\left(\mathcal{P}_{\infty,\pi}\right) \,=\, q_\pi$$

for every $n \in \mathbb{N}$ *and every partition* $\pi \in \mathcal{P}_n\backslash\{\mathbf{0}_{[n]}\}$.

Proof Take any $n' \geq n$, and note the identity

$$\mathcal{P}_{\infty,\pi} \;=\; \bigcup_{\pi' \in \mathcal{P}_{n',\pi}} \mathcal{P}_{\infty,\pi'}\,, \tag{3.4}$$

where, in the right-hand side, the union is over disjoint subsets of $\mathcal{P}_\infty$. Because the Markov chain $\Pi_{|[n]}$ can be obtained as the restriction of $\Pi_{|[n']}$ to $[n]$, its jump rate q_π from $\mathbf{0}_{[n]}$ to $\pi \in \mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}$ coincides with the total jump rate for $\Pi_{|[n']}$ from $\mathbf{1}_{[n']}$ to $\mathcal{P}_{n',\pi}$. In other words, we have

$$q_\pi \;=\; \sum_{\pi' \in \mathcal{P}_{n',\pi}} q_{\pi'}\,. \tag{3.5}$$

This shows that the function

$$\mu : \mathcal{P}_{\infty,\pi} \to q_\pi\,, \qquad \pi \in \mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\} \text{ for some } n \in \mathbb{N}$$

is additive, and we conclude by an easy application of Caratheodory's extension theorem that μ has a unique extension to a measure on $\mathcal{P}_\infty \backslash \{\mathbf{0}_{\mathbb{N}}\}$. The exchangeability is immediate. □

3.3 Poissonian construction

The purpose of this section is to present an explicit construction of exchangeable coalescents from their coagulation rates. To start with, let us observe that the coagulation rate μ of an exchangeable coalescent is a measure μ on $\mathcal{P}_\infty$ which fulfills the following properties:

$$\mu(\{\mathbf{0}_{[\infty]}\}) = 0 \quad \text{and} \quad \mu(\pi \in \mathcal{P}_\infty : \pi_{|[n]} \neq \mathbf{0}_{[n]}) < \infty \quad \text{for every } n \in \mathbb{N}\,, \tag{3.6}$$

and

$$\mu \text{ is exchangeable, that is invariant by the action of permutations.} \tag{3.7}$$

Indeed, (3.6) is plain from Proposition 3.2, whereas (3.7) derives from (3.3) and Proposition 3.2. Conversely, it is easy to show that conditions (3.6) and (3.7) are sufficient for a measure to be the coagulation rate of some exchangeable coalescent.

Lemma 3.3 *Let μ be a measure on $\mathcal{P}_\infty$ which fulfills (3.6) and (3.7). There exists an exchangeable coalescent with coagulation rate μ.*

Proof We merely present a sketch of the proof as precise construction will be provided in Proposition 3.3 below. For every $n \in \mathbb{N}$, we can use the family

$$q_\pi \;=\; \mu\left(\mathcal{P}_{\infty,\pi'}\right)\,, \qquad \pi' \in \mathcal{P}_k \backslash \{\mathbf{0}_{[k]}\} \text{ and } k \leq n$$

as the jump rates of some Markov process $\Pi^{[n]}$ with values in $\mathcal{P}_n$. More precisely, for every $\pi \in \mathcal{P}_n$ with $\#\pi = k$ non-empty blocks and every $\pi' \in \mathcal{P}_k \backslash \{\mathbf{0}_{[k]}\}$, the jump rate of

$\Pi^{[n]}$ from π to $\mathrm{Coag}(\pi, \pi')$ is $q_{\pi'}$, and all the other jump rates are 0. By additivity of μ and (3.4), we see that (3.5) holds. It is easy to check that the restriction of $\Pi^{[n]}$ to $[k]$ is again Markovian, with the same jump rates as $\Pi^{[k]}$. By the same argument based on Kolmogorov's extension theorem that we used just before Definition 2.1, we can construct a unique (in distribution) Markov process $\Pi = (\Pi(t), t \geq 0)$ with values in $\mathcal{P}_\infty$ and started from $\mathbf{0}_{[\infty]}$, such that $\Pi_{|[n]}$ is a version of $\Pi^{[n]}$ for every $n \in \mathbb{N}$.

The exchangeability assumption (3.7) implies that Π is an exchangeable coalescent. Indeed, by the very construction of Markov chains with assigned jump rates, it implies that if τ stands for the first jump time of $\Pi_{|[n]}$, then the random partition $\Pi_{|[n]}(\tau)$ is exchangeable. By iteration and Lemma 3.2, we deduce that for every $t \geq 0$, $\Pi_{|[n]}(t)$ is also an exchangeable random partition, and we conclude that Π is an exchangeable coalescent with coagulation rate μ. □

We now present a more explicit construction which does not require the appeal to Kolmogorov's extension Theorem, and sheds light on the Poissonian structure of exchangeable coalescent. Specifically, let μ be some measure on $\mathcal{P}_\infty$ which fulfills (3.6) and (3.7). Introduce a Poisson random measure $\mathtt{M}$ on $\mathbb{R}_+ \times \mathcal{P}_\infty$ with intensity $dt \otimes \mu(d\pi)$, and for each $n \in \mathbb{N}$, let $\mathtt{M}_n$ be the image of $\mathtt{M}$ by the map $(t, \pi) \to (t, \pi_{|[n]})$. So $\mathtt{M}_n$ is a Poisson measure on $\mathbb{R}_+ \times \mathcal{P}_n$ with intensity $dt \times \mu_n(d\pi)$, where μ_n denotes the measure on $\mathcal{P}_n$ obtained as the image of μ by the restriction map $\pi \to \pi_{|[n]}$. In particular, for every $\pi \in \mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}$, the process $(\mathtt{M}_n([0,t] \times \{\pi\}), t \geq 0)$ that measures the fiber based on π as time passes, is Poisson with intensity $q_\pi = \mu(\mathcal{P}_{\infty,\pi})$ and to different partitions in $\mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}$ correspond independent processes.

We denote by $\big\{(t_i, \pi^{(i)}), i \in \mathbb{N}\big\}$ the family of the atoms of $\mathtt{M}_n$ on $\big(\mathbb{R}_+ \times \mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}\big)$, ranked in increasing order of their first coordinate. We set $\Pi^{[n]}(t) = \mathbf{0}_{[n]}$ for $t \in [0, t_1[$, and then define recursively

$$\Pi^{[n]}(t) \;=\; \mathrm{Coag}(\Pi^{[n]}(t_i-), \pi^{(n)}(t_i))\,, \qquad \text{for every } t \in [t_i, t_{i+1}[\,.$$

Proposition 3.3 *In the notation above, for every $t \geq 0$, the sequence of random partitions $\big(\Pi^{[n]}(t), n \in \mathbb{N}\big)$ is compatible. If we denote by $\Pi(t)$ the unique partition of $\mathcal{P}_\infty$ such that $\Pi_{|[n]}(t) = \Pi^{[n]}(t)$ for every $n \in \mathbb{N}$, then the process $\Pi = (\Pi(t), t \geq 0)$ is a standard exchangeable coalescent with coagulation rate μ.*

Proof Fix $n \geq 2$ and write $(t_1, \pi^{(1)})$ for the first atom of $\mathtt{M}_n$ on $\mathbb{R}_+ \times \big(\mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}\big)$. Plainly, $\Pi^{[n-1]}(t) = \Pi^{[n]}_{|[n-1]}(t)$ for every $t \in [0, t_1[$.

Consider first the case when $\pi^{(1)}_{|[n-1]} \neq \mathbf{0}_{[n-1]}$. Then $(t_1, \pi^{(1)}_{|[n-1]})$ is the first atom of $\mathtt{M}_{n-1}$ on $\mathbb{R}_+ \times \big(\mathcal{P}_{n-1} \backslash \{\mathbf{0}_{[n-1]}\}\big)$, and it follows from (3.1) that $\Pi^{[n-1]}(t) = \Pi^{[n]}_{|[n-1]}(t)$ for every $t \in [t_1, t_2[$. Next, consider the case $\pi^{(1)}_{|[n-1]} = \mathbf{0}_{[n-1]}$. Then $\mathtt{M}_{n-1}$ has no atoms on $[0, t_2[\times \big(\mathcal{P}_{n-1} \backslash \{\mathbf{0}_{[n-1]}\}\big)$, and it follows again from (3.1) that $\Pi^{[n-1]}(t) = \Pi^{[n]}_{|[n-1]}(t) = \mathbf{0}_{[n-1]}$ for every $t \in [0, t_2[$. By iteration, this shows that the restriction of $\Pi^{[n]}$ to $[n-1]$ coincides with $\Pi^{[n-1]}$.

It is immediate from the Poissonian construction that each $\Pi^{[n]}$ is a continuous time Markov chain on $\mathcal{P}_n$. More precisely, for every $\pi \in \mathcal{P}_n$ with $\#\pi = k$ non-empty blocks and every $\pi' \in \mathcal{P}_k \backslash \{\mathbf{0}_{[k]}\}$, the jump rate of $\Pi^{[n]}$ from π to $\mathrm{Coag}(\pi, \pi')$ is given by the intensity of the Poisson process $\mathtt{M}_n([0,\cdot] \times \{\pi'\})$, and we know that the latter equals $q_{\pi'} = \mu\left(\mathcal{P}_{\infty,\pi'}\right)$. Plainly, all the other jump rates are 0, and by an application of Lemma 3.2, we have thus that $\Pi^{[n]}$ is an exchangeable coalescent.

It should now be plain that the process Π which is specified by its restrictions $\Pi_{|[n]} = \Pi^{[n]}$, is a standard exchangeable coalescent with coagulation rate μ. □

3.4 Characterization of coagulation rates

Our next goal is to characterize explicitly the coagulation rates of exchangeable coalescents, that is the measures on $\mathcal{P}_\infty$ which fulfill (3.6) and (3.7). In this direction, we first give two fundamental examples.

First, for every pair $(i,j) \in \mathbb{N}^2$ with $i < j$, we write $\mathtt{K}(i,j)$ for the partition of $\mathbb{N}$ whose blocks consist of the pair $\{i,j\}$ and the singletons $\{k\}$ for $k \neq i,j$. If δ_π stands for the Dirac point mass at $\pi \in \mathcal{P}_\infty$, then the measure

$$\mu^{\mathtt{K}} := \sum_{1 \leq i < j < \infty} \delta_{\mathtt{K}(i,j)}$$

fulfills conditions (3.6) and (3.7). It is immediately seen from the Poissonian construction that $\mu^{\mathtt{K}}$ is the coagulation rate of Kingman's coalescent.

To construct the second example, recall that $\varrho_{\mathbf{s}}$ denotes the distribution of a paint-box based on $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$. Next, we consider mixtures of paint-boxes. Recall from Proposition 1.4 that the map $s \to \varrho_{\mathbf{s}}$ is continuous, and that we denote by $\mathbf{0} = (0,\ldots)$ the mass-partition identical to 0. Consider a sigma-finite measure ν on $\mathcal{P}_{\mathtt{m}}$ such that

$$\nu(\{\mathbf{0}\}) = 0 \quad \text{and} \quad \int_{\mathcal{P}_{\mathtt{m}}} \left(\sum_{i=1}^{\infty} s_i^2 \right) \nu(\mathrm{d}\mathbf{s}) < \infty\,, \tag{3.8}$$

and define a measure on $\mathcal{P}_\infty$ by

$$\varrho_\nu(d\pi) = \int_{\mathbf{s} \in \mathcal{P}_{\mathtt{m}}} \varrho_{\mathbf{s}}(d\pi) \nu(\mathrm{d}\mathbf{s})\,.$$

Lemma 3.4 *The measure ϱ_ν is the coagulation rate of some exchangeable coalescent.*

Proof Each $\varrho_{\mathbf{s}}$ is an exchangeable probability measure on $\mathcal{P}_\infty$. Exchangeability is preserved by mixing, so ϱ_ν is an exchangeable measure on $\mathcal{P}_\infty$. For all $\mathbf{s} \in \mathcal{P}_{\mathtt{m}} \backslash \{\mathbf{0}\}$, the measures $\varrho_{\mathbf{s}}$ assign zero mass to the partition into singletons $\mathbf{0}_{[\infty]}$, so (3.7) holds for the mixture ϱ_ν.

Next, for $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$, we see from the paint-box construction that

$$\varrho_{\mathbf{s}}(\pi_{|[2]} \neq \mathbf{0}_{[2]}) = \sum_{i=1}^{\infty} s_i^2\,.$$

Thus (3.8) ensures that condition (3.6) holds for $n = 2$ and $\mu = \varrho_\nu$. By exchangeability of ϱ_ν, we conclude that in fact (3.6) holds for any $n \in \mathbb{N}$. This establishes our statement, thanks to Lemma 3.3. □

We now state the main result of this section, which claims that every coagulation rate can be obtained as a linear combination of the two preceding examples.

Theorem 3.1 *Consider a measure μ on $\mathcal{P}_\infty$ which fulfills (3.6) and (3.7). Then there exists a unique $\mathtt{c} \geq 0$ and a unique measure ν on $\mathcal{P}_{\mathtt{m}}$ that fulfills (3.8), such that*

$$\mu = \mathtt{c}\mu^{\mathtt{K}} + \varrho_\nu.$$

Specifically, the following holds:

(i) *For μ-almost every $\pi \in \mathcal{P}_\infty$, π possesses asymptotic frequencies $|\pi|$.*

(ii) *The measure on $\mathcal{P}_{\mathtt{m}}$ given by*

$$\nu(\mathrm{d}\mathbf{s}) \;=\; \mathbb{1}_{\{s \neq \mathbf{0}\}} \mu(|\pi|^{\downarrow} \in \mathrm{d}\mathbf{s})$$

fulfills (3.8), and there is the identity

$$\mathbb{1}_{\{|\pi|^{\downarrow} \neq \mathbf{0}\}} \mu(d\pi) \;=\; \varrho_\nu(d\pi)\,, \qquad \pi \in \mathcal{P}_\infty.$$

(iii) *There exists a real number $\mathtt{c} \geq 0$ such that*

$$\mathbb{1}_{\{|\pi|^{\downarrow} = \mathbf{0}\}} \mu(d\pi) \;=\; \mathtt{c}\mu^{\mathtt{K}}(d\pi)\,.$$

Later in the text, we shall refer to $\mathtt{c}$ as the *coefficient of binary coagulation* and to ν as the *measure of multiple coagulations* of the exchangeable coalescent Π.

Proof (i) For every $n \in \mathbb{N}$, introduce the measure

$$\mu_n(d\pi) \;=\; \mathbb{1}_{\{\pi_{|[n]} \neq \mathbf{0}_{[n]}\}} \mu(d\pi)\,, \qquad \pi \in \mathcal{P}_\infty\,.$$

Then μ_n is a finite measure on $\mathcal{P}_\infty$ which is invariant by the action of permutations that coincide with the identity on $[n]$. Let $\vec{\mu}_n$ be the image of μ_n by the n-shift on partitions, namely the map $\pi \to \vec{\pi}$ defined by

$$i \overset{\vec{\pi}}{\sim} j \iff i + n \overset{\pi}{\sim} j + n\,, \qquad i, j \in \mathbb{N}.$$

Then $\vec{\mu}_n$ is a finite exchangeable measure on $\mathcal{P}_\infty$, and by Kingman's Theorem 1.1, $\vec{\mu}_n$ can be expressed as a mixture of paint-boxes:

$$\vec{\mu}_n(d\pi) \;=\; \int_{\mathcal{P}_{\mathtt{m}}} \varrho_{\mathbf{s}}(d\pi) \vec{\mu}_n(|\pi|^{\downarrow} \in \mathrm{d}\mathbf{s})\,. \tag{3.9}$$

As shift does not affect asymptotic frequencies, μ_n-almost every partition has asymptotic frequencies. Since μ can be obtained as the increasing limit of the μ_n, this establishes the first claim.

(ii) Let us write $\{i \sim j\}$ for the event that two integers i and j belong to the same block of π. By (3.9), we have for every $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ that

$$\mu_n(n+1 \sim n+2 \mid |\pi|^{\downarrow} = \mathbf{s}) \;=\; \sum_{k=1}^{\infty} s_k^2 \,.$$

Hence, if we denote by ν_n the restriction to $\mathcal{P}_{\mathtt{m}} \backslash \{\mathbf{0}\}$ of the image measure of μ_n by the map $\pi \to |\pi|^{\downarrow}$, then

$$\mu_n(n+1 \sim n+2) \;\geq\; \int_{\mathcal{P}_{\mathtt{m}}} \left(\sum_{i=1}^{\infty} s_i^2 \right) \nu_n(\mathrm{d}\mathbf{s}) \,.$$

On the one hand, the finite measure ν_n increases as $n \uparrow \infty$ to the measure ν defined in the statement, so

$$\lim_{n\to\infty} \int_{\mathcal{P}_{\mathtt{m}}} \left(\sum_{i=1}^{\infty} s_i^2 \right) \nu_n(\mathrm{d}\mathbf{s}) \;=\; \int_{\mathcal{P}_{\mathtt{m}}} \left(\sum_{i=1}^{\infty} s_i^2 \right) \nu(\mathrm{d}\mathbf{s}) \,.$$

On the other hand, by the exchangeability of μ,

$$\mu_n(n+1 \sim n+2) \;\leq\; \mu(n+1 \sim n+2) \;=\; \mu(1 \sim 2) \;<\; \infty \,,$$

which shows that ν fulfills (3.8).

Finally, fix $k \in \mathbb{N}$ and pick any $\pi^{[k]} \in \mathcal{P}_k \backslash \{\mathbf{0}_{[k]}\}$. We have by monotone convergence

$$\mu(\pi_{|[k]} = \pi^{[k]}, |\pi|^{\downarrow} \neq \mathbf{0}) \;=\; \lim_{n\to\infty} \mu(\pi_{|[k]} = \pi^{[k]}, |\pi|^{\downarrow} \neq \mathbf{0}, \pi_{|\{k+1,\ldots,k+n\}} \neq \mathbf{0}_{\{k+1,\ldots,k+n\}}) \,.$$

In the notation introduced in (i), we see from an obvious permutation that

$$\mu(\pi_{|[k]} = \pi^{[k]}, |\pi|^{\downarrow} \neq \mathbf{0}, \pi_{|\{k+1,\ldots,k+n\}} \neq \mathbf{0}_{\{k+1,\ldots,k+n\}}) \;=\; \vec{\mu}_n(\pi_{|[k]} = \pi^{[k]}, |\pi|^{\downarrow} \neq \mathbf{0}) \,.$$

Applying (3.9) and then letting n tend to ∞, we conclude that

$$\mu(\pi_{|[k]} = \pi^{[k]}, |\pi|^{\downarrow} \neq \mathbf{0}) \;=\; \int_{\mathcal{P}_{\mathtt{m}}} \varrho_{\mathbf{s}}(\pi_{|[k]} = \pi^{[k]}) \nu(\mathrm{d}\mathbf{s}) \,.$$

As k is arbitrary, this establishes (ii).

(iii) Consider $\tilde{\mu}$, the restriction of μ to the event $\left\{1 \sim 2, |\pi|^{\downarrow} = \mathbf{0}\right\}$, which has finite mass. Its image by the 2-shift as defined in (i) for $n = 2$, is an exchangeable finite measure on $\mathcal{P}_\infty$ for which almost every partition has asymptotic frequencies $\mathbf{0}$. Thus it is proportional to the Dirac mass at $\mathbf{0}_{\mathbb{N}}$, the partition of $\mathbb{N}$ into singletons. By exchangeability of μ and the finiteness of $\tilde{\mu}$, the set of partitions π for which there is some $n \geq 3$ in the same block as 1 and 2 must have zero $\tilde{\mu}$ measure. Thus $\tilde{\mu} = \mathtt{c}\delta_{\mathtt{K}(1,2)}$ for some $\mathtt{c} \geq 0$, and by exchangeability, we conclude that $\mathbb{1}_{\{|\pi|^{\downarrow}=\mathbf{0}\}}\mu(d\pi) = \mathtt{c}\mu^{\mathtt{K}}(d\pi)$. □

Remark. One can compute the jump rates q_π of the restricted chains $\Pi_{|[n]}$ explicitly in terms of the rate of binary coagulation $\mathtt{c}$ and the measure of multiple coagulation ν; however, the formulas than can be obtained are rather involved in this general setting (see Schweinsberg [53]). Nonetheless, in the special case of simple coalescents discussed in the next section, expressions for the jump rates become more tractable; see (4.2) below.

3.5 Exchangeable mass-coalescents

In this section, we consider the process of the ranked asymptotic frequencies associated with a standard exchangeable coalescent Π.

3.5.1 Markov property

We start by introducing the operator of coagulation of mass-partitions by a partition of $\mathbb{N}$. Dust plays a the special role in this definition, as microscopic particles of dust can coagulate and form macroscopic masses. In other words, the mass of dust of the coagulation of some mass-partition $\mathbf{s}$ may be strictly smaller than the initial mass of dust of $\mathbf{s}$. Recall that $\tilde{\mathcal{P}}_{\mathtt{m}}$ stands for the space of numerical sequences whose decreasing rearrangement is a mass-partition.

Definition 3.3 *Let $\mathbf{s} \in \tilde{\mathcal{P}}_{\mathtt{m}}$ and $\pi \in \mathcal{P}_\infty$ a partition of $\mathbb{N}$ which possesses asymptotic frequencies. We write $\mathrm{Coag}(s, \pi)$ for the mass-partition obtained by the decreasing rearrangement of the terms of the sequence*

$$s_0|\pi_i| + \sum_{j\in\pi_i} s_j\,, \qquad i \in \mathbb{N}\,,$$

where $s_0 = 1 - \sum_{j=1}^\infty s_j$ denotes the mass of dust for $\mathbf{s}$.

The following elementary lemma makes the connection with Definition 3.1.

Lemma 3.5 *Let $\pi \in \mathcal{P}_\infty$ be a deterministic partition which possesses asymptotic frequencies, and such that the blocks of π with zero asymptotic frequency are either empty or singletons. Write*

$$S \;=\; \bigcup_{i\in\mathbb{N}\,:\,|\pi_i|=0} \pi_i\,,$$

for the set of singletons of π. Suppose also S has an asymptotic frequency and $|S| = 1 - \sum_{i=1}^\infty |\pi_i|$. Let π' be a random exchangeable partition. Then the random partition $\pi'' = \mathrm{Coag}(\pi, \pi')$ possesses asymptotic frequencies, and more precisely, there is the identity

$$|\pi''|^\downarrow \;=\; \mathrm{Coag}(|\pi|, \pi')\,.$$

Proof Thanks to Kingman's Theorem 1.1, we may suppose that π' is given by a paint-box based on some interval-partition $\vartheta \in \mathcal{P}_{\mathrm{I}}$ and an i.i.d. sequence of uniform variables, $U_1, \ldots$ Let I be some interval component of ϑ, and focus on the block $B' := \{i \in \mathbb{N} : U_i \in I\}$ of π' and on the corresponding block of π'':

$$B'' \;:=\; \bigcup_{i\in B'} \pi_i\,.$$

The points in S are the singletons of π; we denote by $\sigma : S \to \mathbb{N}$ the injective map such that for $j \in S$, $\sigma(j)$ is the index of the block $\{j\}$ in π, that is $\{j\} = \pi_{\sigma(j)}$. We also introduce $J := \{j \in \mathbb{N}, |\pi_j| > 0\}$.

For every integer n we have

$$\frac{1}{n}\#(B'' \cap [n]) \;=\; \frac{1}{n} \sum_{j \in S \cap [n]} \mathbb{1}_{\{U_{\sigma(j)} \in I\}} + \sum_{j \in J \cap B'} \frac{1}{n}\#(\pi_j \cap [n])\,.$$

Since $(U_{\sigma(j)}, j \in S)$ form an i.i.d. sequence of uniform variables, we deduce from the law of large numbers and the assumption that S has an asymptotic frequency that when n tends to ∞, the first term in the sum in the right-handle side converges to $|S||I| = |S||B'|$. Since each block π_j $(j \in J)$ has an asymptotic frequency, it follows from Fatou's lemma that

$$\liminf_{n\to\infty} \frac{1}{n}\#(B'' \cap [n]) \;\geq\; |S||B'| + \sum_{j \in J \cap B'} |\pi_j| \;=\; |S||B'| + \sum_{j \in B'} |\pi_j|\,.$$

Next, fix $\eta > 0$, consider the finite set $J(\eta) := \{j \in \mathbb{N}, |\pi_j| > \eta\}$, and introduce

$$B_\eta \;:=\; \bigcup_{j \in J \setminus J(\eta)} \pi_j\,,$$

so that B_η is the complementary set of $S \cup \left(\bigcup_{j\in J(\eta)} \pi_j\right)$. It follows from our assumptions that B_η has an asymptotic frequency which tends to 0 as $\eta \to 0$. On the other hand, there is the obvious inclusion

$$B'' \;\subseteq\; \{j \in S : U_{\sigma(j)} \in I\} \cup B_\eta \cup \left(\bigcup_{j \in B' \cap J(\eta)} \pi_j \right)\,,$$

from which we deduce the lower-bound

$$\limsup_{n\to\infty} \frac{1}{n}\#(B'' \cap [n]) \;\leq\; |S||B'| + |B_\eta| + \sum_{j \in J(\eta) \cap B'} |\pi_j|\,.$$

Letting η tend to 0, we conclude that the block B'' has an asymptotic frequency given by $|B''| = |S||B'| + \sum_{j\in B'} |\pi_j|$. □

The random partition $\Pi(t)$ given by the state of some exchangeable coalescent Π evaluated at time $t \geq 0$ is exchangeable, and we denote by $|\Pi|^\downarrow = \big(|\Pi(t)|^\downarrow, t \geq 0\big)$ the process of the ranked asymptotic frequencies of Π. Here is a first basic property.

Proposition 3.4 *The process of the ranked asymptotic frequencies $|\Pi|^\downarrow$ is Markovian and, more precisely, for every $t, t' \geq 0$, the conditional distribution of $|\Pi|^\downarrow(t + t')$ given $|\Pi|^\downarrow(t) = s$ is that of* $\mathrm{Coag}(s, \Pi(t'))$. *Moreover, this semigroup fulfills the Feller property.*

Proof Let $(\mathcal{F}_t)_{t\geq 0}$ denote the natural filtration of Π, and $\Phi : \mathcal{P}_{\mathtt{m}} \to \mathbb{R}$ some continuous function. The Markov property of Π implies that for every $t, t' \geq 0$, on the event $\Pi(t) = \pi$,

$$\mathbb{E}\left(\Phi(|\Pi(t+t')|^{\downarrow}) \mid \mathcal{F}_t\right) \,=\, \mathbb{E}\left(\Phi(|\mathrm{Coag}(\pi, \Pi(t'))|^{\downarrow})\right) \,.$$

We know from Proposition 1.3 that π fulfills the requirements of Lemma 3.5 with probability one, and as $\Pi(t')$ is exchangeable, this implies that $|\Pi|^{\downarrow}$ is Markovian and has the semigroup given in the statement.

The Feller property follows readily from the continuity of the coagulation operator (cf. Lemma 3.1) and Proposition 1.4. □

3.5.2 Dust in exchangeable mass-coalescents

Recall from Theorem 3.1 that the coagulation rate μ can be expressed in terms of a coefficient $\mathtt{c} \geq 0$ of binary coagulation, and of a measure ν on $\mathcal{P}_{\mathtt{m}}$ that fulfills (3.8) and specifies the rate of multiple coagulations. We conclude this section by considering the question of whether the asymptotic frequencies of $\Pi(t)$ are proper or improper, and in the latter case, we characterize the distribution of the mass of dust in terms of $\mathtt{c}$ and ν. Recall that $(\mathcal{F}_t)_{t\geq 0}$ stand for the natural filtration of Π.

Proposition 3.5 *The random partition $\Pi(t)$ has improper asymptotic frequencies with positive probability if and only if*

$$\mathtt{c} = 0 \quad \textit{and} \quad \int_{\mathcal{P}_{\mathtt{m}}} (1 - s_0)\nu(\mathrm{d}\mathbf{s}) \,<\, \infty \,, \qquad \textit{where } s_0 = 1 - \sum_{i=1}^{\infty} s_i \,.$$

In this case, if we denote by

$$\mathtt{D}(t) \,=\, 1 - \sum_{i=1}^{\infty} |\Pi_i(t)| \,, \qquad t \geq 0 \,,$$

the mass of dust at time t, then the law of $\mathtt{D}(t)$ is specified by its entire moments

$$\mathbb{E}\left(\mathtt{D}(t)^k\right) \,=\, \exp\left(-t \int_{\mathcal{P}_{\mathtt{m}}} (1 - s_0^k)\nu(\mathrm{d}\mathbf{s})\right) \,, \qquad k \in \mathbb{N} \,.$$

Moreover, the multiplicative increments of the process $(\mathtt{D}(t), t \geq 0)$ are independent and stationary, in the sense that for every $t, t' \geq 0$, the conditional distribution of $\mathtt{D}(t+t')$ given $\mathcal{F}_t$ is that of $\mathtt{D}(t)\tilde{\mathtt{D}}(t')$, where $\tilde{\mathtt{D}}(t')$ is independent of $\mathcal{F}_t$ and has the same law as $\mathtt{D}(t')$. This characterizes the finite-dimensional distributions of the process $(\mathtt{D}(t), t \geq 0)$.

Remark. Another way of formulating Proposition 3.5 is to consider the right-continuous version $\xi = (\xi(t), t \geq 0)$ of the process $(-\ln \mathtt{D}(t), t \geq 0)$. Then ξ is a subordinator, with Laplace exponent

$$\Phi(q) \,=\, \int_{\mathcal{P}_{\mathtt{m}}} (1 - s_0^q)\nu(\mathrm{d}\mathbf{s}) \,, \qquad q > 0 \,.$$

This formula for Φ is of the Lévy-Khintchine type. In particular, the drift coefficient is zero, the Lévy measure is the image of the measure of multiple coagulations ν by the map $s \to -\ln s_0$, and the killing rate given by $\mathtt{k} = \nu(s_0 = 0)$.

Proof Recall that an exchangeable partition has proper asymptotic frequencies if and only if none of its block is a singleton; see Proposition 1.3. By exchangeability, we thus see that $\Pi(t)$ has proper asymptotic frequencies a.s. if and only if $\mathbb{P}(\Pi_1(t) = \{1\}) = 0$.

By the Poissonian construction of Section 4.2.3, the event $\Pi_1(t) = \{1\}$ occurs if and only if all the atoms (r, π) of the Poisson random measure $\mathtt{M}$ with $r \leq t$ fulfill $\pi_1 = \{1\}$. Because $\mathtt{M}$ has intensity $dt \otimes \mu(d\pi)$, the latter event has probability $\exp(-at)$, where $a = \mu(\{\pi \in \mathcal{P}_\infty : \pi_1 \neq \{1\}\})$. An easy calculation based on the expression $\mu = \mathtt{c}\mu^{\mathtt{K}} + \varrho_\nu$ for the coagulation rate μ (cf. Theorem 3.1) shows that

$$a \,=\, \infty \mathbb{1}_{\{\mathtt{c}>0\}} + \int_{\mathcal{P}_{\mathtt{m}}} (1 - s_0)\nu(\mathrm{d}\mathbf{s})\,,$$

which establishes our first claim.

From now on, we suppose that $\mathtt{c} = 0$ and $\int (1 - s_0)\nu(\mathrm{d}\mathbf{s}) < \infty$. In order to compute the entire moments of $\mathtt{D}(t)$, we observe from the paint-box construction that for every $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ and $k \in \mathbb{N}$, we have

$$\varrho_{\mathbf{s}}\left(\{\pi \in \mathcal{P} : 1, \ldots, k \text{ are singletons of } \pi\}\right) \,=\, s_0^k\,,$$

where $\varrho_{\mathbf{s}}$ denotes the law of an exchangeable random partition with ranked asymptotic frequencies given by $\mathbf{s}$. The same argument as above shows that the event that $1, \ldots, k$ are singletons of $\Pi(t)$ occurs if and only if $1, \ldots, k$ are singletons of π for all partitions π such that (r, π) is an atom of the Poisson random measure $\mathtt{M}$ for some $r \leq t$. By immediate Poissonian calculations, we get

$$\begin{aligned}
\mathbb{E}\left(\mathtt{D}(t)^k\right) &= \mathbb{P}\left(1, \ldots, k \text{ are singletons of } \Pi(t)\right) \\
&= \exp\left(-t\mu\left(\{\pi \in \mathcal{P} : \exists i \leq k \text{ which is not a singleton of } \pi\}\right)\right) \\
&= \exp\left(-t \int_{\mathcal{P}_{\mathtt{m}}} (1 - s_0^k)\nu(\mathrm{d}\mathbf{s})\right)\,,
\end{aligned}$$

where the last equality stems from the expression $\mu = \varrho_\nu$ for the coagulation rate. Of course, $\mathtt{D}(t)$ takes its values in $[0, 1]$ and thus is determined by its entire moments.

Finally, we check that the multiplicative increments of $\mathtt{D}(\cdot)$ are independent and stationary. In this direction, consider $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ and π a partition which possesses asymptotic frequencies. We write $\mathbf{s}' = \mathrm{Coag}(\mathbf{s}, \pi)$, so that by Definition 3.3, the mass of dust $s'_0 = 1 - \sum_{i=1}^\infty s'_i$ of s' is given by

$$s'_0 \,=\, 1 - \sum_{i=1}^\infty \left(s_0|\pi_i| + \sum_{j \in \pi_i} s_j\right) \,=\, 1 - \sum_{j=1}^\infty s_j - s_0 \sum_{i=1}^\infty |\pi_i| \,=\, s_0\left(1 - \sum_{i=1}^\infty |\pi_i|\right)\,.$$

Combining this observation with Proposition 3.4 completes the proof. □

Chapter 4

Simple coalescents and dual population models

The motivation for introducing Kingman's coalescent was to study the genealogy in the Wright-Fisher model in the limit when the size N of the population tends to ∞, and in the regime when one unit of time corresponds to N generations. The Fleming-Viot process arises in the limit of rescaled Wright-Fisher in the same regime, and can be viewed in some sense (which has a rigorous mathematical interpretation) as the dual of Kingman's coalescent.

Cannings [22, 23] introduced a natural generalization of the model of Wright and Fisher, based on a arbitrary N-tuple $(\xi_1, \ldots, \xi_N)$ of integer-valued variables such that $\xi_1 + \cdots + \xi_N = N$, and which is exchangeable. We should think of ξ_j as the number of children of the j-th individual, so using a sequence of i.i.d. copies of the N-tuple $(\xi_1, \ldots, \xi_N)$, we can construct by an obvious iteration a random population model with fixed size; observe that the Wright-Fisher model corresponds to the case when $(\xi_1, \ldots, \xi_N)$ is a multinomial variable. Kingman has established the robustness of the coalescent by showing that under quite general assumptions (namely that if $\xi_{1,N}$ is the number of children of a typical individual for the population model of size N, then $\mathrm{Var}(\xi_{1,N})$ converges while the higher moments of $\xi_{1,N}$ remain bounded), the genealogy for the Cannings' model converges when the size N of the population tends to ∞ to the coalescent. Kingman's criterion has been improved later by Möhle [42] who pointed out that convergence to the coalescent holds whenever $\mathbb{E}(\xi_{1,N}^3) = o(N\mathbb{E}(\xi_{1,N}^2))$. Then Möhle and Sagitov [44] pointed at different regimes under which the genealogy for Cannings' model converges more generally to an arbitrary exchangeable coalescent. Loosely speaking, multiple coagulations for the ancestral lineages correspond to large birth events, in the sense that at certain exceptional times, a single parent may generate a significant proportion of children at the next generation. The problem of studying limits of rescaled version of Cannings' model has only been addressed very recently by N. Perkowski [45].

The purpose of this chapter is to develop directly the study of infinite population models which are dual to a natural sub-class of exchangeable coalescents. This yields to a generalization of the Fleming-Viot process which can be introduced from various perspectives.

4.1 Simple coalescents

In chapter, we will be interested in an important and natural sub-family of exchangeable coalescents, which was introduced independently by Pitman [47] and Sagitov [52] and bear deep relations with a model for the genealogy of large populations which generalizes the Wright-Fisher model. A key feature is that when a coagulation occurs, all the blocks involved in the coagulation merge as a single block.

Definition 4.1 (i) *Call a partition $\pi \in \mathcal{P}_\infty$* **simple** *if and only if at most one of its blocks is neither empty nor reduced to a singleton.*

(ii) *A mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ is called* **simple** *if and only if it is given by $\mathbf{s} = (x, 0, \ldots)$ for some $0 \leq x \leq 1$.*

(iii) *An exchangeable coalescent Π is called* **simple** *if and only if its coagulation rate μ is supported by simple partitions.*

Simple exchangeable coalescents are better known in the literature as **Λ-coalescents**, where Λ refers to some finite probability measure on $[0, 1]$. We prefer to use a different terminology as the measure Λ has a rather indirect interpretation in terms of coalescence.

We may use the representation given in Theorem 3.1 of the coagulation rate of an exchangeable coalescent Π in the form $\mu = \mathtt{c}\mu^{\mathtt{K}} + \varrho_\nu$ as the sum of a binary coagulation rate and a multiple coagulation rate with rate ν, where $\mathtt{c} \geq 0$ and ν fulfills (3.8). It should be clear that Π is simple if and only if ν is supported by simple mass-partitions.

Later in this text, it will be convenient to use the canonical projection $(x, 0, \ldots) \to x$ and identify the sub-space of simple mass-partitions with the unit interval, and then ν as a measure on $[0, 1]$ with

$$\nu(\{0\}) \,=\, 0 \quad \text{and} \quad \int_{[0,1]} x^2 \nu(\mathrm{d}x) < \infty \,. \tag{4.1}$$

An alternative formulation is that Π is simple if and only if for every $n \in \mathbb{N}$ and every $\pi \in \mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}$, the jump rate q_π of $\Pi_{|[n]}$ from $\mathbf{0}_{[n]}$ to π equals zero, except when π is simple. In that case, we may compute the jump rates explicitly in terms of the rates $\mathtt{c} \geq 0$ and ν. Indeed, for every $2 \leq k \leq n$, if $\pi \in \mathcal{P}_n \backslash \{\mathbf{0}_{[n]}\}$ is simple and has one block with k elements, then

$$q_\pi \,:=\, q_{n,k} \,=\, \mathtt{c}\mathbb{1}_{\{k=2\}} + \int_{]0,1]} x^k (1-x)^{n-k} \nu(\mathrm{d}x) \,. \tag{4.2}$$

We mention that Pitman [47] uses the finite measure

$$\Lambda(\mathrm{d}x) = \mathtt{c}\delta_0 + x^2 \nu(\mathrm{d}x)$$

on $[0, 1]$ to characterize the coagulation rates rather than $\mathtt{c}$ and ν, and in this setting (4.2) becomes

$$q_\pi \,:=\, q_{n,k} \,=\, \int_{[0,1]} x^{k-2} (1-x)^{n-k} \Lambda(\mathrm{d}x) \,.$$

In the rest of this section, we will be interested in the *fixation time*

$$\zeta := \inf\{t \geq 0 : \Pi(t) = \mathbf{1}_{\mathbb{N}}\},$$

where $\mathbf{1}_{\mathbb{N}}$ denote the trivial partition having just one non-empty block. Clearly, this is the absorbing state for any coalescent process. One says that *fixation* occurs when the fixation time ζ is finite. This is equivalent for the coalescent to come down from infinity The following useful bound is due to Schweinsberg [54].

Proposition 4.1 *Set for every integer $n \geq 2$*

$$\varphi(n) := \frac{n(n-1)}{2}\mathsf{c} + \int_{]0,1]}((1-x)^n - 1 + nx)\nu(\mathrm{d}x).$$

Then the expectation of the fixation time is bounded by

$$\mathbb{E}(\zeta) \leq \sum_{n=2}^{\infty} 1/\varphi(n).$$

As a consequence, fixation occurs with probability one provided that the series in the right-hand side converges (this holds in particular when the coefficient c of binary coagulation is not zero).

More precisely, Schweinsberg [54] has proved that, as soon as the measure ν has no atom at 1, the condition of convergence of the series in Proposition 4.1 is also necessary for fixation. In the same vein, this condition is necessary and sufficient for the coalescent Π to come down from infinity, in the sense that $\#\Pi(t) < \infty$ a.s. for every $t > 0$, where $\#\pi$ stands for the number of non-empty blocks of a partition π. In terms of the population model, this means that for any $t > 0$, we can find a finite number of individuals in the initial population which generate the entire population at time t.

It has been pointed out in [13] that the necessary and sufficient condition

$$\sum_{n=2}^{\infty} 1/\varphi(n) < \infty$$

for coming down from ∞ is equivalent to

$$\int_1^{\infty} \Psi^{-1}(q)\mathrm{d}q < \infty \tag{4.3}$$

where

$$\Psi(q) = \mathsf{c}\frac{q^2}{2} + \int(\mathrm{e}^{-qx} - 1 + qx)\nu(\mathrm{d}x), \qquad q \geq 0.$$

The upshot of this reformulation is that Ψ should be viewed as a branching mechanism of a continuous state branching process (see the forthcoming Section 6.2.1) and then (4.3) corresponds to the classical necessary and sufficient condition due to Grey for extinction of the branching process. In this direction, we also mention an interesting result due

to Berestycki *et al.* [6] about the speed at which a simple coalescent comes down from infinity: when (4.3), the number of its blocks at time $t > 0$ is asymptotic to $v_\Psi(t)$ as $t \to 0+$ where $v_\Psi(t)$ is defined as the unique solution to

$$\int_{v_\Psi(t)}^{\infty} \Psi^{-1}(q)\mathrm{d}q = t\,.$$

We further refer to Limic [41] and references therein for extensions to some large classes of general exchangeable coalescents.

The proof of Proposition 4.1 relies on the following technical lemma. Recall the notation (4.2).

Lemma 4.1 *The function φ increases, and for every $n \geq 2$, there is the identity*

$$\varphi(n) \,=\, \sum_{k=2}^{n}(k-1)\binom{n}{k} q_{n,k}\,.$$

Proof For every $0 < x \leq 1$, function $b \to bx - 1 + (1-x)^b$ increases on $b \geq 2$, which implies the first claim. The second follows easily from the definition of $q_{n,k}$ and the binomial formula. □

We can now tackle the proof of Proposition 4.1.

Proof Let $n \in \mathbb{N}$ denote a fixed integer, and consider for every $t \geq 0$ the number of non-empty blocks $\#\Pi_{|[n]}(t)$. It should be plain from the dynamics of the restricted chain $\Pi_{|[n]}$ that the process $\#\Pi_{|[n]}$ is a Markov chain with values in $[n]$ with non-increasing paths, which is absorbed at 1. More precisely, for every $\ell = 2, \ldots, n$ and $k = 2, \ldots, \ell$, when the coalescent chain $\Pi_{|[n]}$ has ℓ blocks and a coagulation involving k of its blocks occurs, then $\#\Pi_{|[n]}$ decreases by $k-1$. Hence the jump rate $r_{\ell,\ell-k+1}$ of $\#\Pi_{|[n]}$ from ℓ to $\ell-k+1$ is given in terms of the jump rates (4.2) of the coalescent by

$$r_{\ell,\ell-k+1} \,=\, \binom{\ell}{k} q_{\ell,k}\,.$$

In other words, the infinitesimal generator $\mathtt{G}^{[n]}$ of $\#\Pi_{|[n]}$ is specified by

$$\mathtt{G}^{[n]} f(\ell) \,=\, \sum_{k=2}^{\ell}\binom{\ell}{k} q_{\ell,k}\,(f(\ell-k+1) - f(\ell))\,, \qquad \ell \in [n]\,.$$

Now assume that the series $\sum_{b=2}^{\infty} 1/\varphi(b)$ converges (since otherwise there is nothing to prove), and define

$$f(\ell) \,:=\, \sum_{k=\ell+1}^{\infty} 1/\varphi(k)\,, \qquad \ell \geq 1\,.$$

Recall Lemma 4.1. Since $1/\varphi$ decreases, we have for $2 \leq k \leq \ell$

$$f(\ell-k+1) - f(\ell) \geq (k-1)/\varphi(\ell)$$

and therefore

$$\mathtt{G}^{[n]}f(\ell) \;\geq\; \frac{1}{\varphi(\ell)}\sum_{k=2}^{\ell}(k-1)\binom{\ell}{k}q_{\ell,k} \;=\; 1\,.$$

The process

$$f(\#\Pi_{|[n]}(t)) - \int_0^t \mathtt{G}^{[n]}f(\#\Pi_{|[n]}(s))ds\,, \qquad t\geq 0$$

is a martingale, and an application of the optional sampling theorem at the absorption time

$$\zeta_n := \inf\left\{t\geq 0 : \#\Pi_{|[n]}(t)) = 1\right\}\,,$$

yields the bound

$$\mathbb{E}(\zeta_n) \;\leq\; \mathbb{E}\left(\int_0^{\zeta_n} \mathtt{G}^{[n]}f(\#\Pi_{|[n]}(s))ds\right) \;=\; f(1)-f(n)\,.$$

Plainly, the sequence $(\zeta_n, n\in\mathbb{N})$ increases and $\lim_{n\to\infty}\zeta_n := \zeta_\infty$ is the time of entire coalescence for Π. Further, the latter obviously has the same distribution as the fixation time ζ of the population model, which proves that $\mathbb{E}(\zeta)\leq f(1)$. □

4.2 The look-down construction

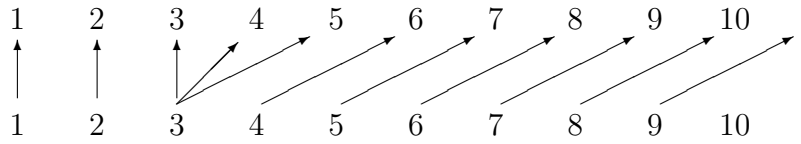
In this section we sketch the presentation of a clever and useful construction due to Donnelly and Kurtz [26, 27] of a population model whose genealogy can then be interpreted in terms of an exchangeable coalescent. We consider a measure μ on $\mathcal{P}_\infty$ which fulfills (3.6) and (3.7), and thus may serve as coagulation rates according to Lemma 3.3. To stick to the framework of this chapter, we may further impose that the coagulation rates are simple, i.e. $\mu = \mathtt{c}\mu^{\mathrm{K}} + \varrho_\nu$ where $\mathtt{c}\geq 0$ is the coefficient of binary coagulations and ν a measure ν on $(0,1]$ with $\int x^2\nu(\mathrm{d}x)<\infty$, though the general case does not yield additional difficulties.

Let us start by specifying the Poissonian construction of coalescents with coagulation rate μ which was presented in Section 3.3. We consider a Poisson point measure $\hat{\mathcal{M}}$ on $\mathbb{R}\times\mathcal{P}_\infty$ with intensity $\mathrm{d}t\otimes\mu(\mathrm{d}\pi)$. Each atom of $\hat{\mathcal{M}}$ corresponds to a coagulation event. More precisely if (t,π) is such an atom and $\Pi(t-)$ the state of the coalescent immediately before time t, then $\Pi(t) = \mathrm{Coag}(\Pi(t-),\pi)$.

Donnelly and Kurtz [26] introduced a population model which evolves in continuous time and whose genealogy is precisely described by a simple exchangeable coalescent with coagulation rate μ. The individuals in the population are identified with integers, it may be convenient to think that each individual has a certain type (for instance the space of types can be the unit interval) which is transmitted to its children in the case of birth events, so that the genealogy can be observed by tracing back types of individuals. As previously we consider a Poisson random measure $\mathcal{M}$ on $\mathbb{R}\times\mathcal{P}_\infty$ with intensity $\mathrm{d}t\otimes\mu(\mathrm{d}\pi)$. It is important to observe that the law of Poisson random measures with such intensity

is preserved by time-reversal $t \to -t$; in this direction $\mathcal{M}$ should be viewed as the image by time reversal of the random measure $\hat{\mathcal{M}}$ used above to construct the coalescent Π.

Each atom of $\mathcal{M}$ now corresponds to a birth event for the population. More precisely if (t, π) is such an atom with π a simple partition which is not the partition into singleton. Let $j_0 < j_1 < \cdots$ the elements of the unique block of π which is not a singleton (either this is a doubleton or an infinite set). We decide that the individuals $j_1 < j_2 < \cdots$ at time t are all the children of the individual j_0 at time $t-$, and that all the other individuals are shifted upwards accordingly, keeping the order immediately before the birth event. For instance, if $j_0 = 3, j_1 = 4, j_2 = 5, j_3 = 16$, ... then the filiation lineage of the birth event at time t can be represented as follows (parents are listed on the bottom line and arrows represent the filiation):



Note that, by the law of large numbers, a proportion x of the population at time t descends from the same parent at generation $t-$.

We stress that the filiation lineages are non-decreasing and if we look backwards in time to recover the genealogy of an individual, the lineage will only be perturbed by finitely many birth events. Indeed the set of instants at which at least one of the first k filiation lineages jumps upwards corresponds to an atom (t, π) such that the restriction of π to $[k]$ differs from the partition into singletons and thus is discrete a.s. thanks to (3.6). So the population model is well-defined. A short moment of reflexion confirms that its genealogy is precisely given by the Poissonian construction of the simple exchangeable coalescent with coagulation rate μ (more precisely, consider the restriction of the population at every generation to the first n individuals). Plainly, it is easy to extend this construction for arbitrary exchangeable coalescents; see [17].

Donnelly and Kurtz [26] pointed at the key fact that if at the initial time genetic types are assigned to individuals in an exchangeable way (i.e. the sequence of the types of the ancestors is exchangeable), and if types are transmitted from parent to children, then the sequence of the types of individuals at any fixed time is still exchangeable. In particular, the partition of the population according to types possesses asymptotic frequencies a.s. Roughly speaking, one gets a Markovian measure-valued process by considering the empirical distribution of the types, which describes the evolution of a population whose genealogy corresponds to the coalescent Π.

4.3 Compositions of bridges, stochastic flows and coalescence

In this section, which is largely excerpt from [11], we shall dwell on a fundamental connexion between coalescence and compositions of independent bridges, which will point at a construction of simple coalescents via a certain stochastic flow on the unit interval. In the next section, time-reversing the flow will then yield another construction of the dual population model.

It is convenient here to think of a mass-partition $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ as the ranked sequence of the masses assigned to atoms by some random probability measure. More precisely, we work on the unit interval $[0,1]$, throw the atoms at random according to the uniform distribution, independently of each other, and let the possible dust be uniformly spread. In other words, let us introduce a sequence $V_1, \ldots$ of i.i.d. uniformly distributed variables on $[0,1]$ and consider the random probability measure

$$b_{\mathbf{s}}(\mathrm{d}x) \;=\; s_0 \mathrm{d}x + \sum_{i=1}^{\infty} s_i \delta_{V_i}(\mathrm{d}x)\,, \qquad x \in [0,1]\,,$$

where δ_a stands for the Dirac point mass at a. The distribution function

$$x \to b_{\mathbf{s}}(x) = b_{\mathbf{s}}([0,x])$$

is a right-continuous increasing function on $[0,1]$ with $b_{\mathbf{s}}(0) = 0$ and $b_{\mathbf{s}}(1) = 1$; the ranked sequence of jump sizes of $b_{\mathbf{s}}$ is given by $\mathbf{s}$, and the jump locations by the i.i.d. uniform variables.

We record this in the following definition.

Definition 4.2 *Let $\mathbf{s} \in \mathcal{P}_{\mathtt{m}}$ be fixed, and consider a sequence $V_1, \ldots$ of i.i.d. uniformly distributed variables on $[0,1]$. We call* **s-bridge** *any process distributed as*

$$b_{\mathbf{s}}(x) \;:=\; s_0 x + \sum_{i=1}^{\infty} s_i \mathbb{1}_{\{V_i \le x\}}\,, \qquad x \in [0,1]\,,$$

where $s_0 = 1 - \sum_{i=1}^{\infty} s_i$. We refer to $\mathbf{s}$ as the jump-sizes of $b_{\mathbf{s}}$. Finally, a process distributed as a mixture of $\mathbf{s}$-bridge (i.e. when one randomizes the mass-partition) is just called a bridge.

Plainly, $b_{\mathbf{s}}(0) = 0$, $b_{\mathbf{s}}(1) = 1$, and the random measure $\mathrm{d}b_{\mathbf{s}}(x)$ is exchangeable, in the sense that its image by any injective map $[0,1] \to [0,1]$ that preserves the Lebesgue measure has the same distribution as $\mathrm{d}b_{\mathbf{s}}(x)$. It follows in particular that the bridge $b_{\mathbf{s}}$ has exchangeable increments, that is for every $n \in \mathbb{N}$, the law of the n-tuple

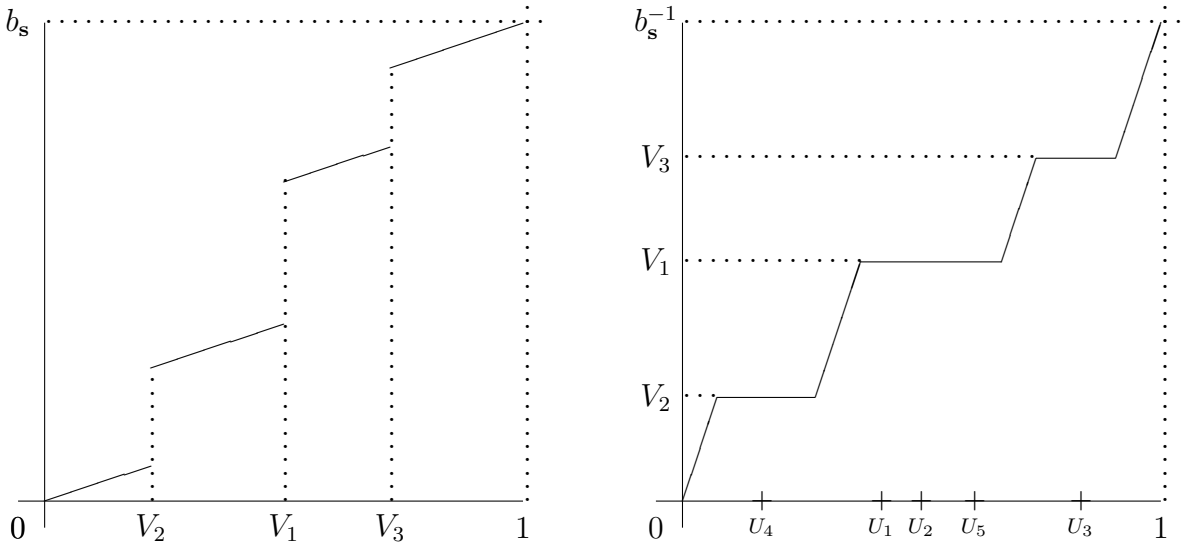
$$(b_{\mathbf{s}}(1/n), b_{\mathbf{s}}(2/n) - b_{\mathbf{s}}(1/n), \ldots, b_{\mathbf{s}}(1) - b_{\mathbf{s}}((n-1)/n))$$

is invariant under permutation. We refer to Kallenberg [37] for fundamental properties of bridges with exchangeable increments.

We will use the notation

$$b_{\mathbf{s}}^{-1}(r) \;:=\; \inf\{v \in [0,1] : b_{\mathbf{s}}(v) > r\}\;, \qquad r \in]0,1[\,,$$

for the right-continuous inverse of a bridge $b_{\mathbf{s}}$. In other words, $b_{\mathbf{s}}^{-1}$ is the quantile function for the distribution function $b_{\mathbf{s}}$; note that $b_{\mathbf{s}}^{-1}$ is continuous if and only if $b_{\mathbf{s}}$ is strictly increasing, which in turn is equivalent to either $s_0 = 1 - \sum_1^\infty s_i > 0$ or $s_i > 0$ for all $i \geq 1$.



s-*bridge* $b_{\mathbf{s}}$*, its inverse* $b_{\mathbf{s}}^{-1}$ *and the associated paint-box* $\pi = (\{1,2,5\},\{3\},\{4\})$

Observe that when U is an independent uniform variable, the conditional law of $b_{\mathbf{s}}^{-1}(U)$ given $b_{\mathbf{s}}$ is $\mathrm{d}b_{\mathbf{s}}$. Kingman's paint-box construction can be conveniently rephrased using the inverse of bridges. Specifically, introduce the usual sequence $U_1, \ldots$ of i.i.d. uniform variables, which we assume to be independent of the bridge $b_{\mathbf{s}}$, and consider the equivalence relation

$$i \overset{\pi}{\sim} j \iff b_{\mathbf{s}}^{-1}(U_i) \;=\; b_{\mathbf{s}}^{-1}(U_j)\,.$$

As conditionally on the bridge $b_{\mathbf{s}}$, $b_{\mathbf{s}}^{-1}(U_1), \ldots$ is a sequence of i.i.d. variables with distribution $b_{\mathbf{s}}$, π is clearly a paint-box with ranked asymptotic frequencies **s**.

In the case of a simple mass-partition $\mathbf{s} = (x, 0, \ldots)$, it is convenient to use the notation $b_x = b_{(x,0,\ldots)}$, so that

$$b_x(r) \;=\; (1-x)r + x\mathbb{1}_{\{V \leq r\}}\,, \qquad r \in [0,1]\,, \tag{4.4}$$

where V is a uniform random variable on $[0,1]$. We then call b_x a *simple bridge*. Plainly, the paint-box π constructed above is simple a.s. if and only if the bridge b is simple.

We shall now establish a few elementary properties of simple bridges which are easy to check and will suffice for our purposes; we mention however, that they remain valid in

fact for arbitrary bridges (cf. [11]). Suppose now that $x < 1$; the continuous inverse of b_x is thus given by

$$b_x^{-1}(r) \;=\; \begin{cases} r/(1-x) & \text{if } \; r \in [0,(1-x)V]\,, \\ V & \text{if } \; r \in](1-x)V, x+(1-x)V[\,, \\ (r-x)/(1-x) & \text{if } \; r \in [(1-x)V+x, 1]\,. \end{cases} \tag{4.5}$$

We make the following key observation.

Lemma 4.2 *Define for every $i \in \mathbb{N}$*

$$U'_i \;:=\; b_x^{-1}(U_j)\,, \qquad j \in \pi_i\,.$$

Then $U'_1, \ldots$ form a sequence of i.i.d. uniform variables which is independent of the exchangeable partition π.

Proof We work first conditionally on $V = v \in]0,1[$ and consider the index of the first uniform variable U_i which falls into the flat part of b_x^{-1},

$$i_0 = \min\{i : (1-x)v < U_i < x + (1-x)V\}\,.$$

The i_0-th block π_{i_0} of the paintbox π is the unique block which is not reduced to a singleton. Condition further on π, so the variables U_j for $j \notin \pi_0$ form an i.i.d. sequence of variables which have the uniform distribution on $]0,(1-x)v[\cup]x+(1-x)v,1[$. Their images by the inverse bridge b_x^{-1} is thus a sequence of i.i.d. uniform variables on $]0,1[$. Because $U'_{i_0} = v$ and V has also the uniform distribution, we conclude that conditionally on the paintbox π, the variables U'_i form a sequence of i.i.d. uniform variables. $\square$

Lemma 4.2 suggests the use the i.i.d. uniform variables $U'_1, \ldots$ as the basis for an independent paint-box. More precisely, let $x' \in]0,1[$, and consider an independent simple bridge $b_{x'}$, that is

$$b_{x'}(r) \;=\; (1-x')r + x' \mathbb{1}_{\{V' \le r\}}\,, \qquad r \in [0,1]\,,$$

where V' is a uniform variable which is independent of V. We write π' for the paint-box defined from the inverse bridge $b_{x'}^{-1}$ and the variables U'_i as above, so that

$$i \overset{\pi'}{\sim} j \iff b_{x'}^{-1}(U'_i) \;=\; b_{x'}^{-1}(U'_j)\,.$$

Note from Lemma 4.2 that the random partitions π and π' are independent.

Corollary 4.1 *Consider the composition of bridges $b^\circ := b_x \circ b_{x'}$ and its continuous inverse $b^{\circ -1} = b_{x'}^{-1} \circ b_x^{-1}$.*

(i) *The random partition π° defined by*

$$i \overset{\pi^\circ}{\sim} j \iff b^{\circ -1}(U_i) \;=\; b^{\circ -1}(U_j)\,,$$

coincides with the coagulation $\mathrm{Coag}(\pi, \pi')$ *of* π *by* π'.

(ii) *If we define*

$$U_i^\circ \,=\, b^{\circ -1}(U_j)\,, \qquad j \in \pi_i^\circ\,,$$

then $U_1^\circ, \ldots$ *is a sequence of i.i.d. variables which is independent of* π, π', *and a fortiori of* π°.

(iii) b° *is a bridge in the sense of Definition 4.2. More precisely,* b° *is an* $\mathbf{s}$*-bridge, where* $\mathbf{s}$ *stands for the ranked sequence of the asymptotic frequencies of* π°.

Proof (i) By definition, i and j belong to the same block of π° if and only if

$$b_{x'}^{-1} \circ b_x^{-1}(U_i) \,=\, b_{x'}^{-1} \circ b_x^{-1}(U_j)\,.$$

Let k and ℓ be the respective indices of the blocks of π which contain i and j, that is $i \in \pi_k$ and $j \in \pi_\ell$. Then $b_x^{-1}(U_i) = U'_k$ and $b_x^{-1}(U_j) = U'_\ell$, and we see that i and j belong to the same block of π° if and only if $b_{x'}^{-1}(U'_k) = b_{x'}^{-1}(U'_\ell)$, that is if and only if k and ℓ belong to the same block of π'. This shows that $\pi^\circ = \mathrm{Coag}(\pi, \pi')$.

(ii) Recall from Lemma 4.2 that π is independent of the variables U'_i and $b_{x'}$. On the other hand, observe that for every $i \in \mathbb{N}$, there is the identity

$$U_i^\circ \,=\, b_{x'}^{-1}(U'_j)\,, \qquad j \in \pi'_i\,.$$

Another application of Lemma 4.2 now shows that the U_i° form an i.i.d. sequence of uniform variables which is independent of π and π'.

(iii) The jump locations of b° are necessarily of the type U_i° for some $i \in \mathbb{N}$. More precisely, U_i° is the location of a jump of b° if and only if the i-th block π_i° is not reduced to a singleton and, in this case, the size of the jump coincides with the asymptotic frequency $|\pi_i^\circ|$ of π_i°. As we know from (ii) that the variables U_i° are independent of π°, this shows that b° is an $\mathbf{s}$-bridge with $\mathbf{s} = |\pi^\circ|^\downarrow$. □

We stress that the argument used for proving Corollary 5.2 can be iterated to deal with compositions of finitely many independent simple bridges. More precisely, if $b_{x_1}, \ldots, b_{x_n}$ are independent simple bridges, and if we set $b := b_{x_1} \circ b_{x_2} \circ \cdots \circ b_{x_n}$, then Corollary 5.2 still holds when we replace b_x by b. We shall often use this straightforward extension of Corollary 5.2 later in the text.

We now apply the preceding observations to dwell on a natural representation of simple exchangeable coalescents in terms of flows of bridges. We first consider the elementary case when the coefficient of binary coagulation is $\mathsf{c} = 0$ and the measure of multiple coagulations ν is a finite measure on $]0,1[$. The Poissonian construction of Section 4.2.3 leads us to introduce a Poisson random measure on $\mathbb{R}\times]0,1[$ with intensity $dt \otimes \nu(\mathrm{d}x)$. Next, conditionally on the Poisson measure, we associate to each atom (t,x) a bridge, denoted by $b^{(t)}$, which is distributed as b_x and such that to different atoms correspond independent bridges. The assumption that ν is finite ensures that for every $t < t'$, there are only finitely many atoms on $]t,t']\times]0,1[$. More precisely, we write $(t_1,x_1), \ldots, (t_N,x_N)$ for these atoms, where N is a Poisson variable with intensity $(t'-t)\nu(]0,1[)$ and

$$t < t_1 < \cdots < t_N \leq t'\,.$$

This enables us to define a random function $B_{t,t'} : [0,1] \to [0,1]$ by

$$B_{t,t'} \;=\; b^{(t_1)} \circ \cdots \circ b^{(t_N)} \,.$$

By conditioning on the number N of atoms of the Poisson measure on $]t,t'] \times]0,1]$ and by applying the extension of Corollary 5.2, we see that $B_{t,t'}$ is a bridge in the sense of Definition 4.2.

Let $U_1, \ldots$ be a sequence of i.i.d. uniform variables which is independent of the family $(B_{t,t'}\,, -\infty < t \leq t' < \infty)$. For every $t \geq 0$, define an exchangeable random partition $\Pi(t)$ by

$$i \overset{\Pi(t)}{\sim} j \iff B_{0,t}^{-1}(U_i) \;=\; B_{0,t}^{-1}(U_j)\,.$$

Lemma 4.3 *The process $(\Pi(t), t \geq 0)$ constructed above is a standard simple exchangeable coalescent with coagulation rate $\mu = \varrho_\nu$.*

Proof This follows immediately from the Poissonian construction of an exchangeable coalescent (cf. Proposition 3.3) and Corollary 5.2. □

Lemma 4.3 provides a nice representation of simple exchangeable coalescents in terms of the composition of simple bridges, but this representation only concerns the rather elementary case when the coagulation rate is finite. We shall now see that this restriction can be removed using an approximation scheme. Specifically, let $\mathsf{c} \geq 0$ and ν be an arbitrary measure on $\mathcal{P}_{\mathbb{m}}$ which fulfills (3.8). Suppose that ν is supported by simple mass-partitions, so it can be identified with a measure on $[0,1]$ which fulfills (4.1). We may then find a sequence $(\nu^{(n)}, n \in \mathbb{N})$ of finite measures on $[0,1]$ such that

$$x^2\nu^{(n)}(\mathrm{d}x) \text{ converges weakly as } n \to \infty \text{ to } c\delta_0 + x^2\nu(\mathrm{d}x)\,. \tag{4.6}$$

For each $n \in \mathbb{N}$, let us denote by

$$B^{(n)} \;=\; (B^{(n)}_{t,t'}\,, -\infty < t \leq t' < \infty)$$

a family of bridges constructed as above from the finite measure $\nu^{(n)}$. Recall also that we consider bridges as random variables with values in the space $L^2([0,1], dr)$.

Theorem 4.1 *In the notation above, the following holds:*
(i) *As $n \to \infty$, $(B^{(n)}_{t,t'}, -\infty < t \leq t' < \infty)$ converges in the sense of weak convergence of finite-dimensional distributions to, say, $(B_{t,t'}\,, -\infty < t \leq t' < \infty)$, which fulfills the following properties:*

(i.a) *For every $t \leq t' \leq t''$, $B_{t,t''} = B_{t,t'} \circ B_{t',t''}$ a.s.*

(i.b) *$B_{t,t'}$ is a bridge and its law only depends on $t' - t$.*

(i.c) *If $t'_1 < t'_2 < \cdots < t'_n$, $B_{t'_1,t'_2}, B_{t'_2,t'_3}, \ldots, B_{t'_{n-1},t'_n}$ are independent.*

(i.d) *$B_{0,0} = \mathrm{Id}$ and $\lim_{t\to 0+} B_{0,t}(r) = r$ in $L^2(dr \otimes d\mathbb{P})$.*

(ii) *For every $t \geq 0$, denote the right-continuous inverse of $B_{0,t}$ by*

$$B_{t,t'}^{-1}(r) := \inf\{v \in [0,1] : B_{t,t'}(v) > r\}\,, \qquad r \in]0,1[\,.$$

Let $U_1, \ldots$ be a sequence of i.i.d. uniform variables which is independent of $(B_{0,t}\,, t \geq 0)$, and define for every $t \geq 0$ the random partition $\Pi(t)$ by

$$i \overset{\Pi(t)}{\sim} j \iff B_{0,t}^{-1}(U_i) = B_{0,t}^{-1}(U_j)\,.$$

Then $(\Pi(t), t \geq 0)$ is a standard simple exchangeable coalescent with coagulation rate $\mu = \mathsf{c}\mu^{\mathsf{K}} + \varrho_\nu$.

(iii) *Furthermore, for each fixed $t \geq 0$, if we define for every $i \in \mathbb{N}$*

$$U_i' := B_{0,t}^{-1}(U_j)\,, \qquad j \in \Pi_i(t)\,,$$

then $U_1', \ldots$ form a sequence of i.i.d. uniform variables which is independent of $\Pi_i(t)$.

As a reference to the properties (i.a–d), we say that the family $(B_{t,t'}\,, -\infty < t \leq t' < \infty)$ is a *flow of bridges* on $[0,1]$. More generally, one can show that there is a bijective correspondence between on the one hand the laws of flows of bridges, and on the other hand the laws of standard exchangeable coalescents; see Bertoin and Le Gall [11].

The proof of Theorem 4.1 follows easily from Lemma 4.3 as soon as we take for granted the following useful result of continuity in distribution that completes Proposition 1.4.

Lemma 4.4 *Consider for each $n \in \overline{\mathbb{N}}$ a random variable $S^{(n)}$ with values in $\mathcal{P}_{\mathtt{m}}$. Let $b^{(n)}$ be a bridge with jump-sizes $S^{(n)}$. The following conditions are equivalent:*

(i) *When $n \to \infty$, $S^{(n)}$ converges in distribution to $S^{(\infty)}$.*

(ii) *When $n \to \infty$, $b^{(n)}$ converges in distribution to $b^{(\infty)}$.*

4.4 Generalized Fleming-Viot processes and duality

In this section, we shall see that the flow of bridges which is used in Theorem 4.1 to construct a simple exchangeable coalescent, can be interpreted in terms of a natural population model. Throughout this section, $(B_{t,t'}, -\infty < t \leq t' < \infty)$ denotes a flow of bridges associated to some simple exchangeable coalescent Π; we also implicitly assume that the coagulation rate $\mu \neq 0$ to avoid the useless discussion of a trivial case.

As we explained in Section 4.1.1, coagulations arise when one studies the genealogy of populations, and for this purpose, one has to work as time goes *backwards*. Therefore, a population model based on a flow of bridges should rather be defined via the *dual* flow, namely

$$\hat{B}_{t,t'} := B_{-t',-t}\,, \qquad -\infty < t \leq t' < \infty\,.$$

Plainly, $\hat{B}_{t,t'}$ is a bridge whose law only depends on $t'-t$, and which converges in probability to Id when $t'-t \to 0$. Furthermore the bridges $\hat{B}_{t_1,t_2}, \ldots, \hat{B}_{t_{n-1},t_n}$ are independent for every $t_1 < \cdots < t_n$, and the following cocycle property holds for every $t < t' < t''$:

$$\hat{B}_{t',t''} \circ \hat{B}_{t,t'} \;=\; \hat{B}_{t,t''}\,.$$

We write ρ_t for the probability measure with distribution function $\hat{B}_{0,t}$, that is

$$\rho_t(\mathrm{d}y) \;=\; d\hat{B}_{0,t}(y)\,, \qquad 0 \le y \le 1.$$

We immediately derive from the cocycle property of the dual flow that $(\rho_t, t \ge 0)$ is a Markov process with values in the space $\mathcal{M}_1$ of the probability measures on $[0,1]$. Recall that $\mathcal{M}_1$ is a compact metric space when endowed with Prohorov's distance. Further, it follows readily from the fact that $B_{0,t}$ has no fixed discontinuities, that $(\rho_t, t \ge 0)$ is in fact a Feller process, and in particular it possesses a càdlàg modification. From now on, we implicitly deal with this càdlàg version. Note also that $\rho_0(\mathrm{d}y) = \mathrm{d}y$ is the Lebesgue measure on $[0,1]$, and if ρ^μ denotes the process ρ started at $\mu \in \mathcal{M}_1$, then ρ^μ can be constructed from the special case $\mu =$ Lebesgue measure on $[0,1]$ by the explicit formula

$$\rho_t^\mu([0,x]) = \rho_t([0,\mu([0,x])])\,, \quad x \in [0,1]\,,\ t \ge 0.$$

The process $(\rho_t, t \ge 0)$ can be interpreted as a population model: we may think of $\rho_t(dr)$ as the size of the progeny at time t of the fraction dr of the initial population. Consider for simplicity the case of when the coagulation rate is finite, that is $\mu = \varrho_\nu$ where ν is a finite measure on $]0,1[$. Recall the discrete Poissonian construction of the flow of bridges that was presented at the beginning of Section 4.4.2. We see that the process $(\rho_t, t \ge 0)$ (or equivalently, $(\hat{B}_{0,t}, t \ge 0)$) is a continuous time Markov chain, and that the jump times of this chain are given by a Poisson process with intensity $\nu(]0,1[)$. More precisely, if t_n is the instant of the n-th jump, then

$$\hat{B}_{0,t_n} \;=\; b_X \circ \hat{B}_{0,t_{n-1}}\,,$$

where b_X is a simple bridge which is independent of $B_{0,t_{n-1}}$, such that its jump size X is a random variable with distribution $\nu(\cdot)/\nu(]0,1[)$, and its jump location U an independent uniform variable. This means that

$$\rho_{t_n} \;=\; (1-X)\rho_{t_{n-1}} + X\delta_Y\,,$$

where conditionally on $\rho_{t_{n-1}}$, X and $Y := \hat{B}^{-1}_{0,t_{n-1}}(U)$ are independent random variables, with Y distributed according to $\rho_{t_{n-1}}$. In terms of the evolution of the population, this means that an individual picked at random in the population at time t_{n-1} generates a proportion X of the population at time t_n. The rest of the population at time t_{n-1} is reduced by a factor $(1-X)$ so that, at time t_n, the total size of the population is still 1. This description bears obvious similarities with that for the evolution of the Moran and the Fleming-Viot processes; see for example Chapter 1 of Etheridge [29].

We can now interpret the coalescent in terms of the genealogy of this population model. More precisely, fix some time $T > 0$, and consider the population at time T,

which is identified as $[0,1]$. Pick a sequence of individuals uniformly at random, that is consider a sequence $U_1, \ldots$ of i.i.d. uniform variables which is independent of the flow $(B_{t,t'}, 0 \leq t \leq t' \leq T)$. Two individuals i and j have the same ancestor $r \in]0,1[$ at the generation $T-t$ if and only if U_i and U_j both belong to the interval $]\hat{B}_{T-t,T}(r-), \hat{B}_{T-t,T}(r)[$. In other words, for each $t \in [0,T]$, we may consider the partition $\Pi(t)$ of $\mathbb{N}$ defined by

$$i \overset{\Pi(t)}{\sim} j \iff \hat{B}^{-1}_{T-t,T}(U'_i) = \hat{B}^{-1}_{T-t,T}(U'_j)\,;$$

where the blocks of the partition consist of the families of individuals which have the same ancestor at the generation $T-t$. Lemma 4.3 shows that $(\Pi(t), 0 \leq t \leq T)$ is a simple exchangeable coalescent with coagulation rate $\mu = \varrho_\nu$. Further, we can use Theorem 4.1 to extend this to situations where ν is an infinite measure on $]0,1]$ with $\int_{]0,1]} x^2 \nu(\mathrm{d}x) < \infty$.

We shall now dwell on the duality between the generalized Fleming-Viot processes and simple exchangeable coalescents. In this direction, we will consider functionals on $\mathcal{M}_1$ of the following special type. Let $p \geq 1$ be an integer and let f be a continuous function on $[0,1]^p$. For every $m \in \mathcal{M}_1$, we set

$$G_f(m) = \int_{[0,1]^p} m(\mathrm{d}x_1) \ldots m(\mathrm{d}x_p)\, f(x_1, \ldots, x_p).$$

Note for instance that when $f(x_1, \ldots, x_p) = g(x_1) \cdots g(x_p)$ for some continuous function $g : [0,1] \to \mathbb{R}$, then $G_f(m) = \langle m, g \rangle^p$. More generally, we denote by Φ_f the function defined on $\mathcal{M}_1 \times \mathcal{P}_p$ by

$$\Phi_f(m, \pi) = \int_{[0,1]^{\#\pi}} m(\mathrm{d}x_1) \ldots m(\mathrm{d}x_{\#\pi})\, f(Y(\pi; x_1, \ldots, x_{|\pi|})),$$

where, if π has $q = \#\pi$ blocks $A_1, \ldots, A_q$ listed in the order of their least elements, we put $Y(\pi; x_1, \ldots, x_q) = (y_1, \ldots, y_p)$, with $y_j = x_i$ if and only if $j \in A_i$. In particular $G_f(m) = \Phi_f(m, \mathbf{0}_p)$ when $\pi = \mathbf{0}_p$ is the partition into singletons. Note also that if $f(x_1, \ldots, x_p) = g(x_1) \cdots g(x_p)$, then $\Phi_f(m, \pi) = \langle m, g^{\#\pi_1} \rangle \cdots \langle m, g^{\#\pi_q} \rangle$ where $\#\pi_i$ is the cardinal of the i-th block π_i of π.

Lemma 4.5 (duality lemma) *For every $m \in \mathcal{M}_1$, we have*

$$\mathbb{E}_m\left(G_f(\rho_t)\right) = \mathbb{E}_m\left(\Phi_f(\rho_t, \mathbf{0}_p)\right) = \mathbb{E}\left(\Phi_f(m, \Pi_{|[p]}(t))\right),$$

where the notation $\mathbb{E}_m$ refers to expectation when the generalized Fleming-Viot process $\rho_\cdot$ starts from $\rho_0 = m$.

Proof Consider first the case when $m = \rho_0$ is the Lebesgue measure. Then $\Phi_f(\rho_t, \mathbf{0}_p)$ has the same law as

$$\int_{[0,1]^p} \mathrm{d}B_{0,t}(x_1) \cdots \mathrm{d}B_{0,t}(x_p) f(x_1, \ldots, x_p) = \int_{[0,1]^p} \mathrm{d}x_1 \cdots \mathrm{d}x_p f(B^{-1}_{0,t}(x_1), \ldots, B^{-1}_{0,t}(x_p))\,.$$

In particular, taking expectations, we get

$$\mathbb{E}\left(\Phi_f(\rho_t, \mathbf{0}_p)\right) = \mathbb{E}\left(f(B^{-1}_{0,t}(V_1), \ldots, B^{-1}_{0,t}(V_p))\right),$$

where $V_1, \ldots, V_p$ are i.i.d. uniform $[0,1]$ variables which are independent of the bridge $B_{0,t}$. In the notation introduced above the statement, we thus have

$$\mathbb{E}\left(\Phi_f(\rho_t, \mathbf{0}_p)\right) \;=\; \mathbb{E}\left(f\left(Y(\Pi_{|[p]}(t); V_1', \ldots, V'_{|\Pi_{|[p]}(t)|})\right)\right)\,, \tag{4.7}$$

where $\Pi_{|[p]}(t)$ stands for the restriction of the partition $\Pi(t)$ to $[p]$. Recall from Theorem 4.1(iii) that $\Pi(t)$ and $(V_1', \ldots)$ are independent and that the latter is a sequence of i.i.d. uniform $[0,1]$ variables. We now see that the right-hand side in (4.7) coincides with

$$\mathbb{E}\left(\Phi_f(\rho_0, \Pi_{|[p]}(t))\right)\,.$$

This establishes the claim in the case when $m = \rho_0$ is the Lebesgue measure, and the general case follows immediately using the special form of the semi-group of generalized Fleming-Viot processes. Indeed, let $b : [0,1] \to [0,1]$ be the distribution function of an arbitrary probability measure on $[0,1]$, so that the distribution function of ρ_t started from $\rho_0(\mathrm{d}x) = \mathrm{d}b(x)$ has the same law as $B_{0,t}(b(\cdot))$. Then $\Phi_f(\rho_t, \mathbf{0}_p)$ has the same law as

$$\begin{aligned} &\int_{[0,1]^p} \mathrm{d}B_{0,t}(b(x_1)) \cdots \mathrm{d}B_{0,t}(b(x_p)) f(x_1, \ldots, x_p) \\ =\; &\int_{[0,1]^p} \mathrm{d}B_{0,t}(y_1) \cdots \mathrm{d}B_{0,t}(y_p) f(b^{-1}(y_1), \ldots, b^{-1}(y_p))\,. \end{aligned}$$

Introducing the function $g(y_1, \ldots, y_p) = f(b^{-1}(y_1), \ldots, b^{-1}(y_p))$, this reduces the calculation to that in the case when ρ_0 is the Lebesgue measure. □

It follows immediately from the duality lemma and Kolmogorov's equation that the infinitesimal generators L of the generalized Fleming-Viot process ρ and L^* of the simple exchangeable coalescent Π are connected by the simple identity

$$L\Phi_f(\cdot, \pi)(m) = L^*\Phi_f(m, \cdot)(\pi), \tag{4.8}$$

which in turn implies the following characterization of the law of the generalized Fleming-Viot process by martingale problem (we refer to [11] for details).

Proposition 4.2 *The law of the process $(\rho_t, t \geq 0)$ is characterized by the following martingale problem. For every integer $p \geq 1$ and every continuous function f on $[0,1]^p$,*

$$G_f(\rho_t) - \int_0^t ds\, LG_f(\rho_s)$$

is a martingale.

Proposition 4.2 is the key to several interesting results on generalized Fleming-Viot processes and simple coalescent; in particular it open the path to the use of stochastic calculus for the study of the latter; we refer to [12, 13] for much more on this topic.

One upshot of the identity (4.8) is that the infinitesimal generator L^* of a exchangeable coalescent is know explicitly as $\Pi_{|[p]}$ is a continuous time Markov chain. This readily yields an explicit expression for the infinitesimal generator L of the generalized Fleming-Viot process that we should have expected from the description of its dynamics.

Theorem 4.2 *Assume that the coefficient of binary coagulations is* $\mathsf{c} = 0$ *and recall that* ν *denotes the measure of multiple coagulations.*

(i) *For every integer* $p \geq 1$ *and every continuous function* f *on* $[0,1]^p$*, we have*

$$LG_f(m) = \int \nu(\mathrm{d}x) \int m(\mathrm{d}a) \Big(G_f\left((1-x)m + x\delta_a\right) - G_f\left(m\right) \Big) .$$

(ii) *Let* $\mathcal{D}$ *stand for the domain of the infinitesimal generator of* L*. Linear combinations of pairs* (G_f, LG_f) *are everywhere dense in* $\{(G, LG) : G \in \mathcal{D}\}$ *in the sense of uniform convergence. In other words, the vector space generated by functionals of the type* G_f *forms a core of* $(L, \mathcal{D})$*.*

Proof (i) We know from (4.8) that

$$LG_f(m) = L^* \Phi_f(m, \cdot)(\mathbf{0}_p)$$

where $\mathbf{0}_p$ is the partition of $[p]$ into singletons, and from Sections 3.4 and 3.5 that the coefficient of binary coagulations is $\mathsf{c} = 0$,

$$L^* \Phi_f(m, \cdot)(\mathbf{0}_p) = \int_{]0,1]} \nu(\mathrm{d}x) \int_{\mathcal{P}_p} \varrho_x(\mathrm{d}\pi) \left(\Phi_f(m, \pi) - \Phi_f(m, \mathbf{0}_p) \right) \tag{4.9}$$

where ϱ_x stands for the law of the paintbox associated to the simple mass-partition $(x, 0, \ldots)$. The same calculation as in the proof of the duality lemma 4.5 shows that

$$\int_{\mathcal{P}_p} \varrho_x(\mathrm{d}\pi) \Phi_f(m, \pi) = \mathbb{E}\left(\Phi_f(m_x, \mathbf{0}_p) \right) ,$$

where m_x denotes a random probability measure on $[0,1]$ given by $m_x(\mathrm{d}u) = \mathrm{d}\left(b_x \circ M(u)\right)$, with b_x an x-bridge in the sense of (4.4) and M the distribution function of m (i.e. $m(\mathrm{d}u) = \mathrm{d}M(u)$). In other words, $m_x = (1-x)m + x\delta_a$ where a is random with law m. Putting the pieces together, this yields the stated formula for $LG_f(m)$.

(ii) Identities (4.8) and (4.9) show that $LG_f(m)$ is a linear combination of functionals of the type $\Phi_f(m, \pi)$ with π partition of $[p]$. Observe that any $\Phi_f(m, \pi)$ can also be expressed in the form G_h for some adequate continuous function $h : [0,1]^p \to \mathbb{R}$, so the vector space generated by functionals of the type G_f is left invariant for the generator L.

On the other hand, recall that for $f(x_1, \ldots, x_p) = g(x_1) \cdots g(x_p)$, we have $G_f(m) = \langle m, g \rangle^p$, and as a consequence of the Stone-Weierstrass theorem, the space of linear combinations of such functionals is dense in the space of continuous functions on $\mathcal{M}_1$. According to a result due to Watanabe, this implies that this space is a core.

□

Chapter 5

The Bolthausen-Sznitman coalescent

The Bolthausen-Sznitman coalescent is a remarkable process which has first appeared in Statistical Physics, more precisely in the study of the Generalized Random Energy Model of Derrida, in connexion with Ruelle's cascades. It has then been studied in depth by Pitman who proved a number of important results. We shall follow here the approach by Bertoin and Le Gall [10] (see also [14]) which relies on stable subordination, as it provides a nice illustration of the preceding lecture on stochastic flows. We will also present the interesting construction via the Random Recursive Tree due to Goldschmidt and Martin [33] which has several simple consequences.

5.1 Stable subordinators and subordination

In this section, we will investigate a remarkable example of flow of bridges based on stable subordinators. Specifically, for every $\alpha \in]0,1]$, let $\varsigma_\alpha = (\varsigma_\alpha(t), t \geq 0)$ be a standard subordinator with index α. This means that ς_α is an increasing process with independent and stationary increments, and its one-dimensional distributions are characterized via their Laplace transforms by

$$\mathbb{E}\left(\exp(-q\varsigma_\alpha(t))\right) \,=\, \exp(-tq^\alpha)\,, \qquad q \geq 0\,.$$

It is well-known that ς_α has no drift and Lévy measure

$$\Lambda_\alpha(\mathrm{d}x) = c_\alpha x^{-1-\alpha}\mathrm{d}x\,, \qquad x > 0$$

whith $c_\alpha = \alpha/\Gamma(1-\alpha)$. Further ς_α fulfills the scaling property, namely for every $a > 0$, the re-scaled process $(a\varsigma_\alpha(a^{-\alpha}t) : t \geq 0)$ has the same law as ς_α.

We write $\Delta_\alpha(t) = \varsigma_\alpha(t) - \varsigma_\alpha(t-)$ for the possible jump of ς_α at time t, so $(t, \Delta_\alpha(t))_{t\geq 0}$ is the family of the atoms of a Poisson point measure on $\mathbb{R}_+ \times]0,\infty[$ with intensity $\mathrm{d}t \otimes \Lambda_\alpha(\mathrm{d}x)$. Following Pitman and Yor, we define the *Poisson-Dirichlet law* with parameter $(\alpha, 0)$ and write $\mathrm{PD}(\alpha, 0)$, as the law of the random mass-partition obtained by ranking in the decreasing order the family $(\Delta_\alpha(s)/\varsigma_\alpha(t) : 0 \leq s \leq t)$. The scaling property ensures that this distribution does not depend on t.

The fundamental property of stable subordinators that we will use in this section is the *subordination scheme*. Specifically, for every $\beta \in]0,1]$, if ς'_β denotes a standard stable subordinator with index β which is independent of ς_α, then it is immediately checked that the compound process $\varsigma_\alpha \circ \varsigma'_\beta$ is a standard stable subordinator with index $\alpha\beta$.

In fact, we shall use a variant of the subordination scheme for bridges. The scaling property implies that the distribution of the process

$$b_\alpha(r) \,:=\, \varsigma_\alpha(rt)/\varsigma_\alpha(t)\,, \qquad r \in [0,1]\,, \tag{5.1}$$

does not depend on t; we call $\mathrm{PD}(\alpha,0)$-bridge any process distributed as b_α. More precisely, the ranked sequence of the jump sizes of b_α is a random mass-partition with the $\mathrm{PD}(\alpha,0)$-distribution, and the locations of these jumps, that is the locations of the jumps of $\varsigma_\alpha(t\cdot)$, form a sequence of i.i.d. uniform variables which is independent of the sequence of the jump sizes. Thus b_α is a bridge in the sense of Definition 4.2. We now state the version of the subordination scheme that will be useful in this section.

Lemma 5.1 *Fix $\alpha, \beta \in]0,1]$, let b_α be a $\mathrm{PD}(\alpha,0)$-bridge and b'_β a $\mathrm{PD}(\beta,0)$-bridge which is independent of b_α. Then the compound process $b_\alpha \circ b'_\beta$ is a $\mathrm{PD}(\alpha\beta,0)$-bridge.*

Proof Let ς_α and ς'_β be two independent standard stable subordinators with indices α and β. Then the process b_α defined by (5.1) for $t = \varsigma'_\beta(1)$ is a $\mathrm{PD}(\alpha,0)$-bridge which is independent of ς'_β, and a fortiori of the $\mathrm{PD}(\beta,0)$-bridge $b'_\beta(r) := \varsigma'_\beta(r)/\varsigma'_\beta(1)$. The claim now follows from the fact that $\varsigma_\alpha \circ \varsigma'_\beta$ is a standard stable subordinator with index $\alpha\beta$. □

We call the paint-box process $\pi^{(\alpha)}$ based on a $\mathrm{PD}(\alpha,0)$ mass-partition, a $\mathrm{PD}(\alpha,0)$-partition. Lemma 5.1 readily yields the following.

Corollary 5.1 *Fix $0 < \alpha, \beta < 1$. Let $\pi^{(\alpha)}$ be a $\mathrm{PD}(\alpha,0)$-partition and $\pi^{(\beta)}$ an independent $\mathrm{PD}(\beta,0)$-partition. Then the exchangeable partition $\mathrm{Coag}(\pi^{(\alpha)}, \pi^{(\beta)})$ is a $\mathrm{PD}(\alpha\beta,0)$-partition.*

Proof Let b_α be a $\mathrm{PD}(\alpha,0)$-bridge. Write $S^{(\alpha)} = (S_1^{(\alpha)}, \ldots)$ for the random mass-partition with $\mathrm{PD}(\alpha,0)$-law given by the ranked sequence of the jump-sizes of b_α, and for every $i \in \mathbb{N}$, let $U_i^{(\alpha)}$ be the location of the jump with size $S_i^{(\alpha)}$. Recall that $U_1^{(\alpha)}, \ldots$ form a sequence of i.i.d. uniform variables which is independent of $S^{(\alpha)}$.

Denote the inverse bridge by

$$b_\alpha^{-1}(r) := \inf\{s \in [0,1] : b_\alpha(s) > r\}\,, \qquad 0 \le r < 1\,,$$

and let $U_1, \ldots$ be another sequence of i.i.d. uniform variables which is independent of b_α. We introduce the $\mathrm{PD}(\alpha,0)$-partition $\pi^{(\alpha)}$ as the paint-box based on b_α^{-1} and the U_i, which is defined by

$$i \overset{\pi^{(\alpha)}}{\sim} j \iff b_\alpha^{-1}(U_i) \,=\, b_\alpha^{-1}(U_j)\,.$$

For every $i \in \mathbb{N}$, define also $U'_i = b_\alpha^{-1}(U_j)$ whenever $U_j \in \pi_i^{(\alpha)}$.

Just as in Lemma 4.2, we now claim that $U'_1, \ldots$ is a sequence of i.i.d. uniform variables which is independent of $\pi^{(\alpha)}$. Indeed, if we write $\vartheta_\alpha = [0,1]\backslash\{b_\alpha(r), r \in [0,1]\}^{\mathrm{cl}}$ for the complementary of the closed range of b_α, then by definition $\pi^{(\alpha)}$ coincides with the paint-box based on ϑ_α and the U_i. The ranked sequence of the lengths of the interval components of ϑ_α is $S^{(\alpha)}$, and thus it follows from Lemma 1.4 that $\pi^{(\alpha)}$ is independent of $(U_i^{(\alpha)}, i \in \mathbb{N})$. Next, recall that $S^{(\alpha)}$ is proper (because the subordinator ς_α has no drift) and $S_i^{(\alpha)} > 0$ for every $i \in \mathbb{N}$ a.s. because the Lévy measure of ς_α is infinite), so all the blocks of $\pi^{(\alpha)}$ have strictly positive asymptotic frequencies. This enables us to define, for every $i \in \mathbb{N}$, $\sigma(i)$ by $|\pi_i^{(\alpha)}| = S^{(\alpha)}_{\sigma(i)}$, so that $\sigma : \mathbb{N} \to \mathbb{N}$ is a random permutation which is independent of $(U_i^{(\alpha)}, i \in \mathbb{N})$. It follows that $(U'_i = U^{(\alpha)}_{\sigma(i)}, i \in \mathbb{N})$ is a sequence of i.i.d. uniform variables which is independent of $\pi^{(\alpha)}$.

The rest of the proof is now straightforward. Let b_β be an independent PD$(\beta, 0)$-bridge and b_β^{-1} its inverse. We define the paint-box $\pi^{(\beta)}$ by

$$i \overset{\pi^{(\beta)}}{\sim} j \iff b_\beta^{-1}(U'_i) = b_\beta^{-1}(U'_j)\,;$$

we know from above that $\pi^{(\alpha)}$ and $\pi^{(\beta)}$ are independent. We can repeat the argument of Corollary 5.2(i) and get that $\mathrm{Coag}(\pi^{(\alpha)}, \pi^{(\beta)}) := \pi^{(\alpha\beta)}$ is given by the paint-box

$$i \overset{\pi^{(\alpha\beta)}}{\sim} j \iff b_\beta^{-1} \circ b_\alpha^{-1}(U_i) = b_\beta^{-1} \circ b_\alpha^{-1}(U_j)\,.$$

We conclude the proof by observing that $b_\beta^{-1} \circ b_\alpha^{-1}$ is the inverse of the bridge $b_\alpha \circ b_\beta$, which is a PD$(\alpha\beta, 0)$-bridge by Lemma 5.1. □

For every $t \geq 0$, let P_t^{BS} be the operator on the space of continuous function $\Phi : \mathcal{P}_\infty \to \mathbb{R}$ defined by

$$P_t^{\mathrm{BS}}\Phi(\pi) := \mathbb{E}\left(\Phi\left(\mathrm{Coag}(\pi, \pi^{(\mathrm{e}^{-t})}\right)\right), \qquad \pi \in \mathcal{P}_\infty\,,$$

where $\pi^{(\mathrm{e}^{-t})}$ stands for a PD$(\mathrm{e}^{-t}, 0)$-partition. Corollary 5.1 combined with the associativity property of the coagulation operator (see Lemma 3.1) shows that the family of operators $(P_t^{\mathrm{BS}}, t \geq 0)$ is a Markovian semigroup. More precisely, it gives the transition probabilities of some exchangeable coalescent process $\Pi^{\mathrm{BS}}(\cdot)$. This semigroup was introduced by Bolthausen and Sznitman [18], which explains the superscript BS in the notation.

5.2 Pitman's sampling formula and applications

The study of the Bolthausen-Sznitman coalescent requires first the development of properties of Poisson-Dirichlet partitions. We shall briefly present some of the most useful in this setting, and refer to the survey [49] for much more in this area, including proofs of the results which are merely stated here.

5.2.1 The two-parameter Poisson-Dirichlet partitions

The next proposition states a few basic properties of PD$(\alpha, 0)$-distributions.

Proposition 5.1 *Let $S = (S_1, \ldots)$ be a* PD$(\alpha, 0)$*-variable.*

(i) *Put $R_n = S_{n+1}/S_n$ for every $n \in \mathbb{N}$. Then the variables $R_1, \ldots$ are independent, and for each $n \in \mathbb{N}$, R_n has the beta$(\alpha n, 1)$ distribution, that is $\mathbb{P}(R_n \leq r) = r^{\alpha n}$ for $r \in [0, 1]$.*

(ii) *The limit*

$$L := \lim_{n\to\infty} n^{1/\alpha} S_n$$

exists a.s. More precisely, if S is given in the form $S_i = \mathtt{a}_i/\varsigma(1)$ with $\mathtt{a}_i$ the i-th largest jump of ς on the time-interval $[0, 1]$, then the variable $L\varsigma(1)$ is a positive and finite constant. In particular, $L > 0$ a.s. and $\mathbb{E}(L^a) < \infty$ for every $a > -\alpha$.

Remarks. • The whole sequence S can be recovered from the independent beta variables $R_1, \ldots$ Indeed, since $S_1 + S_2 + \cdots = 1$ a.s., one has

$$S_1 \,=\, 1/(1 + R_1 + R_1R_2 + R_1R_2R_3 + \cdots) \quad \text{and} \quad S_{n+1} \,=\, S_1R_1 \cdots R_n\,.$$

• Note that, contrary to the case of gamma subordinators, Proposition 5.1(ii) shows that $\varsigma(1)$ is *measurable* with respect to the sequence $(\mathtt{a}_1/\varsigma(1), \mathtt{a}_2/\varsigma(1), \ldots)$.

Proof (i) The tail of the Lévy measure of ς is $\bar{\Lambda}(x) = \Lambda(]x, \infty[) = cx^{-\alpha}$. It follows for the image property of Poisson random measures (see Lemma 6.3) that the family $\{c\mathtt{a}_i^{-\alpha}, i \in \mathbb{N}\}$ can be viewed as that of the atoms of a Poisson random measure on $\mathbb{R}_+$ with intensity the Lebesgue measure, ranked in increasing order. Thus the sequence of the increments $\mathbf{e}_1 := c\mathtt{a}_1^{-\alpha}, \mathbf{e}_2 := c\mathtt{a}_2^{-\alpha} - c\mathtt{a}_1^{-\alpha}, \ldots$ is formed by i.i.d. standard exponential variables. It then follows that the ratios

$$\frac{\mathtt{a}_1^{-\alpha}}{\mathtt{a}_2^{-\alpha}} \,=\, \frac{\mathbf{e}_1}{\mathbf{e}_1 + \mathbf{e}_2}\,,\; \frac{\mathtt{a}_2^{-\alpha}}{\mathtt{a}_3^{-\alpha}} \,=\, \frac{\mathbf{e}_1 + \mathbf{e}_2}{\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3}\,,\; \ldots$$

are mutually independent beta variables with respective parameters $(1, 1), (2, 1), \ldots$; see (1.1). We conclude that

$$S_2/S_1 \,=\, (\mathtt{a}_1^{-\alpha}/\mathtt{a}_2^{-\alpha})^{1/\alpha}, S_3/S_2 \,=\, (\mathtt{a}_2^{-\alpha}/\mathtt{a}_3^{-\alpha})^{1/\alpha}, \ldots$$

are mutually independent beta variables with respective parameters $(\alpha, 1), (2\alpha, 1), \ldots$

(ii) In the notation of part (i), we have $c\mathtt{a}_n^{-\alpha} = \mathbf{e}_1 + \cdots + \mathbf{e}_n$. By the law of large numbers, we deduce that $n^{-1}\mathtt{a}_n^{-\alpha}$ converges to $1/c$ a.s., and we deduce from the representation of $S_n = \mathtt{a}_n/\varsigma(1)$ that

$$\lim_{n\to\infty} n^{1/\alpha} S_n \,=\, c^{1/\alpha}/\varsigma(1)\,.$$

Then, using for $a > 0$ the identity

$$x^{-a} \,=\, \frac{1}{\Gamma(a)} \int_0^\infty \mathrm{e}^{-xt} t^{a-1} dt\,,$$

we get by Tonelli's theorem

$$\begin{aligned}\mathbb{E}\left(\varsigma(1)^{-a}\right) &= \frac{1}{\Gamma(a)}\int_0^\infty \mathbb{E}(\mathrm{e}^{-t\varsigma(1)})t^{a-1}dt \\ &= \frac{1}{\Gamma(a)}\int_0^\infty \exp(-t^\alpha)t^{a-1}dt \\ &= \frac{1}{\alpha\Gamma(a)}\int_0^\infty \exp(-s)s^{a/\alpha-1}ds \\ &= \frac{\Gamma(a/\alpha)}{\alpha\Gamma(a)}\,.\end{aligned}$$

We conclude that

$$\mathbb{E}(L^a) = \frac{\Gamma(1+a/\alpha)}{\Gamma(1-\alpha)^{a/\alpha}\Gamma(1+a)}$$

for every $a > 0$, and this formula can then be extended by analytic continuation to any $a > -\alpha$. □

Proposition 5.1(ii) opens the way to a two-parameter extension of Poisson-Dirichlet partitions.

Definition 5.1 *For every $\alpha \in]0,1[$ and $\theta > -\alpha$, we write* $\mathrm{PD}(\alpha,\theta)$ *and call Poisson-Dirichlet law with parameters (α,θ) the probability distribution on $\mathcal{P}_{\mathtt{m}}$ which is absolutely continuous with respect to* $\mathrm{PD}(\alpha,0)$ *with density $L^\theta/\mathbb{E}(L^\theta)$, where L is the variable that appears in Proposition 5.1(ii).*

Probably the most useful tool for the study of this two-parameter family is the following extension of the residual allocation model which was obtained first for $\mathrm{PD}(0,\theta)$ random mass partitions (see Corollary 1.1).

Proposition 5.2 *Fix $\alpha \in]0,1[$ and $\theta > -\alpha$. Let $\beta_1, \beta_2, \ldots$ be a sequence of independent variables such that β_n has the beta$(1-\alpha, \theta+n\alpha)$ distribution. Put*

$$S_1^* = \beta_1\,,\ S_2^* = (1-\beta_1)\beta_2\,,\ \ldots, S_n^* = \beta_n \times \prod_{i=1}^{n-1}(1-\beta_i),\ \ldots$$

Then the random sequence $S^* = (S_1^*, \ldots)$ is distributed as the size-biased reordering of a $\mathrm{PD}(\alpha,\theta)$ *random mass partition.*

We can now present an important extension of Ewens sampling formula for the two parameter family, which follows easily from description of the residual allocation model above. In this direction, for every $\alpha \in [0,1[$ and $\theta > -\alpha$, it is convenient to call any random exchangeable partition such that the ranked sequence of its asymptotic frequencies is distributed according $\mathrm{PD}(\alpha,\theta)$ a $\mathrm{PD}(\alpha,\theta)$-partition. Alternatively, a $\mathrm{PD}(\alpha,\theta)$-partition is a random partition with the same law as a mixture of paint-boxes based on some $\mathrm{PD}(\alpha,\theta)$-random mass-partition.

Theorem 5.1 *For every $\alpha \in]0,1[$ and $\theta > -\alpha$, write $\mathtt{p}_{\alpha,\theta}$ for the EPPF of an (α,θ)-partition. Pick integers $k \leq n$ and $n_1, \ldots, n_k$ such that $n_1 + \cdots + n_k = n$.*

We have **Pitman sampling formula***: for $\theta = 0$,*

$$\mathtt{p}_{\alpha,0}(n_1, \ldots, n_k) \;=\; \alpha^k \frac{(k-1)!}{(n-1)!} \prod_{i=1}^{k} (1-\alpha)_{(n_i-1)\uparrow} \,.$$

where for every integer $\ell \geq 1$ and real number a,

$$(a)_{0\uparrow} = 1 \quad \textit{and} \quad (a)_{\ell\uparrow} = a(a+1)\cdots(a+\ell-1)\,,$$

while for $\theta \neq 0$

$$\mathtt{p}_{\alpha,\theta}(n_1, \ldots, n_k) \;=\; \alpha^k \frac{(\theta/\alpha)_{k\uparrow}}{(\theta)_{n\uparrow}} \prod_{i=1}^{k} (1-\alpha)_{(n_i-1)\uparrow} \,.$$

The proof follows easily from the residual allocation scheme (Proposition 5.2) and elementary calculations involving the moments of beta variables which are similar to those in the proof of Ewens sampling formula (Theorem 1.3).

5.2.2 Some consequences

Proposition 5.3 *The exchangeable coalescent $\Pi^{\mathrm{BS}}(\cdot)$ based on the Bolthausen-Sznitman semigroup $(P_t^{\mathrm{BS}}, t \geq 0)$ is simple. Its coagulation rate is given by*

$$\mu^{\mathrm{BS}} \;=\; \int_0^1 \varrho_x x^{-2} \mathrm{d}x \,,$$

where ϱ_x is the law of a random exchangeable partition with ranked asymptotic frequencies $(x, 0, \ldots)$. In other words, the coefficient of binary coagulation is $\mathtt{c} = 0$ and the rate of multiple coagulations is given by the measure $\nu(\mathrm{d}x) = x^{-2}\mathrm{d}x$, $x \in]0,1[$; i.e. $\Lambda(\mathrm{d}x) = \mathrm{d}x$.

Proof Fix $n \geq 2$, and consider some partition $\pi^{[n]} \in \mathcal{P}_n$ with $\pi^{[n]} \neq \mathbf{0}_{[n]}$. We have to evaluate the jump rate

$$q_{\pi^{[n]}} \;=\; \lim_{t\to 0} \frac{1}{t} \mathbb{P}\left(\Pi^{\mathrm{BS}}_{|[n]}(t) = \pi^{[n]}\right) \,.$$

In this direction, suppose that the number of non-empty blocks of $\pi^{[n]}$ is $\#\pi^{[n]} = k$ and that these k blocks have respective sizes $n_1, \ldots, n_k$ where $n_1 + \cdots + n_k = n$. Recall from Theorem 5.1 the EPPF for PD$(\alpha, 0)$-partitions; this yields

$$\frac{1}{t} \mathbb{P}\left(\Pi^{\mathrm{BS}}_{|[n]}(t) = \pi^{[n]}\right) \;=\; t^{-1} \mathrm{e}^{-t(k-1)} \frac{(k-1)!}{(n-1)!} \prod_{i=1}^{k} (1-\mathrm{e}^{-t})_{n_i-1\uparrow} \,,$$

with $(a)_{\ell\uparrow} = a(a+1)\cdots(a+\ell-1)$ for $\ell \in \mathbb{N}$ and $(a)_{0\uparrow} = 1$. Now we have $(1-\mathrm{e}^{-t})_{\ell\uparrow} = O(t)$ as $t \to 0+$ whenever $\ell \geq 1$, and thus

$$\prod_{i=1}^{k} (1-\mathrm{e}^{-t})_{n_i-1\uparrow} \;=\; o(t) \,, \qquad t \to 0 \,,$$

except when $\pi^{[n]}$ is simple. In the latter case, if $\pi^{[n]}$ has a block with size $\ell \in \{2, \ldots, n\}$, then we have $k = n - \ell + 1$, and we find

$$\begin{aligned} q_\pi &= \lim_{t\to 0} \frac{1}{t} \mathbb{P}\left(\Pi^{\mathrm{BS}}_{|[n]}(t) = \pi^{[n]}\right) \\ &= \frac{\Gamma(n-\ell+1)\Gamma(\ell-1)}{\Gamma(n)} \\ &= \int_0^1 x^{\ell-2}(1-x)^{n-\ell}\mathrm{d}x\,. \end{aligned}$$

The comparison with (4.2) completes the proof. □

Pitman [47] also discovered a remarkable duality between fragmentation and coagulation operators based on Poisson-Dirichlet variables, which extends considerably Corollary 5.1. To state this result, it is convenient to introduce for every $\alpha \in]0,1[$ and $\theta > -\alpha$ the following notation. For every partition $\gamma \in \mathcal{P}_\infty$, we write $\mathrm{Coag}_{\alpha,\theta}(\gamma)$ for the distribution of $\mathrm{Coag}(\gamma, \pi)$ where π is a $\mathrm{PD}(\alpha, \theta)$-partition. Similarly, we write $\mathrm{Frag}_{\alpha,\theta}(\gamma)$ for the distribution of $\mathrm{Frag}(\gamma, \pi^{(\cdot)})$ where $\pi^{(\cdot)} = (\pi^{(1)}, \ldots)$ is a sequence of independent $\mathrm{PD}(\alpha, \theta)$-partitions. This means that for every integer i, we consider the partition of the i-th block γ_i of γ induced by the i-th term $\pi^{(i)}$ of the sequence $\pi^{(\cdot)}$, that is

$$\pi^{(i)}_{|\gamma_i} = \left(\pi^{(i)}_j \cap \gamma_i, j \in \mathbb{N}\right)\,.$$

As i varies in $[n] = \{1, \ldots, n\}$, the collection $\left\{\pi^{(i)}_j \cap \gamma_i : i, j \in \mathbb{N}\right\}$ of the blocks of these induced partitions forms a partition of B which we denote by $\mathrm{Frag}(\gamma, \pi^{(\cdot)})$ and call the fragmentation of γ by $\pi^{(\cdot)}$.

We may now state:

Theorem 5.2 *Let Γ, Γ' be two random partitions. For every $\alpha, \beta \in]0,1[$ and $\theta > -\alpha\beta$, the following assertions are equivalent:*

(i) *Γ is a $\mathrm{PD}(\alpha, \theta)$-partition and conditionally on $\Gamma = \gamma$, Γ' has the law $\mathrm{Coag}_{\beta,\theta/\alpha}(\gamma)$.*

(ii) *Γ' is a $\mathrm{PD}(\alpha\beta, \theta)$-partition and conditionally on $\Gamma' = \gamma'$, Γ has the law $\mathrm{Frag}_{\alpha,-\alpha\beta}(\gamma')$.*

Proof (i) and (ii) provide two descriptions of the joint law of (Γ, Γ'), and we have to check that they coincide. In this direction, it suffices to show that their joint EPPFs are the same, that is for every $\gamma, \gamma' \in \mathcal{P}_\infty$ and $n \in \mathbb{N}$, the two descriptions yield the same value for the probabilities, say $p^{(\mathrm{i})}$ and $p^{(\mathrm{ii})}$, that $\Gamma_{|[n]} = \gamma_{|[n]}$ and $\Gamma'_{|[n]} = \gamma'_{|[n]}$. We focus on the case when $\gamma_{|[n]}$ is finer than $\gamma'_{|[n]}$, as otherwise this probability is obviously zero for both descriptions.

Suppose that $\gamma_{|[n]}$ has K non-empty blocks with sizes $a_1, \ldots, a_K$ and that $\gamma'_{|[n]} = \mathrm{Coag}(\gamma_{|[n]}, \eta)$ where $\eta \in \mathcal{P}_K$ has $k \leq K$ non-empty blocks with sizes $j_1, \ldots, j_k$. By the very definition of the coagulation operator, we have in the case (i)

$$p^{(\mathrm{i})} = \mathtt{p}_{\alpha,\theta}(a_1, \ldots, a_K)\mathtt{p}_{\beta,\theta/\alpha}(j_1, \ldots, j_k)$$

where $\mathtt{p}_{\alpha,\theta}$ stands for the EPPF of a $\mathrm{PD}_{\alpha,\theta}$-partition.

In this situation, $\gamma'_{|[n]}$ has k non-empty blocks, say $B_1, \ldots, B_k$, with respective sizes $b_1, \ldots, b_k$. There exists a unique k-tuple $\eta^{(\cdot)} = (\eta^{(1)}, \ldots, \eta^{(k)})$ where $\eta^{(\ell)}$ is a partition of B_ℓ with j_ℓ non-empty blocks, such that $\gamma_{|[n]}$ coincides with the family of the blocks of $\eta^{(\ell)}$ for $\ell = 1, \ldots k$. Denote by $c_{\ell,1}, \ldots, c_{\ell,j_\ell}$ the lengths of the non-empty blocks of $\eta^{(\ell)}$, and observe that $(c_{\ell,i}, i = 1, \ldots, j_\ell \text{ and } \ell = 1, \ldots, k)$ is a reordering of $(a_i, i = 1, \ldots, K)$

By the very definition of the fragmentation operator and exchangeability, we have in the case (ii)

$$p^{(\mathrm{ii})} \;=\; \mathtt{p}_{\alpha\beta,\theta}(b_1, \ldots, b_k) \prod_{\ell=1}^{k} \mathtt{p}_{\alpha,-\alpha\beta}(c_{\ell,1}, \ldots, c_{\ell,j_\ell})\,.$$

Applying Pitman's sampling formula (cf. Theorem 5.1), we get the explicit formulas

$$\begin{aligned} p^{(\mathrm{i})} &= \frac{(\theta/\alpha)_{K\uparrow}}{(\theta)_{n\uparrow}} \left(\prod_{i=1}^{K} -(-\alpha)_{a_i\uparrow} \right) \frac{(\theta/(\alpha\beta))_{k\uparrow}}{(\theta/\alpha)_{K\uparrow}} \prod_{\ell=1}^{k} -(-\beta)_{j_\ell\uparrow}\,, \\ p^{(\mathrm{ii})} &= \frac{(\theta/(\alpha\beta))_{k\uparrow}}{(\theta)_{n\uparrow}} \left(\prod_{\ell=1}^{k} -(-\alpha\beta)_{b_\ell\uparrow} \frac{(-\beta)_{j_\ell\uparrow}}{(-\alpha\beta)_{b_\ell\uparrow}} \prod_{i=1}^{j_\ell} -(-\alpha)_{c_{\ell,i}\uparrow} \right)\,. \end{aligned}$$

By obvious cancellations and the relations between the parameters noted above, we see that $p^{(\mathrm{i})} = p^{(\mathrm{ii})}$, which completes the proof. □

Theorem 5.2 has important consequences for the Bolthausen-Sznitman coalescent Π^{BS}. Probably the most striking one is the following description of the time-reversed process as a time-inhomogeneous fragmentation.

Corollary 5.2 *The reversed Bolthausen-Sznitman coalescent* $\big(\Pi^{\mathrm{BS}}(-\ln t), t \in]0,1]\big)$ *is a time-inhomogeneous Markov process on* $\mathcal{P}_\infty$*. Its semigroup can be described as follows. For every* $0 < t \leq t' \leq 1$*, conditionally on* $\Pi^{\mathrm{BS}}(-\ln t) = \pi$*,* $\Pi^{\mathrm{BS}}(-\ln t')$ *is distributed as* $\mathrm{Frag}_{t',-t}(\pi)$*.*

Proof Recall that the semigroup P_t^{BS} of the Bolthausen-Sznitman coalescent can be expressed as the law of the random coagulation operator $\mathrm{Coag}_{\mathrm{e}^{-t},0}(\cdot)$. The claim follows immediately from Theorem 5.2 specified for $\alpha = t', \beta = t/t', \theta = 0$. □

We conclude this section by establishing an interesting identity in law between the process of the asymptotic frequencies of the first block in a Bolthausen-Sznitman coalescent, and the Dirichlet process with parameter 1 (cf. Definition 1.9).

Corollary 5.3 *The process* $\big(|\Pi_1^{\mathrm{BS}}(t)|, t \geq 0\big)$ *has the same finite-dimensional distributions as* $(\gamma(1 - \mathrm{e}^{-t}))/\gamma(1), t \geq 0)$ *where* $\gamma(\cdot)$ *is a standard gamma subordinator (i.e. with parameter* $(1,1)$*).*

Proof On the one hand, we know that $\Pi^{\mathrm{BS}}(-\ln t)$ is a $\mathrm{PD}(t,0)$ partition, so by the residual allocation model described in Proposition 5.2, $|\Pi_1^{\mathrm{BS}}(-\ln t)|$ is a beta$(1-t,t)$ variable. Taking logarithms, we deduce from the representation (1.1) of beta variables as a ratio of gamma variables that $\ln|\Pi_1^{\mathrm{BS}}(-\ln t)|$ can be expressed as $\ln\gamma(1-t)-\ln\gamma(1)$, where $\gamma(\cdot)$ is a standard gamma process, and the latter quantity is independent of $\ln\gamma(1)$.

On the other hand, it follows easily from the description of the reversed process $\Pi^{\mathrm{BS}}(-\ln t)$ as a time-inhomogeneous fragmentation process on $\mathcal{P}_\infty$ that the process

$$\ln|\Pi_1^{\mathrm{BS}}(-\ln t)|\,, \qquad 0<t\leq 1\,,$$

has independent (non-stationary) increments; see Section 3.2.2. We readily derive from (1.1) that $(\ln\gamma(1)-\ln\gamma(1-t), 0<t\leq 1)$ also has independent increments. As for processes with independent increments, one-dimensional distributions determine the multi-dimensional distributions, we conclude from the first part of the proof that the two processes have the same finite-dimensional distributions. □

5.3 Connexions to other models

5.3.1 Random energy models

The initial motivation of Bolthausen and Sznitman for introducing the coalescent process which has thereafter carried their names was to unravel the probabilistic structure that underlies certain predictions made by Mézard, Parisi and Virasoro on the Sherrington-Kirkpatrick spin-glass model. The latter is defined as a random Gibbs measure on the hypercube $S_N=\{-1,1\}^N$, and one of the most important prediction in this field concerns the hierarchy of randomly sampled spins, known as Parisi's ultrametric conjecture. Derrida introduced a simplified related random energy model in which the ultrametric property is granted from the construction.

More precisely, we may view the hypercube S_N as the set of vertices at height N in the binary tree, and then naturally endow S_N with the ultrametric distance on the tree

$$d(\sigma,\tau)=1-N^{-1}\max\{i\leq N:\sigma_i=\tau_i\}\,,$$

where σ and τ denote two generic spin-configurations, i.e. vertices of S_N. One considers a Gaussian vector $(\xi_\sigma:\sigma\in S_N)$ with covariance [1]

$$\mathrm{Cov}(\xi_\sigma,\xi_\tau)=1-d(\sigma,\tau)$$

(more generally, one may replace the right-hand-side by some non-increasing function of the ultrametric distance between σ and τ which is 0 at 0). So two spins which are close together are strongly correlated, while two spins which are far away are nearly

[1] for the Sherrington-Kirkpatrick model, the covariance is given instead by $1-N^{-1}d_H(\sigma,\tau)$, where $d_H(\sigma,\tau)$ is the Hamming distance, namely $\#\{i\leq N:\sigma_i\neq\tau_i\}$.

independent. Derrida's Generalized Random Energy Model (in short, GREM) is based on the random Gibbs measure μ_β on the hypercube defined by

$$\mu_\beta(\sigma) = Z^{-1} \mathrm{e}^{\beta \xi_\sigma}$$

where $Z = \sum_{S_N} \mathrm{e}^{\beta \xi_\sigma}$ is the random partition function and $\beta > 0$ a fixed parameter that should be viewed as the inverse temperature.

Now imagine that given the random Gibbs measure μ_β, we sample k vertices $\sigma_1, \ldots, \sigma_k$ independently according to μ_β, and that we are interested in the genealogical structure of these spins, i.e. in the subtree of the binary tree which is spanned by these k vertices. In this direction, we consider for every $x \in [0,1]$ the partition $\Pi_N^k(x)$ of $[k]$ such that two integers i, j belong to the same block if and only if $d(\sigma_i, \sigma_j) \leq x$. Roughly speaking, when $N \to \infty$, $\Pi_N^k(\cdot)$ converges weakly to some time-change of the Bolthausen-Sznitman coalescent. A first description of the frequencies of the limiting processes was provided by Ruelle [51] in terms of probability cascades; in the modern terminology would rather call it a time-inhomogeneous fragmentation, see Basdevant [3] for details.

We refer to Section 6.2 in Berestycki [5] for a more detailed outline of the connexions between spin glasses, Derrida's GREM and the Bolthausen-Sznitman coalescent, and to the monograph [19] by Bovier for much more on recent developments of this topic.

5.3.2 Random recursive tree

`Hacked from C. Goldschmidt and J. Martin (2005) :`

Random recursive trees and the Bolthausen-Sznitman coalescent, EJP 10

A tree on n vertices labelled $1, 2, \ldots, n$ is called a recursive tree if the vertex labelled 1 is the root and, for all $2 \leq k \leq n$, the sequence of vertex labels in the path from the root to k is increasing. Call a random recursive tree a tree chosen uniformly at random from the possible recursive trees on n vertices. It should be obvious that the following simple algorithm produces a random recursive tree. The vertex 1 is distinguished as the root. We imagine the vertices arriving one by one. For $k \geq 2$, vertex k attaches itself to a vertex chosen uniformly at random from $1, 2, \ldots, k-1$. As a consequence, we see that there are $(n-1)!$ recursive trees on n vertices.

For the purposes of this section, it will be convenient also to define a random recursive tree on a label set $\{\pi_1, \ldots, \pi_j\}$ where $\pi_1, \ldots, \pi_j$ are the blocks of a partition π of $[n]$ for some n, listed in increasing order of least elements. The tree is constructed in the obvious way: π_1 labels the root and π_k is attached to a vertex chosen uniformly at random from those labelled $\pi_1, \ldots, \pi_{k-1}$. Call the weight of a label the number of integers it contains and let $\mathcal{P}_n$ be the set of partitions of $[n]$.

Proposition 5.4 *Suppose T is a random recursive tree on $\pi \in \mathcal{P}_n$. Pick an edge at random, cut it and add the labels below the cut to the label above. Then the resulting tree is a random recursive tree on the new label-set.*

Proof The resulting tree is clearly recursive because its labels still increase along all paths away from the root. So we need to show that it is chosen uniformly from the set of all

recursive trees with the same label-set. Put another way, we will show that if $\#\pi = b$, then each of the $(b-1)!(b-1)$ recursive trees on the label-set π with a single marked edge corresponds to a tree constructed as follows: for some $2 \leq k \leq b$, pick k of the labels, say $\pi_{i_1}, \ldots, \pi_{i_k}$ (taken to be in increasing order), make a recursive tree on $\pi \backslash \{\pi_{i_2}, \ldots, \pi_{i_k}\}$, make another recursive tree on $\{\pi_{i_2}, \ldots, \pi_{i_k}\}$ and then join them together with an edge between the vertices labelled π_{i_1} and π_{i_2} . There are $\binom{b}{k}$ ways of picking the k labels $\pi_{i_1}, \ldots, \pi_{i_k}$. There are $(k-2)!$ ways of arranging the $k-1$ largest into a recursive tree rooted at π_{i_2} . There are $(b-k)!$ ways of arranging the $b-k+1$ other labels into a recursive tree. Clearly each of the trees constructed in this way is distinct and also a recursive tree. The number which can be constructed is

$$\sum_{k=2}^{b} \binom{b}{k} (k-2)!(b-k)! = b! \sum_{k=2}^{b} \frac{1}{k(k-1)} = (b-1)!(b-1)\,.$$

Hence, the claimed correspondence holds. □

This yields a remarkably simple construction of the Bolthausen-Sznitman coalescent. Start with a random recursive tree on $[n]$ and associate an independent exponential random variable with mean 1 to each edge. This exponential time is the time at which the edge is deleted, at which point the labels in the subtree below it are instantaneously added to the label of the vertex above the edge. Then at time $t \geq 0$, the set of labels forms a partition $\Pi^{[n]}(t)$ of $[n]$.

Corollary 5.4 *The process* $(\Pi^{[n]}(t) : t \geq 0)$ *evolves according to the dynamics of the Bolthausen-Sznitman coalescent restricted to* $[n]$.

Proof We need to show that the rate of coalescence of any set of k of the labels is

$$\frac{(k-2)!(b-k)!}{(b-1)!}$$

whenever there are b vertices in the tree. The total rate of events when there are b vertices is $b-1$. The probability that the next event will coalesce a particular $k-$set is worked out in the same way as in the proof of the preceding Proposition. Suppose we start with label-set $\pi = (\pi_1, \ldots, \pi_b)$ and we want the probability that the next event is the coalescence of $\{\pi_{i_1}, \ldots, \pi_{i_k}\}$. There are $(k-2)!$ ways of making a recursive tree on $\{\pi_{i_2}, \ldots, \pi_{i_k}\}$; there are $(b-k)!$ ways of making a recursive tree on the remaining labels. There are $(b-1)!(b-1)$ recursive trees on a label-set of size b with a single marked edge and so the probability that we coalesce $\{\pi_{i_1}, \ldots, \pi_{i_k}\}$ is

$$\frac{(k-2)!(b-k)!}{(b-1)!(b-1)}\,.$$

Hence, the rate at which we coalesce any $k-$set is

$$\frac{(k-2)!(b-k)!}{(b-1)!}\,.$$

The evolution is Markovian because, by the preceding Proposition, the resulting tree is another random recursive tree, this time on $b - k + 1$ labels. □

Because of the recursive way in which the original tree is built, the representations are consistent for different n: $\Pi^{[n]}(t)$ coincides with the restriction of $\Pi^{[n+1]}(t)$ to $[n]$. Thus we are able to define the Bolthausen-Sznitman coalescent $(\Pi^{\mathrm{BS}}(t) : t \geq 0)$ on the whole of $\mathbb{N}$, by means of the cutting procedure applied to an infinite random recursive tree (indexed by $\mathbb{N}$).

We can find a Chinese restaurant process in the construction of the random recursive tree on $[n+1]$. For this purpose, it is easier to imagine the random recursive tree labelled by the set $\{0, 1, \ldots, n\}$ rather than $[n + 1]$. The root (labelled 0) is fixed and does not appear in the permutation. Vertices attached to the root correspond to individuals who start a new table. Thus, individual 1 necessarily starts a new table. If vertex k arrives and attaches to a vertex other than the root (say j) then individual k sits directly to the left of individual j. As vertex k is equally likely to attach to each of the vertices labelled $0, 1, \ldots, k - 1$, individual k is equally likely to sit to the left of any of the individuals $1, \ldots, k - 1$ or to form a new table.

Thus, a random recursive tree on $[n + 1]$ corresponds to a random permutation of $[n]$. Just to give some interpretation to the cutting procedure in the Chinese restaurant, we embellish the usual model as follows. Each customer present in the restaurant has friends arrive at rate 1 and there is also a rate 1 stream of customers who know no-one in the restaurant at their time of arrival. Customers who arrive and have a friend present always sit to the left of that friend. Friendless customers sit at new tables. Stop the process when there have been n arrivals. A meal in the restaurant costs one euro. At rate 1, each customer decides to leave and go home. Whenever he leaves, any of his friends who arrived after him want to leave (and their friends, and so on). (If the person who decided to depart was k then it is the subtree rooted at k in the random recursive tree which departs.) Anyone leaving gives the money for their meal to the person to whose left k sat down on entering the restaurant, so that he can pay for them when he leaves. If k was, in fact, the first person to have sat at that table then he collects all the money for the whole table (plus, of course, the price of his own meal), takes it to the cashier and departs. Note that at any time t, the amount of money in the cash register is the same as the weight of the label at the root in the random recursive tree (i.e. the size of the block containing 1 in the Bolthausen-Sznitman coalescent).

5.3.3 Branching Brownian motion with absorption

We now conclude this section by briefly presenting a recent result of Berestycki *et al.* [9] in which the Bolthausen-Sznitman coalescent appears in describing the genealogy of a branching Brownian motion with absorption. This is closely connected to predictions made by Brunet *et al.* [20, 21] for a population model with selection.

Specifically, consider a spatial particle system where each particle branches at unit rate, independently of the other particles. We suppose that particles live in the half-line $(0, \infty)$

and evolve according to a Brownian motion with drift $-b < 0$ and killed upon entering $(-\infty, 0]$, again independently one of the others and of the branching phenomenon. It has been observed by Kesten that when $b > \sqrt{2}$, then the system started from a single particle in $(0, \infty)$ gets eventually extinct with probability one, while if $b < \sqrt{2}$, then the system survives for ever with positive probability.

For the sake of simplicity, we shall describe only a special case of the main result of [9] where more general settings are treated, as our aim here is merely to provide a flavor of that paper. Now suppose that the drift coefficient depends on a large parameter N, specifically

$$b_N = \sqrt{2 - \frac{2\pi^2}{(\ln N + 3 \ln \ln N)^2}},$$

which thus approaches the critical value $\sqrt{2}$ when $N \to \infty$. We further set

$$\ell_N = 2^{-1/2}(\ln N + 3 \ln \ln N)$$

and assume that at the initial time the branching Brownian motion starts with $N \ln^2 N$ particles located at ℓ_N. Then Theorem 2 of [9] asserts that if one rescales times by a factor $\ln^3 N$ and assigns a mass $1/N$ to the particles, then the resulting process converges weakly to a certain continuous state branching process (see the forthcoming Section 6.2.1 for background) whose genealogy is described by the Bolthausen-Sznitman coalescent. More precisely, fix $t > 0$ and sample n particles labeled $1, \ldots, n$ uniformly at random from the particles at time $t \ln^3 N$. For $0 \leq s \leq 2\pi t$, consider the partition $\Pi_N(s)$ of $[n]$ such that i and j are in the same block of $\Pi_N(s)$ if and only if the particles i and j have the same ancestor at time $(t?s/2\pi) \ln^3 N$. Then as $N \to \infty$, the process $(\Pi_N(s), 0 \leq s \leq 2\pi t)$ converges in distribution the Bolthausen-Sznitman coalescent restricted to $[n]$ and stopped at time $2\pi t$.

Chapter 6

Beta coalescents and branching processes

The beta coalescents form a one-parameter family of simple exchangeable coalescents with coagulation rates governed by the beta $(2-\alpha, \alpha)$-measures for $\alpha \in]0, 2[$ i.e.

$$\Lambda(\mathrm{d}x) = x^2 \nu(\mathrm{d}x) = c x^{1-\alpha}(1-x)^{\alpha-1}\mathrm{d}x\,, \qquad 0 < x < 1\,,$$

where $c > 0$ is some constant. In other words, when the partition has ℓ blocks, the rate at which a given family of $k \leq \ell$ blocks coalesces is

$$\begin{aligned} \lambda_{k,\ell} &= c\int_0^1 x^{k-1-\alpha}(1-x)^{\ell+\alpha-1-k}\mathrm{d}x \\ &= c\frac{\Gamma(k-\alpha)\Gamma(\ell+\alpha-k)}{\Gamma(\ell)}\,. \end{aligned}$$

Note that the case $\alpha = 1$ corresponds to the Bolthausen-Sznitman coalescent.

Our purpose in this final lecture is to provide some insight on their role in a variety of limit theorems related to branching processes. We shall sometimes avoid technicalities and merely present sketches of the proofs, emphasizing guiding ideas rather than rigorous technical arguments on convergence of Markov processes. Technically, the duality between exchangeable coalescents and population models will have a crucial role, as it allows us to transfer questions about convergence of the genealogy in certain discrete models to that of convergence of the population models themselves. The evolution of the latter is characterized by an infinitesimal generator which, roughly speaking specifies the rates of jumps. The fact that often jumps are not absolutely summable is the source of difficulties.

6.1 A branching process with competition

Following Schweinsberg [55], we shall first show that beta-coalescents can describe asymptotically the genealogy of certain population models with fixed size N as $N \to \infty$. Such

population models may be viewed as variations of the Wright-Fisher model, in which exceptionally certain individuals may have a very large offsprings.

Let μ be some finite measure on $\mathbb{N}$, where for every integer k, $\mu(k)$ describes the rate at which each individual in the population gives birth to k children (simultaneously). We consider a population model in continuous time with a fixed size N at any time, in which each individual begets with rate governed by μ and independently of the other individuals. When a birth event occurs, say when k children are born, we instantaneously eliminate uniformly at random k individuals amongst the $N+k$ present, so the total size of the population remains equal to N. One possible justification for the elimination of individuals in excess may be for instance competition between individuals for resources or for space. This model is a variation of that considered by Schweinsberg [55] who rather dealt with super-critical Galton-Watson processes, i.e. branching processes in discrete time.

We use a superscript N in the notation to underline the size of the population (the law of reproduction μ being fixed). We will be interested in the genealogy, and in this direction, we sample n individuals from the generation at the present time and follow their ancestral lineages backwards in time to obtain a coalescent tree. We write thus $\Pi^N_{|[n]}(t)$ for the partition of the n-sample into families having the same ancestor at time t in the past.

Theorem 6.1 *Suppose that the tail-distribution $\bar{\mu}(k) = \mu(\{k+1, k+2, \ldots\})$ of the reproduction law μ is regularly varying with index $-\alpha$ for some $\alpha \in (0,1) \cup (1,2)$, i.e.*

$$\lim_{N\to\infty} \frac{\bar{\mu}(\lfloor xN \rfloor)}{\bar{\mu}(N)} = x^{\alpha}\,, \qquad x > 0\,.$$

Then for every $n \in \mathbb{N}$, the process $(\Pi^N_{|[n]}(t/\bar{\mu}(N)) : t \geq 0)$ converges in distribution as $N \to \infty$ to a beta$(2-\alpha, \alpha)$-coalescent restricted to $[n]$.

Remark: In the case $\alpha < 1$, the limiting coalescents in Schweinsberg [55] are not simple (i.e. they involve simultaneous multiple collisions). The difference with the present statement stems from the fact that Schweinsberg consider Galton-Watson processes in discrete time while we deal here with branching processes in continuous time.

Before tackling the proof of Theorem 6.1, we need the following elementary bounds for the distance in total variation between two probability measures on some discrete set, say $\{x_1, \ldots, x_{n+k}\}$ which occur in this setting. First, we fix an integer p and we denote by

$$\epsilon = \frac{1}{(n+k)^p} \sum_{i_1,\ldots,i_p=1}^{n+k} \delta_{(x_{i_1},\ldots,x_{i_p})}$$

the uniform probability measure on $\{x_1, \ldots, x_{n+k}\}^p$. Next, we consider η the probability measure η on $\{x_1, \ldots, x_{n+k}\}^p$ obtained by first removing k points uniformly at random, and then sampling with replacement p elements amongst the remaining points.

Lemma 6.1 *Denote by $|\epsilon - \eta|$ the distance in total variation between ϵ and η, that is*

$$|\epsilon - \eta| = \sup \left| \int f \mathrm{d}\epsilon - \int f \mathrm{d}\eta \right|$$

where the supremum is taken over the set of functions $f : \{x_1, \ldots, x_{n+k}\}^p \to [-1, 1]$. Then

$$|\epsilon - \eta| = O\left(\frac{k}{n^2} \wedge \frac{1}{n}\right) \qquad \text{as } n \to \infty\,.$$

Proof For every $y_1, \ldots, y_p \in \{x_1, \ldots, x_{n+k}\}$, let

$$\rho(y_1, \ldots, y_p) = (n+k)^p \eta(y_1, \ldots, y_p)$$

denote the density of η with respect to ϵ at $(y_1, \ldots, y_p)$. We shall establish bounds for ρ on three disjoint subsets of $\{x_1, \ldots, x_{n+k}\}^p$, S_1, S_2 and S_3. Specifically, we denote by S_1 the set of p-tuples $(y_1, \ldots, y_p)$ such that the y_i are all distinct, by S_2 the set of p-tuples such that two coordinates are equals and the others are all distinct, and by S_3 the complementary set of $S_1 \cup S_2$. In this direction, it will be convenient to use a slightly abusive notation, denoting by c a number depending only on p which may take different values in different expressions.

Consider first the case when $(y_1, \ldots, y_p) \in S_1$. The probability that the subset of the k points which are removes do not contain any of the y_i is

$$\binom{n+k-p}{k} \Big/ \binom{n+k}{k} = \frac{(n+k-p)!}{(n+k)!} \times \frac{n!}{(n-p)!}$$

and it follows

$$\rho(y_1, \ldots, y_p) = \frac{(1+k/n)^p}{(1+k/n)\cdots(1+k/(n-p+1))}\,.$$

In particular we have the bounds

$$\frac{(1+k/n)^p}{(1+k/(n-p+1))^p} \leq \rho(y_1, \ldots, y_p) \leq 1\,,$$

from which we easily get

$$|\rho(y_1, \ldots, y_p) - 1| \leq c\left(\frac{k}{n^2} \wedge \frac{1}{n}\right)\,, \qquad (y_1, \ldots, y_p) \in S_1\,.$$

As a consequence,

$$\left| \int_{S_1} f \mathrm{d}\epsilon - \int_{S_1} f \mathrm{d}\eta \right| \leq c\left(\frac{k}{n^2} \wedge \frac{1}{n}\right)$$

where f is an arbitrary function bounded in modulus by 1.

Next we consider the case when $(y_1, \ldots, y_p) \in S_2$; calculations similar to those above yield

$$\rho(y_1, \ldots, y_p) = \frac{(1+k/n)^p}{(1+k/n)\cdots(1+k/(n-p+2))}\,,$$

then

$$\frac{(1+k/n)^p}{(1+k/(n-p+2))^{p-1}} \leq \rho(y_1,\dots,y_p) \leq 1+k/n\,,$$

and finally

$$|\rho(y_1,\dots,y_p)-1| \leq ck/n\,, \qquad (y_1,\dots,y_p)\in S_2.$$

Since the cardinal of S_2 is $c(n+k)^{p-1}$, we get

$$\left|\int_{S_2} f\mathrm{d}\epsilon - \int_{S_2} f\mathrm{d}\eta\right| \leq c\frac{k}{n(n+k)}\,.$$

Finally, it is immediately seen that $\epsilon(S_3)\leq (n+k)^{-2}$ and $\eta(S_3)\leq n^{-2}$, which completes the proof of the claim. □

Recall from Section 4.3 that we denote by $\mathcal{M}_1$ the space of probability measures on $[0,1]$, and the notation

$$\Phi_f(m) = \int_{[0,1]^p} m(\mathrm{d}x_1)\cdots m(\mathrm{d}x_p) f(x_1,\dots,x_p)\,,$$

where $f:[0,1]^p\to\mathbb{R}$ is a continuous function and $m\in\mathcal{M}_1$. We also recall that according to the Stone-Weierstrass theorem, the space of linear combinations of such functionals is dense in the space of continuous functions on $\mathcal{M}_1$.

Proof of Theorem 6.1 Imagine that each ancestor is assigned a type τ with values in $[0,1]$, and that types are transmitted from parents to children. We write $\rho^N(t)$ for the empirical distribution of types at time t, i.e.

$$\rho^N(t) = \frac{1}{N}\sum_{i=1}^N \delta_{\tau_i(t)}$$

where $\tau_i(t)$ denotes the type of the i-th individual at time t. It should be plain from the description of the model that $(\rho^N(t): t\geq 0)$ is a Markov chain in continuous time on the finite subspace of $\mathcal{M}_1$ consisting of linear combinations of Dirac point masses where the masses of atoms are multiples of $1/N$.

We also introduce the following notation. Let m be a probability measure as above, say $m = N^{-1}\sum_{i=1}^n \delta_{x_i}$ where $x_1,\dots,x_N$ are the atoms of Nm repeated according to their multiplicity. For every $y\in[0,1]$ and $k\in\mathbb{N}$, we denote by $m(N,k,y)$ the empirical distribution $N^{-1}\sum_{i=1}^N \delta_{x'_i}$ where the x'_i are N atoms obtained by sampling without replacement from the set of $N+k$ points $x_1,\dots,x_N,x_{N+1},\dots,x_{N+k}$ with $x_{N+1},\dots,x_{N+k}=y$.

The mathematical translation of the description of the evolution of the population model is that the infinitesimal generator L^N is given by

$$L^N\Phi_f(m) = \sum_{k=1}^\infty \mu(k)\int_{[0,1]} \left(\mathbb{E}\left(\Phi_f\left(m(N,k,y)\right)\right) - \Phi_f(m)\right) m(\mathrm{d}y)\,.$$

It is easy to derive by an application of Lemma 6.1 that

$$L^N\Phi_f(m) = \sum_{k=1}^{\infty} \mu(k) \int_{[0,1]} \left(\Phi_f\left(\frac{(Nm + k\delta_y}{N+k} \right) - \Phi_f(m) \right) m(\mathrm{d}y) + r(N)\,,$$

where $r(N) = O(N^{-1})$, and even $r(N) = O(N^{-2})$ whenever $\sum_{k=1}^{\infty} k\mu(k) < \infty$.

Now we divide this quantity by $\bar{\mu}(N)$, which corresponds, in terms of the Markov process, to speeding up time by a factor $1/\bar{\mu}(N)$. Assume that the function f which appears in the definition of the functional Φ_f is $\mathcal{C}^1$, and let N tend to ∞. The assumption that the tail $\bar{\mu}$ is regularly varying with index $-\alpha$ means that the sequence of point measures

$$\sum_k \frac{\mu(k/N)}{\bar{\mu}(N)} \delta_{k/N}$$

converges vaguely as $N \to \infty$ to $\alpha x^{-1-\alpha}\mathrm{d}x$, and hence one sees[1] that

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{\bar{\mu}(N)} L^N\Phi_f(m) \\
&= \alpha \int_0^{\infty} \mathrm{d}x\, x^{-\alpha-1} \int_{[0,1]} m(\mathrm{d}y) \left(\Phi_f\left(\frac{m + x\delta_y}{1+x} \right) - \Phi_f(m) \right) \\
&= \alpha \int_0^1 \mathrm{d}z\, z^{-\alpha-1}(1-z)^{\alpha-1} \int_{[0,1]} m(\mathrm{d}y) \left(\Phi_f\left((1-z)m + z\delta_y \right) - \Phi_f(m) \right)\,.
\end{aligned}$$

We now recognize the generator of the generalized Fleming-Viot process corresponding to the measure $\nu(\mathrm{d}z) = \alpha z^{-\alpha-1}(1-z)^{\alpha-1}\mathrm{d}z$.

An application of the duality lemma enables us to conclude that the semi-group of $\Pi^N_{\lfloor [n] \rfloor}(t/\bar{\mu}(N))$ converges to that of a beta$(2-\alpha, \alpha)$ coalescent. □

In the case when the tail distribution $\bar{\mu}$ is regularly varying with index $-\alpha$ with $\alpha > 2$, one can also prove that under a different rescaling of time, the limit genealogy exists and is described by Kingmans coalescent.

6.2 Stable continuous stable branching processes

6.2.1 Background on CSBP

Continuous State Branching Processes, in short CSBP, are a class of time-homogeneous Markov processes $(X_t, t \geq 0)$ with values in $\mathbb{R}_+$, which have been introduced by Jirina as

[1] Checking the convergence of the integral near $x = 0$ requires some care. More precisely, one has to observe that

$$\int_{[0,1]} \left(\Phi_f\left(\frac{m + y\delta_y}{1+x} \right) - \Phi_f(m) \right) m(\mathrm{d}y) = O(x^2)\,,$$

this bound follows easily from the definition of the functional Φ_f.

the weak limits of rescaled Galton-Watson processes. Their semigroup can be characterized via its Laplace transform as follows : For every $q > 0$ and $a \in \mathbb{R}_+$,

$$\mathbb{E}\left(\mathrm{e}^{-qX_t} \mid X_0 = a\right) \;=\; \exp\left\{-au_t(q)\right\} \tag{6.1}$$

where the function $u_t(q)$ solves

$$\frac{\partial u_t(q)}{\partial t} \;=\; -\Psi(u_t(q)) \quad , \quad u_0(q) \;=\; q\,, \tag{6.2}$$

and $\Psi : [0,\infty) \to \mathbb{R}$ is a function known as the branching mechanism.

To emphasize the role of the initial value, let us henceforth write $X_t = X(t,a)$ when the CSBP starts from $X_0 = a \in \mathbb{R}_+$. By the fundamental branching property of CSBP, if $X'(\cdot,b)$ is independent of $X(\cdot,a)$ and has the same distribution as $X(\cdot,b)$, then $X(\cdot,a) + X'(\cdot,b)$ has the same law as $X(\cdot,a+b)$. Invoking Kolmogorov's theorem, we can thus construct a process $(X(t,a), t \geq 0$ and $a \geq 0)$ such that $X(\cdot,0) = 0$ and, for every $a, b \geq 0$, $X(\cdot,a+b) - X(\cdot,a)$ is independent of the family of processes $(X(\cdot,c), 0 \leq c \leq a)$ and has the law of a CSBP with branching mechanism Ψ started from b.

In particular, for each fixed $t \geq 0$, the process $x \to X(t,x)$ has independent and homogeneous increments with values in $[0,\infty)$. We may (and will) choose its right-continuous modification which is then a subordinator. We see from (6.1) that its Laplace exponent is the function $q \to u_t(q)$. The semigroup identity for CSBP

$$u_{t+s}(q) \;=\; u_t\left(u_s(q)\right) \tag{6.3}$$

points at the connection with Bochner's subordination. Indeed, (6.3) implies that the subordinator $X(t+s,\cdot)$ has the same distribution as the compound process $X'\left(s, X(t,\cdot)\right)$ where $X'(s,\cdot)$ is an independent copy of $X(s,\cdot)$. In this direction, it may be convenient to think of the bivariate process $X(t,a)$ as a measure valued process $(X(t,\mathrm{d}a) : t \geq 0)$ where the notation $X(t,\mathrm{d}a)$ refers to the Stieltjes-measure associated to the increasing process $X(t,\cdot)$. Recall from the Lévy-Itô decomposition that the latter can be described in terms of the Lebesgue measure (corresponding to the drift) and a Poisson point measure (corresponding to the jumps).

6.2.2 Stable CSBP and beta Fleming-Viot processes

In the special case when the branching mechanism is $\Psi(q) = q \ln q$ (which is known as Neveu's branching mechanism), one gets by solving (6.2) that

$$u_t(q) = q^{\mathrm{e}^{-t}}\,.$$

Hence the process $X(t,\cdot)$ is a standard stable subordinator with index e^{-t}. In particular the rescaled measure-valued process

$$X(t,1)^{-1}X(t,\mathrm{d}a)\,, \qquad a \in [0,1]$$

is simply the generalized Fleming-Viot process that is dual to the Bolthausen-Sznitman coalescent. This remarkable feature has its roots in [10] and provided the main impulse

for the analysis of the relations between exchangeable coalescents and stochastic flows of bridges.

In this section, we shall present an extension of this result due to Birkner *al.* [16] which connects more generally stable CSBP with beta-generalized Fleming-Viot processes up-to a time substitution. We will henceforth focus on the (stable) case

$$\Psi(q) = cq^\alpha = c' \int_0^\infty (\mathrm{e}^{-qx} - 1 + qx) x^{-\alpha-1} \mathrm{d}x$$

for some $\alpha \in]1, 2[$. Solving the differential equation (6.2) is easy and one gets

$$u_t(q) = \left(q^{-\alpha} + c''t\right)^{-1/\alpha} .$$

Viewing $q \to u_t(q)$ as a Laplace exponent, we observe that $u_t(\infty) = (c''t)^{-1/\alpha}$. In particular the drift coefficient is zero and the Lévy measure has finite total mass; in other words, the subordinator $X(t, \cdot)$ is a compound Poisson process. Observe that $\mathbb{P}(X(t, 1) = 0) > 0$, i.e. the probability of extinction at time t is strictly positive. Note also that

$$\partial u_t(q)/\partial q_{|q=0} = 1 ,$$

i.e. $\mathbb{E}(X(t, a) = a$ and the process $X(\cdot, a)$ is a martingale.

It is also immediately seen that if G denotes the infinitesimal generator and $f_q(x) = \mathrm{e}^{-qx}$, then

$$\begin{aligned} Gf_q(a) &= \lim_{t\to 0+} t^{-1}\mathbb{E}(\mathrm{e}^{-qX(t,a)} - \mathrm{e}^{-qa}) \\ &= \lim_{t\to 0+} t^{-1}(\mathrm{e}^{-au_t(q)} - \mathrm{e}^{-qa}) \\ &= a\Psi(q)\mathrm{e}^{-qa} \\ &= c'a \int_0^\infty (f_q(a + y) - f_q(a) + yf'_q(a))y^{-\alpha-1}\mathrm{d}x \end{aligned}$$

and this identity can then be extended to smooth functions by linearity and density. In the framework of stochastic calculus, the meaning of this identity is that the stable(α) CSBP is a pure jump martingale with only positive jumps, and the rate of jumps of size $x > 0$ when the process is at a is $c'ax^{-\alpha-1}\mathrm{d}x$. In other words, the predictable compensator[2] of the point measure of its jumps $\Delta_s X(s, a) = X(s, a) - X(s-, a)$, i.e.

$$\sum \delta_{(s,\Delta_s X(s,a))}(\mathrm{d}s, \mathrm{d}x) ,$$

is

$$c'X(t, a) \left(\int_0^\infty \mathrm{d}y y^{-\alpha-1} \delta_y(\mathrm{d}x) \right) \mathrm{d}t .$$

[2]This means that for every process $(H(t, x) : t \geq 0, x \geq 0)$ with values in $\mathbb{R}_+$, which is adapted and left-continuous in the first variable, the process

$$t \to \sum_{s\leq t} H(s, \Delta_s X(s, a)) - c' \int_0^t \mathrm{d}s X(s, a) \int_0^\infty \mathrm{d}y y^{-\alpha-1} H(s, y)$$

is a martingale whenever each term in the difference has a finite expectation.

More generally, the measure-valued process $(X(t, \mathrm{d}a) : t \geq 0)$ is a pure-jump martingale, in the sense that for every continuous function $f : \mathbb{R}_+ \to \mathbb{R}$ with compact support, the real-valued process $\langle X(t, \mathrm{d}a), f\rangle$ is a pure-jump martingale. Furthermore, the jumps $\Delta_t X(t, \mathrm{d}a) = X(t, \mathrm{d}a) - X(t-, \mathrm{d}a)$ with value $y\delta_b(\mathrm{d}a)$ occur when $X(t-, \mathrm{d}a) = m(\mathrm{d}a)$ with intensity

$$c' y^{-1-\alpha} \mathrm{d}y m(\mathrm{d}b)\mathrm{d}t\,,$$

and no other types of jump occurs. Technically, this means that the predictable compensator of the point measure of its jumps

$$\sum \delta_{(t,\Delta\langle X(t,\mathrm{d}a),f\rangle)}(\mathrm{d}s, \mathrm{d}x)$$

is

$$c' \left(\int_0^\infty \mathrm{d}y y^{-\alpha-1} \int X(t, \mathrm{d}a)\delta_{yf(a)}(\mathrm{d}x) \right) \mathrm{d}t\,.$$

We may now state the following result which is excerpt from Birkner *et al.* [16].

Theorem 6.2 *Suppose $1 < \alpha < 2$. Fix the initial size of the population $a = 1$ and introduce for every $t \geq 0$ a random probability measure on $[0, 1]$ as the ratio*

$$R(t, \mathrm{d}x) = X(t, \mathrm{d}x)/X(t, 1)\,, \qquad x \in [0, 1]\,.$$

The random function

$$t \to T_t := \int_0^t X(s, a)^{1-\alpha} \mathrm{d}s$$

is a bijection on $\mathbb{R}_+$ a.s., and if we denote by T^{-1} its inverse, then the probability valued process $(R(T_t^{-1}, \mathrm{d}x) : t \geq 0)$ is a generalized Fleming-Viot process dual to a beta$(2-\alpha, \alpha)$ coalescent.

Sketch of the proof We start by observing that for any $0 < a < b < 1$, the process $X(t, [a, b[)/X(t, 1)$ is a pure jump martingale. Indeed, if we set

$$Y_t = X(t, [a, b[) \text{ and } Y'_t = X(t, 1) - X(t, [a, b[) = X(t, [0, 1]\backslash[a, b[)\,,$$

then Y and Y' are independent by the branching property. Each process is a pure jump martingale; and more precisely, when $Y_t = y$ and $Y'_t = y'$, Y makes a jump, say of size $\mathrm{d}z$, at time t with intensity $cyz^{-1-\alpha}\mathrm{d}z\mathrm{d}t$, while this occurs for Y' with intensity $cy'z^{-1-\alpha}\mathrm{d}z\mathrm{d}t$. Now consider the mean effect of such a jump for the ratio

$$\frac{X(t, [a, b[)}{X(t, 1)} = \frac{Y_t}{Y_t + Y'_t}\,.$$

If Y has a jump of size z, the increment of the ratio is $y'z/[(y + y')(y + y' + z)]$, while if Y' has a jump of size z, the increment of the ratio is $-yz/[(y + y')(y + y' + z)]$. Taking into account the intensities of jumps, we find that the expected value of the increment is 0 and hence $Y/(Y + Y')$ is indeed a (pure jump) martingale.

It follows that more generally, for every bounded measurable function $f : [0,1] \to \mathbb{R}$, the process $\langle X(t,\mathrm{d}x), f\rangle / X(t,1)$ is also a pure jump martingale. The time substitution by T^{-1} preserves the martingale property, so measure-valued process $R(t,\mathrm{d}x)$ is a martingale. We can compute the predictable compensator of the point measure of its jumps from that of $X(t,\mathrm{d}x)$. Recall that, roughly speaking, if $T_t^{-1} = s$ and $X(s-,\mathrm{d}x) = \mu(\mathrm{d}x)$, then $X(s,\mathrm{d}x) - \mu(\mathrm{d}x) = z\delta_a(\mathrm{d}x)$ with rate $cz^{-1-\alpha}\mu(\mathrm{d}a)\mathrm{d}z$. By Itô formula, this implies that is Φ is a smooth functional on the space of probability measures on $[0,1]$, then

$$\Phi(R_t) - c' \int_0^t \mathrm{d}s \int_0^\infty \mathrm{d}z z^{-1-\alpha} \int_{[0,1]} X(s,\mathrm{d}x) \left(\Phi\left(\frac{X(s,1)}{X(s,1)+z} R_s + \frac{z}{X(s,1)+z} \delta_x \right) - \Phi(R_s) \right)$$

is a martingale where $R_t = R(t,\mathrm{d}x)$. The change of variables $y = z/(X(s,1)+z)$ now yields that

$$\Phi(R_t) - c' \int_0^t \mathrm{d}s X(s,1)^{1-\alpha} \int_0^1 \mathrm{d}y y^{-1-\alpha}(1-y)^{\alpha-1} \int_{[0,1]} R(s,\mathrm{d}x) \left(\Phi\left((1-y)R_s + y\delta_x \right) - \Phi(R_s) \right)$$

is a martingale.

Finally the time substitution by T_t^{-1} enables us to see that if $R'_t = R(T_t^{-1},\mathrm{d}x)$, then

$$\Phi(R'_t) - c \int_0^t \mathrm{d}s \int_0^1 \mathrm{d}y y^{-1-\alpha}(1-y)^{\alpha-1} \int_{[0,1]} R'(s,\mathrm{d}x) \left(\Phi\left((1-y)R'_s + y\delta_x \right) - \Phi(R'_s) \right)$$

is a martingale. We recognize the martingale problem for the beta$(2-\alpha,\alpha))$ generalized Fleming-Viot process, which completes the proof as we know that this martingale problem is well-posed. □

In the same direction, Berestycki *et al.* [8] pointed at the fact that in some sense, the genealogy of a beta-coalescent can be embedded into a continuous random tree with stable branching mechanism, by deriving the Donnelly-Kurtz lookdown process from a stable continuous random tree. We also stress that the main result of Birnker *et al.* (cf. Theorem 1.1 there) contains much more than what has been presented here; in particular these authors deal with the case $\alpha < 1$ which present specific difficulties due to the fact that the stable CSPB then explodes a.s.

6.2.3 Beta Fleming-Viot processes and stable CSBP

In a converse direction, one can show that stable CSBP can also arise as limits of Beta Fleming-Viot processes when time shrinks and space expands in an appropriate regime.

Proposition 6.1 *Let $\alpha \in]1,2[$ and $(\rho_t : t \geq 0)$ be a generalized Fleming-Viot process which is dual to a beta coalescent with parameter $(2-\alpha,\alpha)$. For every $\varepsilon > 0$, denote by $X^\varepsilon(t,\mathrm{d}x)$ the image of the restriction of $\varepsilon^{-1}\rho_{\varepsilon^{\alpha-1}t}$ to $[0,\varepsilon]$ by the dilation $x \to x/\varepsilon$. In other words, for every bounded measurable set $A \subseteq [0,1]$, we define*

$$X^\varepsilon(t,A) = \frac{1}{\varepsilon} \rho_{\varepsilon^{\alpha-1}t}(\varepsilon A)\,.$$

When $\varepsilon \to 0+$, the measure-valued process $(X^\varepsilon(t, \mathrm{d}x) : t \geq 0)$ converges in distribution to the process of random measures on $[0, 1]$ induced by an α-stable CSBP .

In fact, similar limit theorems hold for a fairly large class of generalized Fleming-Viot processes whose prototype are the beta's; see [13] for the complete story, including a proof that relies on stochastic calculus. Alternatively, one can also derive Proposition 6.1 from a tedious analysis of the asymptotic behavior of the infinitesimal generator of generalized Fleming-Viot processes (cf. Theorem 4.2). Let us just provide a rough idea for explaining Proposition 6.1.

Fix $\varepsilon > 0$ and consider the generalized Fleming-Viot process speeded up by a factor $\varepsilon^{\alpha-1}$, that is $\rho_{\varepsilon^{\alpha-1}t}$. We know from Theorem 4.2 that the latter is a martingale measure and has jump $y(\delta_a - \rho_{\varepsilon^{\alpha-1}t})$ at time t with intensity

$$c\varepsilon^{\alpha-1}\mathrm{d}t\rho_{\varepsilon^{\alpha-1}t}(\mathrm{d}a)y^{-1-\alpha}(1-y)^{1-\alpha}\mathrm{d}y\,, \qquad 0 < y < 1\,.$$

In the notation of Proposition 6.1, such a jump of $\rho_{\varepsilon^{\alpha-1}t}$ corresponds to a jump of $X^\varepsilon(t, \cdot)$ of size

$$z\left(\delta_b - X^\varepsilon(t, \cdot)\right)\,,$$

where $b = a/\varepsilon$ and $z = y/\varepsilon$, and the intensity at which this jump occurs can be re-expressed as

$$c\varepsilon^{\alpha-1}\mathrm{d}t\varepsilon X^\varepsilon(t, \mathrm{d}b)y^{-1-\alpha}(1-y)^{1-\alpha}\mathrm{d}y\,, \qquad 0 < y < 1\,.$$

The change of variables $y = z\varepsilon$ finally yields the expression

$$c\mathrm{d}tX^\varepsilon(t, \mathrm{d}b)z^{-1-\alpha}(1-\varepsilon z)^{1-\alpha}\mathrm{d}z\,, \qquad 0 < z < 1/\varepsilon\,.$$

When $\varepsilon \to 0+$, the factor $(1-\varepsilon z)^{1-\alpha}$ tends to 1, and we recognize the intensity of jump of a stable(α)-CSBP. Of course, this is only an informal justification, the rigorous argument requires further substantial explanations involving in particular tightness properties.

Proposition 6.1 can be used to derive precise information on the sizes of blocks in a beta$(2-\alpha, \alpha)$-coalescent in small time. We denote by $\lambda_1(dr)$ the Lévy measure of the subordinator $(X(1, x), x \geq 0)$. This means that for every $q \geq 0$ we have

$$\mathbb{E}(\exp -qX(1, x)) = \exp(-xu_1(q)) = \exp(-x\left(q^{-\alpha} + c''t\right)^{-1/\alpha})\,.$$

and

$$u_1(q) = \int (1 - \mathrm{e}^{-qr})\lambda_1(dr)\,.$$

Theorem 6.3 *For every $t \geq 0$ and $r \in [0, \infty]$, denote by $N_t(]0, r[)$ the number of blocks at time t with frequencies less than r in a Λ-coalescent started from the partition of $\mathbb{N}$ in singletons. Then,*

$$\sup_{x \in]0,\infty[} \left|\varepsilon N_{\varepsilon^{\alpha-1}}(]0, \varepsilon x[) - \lambda_1(]0, x[)\right| \to 0 \qquad \text{as } \varepsilon \to 0$$

in probability.

This result relies mainly on Proposition 6.1 and exchangeability; see [13] for the proof, and also [7] where similar results where convergence is reinforced to the almost sure sense.

6.3 Further properties related to the allelic partition

One of most remarkable applications of the coalescent of Kingman is that it provides a deep explanation of the important fact that the allelic partition induced by rare neutral mutations in a variety of population models with large sizes can be described by a Poisson-Dirichlet or GEM partition. Unfortunately, there is no such precise description of the distribution of the allelic partition for other exchangeable coalescent processes (except the degenerate star-shaped coalescent); see Möhle [43]. We now conclude this section by mentioning briefly some advances which have been established recently by several authors in this field for beta coalescents. Proofs may be tedious; we shall merely state the results and refer to the papers for complete arguments and further material.

Let us first recall the framework. We consider a beta$(2-\alpha, \alpha)$ coalescent with $1 < \alpha < 2$, and decide to fix the constant c to be $1/(\Gamma(\alpha)\Gamma(2-\alpha))$, so that the finite measure Λ is now a probability. We can think of the coalescent process as a random tree, where the leaves correspond to the integers and such that the tree reduced to the first n leaves is simply given by the coalescing ancestral lineages of the first n individuals. Now imagine that one superposes to the tree structure a Poisson point process of marks with rate $\theta > 0$, which we think of as neutral mutations.

Cutting the coalescent tree at each mark induces the allelic partition of the set of leaves. We call a connected component an allelic family. For every integer k, let us denote by $N_k(n)$ the number of allelic families of size k on the coalescent tree with n leaves. Berestycki *et al.* [8] have obtained a precise estimate of $N_k(n)$ when n tends to ∞:

Theorem 6.4 *Suppose neutral mutations occur with rate $\theta > 0$ on the beta$(2-\alpha, \alpha)$ coalescent tree, where $1 < \alpha < 2$. Then for every $k \in \mathbb{N}$, the number of allelic families of size k on the reduced tree with n leaves fulfills*

$$\lim_{n\to\infty} n^{\alpha-2} N_k(n) = \theta\alpha(\alpha-1)^2 \frac{\Gamma(k+\alpha-2)}{k!}\,,$$

where the convergence holds in probability.

In the case of Kingman coalescent, we know that the allelic partition is distributed according to the Poisson-Dirichlet law with parameter $(0, 2\theta)$, and it is well-known that for any $k \in \mathbb{N}$, the k-tuple $(N_1(n), \ldots, N_k(n)$ converges in distribution to $(N_1, \ldots, N_k)$ where $N_1, \ldots, N_k$ are independent Poisson variables such that N_i has parameter $2\theta/i$

It is interesting to compare Theorem 6.4 with the result obtained by Basdevant and Goldschmidt [4] for the Bolthausen-Sznitman coalescent, which corresponds to the boundary case $\alpha = 1$. In that case, one has for $k = 1$

$$\lim_{n\to\infty} \frac{\ln n}{n} N_1(n) = \theta\,,$$

while for $k \geq 2$

$$\lim_{n\to\infty} \frac{(\ln n)^2}{n} N_k(n) = \frac{\theta}{k(k-1)}\,,$$

where in both cases, the convergence holds in probability.

Berestycki *et al.* also obtained another result in the same vein, for the so-called site frequency spectrum. Imagine now that each mutation affects a different locus on the DNA, and denote for every integer k by $M_k(n)$ the number of mutations affecting exactly k individuals in the n sample. Then, even though $N_k(n)$ and $M_k(n)$ can be quite different, one also has that

$$\lim_{n\to\infty} n^{\alpha-2} M_k(n) = \theta\alpha(\alpha-1)^2 \frac{\Gamma(k+\alpha-2)}{k!}\,,$$

where the convergence holds in probability.

Another natural notion in this setting is that of segregating sites, i.e. sites at which there exists more than one allele in our sample of chromosomes. A quantity of interest is the number of segregating sites, $S(n)$. In other words, $S(n)$ counts mutations along the skeleton of the coalescent tree, discarding any mutation which affects all individuals. It is easily seen that this quantity is closely related to the total length of the coalescent tree cut at the most recent common ancestor. More precisely, one observes first that when $1 < \alpha < 2$, fixation occur for beta$(2-\alpha, \alpha)$-coalescents (i.e. they come down from infinity) by checking Schweinsberg's criterion. In particular there is a MRCA, and the coalescent tree logged to the MRCA has a finite height but infinite total length. An application of the law of large numbers for Poisson processes then shows that the number of segregating sites $S(n)$ is close to the rate of mutation times the total length of the tree reduced to the first n leaves.

In the case of the Bolthausen-Sznitman coalescent, Drmota *et al.* [28] have proved that $n^{-1}\ln nS(n)$ converges in probability to the rate of mutation θ, and have also been able to determine the fluctuations.

In the case of beta$(2-\alpha, \alpha)$ coalescents, the first order asymptotic was obtained by Berestycki *et al.* [7] : $n^{\alpha-2}S(n)$ converges in probability to $\alpha(\alpha-1)\Gamma(\alpha)/(2-\alpha)$. Essentially, this can be derived from Theorem 6.3 specialized for $x = \infty$. Fluctuations have then been characterized by Delmas *et al.* [24]. See also the references cited there, in particular [32, 35].

Appendix : Background on Poisson random measures

For the reader's convenience, we recall here some basic facts about Poisson random measures.

Let E be a Polish space and μ a sigma-finite measure on E. We call a random measure $\mathtt{M}$ on E *a Poisson measure with intensity* μ if $\mathtt{M}$ fulfills the following requirements. For every Borel subset B of E with $\mu(B) < \infty$, $\mathtt{M}(B)$ has a Poisson distribution with parameter $\mu(B)$, and if $B_1, \ldots, B_n$ are disjoint Borel sets, the variables $\mathtt{M}(B_1), \ldots, \mathtt{M}(B_n)$ are independent. Plainly, $\mathtt{M}$ is then a sum of Dirac point masses, that is we can express $\mathtt{M}$ in the form

$$\mathtt{M} \,=\, \sum_{i\in I} \delta_{\mathtt{a}_i} \tag{6.4}$$

where δ_x stands for the Dirac point mass at x. The $\mathtt{a}_i$ will be referred to as the *atoms* of $\mathtt{M}$. If $\mu(E)$ is finite, then the set of atoms is finite a.s., and more precisely its cardinal $N := \mathtt{M}(E)$ follows the Poisson law with parameter $\mu(E)$. In the representation (6.4), we may then choose $\mathtt{a}_1, \ldots$ to be a sequence of i.i.d. variables with common law $\mu(\cdot)/\mu(E)$ and independent of N, and $I = \{1, \ldots, N\}$. If $\mu(E) = \infty$, then there are infinitely many atoms a.s., so we may take $I = \mathbb{N}$. Furthermore, a similar description of the atoms can be obtained using the elementary superposition property of Poisson measure (cf. Lemma 6.3 below) and expressing the intensity measure in the form $\mu = \sum_{n=1}^{\infty} \mu_n$, where each μ_n is a finite measure.

We now recall three key formulas for the computation of moments, Laplace transforms and distributions related to Poisson point measures. In this direction, we use the convention $\exp(-\infty) = 0$. Consider a measurable map $f : E \to \mathbb{R}$ and write

$$\langle \mathtt{M}, f\rangle \,:=\, \int_E f(x)\mathtt{M}(\mathrm{d}x) \,=\, \sum_{i\in I} f(\mathtt{a}_i)\,.$$

Lemma 6.2 (i) *Suppose either $f \geq 0$ or $f \in L^1(\mu)$. Then we have the* **first moment formula***:*

$$\mathbb{E}\left(\langle \mathtt{M}, f\rangle\right) \,=\, \int_E f(x)\mu(\mathrm{d}x)\,.$$

(ii) *Suppose either $f \geq 0$ or $1 - \mathrm{e}^{-f} \in L^1(\mu)$. Then we have the* **Campbell formula***:*

$$\mathbb{E}\left(\exp\left(-\langle \mathtt{M}, f\rangle\right)\right) \,=\, \exp\left(-\int_E (1 - \mathrm{e}^{-f(x)})\mu(\mathrm{d}x)\right)\,.$$

(iii) *Let E_{p} denote the space of point measures on E, and $G : E \times E_{\mathrm{p}} \to \mathbb{R}_+$ be some measurable functional. Then we have the* **Palm formula***:*

$$\mathbb{E}\left(\langle \mathtt{M}, fG(\cdot, \mathtt{M})\rangle\right) \,:=\, \mathbb{E}\left(\sum_{i\in I} f(\mathtt{a}_i)G(\mathtt{a}_i, \mathtt{M})\right) \,=\, \int_E \mathbb{E}(G(x, \delta_x + \mathtt{M}))f(x)\mu(\mathrm{d}x)\,.$$

Loosely speaking, the Palm formula can be interpreted as follows. Suppose that the intensity measure μ gives no mass to some point x. Then the conditional distribution of $\mathtt{M}$, given that $\mathtt{M}$ has an atom at x, is that of $\delta_x + \mathtt{M}$. Of course, such conditioning is only formal as the probability for $\mathtt{M}$ having an atom at x is 0.

Proof The first moment and Campbell formulas are immediate consequences of the description of the distribution of the atoms of $\mathtt{M}$. Let us now establish the Palm formula. Plainly, it suffices to consider functionals G which depend only on the point measure $\mathtt{M}$ and which are of exponential type, that is

$$G(\mathtt{M}) \;=\; \exp\left(-\langle \mathtt{M}, g\rangle\right) \;,$$

where $g : E \to \mathbb{R}_+$ stands for a generic measurable function. The Campbell formula then gives for every $q \geq 0$

$$\mathbb{E}\left(\exp(-q\sum_{i\in I} f(\mathtt{a}_i))G(\mathtt{M})\right) \;=\; \exp\left(-\int_E (1-\mathrm{e}^{-(qf(x)+g(x))})\mu(\mathrm{d}x)\right) \;.$$

Taking the derivative in the variable q at 0 yields

$$\begin{aligned}\mathbb{E}\left(G(\mathtt{M})\sum_{i\in I} f(\mathtt{a}_i)\right) &= \int_E \mu(\mathrm{d}x) f(x)\mathrm{e}^{-g(x)} \exp\left(-\int_E (1-\mathrm{e}^{-g(y)})\mu(\mathrm{d}y)\right)\\ &= \int_E \mu(\mathrm{d}x) f(x)\mathbb{E}\left(G(\delta_x+\mathtt{M})\right) \;,\end{aligned}$$

where the last equality stems from the Campbell formula. □

Next, we recall the effect of some simple transformations for Poisson random measure.

Lemma 6.3 • (Superposition) *Let μ' be another sigma-finite measure on E and $\mathtt{M}'$ a Poisson point measure with intensity μ' which is independent of $\mathtt{M}$. Then $\mathtt{M}+\mathtt{M}'$ is a Poisson point measure with intensity $\mu+\mu'$.*

• (Image) *Let E' be another Polish space, $f : E \to E'$ a measurable map, and μ' the image measure of μ by f. Then the image of $\mathtt{M}$ by f is a Poisson random measure on E' with intensity μ'.*

• (Marking) *Let ρ be some probability measure on E', and $\mathtt{z}_1, \ldots$ a sequence of i.i.d. variables with law ρ, which is independent of $\mathtt{M}$. Each $\mathtt{z}_i$ is viewed as a mark attached to the atom $\mathtt{a}_i$. Then*

$$\sum_{i\in I} \delta_{(\mathtt{a}_i,\mathtt{z}_i)}$$

is a Poisson random measure on $E \times E'$ with intensity $\mu \otimes \rho$.

• (Change of Probability) *Let $f : E \to \mathbb{R}$ be a measurable function such that $\mathrm{e}^f - 1 \in L^1(\mu)$. Consider the probability $\tilde{\mathbb{P}}$ which is absolutely continuous with respect to $\mathbb{P}$, with density*

$$\exp\left(\langle \mathtt{M}, f\rangle - \langle \mathrm{e}^f - 1, \mu\rangle\right) \;.$$

Then under $\tilde{\mathbb{P}}$, $\mathtt{M}$ is a Poisson random measure with intensity $\mathrm{e}^{f(x)}\mu(\mathrm{d}x)$.

Proof The first three statements are straightforward (see for example Kingman [40]); we shall only provide details for the change of probability property. For every measurable function $g : E \to \mathbb{R}_+$, we deduce from the Campbell formula in Lemma 6.2 that

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\exp\left(-\langle \mathtt{M}, g\rangle\right)\right) &= \mathbb{E}\left(\exp\left(\langle f-g, \mathtt{M}\rangle - \langle \mathrm{e}^f - 1, \mu\rangle\right)\right) \\
&= \exp\left(-\int_E \left(1-\mathrm{e}^{f(x)-g(x)}\right)\mu(\mathrm{d}x) - \int_E \left(\mathrm{e}^{f(x)}-1\right)\mu(\mathrm{d}x)\right) \\
&= \exp\left(-\int_E \left(1-\mathrm{e}^{-g(x)}\right)\mathrm{e}^{f(x)}\mu(\mathrm{d}x)\right) .
\end{aligned}$$

The claim follows. □

References

[1] Aldous, D. J. Exchangeability and related topics. *École d'été de probabilités de Saint-Flour, XIII*, Lecture Notes in Math. **1117**, pp. 1-198, Berlin, Springer, 1985.

[2] Arratia, R., Barbour, A. D., and Tavaré, S. *Logarithmic Combinatorial Structures : a Probabilistic Approach.* Monographs in Mathematics, Zürich, European Mathematical Society, 2003.

[3] Basdevant, A.-L. Ruelle's probability cascades seen as a fragmentation process. *Markov Process. Related Fields* **12** (2006), 447-474.

[4] Basdevant, A. L., and Goldschmidt, C. Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent. *Electronic J. Probab.* **13** (2008), 486-512.

[5] Berestycki, N. *Recent Progress in Coalescent Theory*. *Ensaios Matemáticos* Vol. 16, 2009.

[6] Berestycki, J., Berestycki, N. and Limic, V. The Λ-coalescent speed of coming down from infinity. *Ann. Probab.* **38** (2010), 207-233.

[7] Berestycki, J., Berestycki, N. and Schweinsberg, J. Small-time behavior of beta coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2008), 214-238.

[8] Berestycki, J., Berestycki, N. and Schweinsberg, J. Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35** (2007), 1835-1887.

[9] Berestycki, J., Berestycki, N. and Schweinsberg, J. The genealogy of branching Brownian motion with absorption. Preprint available at `http://arxiv.org/abs/1001.2337v1`

[10] Bertoin, J. and Le Gall, J.-F. The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Relat. Fields* **117** (2000), 249-266.

[11] Bertoin, J. and Le Gall, J.-F. Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields* **126** (2003), 261-288.

[12] Bertoin, J. and Le Gall, J.-F. Stochastic flows associated to coalescent processes, II: Stochastic differential equation. *Ann. Inst. Henri Poincaré, Probabilités et Statistiques* **41** (2005), 307-333.

[13] Bertoin, J. and Le Gall, J.-F. Stochastic flows associated to coalescent processes, III: Limits theorems. *Illinois J. Math.* **50** (2006), no. 1-4, 147–181

[14] Bertoin, J. and Pitman, J. Two coalescents derived from the ranges of stable subordinators. *Elect. J. Probab.* **5** (2000), 1-17. Available via:

`http://www.math.washington.edu /~ejpecp/ejp5contents. html`.

[15] Billingsley, P. *Convergence of Probability Measures*. Second edition. New York, John Wiley & Sons, 1999.

[16] Birkner, M., Blath, J., Capaldo, M., Etheridge, A. M., Möhle, M., Schweinsberg, J., and Wakolbinger, A. Alpha-stable branching and beta-coalescents *Electron. J. Probab.* **10** (2005), 303-325. Available via :

`http://www.math.washington.edu/ ejpecp/viewarticle.php?id=1496 &layout=abstract`

[17] Birkner, M., Blath, J., Möhle, M., Steinrücken, M. and Tams, J. A modified lookdown construction for the Xi-Fleming-Viot process with mutation and populations with recurrent bottlenecks.

[18] Bolthausen, E. and Sznitman, A.-S. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Physics* **197** (1998), 247-276. *Alea* **6** (2009), 25-61.

[19] Bovier, A. *Statistical Mechanics of Disordered Systems A Mathematical Perspective*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2006.

[20] Brunet, E., Derrida, B., Mueller, A. H. and Munier, S. Noisy traveling waves: effect of selection on genealogies. *Europhys. Lett.* **76** (2006), 1-7.

[21] Brunet, E., Derrida, B., Mueller, A. H. and Munier, S. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* **76** (2007), 041104.

[22] Cannings, C. The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. *Adv. Appl. Prob.***6** (1974), 260-290.

[23] Cannings, C. The latent roots of certain Markov chains arising in genetics: a new approach, II. Further haploid models. *Adv. Appl. Prob.***7** (1975), 264-282.

[24] Delmas, J.-F., Dhersin, J.-S., Siri-Jegousse, A. Asymptotic results on the length of coalescent trees. *Ann. Appl. Probab.* **18** (2008), 997-1025.

[25] Dong, R., Goldschmidt, C. and Martin, J. B. Coagulation-fragmentation duality, Poisson-Dirichlet distributions and random recursive trees. *Ann. Appl. Probab.* **16** (2006), 1733–1750.

[26] Donnelly, P. and Kurtz, T. G. Particle representations for measured-valued population models. *Ann. Probab.* **27** (1999), 166-205.

[27] Donnelly, P. and Kurtz, T. G. Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.* **9** (1999), 1091-1148.

[28] Drmota, M., Iksanov, A., Möhle, M. and Rösler, U. Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stochastic Process. Appl.* **117** (2007), 14041421.

[29] Etheridge, A. M. *An Introduction to Superprocesses*. University Lecture Series vol. 20. Providence, American Mathematical Society, 2000.

[30] Ethier, S. N. and Kurtz, T. G. *Markov Processes. Characterization and Convergence.* New York, Wiley, 1986.

[31] Ewens, W. J. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3** (1972), 87-112.

[32] Gnedin, A. and Yakubovich, Y. On the number of collisions in Λ-coalescents. *Electron. J. Probab.* **12** (2007), 1547-1567.

[33] Goldschmidt, C. and Martin, J. B. Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **10** (2005), 718-745. Available via :

`http://www.math.washington.edu/ ejpecp/viewarticle.php?id=1521 &layout=abstract`.

[34] Hoppe, F. M. Pólya-like urns and the Ewens sampling formula. *J. Math. Biol.* **20** (1984), 91-94.

[35] Iksanov, A. and Möhle, M. On a random recursion related to absorption times of death Markov chains. *Adv. in Appl. Probab.* **40** (2008), 206-228.

[36] Jacod, J. and Shiryaev, A. N. *Limit Theorems for Stochastic Processes*. Berlin, Springer, 1987.

[37] Kallenberg, O. Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **27** (1973), 23-36.

[38] Kingman, J. F. C. The population structure associated with the Ewens sampling formula. *Theoretical Population Biol.* **11** (1977), 274-283.

[39] Kingman, J. F. C. The coalescent. *Stochastic Process. Appl.* **13** (1982), 235-248.

[40] Kingman, J. F. C. *Poisson Processes*. Oxford Studies in Probability **3**. Oxford, Clarendon Press, 1993.

[41] Limic, V. On the speed of coming down from infinity for Ξ-coalescent processes. *Electron. J. Probab.* **15** (2010), 217-240. Available via :

`http://www.math.washington.edu/ ejpecp/viewarticle.php?id=2057&layout=abstract.`

[42] Möhle, M. Robustness results for the coalescent, *J. Appl. Probab.* **35** (1998), 438-447.

[43] Möhle, M. On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12** (2006), 35-53.

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

[45] Perkowski, N. (2009). Stochastic population models: measure-valued and partition-valued formulations. Master diploma.

[46] Pfaffelhuber, P. and Wakolbinger, A. The process of most recent common ancestors in an evolving coalescent. *Stoch. Processes Appl.* **116** (2006), 1836-1859.

[47] Pitman, J. Coalescents with multiple collisions. *Ann. Probab.* **27** (1999), 1870-1902.

[48] Pitman, J. *Combinatorial Stochastic Processes. École d'été de Probabilités de St-Flour*, Lect. Notes in Maths **1875**, Berlin, Springer, 2006. Available via :

`http://stat-www.berkeley.edu /users/pitman/`

[49] Pitman, J. and Yor, M. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** (1997), 855-900.

[50] Revuz, D. and Yor, M. *Continuous Martingales and Brownian Motion*, 3rd edn. Berlin, Springer, 1999.

[51] Ruelle, D. A mathematical reformulation of Derrida's REM and GREM. *Comm. Math. Physics* **108** (1987), 225-239.

[52] Sagitov, S. The general coalescent with asynchronous mergers of ancester lines. *J. Appl. Prob.* **36** (1999), 1116-1125.

[53] Schweinsberg, J. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5-12** (2000), 1-50. Available via :

`http://www.math.washington.edu/ ejpecp/viewarticle.php?id=1270 &layout=abstract`

[54] Schweinsberg, J. A necessary and sufficient condition for the Lambda-coalescent to come down from infinity. *Electr. Comm. Probab.* **5-1** (2000), 1-11. Available via :

`http://www.math.washington.edu/ ejpecp/ECP/viewarticle.php?id=1596 &layout=abstract`

[55] Schweinsberg, J. Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.* **106** (2003), 107-139.

[56] Simon, D. and Derrida, B. Evolution of the most recent common ancestor of a population with no selection. *J. Stat. Mech.* (2006), P05002.