UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

Population models with selection and the Bolthausen-Sznitman coalescent

TITULACIÓN POR TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR (A) EN CIENCIAS

PRESENTA:
Alejandro Hernández Wences

Arno Siri-Jégousse
IIMAS, UNAM

Gerónimo Uribe Bravo, Carlos Díaz Avalos
IMATE, UNAM, IIMAS, UNAM

Ciudad Universitaria, CD. MX, Agosto del 2022.

Introduction

The general theory of coalescent processes aims to provide a rigorous mathematical framework that can be used to model natural phenomena where a collection of particles may fuse together and form new particles as the system evolves over time. It has a variety of applications in distinct disciplines such as Physics and Biology. In the biological realm, particularly in the field of Population Genetics, it is used to model the parental relationships of a given population as we track the ancestry of individuals backwards in time, thus leading to the construction of a genealogical tree. In this interpretation the fusion of particles occurs at the time when a set of individuals meets a common ancestor in the past. Once we have a suitable coalescent model describing the genealogy of a population, we can use it to study questions of biological relevance such as determining the time needed to reach the last common ancestor of the population, the expected genetic diversity for neutral positions of the genome, or whether natural selection has played an important role in the evolution of the population. From a mathematical perspective, coalescent processes are Markov processes that take values in the space of partitions of $\mathbb{N}$. In Chapter 1 we follow Bertoin 2006 to layout the foundational concepts of the general theory of exchangeable coalescent processes, and also present some of the main results thereof, including Kingman's representation for random exchangeable partitions, the Poissonian construction of exchangeable coalescent processes, and the characterization of their coagulation rates. At the end of the chapter we leave the general setting to focus on the particular case of simple coalescents (Pitman 1999; Sagitov 1999; Schweinsberg 2000b), and formally define the well known family of Beta coalescents of which the Bolthausen-Sznitman coalescent (BSC) (Bolthausen and Sznitman 1998) is a member. The Bolthausen-Sznitman coalescent is a well known example of a simple coalescent process where multiple particles may fuse in a single event; it was first introduced in the study of spin glasses in physics (Bolthausen and Sznitman 1998) but was rapidly adopted for the study of genealogical trees. It has been described as the limit process for the genealogies of different population evolution models, including models where the reproductive success of the individuals is determined by a fitness function (i.e. the population is under the pressure of natural selection), both in discrete and continuous time (J. Berestycki, N. Berestycki, and Schweinsberg 2013; Birkner, Blath, et al. 2005; Cortines and Mallein 2017; Freund 2020; Huillet and Möhle 2021; Schweinsberg 2003).

In Chapter 2 we layout a general scenery of population evolution models with discrete generations, and characterize their genealogies in terms of coalescent processes. We begin with Cannings' models (Cannings 1974, 1975) which presently are the most widely studied constant-size neutral models due to their manageability and consequent applicability. The defining characteristics of Cannings' models are 1) that the collection of offspring sizes is symmetric, and in particular that all the parents have the same number of children in distribution; and 2) that the reproduction events are i.i.d over time. We then pass to *multinomial models* (Cortines and Mallein 2017; Huillet and Möhle 2021) which share some of the basic intuitions of Cannings' models, along with their manageability, while at the same time allow for "non-symmetrical" offspring distributions. We provide a new criterion for the weak convergence of their genealogies to general Ξ-coalescents, as the total population size N tends to infinity; a criterion that reminisces the homologous criterion provided by Möhle and Sagitov 2001 for the neutral Cannings' models, and whose applicability we show with examples in Sections 2.2.3 and 2.3 where Beta coalescents and the Bolthausen-Sznitman coalescent appear in the limit. In Section 2.3 we introduce the family of exponential models (Brunet and Derrida 1997, 2012) which explicitly incorporate the effect of natural selection by assigning fitness levels to its individuals, determining their reproductive success in the next generation. Following Cortines and Mallein 2017, we prove that the genealogies of the exponential models can be written in terms of Multinomial models, allowing us to describe their weak limit as $N \to \infty$. We find that, under strong selection regimes, the genealogy of these models is once again described by the Bolthausen-Sznitman coalescent, although we also find a novel weak selection regime in which the limit genealogy is a discrete-time Poisson-Dirichlet coalescent. Finally, by the end of the chapter, we briefly present a model in continuous time (Schweinsberg 2017b) in which the population is once again under the effect of mutation and natural selection, and whose limit genealogy is given by the Bolthausen-Sznitman coalescent.

We note that the Bolthausen-Sznitman coalescent appears as the limit genealogy of the three models described in this Chapter 2. Of particular importance is the case of the exponential models and the continuous-time model of Schweinsberg 2017b, which give rigorous examples of a population undergoing natural selection whose limiting genealogy is given by this particular coalescent. This gives further evidence for the intuition that this coalescent can serve as a new null model for the genealogy of rapidly adapting populations, an intuition that has become somewhat widespread in later times (Brunet and Derrida 2012; Cortines and Mallein 2017; Neher and Hallatschek 2013; Schweinsberg 2017b).

Finally, in Chapter 3, we introduce the Site Frequency Spectrum (SFS) of a coalescent process, and its biological interpretation as a measure of the genetic diversity present in a population. The latter, being closely related to the structure of the underlying genealogical tree, motivates the study of the SFS for different coalescent processes, as well as its ubiquitous use as a model selection tool to infer the genealogy of a population from present-day genetic data (Eldon et al. 2015; Freund and Siri-Jégousse 2021; Koskela 2018). We then describe the Random Recursive Tree (RRT) construction of the BSC (Goldschmidt and Martin 2005). This construction allows us to derive explicit (easy-to-compute) formulas for the first and second moments of the SFS for the BSC, leading to corresponding asymptotics as the initial number of particles n tends to infinity, and also allows us to characterize the joint distribution of the lengths of branches associated to families of size b for $n/2 < b < n$ (Kersting, Siri-Jégousse, and H. Wences 2021).

Contents

Chapter 1

Main Concepts in Coalescent Theory

In this chapter we mostly follow Bertoin 2006, filling in the proofs whenever necessary, to layout the general theory of coalescent processes.

1.1 Random Exchangeable Partitions

1.1.1 Basic Definitions

In this section we will define the basic mathematical structures that will help us represent and study coalescent processes.

Definition 1.1.1. *Let A be a subset of $\mathbb{N}$ and π be a countable collection of nonempty subsets of A. We call π a partition of A if*

- $B_i \cap B_j = \emptyset \qquad$ *for all B_i and B_j in π*
- $\cup_{i\geq 1} B_i = A$.

We call $\{B \subset \mathbb{N} : B \in \pi\}$ the blocks of π, and denote by $\pi(k)$ the block that contains the element k.

Given a partition π of $\mathbb{N}$ we can define an equivalence relation in $\mathbb{N}$ by setting

$$i \overset{\pi}{\sim} j \iff \pi(i) = \pi(j).$$

Conversely, given an equivalence relation in $\mathbb{N}$ we can define a partition π whose blocks are the corresponding equivalence classes. Given a set $A' \subset A$ we define the *restriction* of π to A' as

$$\pi\big|_{A'} \coloneqq \{B \cap A' : B \in \pi\}.$$

Also, given a set $A \subset \mathbb{N}$ we will denote the collection of all partitions of A by $\mathscr{P}_A$. We will typically work with the sets $\{1, \ldots, n\}$ so we will denote them by $[n]$, and write $\mathscr{P}_n$ instead of $\mathscr{P}_{[n]}$, $\mathscr{P}_\infty$ instead of $\mathscr{P}_{\mathbb{N}}$, and $\pi\big|_n$ instead of $\pi\big|_{[n]}$. Finally for any pair $n \leq m$ and $\pi \in \mathscr{P}_n$ we will denote by $\mathscr{P}_m(\pi)$ the set of all partitions $\pi' \in \mathscr{P}_m$ such that $\pi'\big|_n = \pi$, i.e.

$$\mathscr{P}_m(\pi) \coloneqq \left\{\pi' \in \mathscr{P}_m \colon \pi'\big|_n = \pi\right\}, \quad n \leq m, \forall \pi \in \mathscr{P}_n. \tag{1.1}$$

The following is a well known characterization of the space $\mathscr{P}_\infty$ in terms of sequences of partitions of the form $\{\pi_n \in \mathscr{P}_n, n \in \mathbb{N}\}$.

Definition 1.1.2. *A sequence of partitions $(\pi_n)_{n\in\mathbb{N}}$ with $\pi_n \in \mathscr{P}_n$ is compatible if $\pi_n\big|_k = \pi_k$ for all $k \leq n$, $n \in \mathbb{N}$.*

Lemma 1.1.1. *A sequence of partitions $(\pi_n)_{n\in\mathbb{N}}$ is compatible if and only if there exists a partition $\pi_\infty \in \mathscr{P}_\infty$ such that $\pi_\infty\big|_n = \pi_n$ for all $n \in \mathbb{N}$.*

Let us now formalize the measurable space in which we will define random partitions. We will consider the set $\mathscr{P}_\infty$ and endow it with a distance function which will allow us to define a Borel σ-algebra.

Definition 1.1.3. *We define a distance δ in $\mathscr{P}_\infty$ by*

$$\delta(\pi_1, \pi_2) = 1/\max\left\{n \in \mathbb{N} : \pi_1\big|_n = \pi_2\big|_n\right\}.$$

Theorem 1.1.2. *$(\mathscr{P}_\infty, \delta)$ is a compact metric space.*

Proof. Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of partitions in $\mathscr{P}_\infty$ and let $\pi^1 := \{1\}$. There exists a partition $\pi^2 \in \mathscr{P}_2$ such that $\pi_n\big|_2 = \pi^2$ for an infinite number of $n \in \mathbb{N}$; let ℓ_1 be one of such indexes. Then, recursively, for every $k \in \mathbb{N}$ we can choose a partition $\pi^k \in \mathscr{P}_k$ such that $\pi^k\big|_j = \pi^j$ for all $j \leq k$ and an index ℓ_{k-1} such that $\pi_{\ell_{k-1}}\big|_k = \pi^k$. The sequence of partitions $\left(\pi^k\right)_{k\in\mathbb{N}}$ is compatible so there exists a partition $\pi^\infty \in \mathscr{P}_\infty$ such that $\pi^\infty\big|_k = \pi^k$ for every $k \in \mathbb{N}$ and, by construction, $\pi_{\ell_k} \xrightarrow{\delta} \pi^\infty$ as $k \to \infty$. $\square$

Definition 1.1.4. *Let $\mathscr{B}(\mathscr{P}_\infty)$ be the Borel σ-algebra in $\mathscr{P}_\infty$ induced by δ. A random partition Π is a random element of $(\mathscr{P}_\infty, \mathscr{B}(\mathscr{P}_\infty))$.*

We will only be concerned with a particular type of random partitions, **exchangeable random partitions**. This type of partitions has a nice representation reminiscent of de Finetti's theorem for exchangeable random sequences. Similar to the context of de Finetti's theorem in which permutations of random sequences are defined, let us first define permutations of partitions.

Definition 1.1.5. *A finite permutation of $\mathbb{N}$ is a bijective function $\sigma : \mathbb{N} \to \mathbb{N}$ with the property that there exists an integer N such that $\sigma(j) = j$ for all $j \geq N$.*

In the following when we refer to a "permutation" we mean a "finite permutation" unless otherwise stated.

Definition 1.1.6. *Let σ be a permutation of $\mathbb{N}$ and π be a partition in $\mathscr{P}_\infty$. We define $\sigma(\pi)$ the permutation of π given by σ to be the partition*

$$\sigma(\pi) := \left\{\sigma^{-1}(B) : B \in \pi\right\}.$$

Note that the blocks of $\sigma(\pi)$ are given by the inverse images of the blocks of π under σ and not by $\{\sigma(B) : B \in \pi\}$; actually, one should not expect that if $i \overset{\pi}{\sim} j$ then $\sigma(i) \overset{\sigma(\pi)}{\sim} \sigma(j)$, but rather that $\sigma^{-1}(i) \overset{\sigma(\pi)}{\sim} \sigma^{-1}(j)$.

Proposition 1.1.3. *Let σ be a permutation of $\mathbb{N}$, then the map $\pi \mapsto \sigma(\pi)$ is continuous and, thus, measurable.*

Proof. By definition there is an $N \in \mathbb{N}$ such that for all $j \geq N$ we have $\sigma(j) = j$. If $\delta(\pi, \pi') \leq 1/M$ with $M \geq N$ we have that $\pi\big|_M = \pi'\big|_M$ and $\sigma(\pi\big|_M) = \sigma(\pi'\big|_M)$. Since $\sigma(\pi\big|_M) = \sigma(\pi)\big|_M$ and $\sigma(\pi'\big|_M) = \sigma(\pi')\big|_M$ we see that $\delta(\sigma(\pi), \sigma(\pi')) \leq 1/M$. □

Definition 1.1.7. *Let Π be a random partition. We say that Π is an exchangeable random partition if for every permutation σ we have*

$$\Pi \stackrel{d}{=} \sigma(\Pi).$$

Note that in particular for all $A \in \mathscr{B}(\mathscr{P}_\infty)$ we have

$$\mathbb{P}(\Pi \in A) = \mathbb{P}(\sigma(\Pi) \in A) = \mathbb{P}(\Pi \in \sigma^{-1}(A)).$$

When working with exchangeable random partitions it will be often the case that their distribution will be specified in terms of their asymptotic frequencies (see Kingman's representation Theorem 1.1.6 below), here we define what we mean by an asymptotic frequency.

Definition 1.1.8. *Let B be any subset of $\mathbb{N}$, and π be any partition of $\mathbb{N}$.*

- *We say that a set B has an asymptotic frequency if the following limit exists*

  $$|B| := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_B(i)$$

  where $\mathbb{1}_B$ is the usual indicator function for B.

- *We say that π has asymptotic frequencies if π is such that $|B|$ exists for all $B \in \pi$. In this case we define $|\pi|^\downarrow$ as the sequence of asymptotic frequencies of π written in decreasing order, and write*

  $$|\pi|^\downarrow = (|\pi|^\downarrow_1, \cdots).$$

  Note that by definition we have $\sum |\pi|^\downarrow_i \leq 1$.

Notice that if π has asymptotic frequencies then for every finite permutation σ of $\mathbb{N}$ we have $|\pi|^\downarrow = |\sigma(\pi)|^\downarrow$ since for every block B of π the block $\sigma^{-1}(B)$ of $\sigma(\pi)$ has the same asymptotic frequency as B. Hence, Definition 1.1.7 of random exchangeable partitions can intuitively be interpreted as saying that the distribution of Π is determined by the distribution of its block sizes $|\Pi|^\downarrow$, and not by their particular composition. In the following Section 1.1.2 we describe how an exchangeable random partition may be constructed from a set of asymptotic frequencies.

1.1.2 Paint-Box Construction

We begin by formalizing the space on which the asymptotic frequencies of a partition $\pi \in \mathscr{P}_\infty$ live.

Definition 1.1.9. *Let $\rho = (\rho_1, \rho_2, \cdots)$ be a sequence of real numbers in $[0,1]$ such that*

- $\rho_i \geq \rho_j$ *for all* $i \leq j$
- $\sum_{i=0}^{\infty} \rho_i \leq 1.$

Such a sequence is called a mass partition. Let $\mathscr{P}_{[0,1]}$ be the space of all mass partitions and endow it with the supremum norm in ℓ_1 and its corresponding Borel σ-algebra $\mathscr{M}$.

If ρ is a mass-partition we define ρ_0 as

$$\rho_0 := 1 - \sum_{i=1}^{\infty} \rho_i.$$

We call ρ_0 the *dust* of ρ. We say that ρ is *proper* if $\rho_0 = 0$ and *improper* otherwise. We will interpret a mass partition as the sequence of strictly positive asymptotic frequencies of a partition π, and ρ_0 as the asymptotic frequency of the set formed by the union of all the blocks of π whose asymptotic frequency is equal to zero (thus justifying the term "dust").

We now describe the **paint-box construction** for random exchangeable partitions. Given a mass partition $\rho = (\rho_1, \rho_2, \cdots)$ we can construct a countable sequence of open intervals $(I_i)_{i\in\mathbb{N}}$ such that

- $I_i \bigcap I_j = \emptyset$ for all $i \neq j$
- $\mathsf{Leb}(I_i) = \rho_i$ for all $i \in \mathbb{N}$
- $\rho_0 = \mathsf{Leb}\big([0,1] \setminus \bigcup I_i\big)$

with Leb being the Lebesgue measure on $[0,1]$. We call such a collection of intervals an *interval representation* of ρ. Conversely, given an open set $\mathcal{U}$ in $[0,1]$ we can find a countable sequence of open intervals $(I_i)_{i\in\mathbb{N}}$ such that

- $\bigcup_{\mathbb{N}} I_i = \mathcal{U}$
- $\sum_{\mathbb{N}} \mathsf{Leb}(I_i) = \mathsf{Leb}(\mathcal{U}) \leq 1$
- $\mathsf{Leb}(I_i) \geq \mathsf{Leb}(I_j)$ for all $i \geq j$

so we can construct a mass partition ρ given by $(\mathsf{Leb}(I_1), \mathsf{Leb}(I_2), \cdots)$. We will now use an interval representation $(I_i)_{i\in\mathbb{N}}$ of a mass partition ρ in order to construct an exchangeable random partition Π. Let A_0 be

$$A_0 = [0,1] \setminus \bigcup I_i.$$

Also, consider a sequence of numbers $(u_i)_{i\in\mathbb{N}} \in [0,1]^{\mathbb{N}}$ and construct π by defining the blocks

$$B_k = \{j \in \mathbb{N} : u_j \in I_k\}, \; k \in \mathbb{N}$$

and setting

$$\pi = \{B_k : k \in \mathbb{N}\} \bigcup \{\{j\} : u_j \in A_0\}.$$

In other words, all the indices $j \in \mathbb{N}$ such that u_j falls in A_0 become singletons whereas all the indices such that u_j falls in I_k constitute the block B_k. Clearly this procedure generates a partition $\pi \in \mathscr{P}_\infty$ for each sequence $(u_1, u_2, \cdots)$, that is, we have a map $h\colon [0,1]^{\mathbb{N}} \to \mathscr{P}_\infty$ which is easily seen to be measurable; thus, if $(U_i)_{i\in\mathbb{N}}$ is a sequence of independent uniformly distributed random variables then $\Pi := h(U_1, U_2, \cdots)$ is a random partition. In order to see that Π is exchangeable we just need to note that for every permutation σ we have

$$(U_1, U_2, \cdots) \stackrel{\mathrm{d}}{=} (U_{\sigma(1)}, U_{\sigma(2)}, \cdots)$$

since $(U_i)_{i\in\mathbb{N}}$ are independent and identically distributed. Also, note that if

$$h(u_1, u_2, \cdots) = \pi$$

then

$$h(u_{\sigma(1)}, u_{\sigma(2)}, \cdots) = \sigma(\pi)$$

and thus $\Pi \stackrel{d}{=} \sigma(\Pi)$, which is the condition for being a random exchangeable partition.

In the next section we formally state Kingman's representation theorem Kingman 1978 which says that any random exchangeable partition can be constructed from paint-box procedures if we randomize, and give an appropriate distribution to the mass-partition ρ.

1.1.3 Kingman's Representation

Lemma 1.1.4. *Let Π be an exchangeable random partition, then Π has asymptotic frequencies almost surely.*

Proof. Recall the notation $\pi(k)$ which gives the block of π that contains the element k (Definition 1.1.1). Fix an index $j \in \mathbb{N}$ and for all $i \neq j$ define the random variable

$$\delta_j^\Pi(i) = \begin{cases} 1 & \text{if } \Pi(i) = \Pi(j) \\ 0 & \text{otherwise,} \end{cases}$$

then $\left(\delta_j^\Pi(i)\right)_{i\neq j}$ is an exchangeable random sequence. Indeed, notice that the exchangeability of Π ensures that for every permutation σ of $\mathbb{N}$ with $\sigma(j) = j$, and any collection of zero-one digits $d_1, \cdots, d_k$ we have

$$\begin{aligned} \mathbb{P}(\delta_j^\Pi(i_1) = d_1, \cdots, \delta_j^\Pi(i_k) = d_k) &= \mathbb{P}(\delta_j^{\sigma(\Pi)}(i_1) = d_1, \cdots, \delta_j^{\sigma(\Pi)}(i_k) = d_k) \\ &= \mathbb{P}(\delta_j^\Pi(\sigma(i_1)) = d_1, \cdots, \delta_j^\Pi(\sigma(i_k)) = d_k) \end{aligned}$$

where the second equality holds since $\sigma(j) = j$ and, therefore, $\sigma(\Pi)(j) = \sigma^{-1}(\Pi(j))$, and $\sigma(\Pi)(i) = \sigma(\Pi)(j)$ if and only if $\Pi(\sigma(i)) = \Pi(j)$. By de Finetti's theorem, the limit

$$\begin{aligned} |\Pi(j)| &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_j^\Pi(i) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n], i\neq j} \delta_j^\Pi(i) \end{aligned}$$

exists almost surely. Since the choice of j was arbitrary we conclude that Π has asymptotic frequencies almost surely. □

Lemma 1.1.5. *If Π is an exchangeable random partition then all the blocks of Π are either singletons, or have an infinite number of elements almost surely.*

Proof. We will prove this by contradiction. Denote by $\#\Pi(j)$ the cardinality of the block $\Pi(j)$. Let $j \in \mathbb{N}$ and assume that $\mathbb{P}(\#\Pi(j) = m) > 0$ for some $m \geq 2$. We will construct a probability measure $\mathbb{P}'$ on $\mathbb{N}^{m-1}$ which will turn out to be "uniform", thus leading to a contradiction. For every $(m-1)$-tuple of integers $(n_1, \cdots, n_{m-1})$ define

$$\mathbb{P}'(n_1, \cdots, n_{m-1}) = \frac{\mathbb{P}(\Pi(j) = \{j, n_1, \cdots, n_{m-1}\})}{\mathbb{P}(\#\Pi(j) = m)}.$$

$\mathbb{P}'$ is easily checked to be a probability measure. Now, if $(\ell_1, \cdots, \ell_{m-1})$ is another collection of integers and σ is a permutation such that $\sigma(n_i) = \ell_i$ and $\sigma(k) = k$ for all $k \notin \{n_1, \cdots, n_{m-1}\}$, then, by the exchangeability of Π, it follows that

$$\begin{aligned}\mathbb{P}'(n_1, \cdots, n_{m-1}) &= \frac{\mathbb{P}(\Pi(j) = \{j, n_1, \cdots, n_{m-1}\})}{\mathbb{P}(\#\Pi(j) = m)} \\ &= \frac{\mathbb{P}(\sigma(\Pi)(j) = \{j, n_1, \cdots, n_{m-1}\})}{\mathbb{P}(\#\Pi(j) = m)} \\ &= \frac{\mathbb{P}(\Pi(j) = \{j, \sigma(n_1), \cdots, \sigma(n_{m-1})\})}{\mathbb{P}(\#\Pi(j) = m)} \\ &= \frac{\mathbb{P}(\Pi(j) = \{j, \ell_1, \cdots, \ell_{m-1}\})}{\mathbb{P}(\#\Pi(j) = m)} \\ &= \mathbb{P}'(\ell_1, \cdots, \ell_{m-1}).\end{aligned}$$

Since this is true for any collection of integers $(\ell_1, \cdots, \ell_{m-1})$ it follows that all the elements of $\mathbb{N}^{m-1}$ have the same probability under $\mathbb{P}'$, which is impossible since $\mathbb{P}'$ is a probability measure and $\mathbb{N}^{m-1}$ is an infinite set. Since the choice of j and m was arbitrary, for all j and $m > 1$ we have

$$\mathbb{P}(\#\Pi(j) = m) = 0.$$

□

The preceding lemmas tell us a lot about the structure of exchangeable random partitions. Almost surely, they take values in a set that is much smaller than all of $\mathscr{P}_\infty$, particularly we may assume that they take values on the measurable set

$$\{\pi \in \mathscr{P}_\infty : \forall B \in \pi,\ |B| \text{ exists and } (|B| = 0 \iff B \text{ is a singleton })\}.$$

Moreover, using the same techniques as in the lemmas above, it is easy to show that, with probability one, the set $\{i \in \mathbb{N} \colon |\pi(i)| = 0\}$ has an asymptotic frequency and is either empty or has an infinite number of elements. We now state without proof Kingman's representation theorem which combines the above two lemmas in order to describe all the possible distributions for random exchangeable partitions.

Theorem 1.1.6 (Kingman's Representation). *Let Π be an exchangeable random partition. Then, there exists a probability measure Q on $\mathscr{P}_{[0,1]}$ such that*

$$\mathbb{P}(\Pi \in A) = \int_{\mathscr{P}_{[0,1]}} \varrho_\rho(A)\ Q(d\rho) \quad , \quad \forall A \in \mathscr{B}(\mathscr{P}_\infty),$$

where we use the notation ϱ_ρ for the probability measure on $\mathscr{P}_\infty$ induced by the paint-box construction directed by ρ introduced in Section 1.1.2.

Finally we give a simple condition for the weak convergence of exchangeable partitions in terms of the weak convergence of their underlying mass-partitions.

Proposition 1.1.7. *For each $k \in \mathbb{N}$ let $\Pi^{(k)}$ be a random exchangeable partition with random asymptotic frequencies $\rho^{(k)} := |\Pi_k|^{\downarrow}$. Then the following are equivalent*

I. *As $k \to \infty$, $\rho^{(k)}$ converges weakly on $\mathscr{P}_{[0,1]}$ to $\rho^{(\infty)}$.*

II. *As $k \to \infty$, $\Pi^{(k)}$ converges weakly on $\mathscr{P}_\infty$ to $\Pi^{(\infty)}$.*

Proof. **I⇒II.** Since the space $\mathscr{P}_{[0,1]}$ is metric and compact we may rather assume, by Skorohod's representation theorem, that

$$\rho^{(k)} \overset{a.s.}{\to} \rho^{(\infty)}. \tag{1.2}$$

For any $k \in \mathbb{N} \cup \{\infty\}$ let

$$I_j^{(k)} := \left[\sum_{i=0}^{j-1} \rho_i^{(k)}, \sum_{i=0}^{j} \rho_i^{(k)}\right)$$

be the natural interval representation of $\rho^{(k)}$, and let $(U_j)_{j\in\mathbb{N}}$ be the shared i.i.d. uniform random variables in the paint-box construction of all the partitions $\Pi^{(k)}$. Then (1.2) implies

$$\mathbb{1}_{\left\{U_i \in I_j^{(k)}\right\}} \overset{a.s.}{\to} \mathbb{1}_{\left\{U_i \in I_j^{(\infty)}\right\}}, \quad \forall i, j \in \mathbb{N},$$

which in turn implies

$$\delta\left(\Pi^{(k)}, \Pi^{(\infty)}\right) \overset{a.s.}{\to} 0.$$

II⇒ I. The space $\mathscr{P}_{[0,1]}$ being compact ensures that any subsequence of $\left(\rho^{(k)}\right)_{k\in\mathbb{N}}$ contains a further subsequence $\left(\rho^{(k_j)}\right)_{j\in\mathbb{N}}$ that converges weakly to some random element $\tilde{\rho}$ in $\mathscr{P}_{[0,1]}$; the implication I⇒II then gives $\tilde{\rho} \overset{d}{=} |\Pi^{(\infty)}|^{\downarrow} = \rho^{(\infty)}$. Since the existence of the weakly convergent sub-subsequence and the identification of the limit hold for any starting subsequence $\left(\rho^{(k)}\right)_{k\in\mathbb{N}}$ we conclude I by an application of Theorem 2.6 in Billingsley 1999. □

1.2 Exchangeable Coalescent Processes

1.2.1 Basic Definitions

Exchangeable coalescents are going to be defined as a family of stochastic processes in continuous time and taking values in $\mathscr{P}_\infty$. The evolution of these processes will be determined by a binary operator defined in $\mathscr{P}_\infty$, the coagulation operator, whose increments will be stationary over time. In order to define the coagulation operator Coag we first need to introduce the ordering of the blocks of a partition π by increasing order of their smallest element. That is, for a partition π we construct an ordered sequence of blocks $(\pi_1, \pi_2, \cdots)$ such that

- $\pi_j \in \pi$ for all $j \in \mathbb{N}$
- $\pi = \bigcup_{i=1}^{\infty} \{\pi_i\}$
- $\min\{k : k \in \pi_j\} \geq \min\{k : k \in \pi_i\}$ for all $j \geq i$.

From now on, when we refer to the kth block of π we mean the kth block under the order just described. Also, we will sometimes use the notation $[\pi]_k$ instead of π_k to emphasize that we are referring to the kth block of π, specially when the notation for the partition within the brackets is large. Finally, we recall the notation $\#A$ for the cardinality of a set A; in particular, if π is a partition, $\#\pi$ gives the number of blocks of π.

Definition 1.2.1 (Coagulation). *Let $\pi' \in \mathscr{P}_m$ with $m \in \mathbb{N} \cup \{\infty\}$. If $\pi \in \mathscr{P}_n$ is such that $|\pi| \leq m$, then the pair (π, π') is called an admissible pair, and we define the coagulation of π and π' as*

$$\mathrm{Coag}(\pi, \pi') = (\hat{\pi}_1, \hat{\pi}_2, \cdots),$$

where $\hat{\pi}_k$ is given by

$$\hat{\pi}_k := \bigcup_{j \in \pi'_k} \pi_j,$$

where π_j is set to $\emptyset$ if $j > |\pi|$.

The following properties of the operator Coag are easily proved.

Lemma 1.2.1. *Let (π, π') and (π', π'') be admissible pairs.*

I. *The operator* Coag *is associative*

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

II. *The operator* Coag *commutes with the restriction operation*

$$\mathrm{Coag}(\pi, \pi')\big|_n = \mathrm{Coag}(\pi\big|_n, \pi') = \mathrm{Coag}(\pi\big|_n, \pi'\big|_n) \tag{1.3}$$

for any $n \in \mathbb{N}$.

III. *The operator* Coag *is Lipschitz-continuous.*

The following theorem states that the operator Coag preserves the exchangeability property of random partitions.

Theorem 1.2.2. *Let Π and Π' be two independent exchangeable random partitions. Then $\hat{\Pi} = \mathrm{Coag}(\Pi, \Pi')$ is an exchangeable partition.*

Proof. Let σ be any permutation, then the blocks of $\sigma\left(\hat{\Pi}\right)$ are given by $\left(\sigma^{-1}(\hat{\Pi}_k)\right)_{k\in\mathbb{N}}$ where

$$\sigma^{-1}(\hat{\Pi}_k) = \sigma^{-1}\left(\bigcup_{j\in\Pi'_k} \Pi_j\right) = \bigcup_{j\in\Pi'_k} \sigma^{-1}(\Pi_j).$$

Note that in general we should not expect that $\sigma^{-1}(\Pi_j) = [\sigma(\Pi)]_j$ so it is not true that $\sigma(\hat{\Pi}) = \text{Coag}(\sigma(\Pi), \Pi')$. However, we can define a map from $\mathscr{P}_\infty$ to the set of all finite permutations of $\mathbb{N}$, $\pi \to \widehat{\sigma_\pi}$, where $\widehat{\sigma_\pi}$ is given by

$$\sigma^{-1}(\pi_j) = [\sigma(\pi)]_{\widehat{\sigma_\pi}(j)}, \quad \forall j \in \mathbb{N}.$$

Since there exists an integer M such that $\sigma(k) = k$ for all $k \geq M$ it follows that $\sigma^{-1}(\pi_k) = \pi_k$ for all $k \geq M$, therefore $\widehat{\sigma_\pi}(k) = k$ for all $k \geq M$ which proves that $\widehat{\sigma_\pi}$ is indeed a permutation. Furthermore, since the latter is true for every partition π, the map just described takes values on the set of permutations of the first $m-1$ integers, which is finite. Now let $\widehat{\sigma_\Pi}$ be its composition with Π. Then $\widehat{\sigma_\Pi}$ is independent of Π' since Π is, and $\widehat{\sigma_\Pi}$ induces a discrete probability measure on the set of all possible permutations. Moreover, for every $k \in \mathbb{N}$ we have

$$\begin{aligned}
\sigma^{-1}\left(\hat{\Pi}_k\right) &= \bigcup_{j\in\Pi'_k} \sigma^{-1}(\Pi_j) \\
&= \bigcup_{j\in\Pi'_k} [\sigma(\Pi)]_{\widehat{\sigma_\Pi}(j)} \\
&= \bigcup_{j\in\widehat{\sigma_\Pi}^{-1}\left(\Pi'_k\right)} [\sigma(\Pi)]_j.
\end{aligned}$$

For every $k \in \mathbb{N}$ there is a unique $\ell \in \mathbb{N}$ such that $[\widehat{\sigma_\Pi}(\Pi')]_\ell = \sigma_\Pi^{-1}\left(\Pi'_k\right)$ and vice versa, thus

$$\sigma(\hat{\Pi}) = \text{Coag}(\sigma(\Pi), \widehat{\sigma_\Pi}(\Pi')).$$

Let $\mathcal{A}$ be the finite range of $\widehat{\sigma_\Pi}$, then, by the independence of $(\Pi, \widehat{\sigma_\Pi})$ and Π', and the exchangeability of Π and Π', for any measurable sets $A, B \in \mathscr{P}_\infty$ we have

$$\begin{aligned}
\mathbb{P}\Big(\sigma(\Pi) \in A, \widehat{\sigma_\Pi}(\Pi') \in B\Big) &= \sum_{\sigma'\in\mathcal{A}} \mathbb{P}\Big(\sigma(\Pi) \in A, \widehat{\sigma_\Pi} = \sigma', \sigma'(\Pi') \in B\Big) \\
&= \sum_{\sigma'\in\mathcal{A}} \mathbb{P}\Big(\sigma(\Pi) \in A, \widehat{\sigma_\Pi} = \sigma'\Big)\mathbb{P}\Big(\sigma'(\Pi') \in B\Big) \\
&= \sum_{\sigma'\in\mathcal{A}} \mathbb{P}\Big(\sigma(\Pi) \in A, \widehat{\sigma_\Pi} = \sigma'\Big)\mathbb{P}\Big(\Pi' \in B\Big) \\
&= \mathbb{P}\Big(\sigma(\Pi) \in A\Big)\mathbb{P}\Big(\Pi' \in B\Big) \\
&= \mathbb{P}\Big(\Pi \in A\Big)\mathbb{P}\Big(\Pi' \in B\Big) \\
&= \mathbb{P}\Big(\Pi \in A, \Pi' \in B\Big)
\end{aligned}$$

thus proving that $(\sigma(\Pi), \widehat{\sigma_\Pi}(\Pi')) \stackrel{d}{=} (\Pi, \Pi')$. Finally, it follows that

$$\sigma\big(\operatorname{Coag}(\Pi, \Pi')\big) = \operatorname{Coag}(\sigma(\Pi), \widehat{\sigma_\Pi}(\Pi')) \stackrel{d}{=} \operatorname{Coag}(\Pi, \Pi').$$

□

Remark. *If we replace* Π *in the Theorem 1.2.2 by a partition* π*, then* $\operatorname{Coag}(\pi, \Pi')$ *is also an exchangeable random partition.*

Of course if $(\Pi^1, \dots, \Pi^m)$ is a collection of independent exchangeable partitions then, using the preceding theorem in an induction argument, we see that the sequential coagulation $\mathrm{CO}_{i=1}^m \Pi^i$ (defined by $\mathrm{CO}_{i=1}^2 \Pi^i = \operatorname{Coag}(\Pi^1, \Pi^2)$ and $\mathrm{CO}_{i=1}^k \Pi^i = \operatorname{Coag}\left(\mathrm{CO}_{i=1}^{k-1} \Pi^i, \Pi^k\right)$) is also an exchangeable partition. This gives our primary tool to construct exchangeable coalescent processes.

Definition 1.2.2. *Let* $\mathit{\Pi} = (\mathit{\Pi}_t)_{t\geq 0}$ *be a Markov process in continuous time with values in* $\mathscr{P}_m$ *for some* $m \in \mathbb{N} \cup \{\infty\}$*.* $\mathit{\Pi}$ *is an exchangeable coalescent if* $\mathit{\Pi}_0$ *is an exchangeable partition, and the transition kernels of* $\mathit{\Pi}$ *satisfy*

$$\mathbb{P}(\mathit{\Pi}_{t+h} \in A | \mathit{\Pi}_t = \pi) = \mathbb{P}(\operatorname{Coag}(\pi, \tilde{\Pi}_h) \in A)$$

where A *is any measurable set in* $\mathscr{P}_m$ *and* $\tilde{\Pi}_h$ *is an exchangeable random partition whose law depends only on* h*. We call the collection* $(\tilde{\Pi}_h)_{h\geq 0}$ *the stationary increments of* $\mathit{\Pi}$*. Also, if* $\mathit{\Pi}_0 = \mathbf{0}_m := \{\{1\}, \dots, \{m\}\}$ *we call* $\mathit{\Pi}$ *a standard exchangeable coalescent.*

Because the values of $(\mathit{\Pi}_t)_{t>0}$ are determined by the stationary increments $(\tilde{\Pi}_h)_{h\in\mathbb{R}^+}$ in a way that resembles the definition of Lévy processes, coalescent processes may be loosely interpreted as Lévy processes where the binary operation is Coag in the set $\mathscr{P}_n$, instead of the usual sum operation in $\mathbb{R}$.

Lemma 1.2.3. *If* $\mathit{\Pi}$ *is a standard exchangeable coalescent then*

$$\tilde{\Pi}_h \stackrel{d}{=} \mathit{\Pi}_h, \quad \forall h > 0.$$

Proof. Since $\mathit{\Pi}_0 = (\{1\}, \{2\}, \cdots)$, for any measurable set A we have

$$\mathbb{P}(\mathit{\Pi}_h \in A) = \mathbb{P}(\operatorname{Coag}(\mathit{\Pi}_0, \tilde{\Pi}_h) \in A) = P(\tilde{\Pi}_h \in A).$$

□

If $\mathit{\Pi}$ is an exchangeable coalescent with values in $\mathscr{P}_m$ then $(\operatorname{Coag}(\Pi, \mathit{\Pi}_t))_{t\geq 0}$ is also an exchangeable coalescent whenever Π is an exchangeable partition. In particular, if $\mathit{\Pi}$ is standard then $\operatorname{Coag}(\Pi, \mathit{\Pi}_0) = \Pi$, so $(\operatorname{Coag}(\Pi, \mathit{\Pi}_t)_{t\geq 0}$ is an exchangeable coalescent that starts at Π, and whose probability kernels are determined by the stationary increments $(\mathit{\Pi}_h)_{h\in\mathbb{R}^+}$. For this reason we will only consider standard coalescents from now on.

Theorem 1.2.4. *If* $\mathit{\Pi}$ *takes values on* $\mathscr{P}_\infty$ *and is such that* $\mathit{\Pi}\big|_n$ *is an exchangeable coalescent for every* $n \in \mathbb{N}$*, then* $\mathit{\Pi}$ *is an exchangeable coalescent.*

Proof. To prove this theorem we just note that for every $t > 0$ the collection of exchangeable partitions $\{\Pi_t\big|_n : n \in \mathbb{N}\}$ is consistent and, therefore, Π_t is an exchangeable partition. Indeed, let σ be a permutation such that $\sigma(j) = j$ for all $j > n$, then the exchangeability of $\Pi_t\big|_m$ for $m > n$ yields, for any $\pi \in \mathscr{P}_m$,

$$\begin{aligned}\mathbb{P}\left(\sigma(\Pi_t) \in \mathscr{P}_\infty(\pi)\right) &= \mathbb{P}\left(\sigma(\Pi_t\big|_m) = \pi\right) = \mathbb{P}\left(\Pi_t\big|_m = \pi\right)\\ &= \mathbb{P}\left(\Pi_t \in \mathscr{P}_\infty(\pi)\right).\end{aligned}$$

Also, for any $A \in \mathscr{B}(\mathscr{P}_\infty)$, if $A\big|_n := \{\pi\big|_n : \pi \in A\}$, we have

$$\Big\{\Pi_{t+h} \in A\Big\} \bigcap \Big\{\Pi_t = \pi\Big\} = \bigcap_{n\in\mathbb{N}} \Big\{\Pi\big|_n(t+h) \in A\big|_n \bigcap \Pi\big|_n(t) = \pi\big|_n\Big\}$$

and, similarly,

$$\Big\{\operatorname{Coag}(\pi, \Pi_h) \in A\Big\} = \bigcap_{n\in\mathbb{N}} \Big\{\operatorname{Coag}(\pi, \Pi_h)\big|_n \in A\big|_n\Big\}.$$

Therefore

$$\begin{aligned}\mathbb{P}(\Pi_{t+h} \in A|\Pi_t = \pi) &= \lim_{n\to\infty} \mathbb{P}\Big(\Pi\big|_n(t+h) \in A\big|_n \Big| \Pi\big|_n(t) = \pi\big|_n\Big)\\ &= \lim_{n\to\infty} \mathbb{P}\Big(\operatorname{Coag}(\pi\big|_n, \Pi\big|_n(h)) \in A\big|_n)\\ &= \lim_{n\to\infty} \mathbb{P}\Big(\operatorname{Coag}(\pi, \Pi_h)\big|_n \in A\big|_n)\\ &= \mathbb{P}(\operatorname{Coag}(\pi, \Pi_h) \in A).\end{aligned}$$

So Π is an exchangeable coalescent with increments $(\Pi)_t$. □

Theorem 1.2.5. *The semigroup of an exchangeable coalescent is Feller.*

Proof. It is sufficient to prove that, for every continuous function $\phi \in C(\mathscr{P}_\infty)$, the map

$$\pi \to \mathbb{E}\left[\phi(\operatorname{Coag}(\pi, \tilde{\Pi}_t))\right]$$

is continuous for every t, and that

$$\lim_{t\to 0} \mathbb{E}\left[\phi(\operatorname{Coag}(\pi, \tilde{\Pi}_t))\right] = \phi(\pi).$$

Both follow easily from the Lipschitz-continuity of the operator Coag (see Lemma 1.2.1). □

1.2.2 Coagulation Rates and Poissonian Construction

Let Π be an exchangeable coalescent taking values in $\mathscr{P}_\infty$. Since $\Pi\big|_n$ takes values on the finite set $\mathscr{P}_n$, its trajectories are entirely determined by its jumping rates

$$\mathbf{Q}^{(n)}_{\pi',\pi} = \lim_{t\to 0} \frac{1}{t} \mathbb{P}\Big(\Pi\big|_n(t) = \pi \Big| \Pi\big|_n(0) = \pi'\Big).$$

If (π', π) is a pair of admissible partitions in $\mathscr{P}_n$ and π is distinct from $\mathbf{0}_n$, we have

$$\begin{aligned}\mathbf{Q}^{(n)}_{\pi',\mathrm{Coag}(\pi',\pi)} &= \lim_{t\to 0}\frac{1}{t}\mathbb{P}\Big(\Pi\big|_n(t) = \mathrm{Coag}(\pi',\pi)\Big|\Pi\big|_n(0) = \pi'\Big)\\ &= \lim_{t\to 0}\frac{1}{t}\mathbb{P}\big(\,\mathrm{Coag}(\pi',\Pi_t) = \mathrm{Coag}(\pi',\pi)\big)\\ &= \lim_{t\to 0}\frac{1}{t}\mathbb{P}\big(\Pi\big|_n(t) = \pi\big)\\ &= \mathbf{Q}^{(n)}_{\mathbf{0}_n,\pi}.\end{aligned}$$

In other words, $\mathbf{Q}^{(n)}_{\mathbf{0}_n,\pi}$ is the jumping rate of $\Pi\big|_n$ from π' to $\mathrm{Coag}(\pi',\pi)$. On the other hand, if π cannot be written in the form $\pi = \mathrm{Coag}(\pi',\pi'')$ for any $\pi'' \in \mathscr{P}_n$, then

$$\mathbb{P}\Big(\Pi\big|_n(t) = \pi\Big|\Pi\big|_n(0) = \pi'\Big) = 0$$

so $\mathbf{Q}^{(n)}_{\pi',\pi} = 0$. The last two results combined tell us that the set

$$\{\mathbf{Q}^{(n)}_{\mathbf{0}_n,\pi} : \pi \in \mathscr{P}_n \setminus \mathbf{0}_n, n \in \mathbb{N}\}$$

completely determines the trajectories of $\Pi\big|_m$ for every $m \in \mathbb{N}$, and, thus, the trajectories of Π. To ease notation from now on we will write q_π instead of $\mathbf{Q}^{(n)}_{\mathbf{0}_n,\pi}$.

Theorem 1.2.6. *Recall the notation for $\mathscr{P}_\infty(\pi)$ in (1.1). The set $\{q_\pi : \pi \in \mathscr{P}_n \setminus \mathbf{0}_n, n \in \mathbb{N}\}$ determines a unique measure μ on $\mathscr{P}_\infty$ such that $\mu(\{\mathbf{0}_\infty\}) = 0$, and*

$$\mu(\mathscr{P}_\infty(\pi)) = q_\pi$$

for every $\pi \in \mathscr{P}_n \setminus \mathbf{0}_n, n \in \mathbb{N}$. We call μ the coagulation rate of Π.

Proof. The idea of the proof is to use Caratheodory's extension theorem in order to construct a measure on $\mathscr{P}_\infty \setminus \mathbf{0}_\infty$ and then define $\mu(\mathbf{0}_\infty) := 0$. Towards this, note that the set

$$\mathscr{S} := \{\mathscr{P}_\infty(\pi) : \pi \in \mathscr{P}_n \setminus \mathbf{0}_n,\ \ n \in \mathbb{N}\}$$

is a semiring. Define a measure $\hat{\mu}$ in $\mathscr{S}$ by

$$\hat{\mu}(\mathscr{P}_\infty(\pi)) = q_\pi.$$

Note that if $\pi \in \mathscr{P}_m$ and $n \geq m$, then

$$\mathscr{P}_\infty(\pi) = \bigcup_{\pi' \in \mathscr{P}_n(\pi)} \mathscr{P}_\infty(\pi').$$

Hence, in order to prove that $\hat{\mu}$ is finitely additive we need to verify that

$$\hat{\mu}(\mathscr{P}_\infty(\pi)) = \sum_{\pi' \in \mathscr{P}_n(\pi)} \hat{\mu}(\mathscr{P}_\infty(\pi')).$$

Observe that

$$\{\Pi\big|_m = \pi\} = \bigcup_{\pi' \in \mathscr{P}_n(\pi)} \{\Pi\big|_n = \pi'\},$$

thus, the rate at which $\mathbf{0}_m$ jumps to π, equals the rate at which $\mathbf{0}_n$ jumps to $\bigcup_{\pi' \in \mathscr{P}_n(\pi)}\{\Pi\big|_n = \pi'\}$. Since the sets on the last union are pairwise disjoint, the rate at which the latter occurs is

$$\sum_{\pi' \in \mathscr{P}_n(\pi)} q_{\pi'} = \sum_{\pi' \in \mathscr{P}_n(\pi)} \hat{\mu}(\mathscr{P}_\infty(\pi'));$$

so $\hat{\mu}$ is finitely additive. To see that $\hat{\mu}$ is infinitely additive note that for any partition $\pi \in \mathscr{P}_n$, $\mathscr{P}_\infty(\pi)$ is closed and, by Theorem 1.1.2, compact. Therefore, if there exists a collection of partitions $(\pi_i)_{i\in\mathbb{N}}$ such that $(\mathscr{P}_\infty(\pi_i))_{i\in\mathbb{N}}$ are pairwise disjoint and $\mathscr{P}_\infty(\pi) = \cup_{i=1}^\infty \mathscr{P}_\infty(\pi_i)$, then, since $(\mathscr{P}_\infty(\pi_i))_{i\in\mathbb{N}}$ is an open cover of $\mathscr{P}_\infty(\pi)$, it must be the case that $\mathscr{P}_\infty(\pi_i) = \emptyset$ for all but finitely many $i \in \mathbb{N}$. By the finite additivity of $\hat{\mu}$ we then have the equality $\hat{\mu}(\mathscr{P}_\infty(\pi)) = \sum_{i=1}^\infty \hat{\mu}(\mathscr{P}_\infty(\pi_i))$. By Caratheodory's theorem, $\hat{\mu}$ can be uniquely extended to a measure μ on $\mathscr{P}_\infty \setminus \mathbf{0}_\infty$, and setting $\mu(\mathbf{0}_\infty) \coloneqq 0$ finishes the proof of the theorem. □

Remark. *Note that* $\sum_{\pi\in\mathscr{P}_n\setminus\mathbf{0}_n} \mu(\mathscr{P}_\infty(\pi)) < \infty$ *for every* $n \in \mathbb{N}$ *and, therefore,* μ *is a* σ*-finite measure.*

Proposition 1.2.7. *The coagulation rate* μ *of an exchangeable coalescent is invariant under permutations.*

Proof. Note that if $\pi \neq \mathbf{0}_n$ then for every permutation σ we have

$$\begin{aligned}
\mathbf{Q}^{(n)}_{\mathbf{0}_n,\pi} &= \lim_{t\to 0} \frac{1}{t}\mathbb{P}(\Pi\big|_n(t) = \pi) \\
&= \lim_{t\to 0} \frac{1}{t}\mathbb{P}(\sigma(\Pi\big|_n(t)) = \pi) \\
&= \lim_{t\to 0} \frac{1}{t}\mathbb{P}(\Pi\big|_n(t) = \sigma^{-1}(\pi)) \\
&= \mathbf{Q}^{(n)}_{\mathbf{0}_n,\sigma^{-1}(\pi)}.
\end{aligned}$$

Starting with $\sigma(\pi)$ instead of π above, we see that $\mathbf{Q}^{(n)}_{\mathbf{0}_n,\sigma(\pi)} = \mathbf{Q}^{(n)}_{\mathbf{0}_n,\pi}$. Thus, for any permutation σ and any measurable set A, we have

$$\mu(A) = \mu(\sigma(A)).$$

□

Theorem 1.2.8. *A measure* μ *on* $\mathscr{P}_\infty$ *is the coagulation rate of an exchangeable coalescent if and only if it satisfies*

- $\sum_{\pi\in\mathscr{P}_n\setminus\mathbf{0}_n} \mu(\mathscr{P}_\infty(\pi)) < \infty$ *for all* $n \in \mathbb{N}$,
- $\mu(\sigma(A)) = \mu(A)$ *for any measurable set* A *and permutation* σ*, and*
- $\mu(\mathbf{0}_\infty) = 0$.

The proof of the forward implication is given by Theorem 1.2.6 and Proposition 1.2.7 above. The proof of the backward implication proceeds by giving a Poissonian construction of an exchangeable coalescent Π from such a measure μ, which we provide next.

Let M be a Poisson random measure on $\mathbb{R}^+ \times \mathscr{P}_\infty$ with intensity $\mathsf{Leb} \otimes \mu$. For each $n \in \mathbb{N}$, define a random measure M_n on $\mathbb{R}^+ \times \mathscr{P}_n$ by

$$M_n([0,t] \times \pi) \coloneqq M([0,t] \times \mathscr{P}_\infty(\pi))$$

and note that M_n is a Poisson random measure on $\mathbb{R}^+ \times \mathscr{P}_n$ with intensity $\mathsf{Leb} \otimes \mu_n$, where μ_n is the measure on $\mathscr{P}_n$ given by

$$\mu_n(\pi) = \mu(\mathscr{P}_\infty(\pi)).$$

Since $\mu_n(\mathscr{P}_n \setminus \mathbf{0}_n) < \infty$, M_n has a finite number of atoms in $[0,t] \times \mathscr{P}_n \setminus \mathbf{0}_n$ with probability one. Also, if (t_1, π^1) and (t_2, π^2) are two atoms of M_n in $[0,t] \times \mathscr{P}_n \setminus \mathbf{0}_n$, then $t_1 \neq t_2$ with probability one; that is, the atoms of M_n in $[0,t] \times \mathscr{P}_n \setminus \mathbf{0}_n$ occur at different times with probability one. Therefore, for every $n \in \mathbb{N}$ the atoms of M_n in $[0,t] \times \mathscr{P}_n \setminus \mathbf{0}_n$ can be ordered according to their first (time) coordinate and we may define the sequence of random vectors $((T_i, \Pi_i))_{i\in\mathbb{N}}$ given by this ordering. Using the latter, we consider the process Π^n with values in $\mathscr{P}_n$ such that for every $t \geq 0$, $\Pi^n(t)$ is given by the ordered coagulation

$$\Pi^n(t) = \underset{\{i:0<T_i\leq t\}}{\mathrm{CO}} \Pi_i.$$

From now on we will write $\mathrm{CO}_{0<T_i\leq t} \Pi_i$ instead of $\mathrm{CO}_{\{i:0<T_i\leq t\}} \Pi_i$.

By standard Poisson random measure arguments, it is easily seen that the Poisson random measure M_n can be constructed in the following way: let $\widehat{\mu_n} \coloneqq \mu_n(\mathscr{P}_n \setminus \mathbf{0}_n)$ and consider a sequence of i.i.d. random partitions $(\Pi_i)_{i\in\mathbb{N}}$ with values in $\mathscr{P}_n \setminus \mathbf{0}_n$ and law $\mu_n(\pi)/\widehat{\mu_n}$. Let P be an independent Poisson process in $\mathbb{R}^+$ with parameter $\widehat{\mu_n}$, and denote its jumping times by $(T_i)_{i\in\mathbb{N}}$. Finally, define the atoms of M_n in $\mathbb{R}^+ \times \mathscr{P}_n \setminus \mathbf{0}_n$ to be the points $((\Pi_i, T_i))_{i\in\mathbb{N}}$.

Lemma 1.2.9. *The process Π^n with values in $\mathscr{P}_n$ given by:*

$$\begin{aligned} \Pi^n(0) &= \mathbf{0}_n \\ \Pi^n(t) &= \underset{0<T_i\leq t}{\mathrm{CO}} \Pi_i \end{aligned}$$

is a standard exchangeable coalescent.

Proof. It is clear that if A is any measurable set and $t_1, \cdots, t_n \in [0,t]$ then

$$\mathbb{P}\big(\Pi^n(t+h) \in A | \Pi^n(t)\big) = \mathbb{P}\big(\Pi^n(t+h) \in A | \Pi^n(t), \Pi^n(t_1), \cdots, \Pi^n(t_n)\big)$$

so Π^n is a Markov process. Now, consider the sequence of atoms $((T_i, \Pi_i))_{i\in\mathbb{N}}$ of M_n in $\mathbb{R}^+ \times \mathscr{P}_n \setminus \mathbf{0}_n$. Since μ is invariant under permutations, then μ_n is also invariant under permutations. Thus, for every $i \in \mathbb{N}$ and any permutation σ we have $\mathbb{P}(\Pi_i = \pi) = \mathbb{P}(\Pi_i = \sigma(\pi))$, that is, Π_i is an exchangeable partition. Also, by the construction of the Poisson random measure M_n described above, we see that the random partitions $(\Pi_i)_{i\in\mathbb{N}}$ are independent and identically distributed. We also have that

$$M_n((0,h] \times \mathscr{P}_n \setminus \mathbf{0}_n) \overset{d}{=} M_n((t,t+h] \times \mathscr{P}_n \setminus \mathbf{0}_n)$$

since M_n is a Poisson random measure. Therefore

$$\underset{t<T_i\leq t+h}{\mathrm{CO}} \Pi_i \overset{d}{=} \underset{0<T_i\leq h}{\mathrm{CO}} \Pi_i.$$

Now, since $\Pi^n(t+h)$ is given by

$$\Pi^n(t+h) = \mathrm{Coag}\left(\Pi^n(t), \underset{t<T_i\leq t+h}{\mathrm{CO}} \Pi_i\right),$$

we only need to show that $\mathrm{CO}_{0<T_i\leq h}\,\Pi_i$ is an exchangeable partition. The latter follows from noting that for any finite collection of indices $J \subset \mathbb{N}$, the partition

$$\underset{i\in J}{\mathrm{CO}}\, \Pi_i$$

is exchangeable since the partitions $(\Pi_i)_{i\in J}$ are independent and exchangeable (Theorem 1.2.2). □

Lemma 1.2.10. *For any fixed $t>0$, the sequence of partitions $(\Pi^n(t))_{n\in\mathbb{N}}$ is consistent.*

Proof. For any pair of integers $n>m$, let $((T_i,\Pi_i))_{i\in\mathbb{N}}$ be the atoms of M_n and $((T_{k_i},\Pi_{k_i}))_{i\in\mathbb{N}}$ be the subsequence of $((T_i,\Pi_i))_{i\in\mathbb{N}}$ such that $\Pi_i\big|_m \neq \mathbf{0}_n$. Then we note that the atoms of M_m are given by $\left((T_{k_i},\Pi_{k_i}\big|_m)\right)_{i\in\mathbb{N}}$ and:

$$\begin{aligned}
\Pi^n\big|_m(t) &= \left(\underset{0<T_i\leq t}{\mathrm{CO}} \Pi_i\right)\Big|_m \\
&= \underset{0<T_i\leq t}{\mathrm{CO}} \Pi_i\big|_m \\
&= \underset{0<T_{k_i}\leq t}{\mathrm{CO}} \Pi_{k_i}\big|_m \\
&= \Pi^m(t).
\end{aligned}$$

□

Proof of Theorem 1.2.8. Lemmas 1.2.9 and 1.2.10 combined with Lemma 1.1.1 and Theorem 1.2.4, show that the sequence $(\Pi^n)_{n\in\mathbb{N}}$ determines a unique (in law) exchangeable coalescent Π in $\mathscr{P}_\infty$. Since Π^n is a Markov chain for every n, and since for every $\pi\in\mathscr{P}_n\setminus\mathbf{0}_n$ we have $\mathbb{P}(\Pi^n(T_1)=\pi)=\mu(\mathscr{P}_\infty(\pi))/\mu(\mathscr{P}_n\setminus\mathbf{0}_n)$, it follows that

$$\alpha_\pi = \lim_{t\to 0}\frac{1}{t}\mathbb{P}\big(\Pi^n(t)=\pi\big) = \mu\left(\left\{\pi'\in\mathscr{P}_\infty : \pi'\big|_n = \pi\right\}\right),$$

so Π has coagulation rate μ. □

We now provide a construction of coagulation rates reminiscent of Kingman's representation for exchangeable partitions. Let us first describe the two types of coagulation rates that will be the basis of our construction. For each pair of integers i,j consider the partition $\pi_{i\sim j}$ given by the block $\{i,j\}$ and the singletons $\{\{k\}: k\neq i, k\neq j\}$. *Kingman's coagulation rate* μ_K is the measure on $\mathscr{P}_\infty$ given by atoms of size one at the points $\{\pi_{i\sim j} : 1\leq i<j<\infty\}$.

Note that the coalescent process determined by this coagulation rate evolves through coagulations of exactly two blocks at a time.

For the second type of coagulation rate consider any measure ν on $\mathscr{P}_{[0,1]}$ such that $\nu(\mathbf{0}) = 0$ and

$$\int_{\mathscr{P}_{[0,1]}} \sum_{i=1}^{\infty} \rho_i^2 \ \nu(d\rho) < \infty. \tag{1.4}$$

Then, using the measures $(\varrho_\rho,\ \rho \in \mathscr{P}_{[0,1]})$ given by the paintbox construction introduced in Section 1.1.2, define the measure μ_ν on $\mathscr{P}_\infty$ by:

$$\mu_\nu(\cdot) := \int_{\mathscr{P}_{[0,1]}} \varrho_\rho(\cdot) \ \nu(d\rho).$$

It follows that μ_ν is a coagulation rate. To see this, notice that ϱ_ρ is invariant under permutations and, hence, μ_ν is also invariant under permutations. Also, $\mu_\nu(\mathbf{0}_\infty) = 0$ since $\nu(\mathbf{0}) = 0$. Finally, note that $\mu_\nu(\mathscr{P}_\infty(\mathscr{P}_n \setminus \mathbf{0}_n)) < \infty$ for every $n \in \mathbb{N}$, since for every $\pi \in \mathscr{P}_n \setminus \mathbf{0}_n$ there exists i, j such that $i \overset{\pi}{\sim} j$ and

$$\begin{aligned} \mu_\nu(\mathscr{P}_\infty(\pi)) &\leq \mu_\nu\big(\mathscr{P}_\infty\big(\{\pi : i \overset{\pi}{\sim} j\}\big)\big) \\ &= \int_{\mathscr{P}_{[0,1]}} \sum_{k=1}^{\infty} \rho_k^2 \ \nu(d\rho) < \infty. \end{aligned}$$

Since μ_K and μ_ν are coagulation rates it is easily seen that for every $c > 0$, the measure $\mu := c\mu_k + \mu_\nu$ is again a coagulation rate. The following theorem states that all coagulation rates can be constructed in this way.

Theorem 1.2.11. *Let μ be any coagulation rate. There exists a constant $c > 0$ and a measure ν in $\mathscr{P}_{[0,1]}$ that satisfies $\nu(\mathbf{0}) = 0$ and (1.4) such that*

$$\mu = c\mu_K + \mu_\nu.$$

The proof of this theorem follows by a clever application of Kingman's representation theorem for exchangeable partitions, we refer the reader to Theorem 4.2 in Bertoin 2006.

The class of coalescents with multiple groups of coagulating blocks are also called coalescents with simultaneous multiple collisions, or Ξ-coalescents (in this context the general measure ν is replaced by the finite measure Ξ that satisfies $\nu(d\rho) = \frac{1}{\sum \rho_i^2}\Xi(d\rho)$, see Sagitov 1999; Schweinsberg 2000b); examples of this class include the Poisson-Dirichlet(α, θ) coalescent (Bertoin 2008; Möhle 2010), in which Ξ is set to be the Poisson-Dirichlet(α, θ) probability measure on $\mathscr{P}_{[0,1]}$, and whose genealogies are typically star-shaped (Möhle 2010); the symmetric coalescent, in which the measure Ξ is supported on the set

$$\left\{(\rho_1, \rho_2, \dots) \in \mathscr{P}_{[0,1]} \colon \rho_1 = \cdots = \rho_k = 1/k, k \geq 1\right\},$$

and describe the limiting genealogies of populations undergoing recurrent bottlenecks (González Casanova, Miró Pina, and Siri-Jégousse 2020); and the class of Ξ coalescents with $\Xi = \Xi' \circ \phi$ where ϕ is the map $\phi\colon \mathscr{P}_{[0,1]} \to \mathscr{P}_{[0,1]}$ given by $\phi(\rho_1, \rho_2, \dots) = (\rho_1/2, \rho_1/2, \rho_2/2, \rho_2/2, \dots)$, which describe the genealogies of a wide class of diploid populations (Birkner, Liu, and Sturm 2018). In the following section we introduce the class of simple coalescents which, by reason of their manageability and wide applicability, have been the most widely studied class of coalescent processes.

1.2.3 Simple Coalescents

The Poissonian construction given in Section 1.2.2 tells us that we can intuitively think of an exchangeable coalescent as a process where one selects a number of time points $(T_h)_{h\in\mathbb{N}}$ according to a Poisson process on $\mathbb{R}^+$ and then for each time point T_h one picks an "increment" Π_h according to some distribution μ in $\mathscr{P}_\infty$. The coalescent process at time t is then constructed through the sequential coagulation prescribed by all the increments that occur before time t. Until now we have considered the general case in which the increments $(\Pi_h)_{h\in\mathbb{N}}$ may prescribe the simultaneous coalescence of multiple groups of blocks. We now focus on processes whose infinitesimal increments prescribe the coalescence of at most one group of blocks at a time; we call these coalescent processes *simple coalescents* or Λ-*coalescents* (Pitman 1999; Sagitov 1999, also, see below for the appearance of the measure Λ). Due to their simpler nature and wide applicability, these processes are the most widely studied coalescent processes (see Section 1.3). The following definitions make this intuition precise.

Definition 1.2.3. *Let $\pi \in \mathscr{P}_\infty$ be a partition, and Π be an exchangeable coalescent.*

- *We say that π is a simple partition if all its blocks, except possibly one, are singletons.*
- *We say that Π is simple if its coagulation rate is supported by simple partitions.*

We note that if μ is the coagulation rate of a simple coalescent then the image measure of μ on $\mathscr{P}_{[0,1]}$ under the map $\pi \to |\pi|^\downarrow$ is supported on mass partitions of the form $\rho = (p, 0, 0, \cdots)$, $p \in [0,1]$. Therefore, if ν is the measure on $\mathscr{P}_{[0,1]}$ such that $\mu = c\mu_K + \mu_\nu$, then ν is also supported on mass partitions of this form. Furthermore, making a slight abuse of notation, by equation (1.4) the measure ν satisfies

$$\int_0^1 p^2 \nu(dp) < \infty;$$

therefore, we can write ν in terms of a finite measure Λ on $[0,1]$ through the equation

$$\nu(A) = \int_A \frac{\Lambda(dp)}{p^2}, \quad \forall A \in \mathscr{B}([0,1]). \tag{1.5}$$

The above discussion can also be read in reverse, that is, for any finite measure Λ in $[0,1]$ such that $\Lambda(0) = 0$ and $\int_{[0,1]} p^{-2}\ \Lambda(dp) < \infty$, we can construct a measure ν on $\mathscr{P}_{[0,1]}$ which corresponds to a simple coalescent via equation (1.5). For this reason, from now on we will work with the measures Λ instead of ν and, by a slight abuse of notation, we will say that the coagulation rate μ of a simple coalescent is given by $\mu = c\mu_K + \mu_\Lambda$. Given the coagulation rate of a simple coalescent $\mu = c\mu_K + \mu_\Lambda$ we interpret c as the intensity with which pairs of blocks coalesce (Kingman's part), and μ_Λ as the intensity with which a proportion p of all the blocks is coalesced instead. Also, if we consider the restriction of Π to $\mathscr{P}_n$, then, for any simple partition $\pi \in \mathscr{P}_n \setminus \mathbf{0}_n$ such that its non-singleton block has k elements ($(2 \le k \le n)$), we have

$$\alpha_\pi = c\mathbb{1}_{k=2} + \int_{[0,1]} p^{k-2}(1-p)^{n-k}\ \Lambda(dp);$$

so if we define $\lambda_{n,k} \coloneqq c\mathbb{1}_{k=2} + \int_{[0,1]} p^{k-2}(1-p)^{n-k} \ \Lambda(dp)$, then $\lambda_{n,k}$ gives the intensity with which any particular collection of k blocks coalesce whenever there are $n \geq k$ blocks. Moreover, the intensities $\lambda_{n,k}$ satisfy the recursion

$$\lambda_{n,k} = \lambda_{n+1,k} + \lambda_{n+1,k+1} \tag{1.6}$$

since the rate at which a collection of k blocks coalesce when there are n blocks equals the rate, when there are $n+1$ blocks, at which they coalesce along with the $(n+1)th$ block plus the rate at which they coalesce excluding the $(n+1)th$ block; more precisely, if B is the non-singleton block of a simple partition $\pi \in \mathscr{P}_n$, and π', π'' are the simple partitions in $\mathscr{P}_{n+1}$ with non-singleton blocks B and $B \cup \{n+1\}$ respectively, then, by the additivity of the coagulation rate, we have

$$\alpha_\pi = q_{\pi'} + q_{\pi''},$$

so (1.6) follows.

Furthermore, if A_k is the set of all simple partitions in $\mathscr{P}_n$ such that their non-singleton element has k elements ($2 \leq k \leq n$), then $\sum_{\pi \in A_k} \alpha_\pi = (\#A_k)\, \lambda_{n,k} = \binom{n}{k}\lambda_{n,k}$ gives the rate at which a coalescence of exactly k blocks occurs whenever there are $n \geq k$ blocks; and

$$\lambda_n \coloneqq \sum_{k=1}^{n} \binom{n}{k} \lambda_{n,k}$$

gives the total coagulation rate.

In the following examples we introduce some of the most widely studied simple coalescents, see Section 1.3 for a brief summary of the literature on these and related coalescents.

Example 1 (Kingman's coalescent)**.** *This is the primordial example first introduced by Kingman 1982 where Λ is the zero measure and $c > 0$. This is the most widely studied coalescent due to its simplicity and its ubiquitous appearance in applications, in particular as the genealogy of neutral populations.*

Example 2 (Beta(a,b)-coalescent)**.** *In this two-parameter class of simple coalescents the measure Λ is set to*

$$\Lambda(dp) = \frac{p^{a-1}(1-p)^{b-1}}{\mathsf{B}(a,b)},$$

and c is set to zero. For instance, the coagulation rates $\lambda_{n,k}$ are given by

$$\lambda_{n,k} = \int_0^1 \frac{p^{a+k-2-1}(1-p)^{n-k+b-1}}{\mathsf{B}(a,b)} dp = \frac{\mathsf{B}(a+k-2, b+n-k)}{\mathsf{B}(a,b)}.$$

In the next example we define the subclass of Beta$(2-\alpha, \alpha)$ coalescents which has gained notable attention due to their ubiquitous appearance as the limit genealogies of various population evolution models; and to their adequacy in modelling applications, stemming from their one-parameter definition and the wide variety of dynamics that they can model, raging from neutral evolution (Kingman's coalescent) to strong selection (Bolthausen-Sznitman coalescent), see Figure 1.2.3.

Example 3 (Beta($2-\alpha$,α)-coalescents)**.** *This is the particular case of the general Beta(a,b)-coalescent in which $a = 2-\alpha$ and $b = \alpha$, i.e.*

$$\Lambda(dp) = \frac{p^{1-\alpha}(1-p)^{\alpha-1}}{\Gamma(2-\alpha)\Gamma(\alpha)}dp.$$

They were first introduced in Schweinsberg 2003 as the limiting genealogy of supercritical Galton-Watson processes. In the following example we introduce an important example, the Bolthausen-Sznitman coalescent.

Example 4 (Bolthausen-Sznitman coalescent (BSC))**.** *The Bolthausen-Sznitman coalescent (Bolthausen and Sznitman 1998) is the Beta$(2-\alpha,\alpha)$ coalescent of parameter $\alpha = 1$. This process has gained significant attention due to its apparent universality as the underlying genealogy of populations undergoing natural selection (J. Berestycki, N. Berestycki, and Schweinsberg 2013; Brunet and Derrida 2012; Cortines and Mallein 2017; Neher and Hallatschek 2013; Schweinsberg 2017b).*

One can check that as the parameter α tends to 2 the coagulation rate of the Beta$(2-\alpha,\alpha)$-coalescent converges weakly to δ_0 so that the extreme case $\alpha = 2$ corresponds to Kingman's coalescent. In Figure 1.2.3 below we observe how the topology of the corresponding trees interpolates between that of the Bolthausen-Sznitman coalescent, with large external branches (i.e. branches associated to blocks of size 1) and multiple collisions, and that of Kingman's coalescent, with small external branches and pair-wise collisions. As mentioned above, the Bolthausen-Sznitman coalescent stands as the main model that describes the genealogies of populations undergoing natural selection, while Kingman's coalescent is of course the null model for populations under neutral evolution. This, together with their manageability, are some of the reasons that motivate the vast studies made on this subclass of simple coalescents.

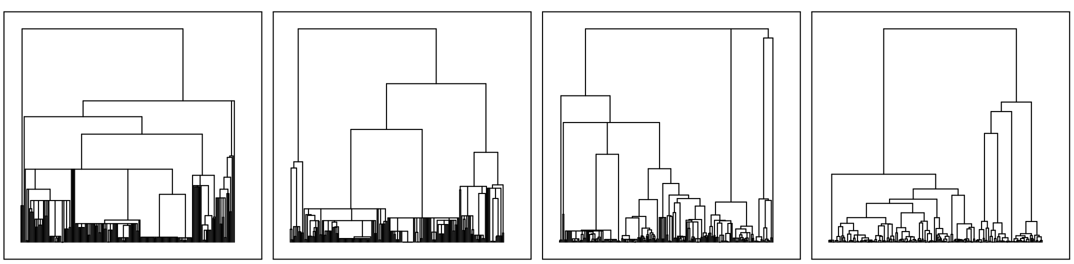

Figure 1.1: Simulations of Beta coalescents where the parameter α has been interpolated between 1 and 2; from left to right: $\alpha = 1$ (BSC), $\alpha = 1.25$, $\alpha = 1.75$, and $\alpha = 2$ (Kingman).

1.3 Overview of Applications

The study of coalescent processes focuses on two main questions motivated by biology. On the one hand, coalescent processes are the natural mathematical model to study the genealogy of population models. In fact, a lot of effort is made to establish a parity between coalescent processes and population models with varying biological assumptions, such as constant or varying population size, the presence of mutation and/or natural selection, genetic drift, spatial constraints, dormancy/latency, etc. On the other hand, the theoretical characterization of different functionals on coalescent processes, such as the tree height, tree length, the size of external and internal branches or, more conveniently, the size of branches with exactly $b \in \mathbb{N}$ descendants, and the coming down from infinity, among others, typically inform the design of new inference algorithms in population genetics. In the next subsections we provide an (incomplete) summary of the results obtained so far in these two directions. Further ahead, in Chapter 2, we develop the general theory, and some applications, that connect discrete-time and constant-size population models with coalescent processes via asymptotic weak limits of their genealogies as the total population size $N \to \infty$. Later, in Chapter 3, we introduce the Site Frequency Spectrum (SFS), a functional that models the expected (neutral) genetic diversity present in a genealogy, and that is the standard model selection tool to infer the genealogical past of a present-day population from genetic data. We then provide complete first and second moment characterizations of the SFS for the Bolthausen-Sznitman coalescent (Kersting, Siri-Jégousse, and H. Wences 2021), along with other related results. Such a characterization had only been accomplished for Kingman's coalescent until now (Fu 1995), evincing the difficulty, often stemming from their deep combinatoric nature, of studying coalescent processes and their functionals. As mentioned earlier in Example 4, the Bolthausen-Sznitman coalescent is an important coalescent in the literature due to its apparent universality as the underlying genealogy of populations undergoing natural selection.

1.3.1 Genealogies of Population Models (and foreword to Chapter 2)

Coalescent processes appear as the natural limiting processes describing the genealogy of population evolution models forward in time; in fact, much of the research made on these models focuses on establishing this connection, often as weak limits under appropriate time scales and/or time changes. A primordial example of forwards-in-time models are Cannings' models (Cannings 1974, 1975), which are able to accommodate most biological premises that fall under the assumption of neutral evolution, and are in fact the most general class of models for discrete neutral populations of constant size. The key assumption of these models is that the distribution of the number of descendants of the parents in any generation should be invariant under permutations, i.e. under any labelling scheme of the parents; thus formalizing the idea that parents are "indistinguishable" in distribution and, consequently, all have the same fitness.

Due to their wide applicability and manageability, Cannings' models are among the most widely studied population evolution models. The genealogy of Cannings' models was first described in whole generality, and in terms of coalescent processes, in the foundational work of Möhle and Sagitov 2001 which we describe in Section 2.1, where we also use the theory developed in the present chapter in order to simplify the proofs whenever possible, allowing us to recycle them in Section 2.2 where we develop, for the first time, a general theory for the genealogy of multinomial models. The coalescents that appear as the genealogies of the general Cannings' model include Kingman's coalescent, the general Λ-coalescent, and both discrete-time and continuous-time Ξ-coalescents (see Theorem 2.1.1). Moreover, a natural example of Cannings' models constructed from supercritical branching processes was introduced and studied by Schweinsberg 2003 where Kingman's coalescent, the class of Beta$(2-\alpha,\alpha)$ coalescents, and discrete-time Poisson-Dirichlet Ξ-coalescents appear (see Section 2.1.1). The genealogies of diploid Cannings' models has also been described in Birkner, Liu, and Sturm 2018 where continuous-time Ξ-coalescents appear. Cannings' models with varying population size have also been addressed in Kaj and Krone 2003 where time-changed Kingman's coalescents appear in the limit, also in Freund 2020 leading to general Λ-coalescents in the limit, and in González Casanova, Miró Pina, and Siri-Jégousse 2020 where Ξ-coalescents appear.

The class of discrete-time multinomial models (Cortines and Mallein 2017; Huillet and Möhle 2021), on the other hand, can be regarded as a simpler parametrization of Cannings' models (and/or Wright-Fisher models), "inheriting" their manageability but, nonetheless, allowing for the incorporation of more complex population dynamics, such as the occurrence of asymmetric offspring distributions, expected to occur in the important case of populations undergoing natural selection. The techniques used to study the genealogy of Cannings' models, in particular the work of Möhle and Sagitov 2001, can be easily adapted for the multinomial model in its whole generality; we provide such a general adaptation for the first time in Section 2.2. The latter, in turn, can be simplified, and specialized, to the case where rare (but recurrent) large reproductive events occur in the population; the first results in this direction can be found in Cortines and Mallein 2017 which we describe and expand in Section 2.2.2. Models where large reproductive events occur include, on the one hand, the neutral model of Huillet and Möhle 2021 which can be regarded as the multinomial analog of Schweinsberg 2003 (finding also Kingman's coalescent, the family of Beta$(2-\alpha,\alpha)$ coalescents, and the Poisson-Dirichlet coalescents in the limit). On the other, the *exponential models* of Brunet and Derrida 1997 that incorporate the effect of mutation and natural selection, and that were first generalized by Cortines and Mallein 2017, and then further expanded on in our joint (yet unpublished) work with Emmanuel Schertzer presented in Section 2.3. In the exponential models it is particularly the Bolthausen-Sznitman coalescent that appears as the limit genealogy, specially for populations under strong natural selection, but recently we have proved that discrete-time Poisson-Dirichlet Ξ-coalescents may also appear under somewhat weaker selection (see Theorem 2.3.2). Finally, an example of a multinomial model where the vector of family frequencies are constructed from Poisson-Dirichlet(α,θ) random mass partitions is studied in Cortines and Mallein 2017 where Beta$\big(1-\frac{\theta}{\alpha},1+\frac{\theta}{\alpha}\big)$ and Kingman's coalescents appear; in Section 2.2.3 we provide a generalization of these results, and note that the martingale techniques that we use can be adapted for more general examples in future work.

On the other hand, coalescent processes of course also appear as the limit genealogy of continuous-time population evolution models, including the primordial Moran models (e.g. Bertoin and Le Gall 2003; Huillet and Möhle 2013), various types of branching processes (Bertoin and Le Gall 2000, 2006; Birkner, Blath, et al. 2005; Foucart, Ma, and Mallein 2019; Kersting, Schweinsberg, and Wakolbinger 2014), and also models that incorporate the effect of mutation and natural selection such as J. Berestycki, N. Berestycki, and Schweinsberg 2013; Schweinsberg 2017b, where it is again the Bolthausen-Sznitman coalescent that appears in the context of selection (see Section 2.4 for a brief description of the model and the main results in Schweinsberg 2017b).

1.3.2 Functionals of Coalescent Processes (and foreword to Chapter 3)

The study of functionals on coalescent processes is motivated by applications in biology, in particular: 1) founding rigorous mathematical basis for the development of inference algorithms for the genealogy of present-day populations, and 2) the inference of distinct aspects of the evolutionary past of a population, such as the Time to the Most Recent Common Ancestor (TMRCA), i.e. the absorption time of the coalescent, the coalescence time of two randomly chosen individuals, the occurrence of important variations in the population size in past generations, the strength of selection, the rate of mutation, the presence of subclones/variants in cancer, virus, and/or bacterial populations, etc.

A key biological heuristic for the inference of the genealogy of a population is the presence of neutral mutations that occur to distinct individuals as the population evolves over time. These mutations, by reason of their assumed innocuity and inheritability, record information on the shape of the underlying genealogy without shaping its structure; this information can then be read in the genome of present-day individuals and used to infer their evolutionary past. This heuristic is modelled in mathematical population genetics through Poisson point processes (PPP) constructed on top of coalescent processes, the marks of the PPP signify neutral mutations that occur in particular individuals/branches, independently from one-another and from future and past generations. There are two main variations in the next step of modelling, in the first, the *infinite alleles model*, mutations are assumed to fall on the same genomic site but create a new allele every time, in this case, the external branches only keep the information of the first mutation that they encounter as they traverse the tree towards the root, masking all the other mutations that occurred above. In the second model, the *infinite sites model*, mutations are assumed to fall in a new genomic site every time; thus, being all inherited according to the shape of the coalescent tree, all their information is kept in the external branches. In any case, information on the number of shared mutations between external branches can be used to infer distinct characteristics of the topology of the tree. This motivates the study of the lengths of internal and external branches for a wide variety of coalescent process, and, more thoroughly, but also with increasing difficulty, the study of the *Site Frequency Spectrum* (SFS), and the *Allele Frequency Spectrum* (AFS), defined as the random vector $(SFS_{n,b})_{1\leq b\leq n}$, $n \in \mathbb{N}$, (resp. $(AFS_{n,b})_{1\leq b\leq n}$) giving the number of mutations shared by exactly b external branches in a coalescent processes started with n particles.

Results have been obtained for particular coalescents and subclasses of coalescents, while others are given in full generality. Goldschmidt and Martin 2005 give a new representation of the Bolthausen-Sznitman coalescent in terms of pruning-merge procedures preformed on random recursive trees; they use this construction to give asymptotics on the block sizes, and the number of blocks, in the final coagulation event of this particular coalescent. Their work motivates the results in Iksanov and Möhle 2007 and Drmota et al. 2009 where the number of cuts needed to isolate the root of a random recursive tree is studied, and also is the basis for our joint work together with Götz Kersting and Arno Siri-Jégousse on the SFS of the Bolthausen-Sznitman coalescent discussed further below. Kersting, Pardo, and Siri-Jégousse 2014 provide asymptotics on the total internal and external branch lengths of the Bothausen-Sznitman coalescent, extending some results in Dhersin and Möhle 2013 for this particular coalescent, where, however, a recursion for the joint moments of external branch lengths for general Λ-coalescents is provided. Moreover, Drmota et al. 2007 give asymptotic results for the total branch length of the Bolthausen-Sznitman coalescent, whereas Gnedin and Yakubovich 2007, Iksanov and Möhle 2008, Iksanov, Marynych, and Möhle 2009, and Gnedin, Iksanov, Marynych, and Möhle 2014, provide asymptotics for the number of collisions, and the total and external branch lengths for Beta(a, b)-coalescents of parameters $(a \in (0,1), b > 0)$, $(a \in (0,2), b = 1)$, $(a = 2, b > 0)$, and $(a = 1, b > 0)$, respectively. Also J. Berestycki, N. Berestycki, and Schweinsberg 2008 establish an a.s. limit theorem for the number of blocks at small times, and related results for the block sizes, in Beta$(2 - \alpha, \alpha)$-coalescents; whereas Delmas, Dhersin, and Siri-Jégousse 2008 give asymptotic distributions for their lengths. Beta$(2 - \alpha, \alpha)$ coalescents are also studied in Dhersin, Freund, et al. 2013 who characterize the limit of the length of a randomly chosen external branch, and also in Kersting 2012 where the asymptotic distribution of their total branch length is given. J. Berestycki, N. Berestycki, and Limic 2014 give weak laws of large numbers for the total number of segregating sites, for both the infinite sites model and the infinite alleles model, in the general Λ-coalescent; they strengthen this result to strong laws whenever Λ is regularly varying at zero of order $\alpha \in (1,2)$ (i.e. $\Lambda(dx) = f(x)dx$ where $f(x) \sim Ax^{1-\alpha}$ as $x \to 0$ for some $1 < \alpha < 2$ and $A > 0$). Recently, in Diehl and Kersting 2019, L^1-laws of large numbers for the total tree length and the total length of external branches for general Λ-coalescents, including Λ-regularly varying at zero of order $\alpha = 1$, were provided.

Asymptotic results have also been obtained in the more thorough direction of the SFS and the AFS, but only for small families, i.e. for fixed family sizes b as the total initial number of particles n tends to ∞. J. Berestycki, N. Berestycki, and Schweinsberg 2007 give in-probability asymptotic sampling formulae for the SFS and AFS in Beta$(2 - \alpha, \alpha)$-coalescents, whereas Basdevant and Goldschmidt 2008 provide weak laws of large numbers for the AFS of the Bolthausen-Sznitman coalescent. J. Berestycki, N. Berestycki, and Limic 2014 give asymptotic a.s. sampling formulae for the SFS and the AFS of Λ-coalescents when Λ is regularly varying at zero of order $\alpha \in (1,2)$; while Diehl and Kersting 2019 give similar in-probability asymptotics for Λ-coalescents that are regularly varying at zero of order $\alpha = 1$, covering the Bolthausen-Sznitman coalescent.

Exact results for the SFS and AFS in the finite n case have also been derived, most of the time in terms of computationally intensive algorithms, hindering their applicability for large populations. Ewens 1972 gives the celebrated sampling formula for the AFS of Kingman's coalescent. Later, Fu 1995 provides the expected SFS in Kingman's coalescent. Twenty years later, Kukla and Pitters 2015 study the spectral decomposition of the jump matrix of the Kingman and Bolthausen-Sznitman coalescents; shortly afterwards, Spence, Kamm, and Song 2016 also derive such decompositions for general Λ- and Ξ-coalescents and, from this, manage to obtain an expression for the expected SFS in this general setting. These expressions are given in terms of matrix operations which in the case of the Bolthausen-Sznitman coalescent result in an algorithm requiring on the order of n^2 computations. In Hobolth, Siri-Jégousse, and Bladt 2019 another expression in terms of matrix operations is given for the SFS and other functionals for general Λ-coalescents, both in expected value (and higher moments) and in distribution. These expressions, however, are deduced from the theory of phase-type distributions, in particular distributions of rewards constructed on top of coalescent processes, and also require vast computations for large population sizes.

In Section 3 we present our joint work together with Götz Kersting and Arno Siri-Jégousse (Kersting, Siri-Jégousse, and H. Wences 2021) in which we provide a complete, easy-to-compute, first and second moment characterization of the SFS, both for finite and infinite population sizes, in the Bolthausen-Sznitman coalescent. We remark that analog results, in terms of exact easy-to-compute formulas, had only been obtained for Kingman's coalescent (Fu 1995) so far. Our formulas give corresponding L^2-laws of large numbers for the SFS which generalize and strengthen those in Diehl and Kersting 2019 for this particular coalescent. Notably, our results describe the complete SFS, including small, but also large family sizes, i.e. families that take up a positive fraction of the population as $n \to \infty$; thus providing new insights into the top of the genealogical tree. We believe that our results could serve as a rigorous mathematical basis for the development of new inference algorithms for this type of genealogy, which would suggest the presence of natural selection in biological populations (Desai, Walczak, and Fisher 2013; Melissa et al. 2021; Neher and Hallatschek 2013).

Finally, other functionals and important theoretical questions on coalescent processes have also been addressed. For example, in Schweinsberg 2000a a general criterion for the Λ-coalescent to come down from infinity is proved, whereas in J. Berestycki, N. Berestycki, and Limic 2010 the speed of this coming down from infinity is described. Limic 2010 provides an analog criterion for general Ξ-coalescents. Moreover, Gnedin, Iksanov, and Marynych 2011 study Λ-coalescents with a positive dust component, for example covering the $\mathrm{Beta}(a,b), a > 1, b > 0$, coalescents, and obtain limit distributions of the absorption time and the number of collisions.

Chapter 2

Population Evolution Models With Selection

In Chapter 1 we provided a general theory for coalescent processes, and motivated its study as the processes describing the limit genealogy population models (Section 1.3). In the present chapter we provide a general theory, with examples, on the limit genealogy for discrete-time and constant-size population evolution models. In Section 2.1 we introduce Cannings' models and, following the primary work of Möhle and Sagitov 2001, provide sufficient conditions for their genealogy to converge weakly to general Λ- and Ξ-coalescents, both in continuous and discrete time. In our exposition we use the theory developed in Chapter 1 in order to simplify the proofs in Möhle and Sagitov 2001, specially for Theorem 2.1.2 below. This allows us to recycle them in Section 2.2 where we introduce and study the multinomial models and, in Theorems 2.2.1 and 2.2.5, provide a new general criteria for the weak convergence their genealogies. Multinomial models resemble Cannings' models and in fact share their simplicity and manageability; however, they are capable of describing a wider range of offspring distributions, in particular non-symmetrical ones (see e.g. Section 2.3).

In Section **??** we briefly describe the Cannings' model constructed from supercritical Galton-Watson processes studied by Schweinsberg 2000b, and in Section 2.2.1 we describe its multinomial homologue introduced by Huillet and Möhle 2021. These two examples contain regimes that conceptually fall into the class of models studied in Section 2.2.2, where we specialize to the case of multinomial models whose genealogies are described by Λ-coalescents by assuming the occurrence of rare but recurrent large reproductive events stemming from single individuals along the generations; these large reproductive events may occur by mere chance, or, as seen in Section 2.3, may be a consequence of the strength of selection. We formalize this heuristic in our conditions (2.12) for Theorem 2.2.8 which generalize and ease the applicability of those given in Cortines and Mallein 2017. Under such conditions we prove the weak convergence of the genealogy to Kingman's and general Λ-coalescents.

In Theorem 2.2.11 of Section 2.2.3 we generalize Theorem 3.3 in Cortines and Mallein 2017, where the family frequencies in the multinomial model are given as renormalized Poisson-Dirichlet(α, θ) N-size biased picks; we apply the results of Section 2.2.2 and prove that the genealogy converges either to the Beta$(1-\frac{\theta}{\alpha}, 1+\frac{\theta}{\alpha})$ or to Kingman's coalescent, depending on whether $\theta \in (-\alpha, \alpha)$ or $\theta \geq \alpha$.

In Section 2.3 we introduce the exponential models of Brunet and Derrida 1997 in terms of discrete-time and constant-size populations undergoing mutation and natural section. In our joint work together with Emmanuel Schertzer, and building on Cortines and Mallein 2017, we use the theory developed in Section 2.2, including some examples of Section 2.2.3, to show that under strong selection regimes the limit genealogy of these models is given by the Bolthausen-Sznitman coalescent, whereas under mild selection regimes the limit genealogy is a discrete-time Poisson-Dirichlet coalescent (Theorem 2.3.2). We also provide asymptotics on the speed of selection for these two regimes in Theorem 2.3.12.

Finally, we end this chapter with a brief introduction to the continuous-time and constant-size model studied by Schweinsberg 2017a,b where again the Botlhausen-Sznitman coalescent appears in the context of natural selection.

2.1 Cannings' (Neutral) Models

Consider a population of constant size N evolving in discrete time. For every $t \in \mathbb{N}$ and $N \in \mathbb{N}$, let $\xi_N^{(t)} := \left(\xi_{1,N}^{(t)}, \cdots, \xi_{N,N}^{(t)}\right)$ be the offspring sizes of generation t, i.e. $\xi_{i,N}^{(t)}$ gives the number of children of the ith individual in the next generation $t+1$. In particular, the population size being fixed at N, the random vector $\xi_N^{(t)}$ must satisfy $\xi_{1,N}^{(t)} + \cdots + \xi_{N,N}^{(t)} \overset{a.s.}{=} N$. To construct the parental relations between generations t and $t+1$ we assume that parents choose children in the following way: the first individual of generation t chooses $\xi_{1,N}^{(t)}$ children among $\{1, \ldots, N\}$ uniformly at random without replacement and, continuing in the same way, the ith parent chooses $\xi_{i,N}^{(t)}$ children among the remaining (unpicked) children uniformly at random without replacement. Let $\tilde{\Pi}_{t+1}^{(N)}$ be the random partition of N given by

$$i \overset{\tilde{\Pi}_{t+1}^{(N)}}{\sim} j \iff i \text{ and } j \text{ share the same parent in generation } t$$

so that, for any $T \in \mathbb{N}$, the genealogy of the population from generation T backwards in time is identical in distribution to the inhomogeneous coalescent process

$$\varPi_t^{(N)} := \begin{cases} \mathbf{0}, & t = 0 \\ \mathrm{CO}_{h=1}^{t} \tilde{\Pi}_{T-h}^{(N)}, & t \in \{1, \cdots, T\}. \end{cases}$$

At this point we impose Cannings' (Cannings 1974, 1975) conditions on the law of $\left(\xi_N^{(t)}\right)_{t\geq 0}$, which will allow for the characterisation of the weak limit of the underlying genalogy as $N \to \infty$ in terms of exchangeable coalescent processes; these conditions are

1. **Static environment:** the sequence of offspring sizes $\left(\xi_N^{(t)}\right)_{t\geq 0}$ is i.i.d.,

2. **Neutrality:** $\xi_N^{(1)}$ is an exchangeable random vector.

Remark. *As we will see in Section 2.2, the neutrality assumption is not necessary for the study of their (exchangeable) genealogies, as long as the parental relationships are constructed in an exchangeable way, i.e. if the partition that groups children according to their parents is a random exchangeable partition.*

Note that under these assumptions the coagulation increments $\left(\tilde{\Pi}_t^{(N)}\right)_{t\geq 0}$ are i.i.d. so that we may write $\left(\Pi_t^{(N)}\right)_{t\geq 0}$ as

$$\Pi_t^{(N)} = \begin{cases} \mathbf{0}, & t = 0 \\ \mathrm{CO}_{h=1}^t \tilde{\Pi}_h^{(N)}, & t \in \{1, \cdots, T\}. \end{cases}$$

where the $\left(\tilde{\Pi}_h^{(N)}\right)_{h\geq 0}$ are i.i.d. exchangeable random partitions. This in turn allows us to construct the genealogy of the processes indefinitely backwards in time (very large T) which simplifies the time-scaling needed in order to obtain a non trivial weak limit for the genealogy as $N \to \infty$. This model, and variations of this model, including models with varying population size, have been extensively studied in the literature (Freund 2020; Huillet and Möhle 2013; Möhle and Sagitov 2001; Schweinsberg 2003).

As demonstrated in Möhle and Sagitov 2001, the time scale needed to obtain a non-trivial limit for the genealogy process $\left(\Pi_t^{(N)}\right)_{t\geq 0}$ is the quantity c_N^{-1} where

$$\begin{aligned} c_N &:= \mathbb{P}\,(\text{two randomly chosen individuals have the same parent}) \\ &= \mathbb{P}\left(1 \overset{\tilde{\Pi}_1^{(N)}}{\sim} 2\right). \end{aligned}$$

Note that c_N can be computed as

$$c_N = \frac{\mathbb{E}\left[(\xi_{1,N})_2\right]}{N-1}$$

where we have used the notation $(a)_b = a(a-1)\cdots(a-b+1)$ for $a > 0$ and $b \in \mathbb{Z}^+ \cup 0$. The quantity c_N^{-1} is also equal to $\mathbb{E}\left[T_2^{(N)}\right]$ where $T_2^{(N)}$ is the time needed (in generations) for two randomly chosen individuals to meet their most recent common ancestor (MRCA), i.e.

$$T_2^{(N)} \overset{d}{=} T_{i,j}^{(N)} := \inf_{t\geq 1}\left\{i \overset{\Pi_t^{(N)}}{\sim} j\right\}, \quad \forall i, j \in \{1, \ldots, N\}.$$

Indeed, the parental relations among any two consecutive generations being i.i.d. entails that $T_2^{(N)}$ is geometrically distributed with parameter c_N.

In their Theorem 2.1 Möhle and Sagitov 2001 give the most general result for the weak convergence of the genealogy of $n \in \mathbb{N}$ randomly chosen individuals

$$\left(\Pi_{[t/c_N]}^{(N,n)}\right)_{t\geq 0} := \left(\Pi_{[t/c_N]}^{(N)}\Big|_n\right)_{t\geq 0}$$

to coalescents with multiple collisions; mainly

Theorem 2.1.1 (Theorem 2.1 in Möhle and Sagitov 2001)**.** *Let* $\mathbf{Q}$ *be the coagulation rate matrix of the* Ξ*-coalescent with values in* $\mathscr{P}_n$ *given by*

$$\mathbf{Q}_{\pi,\pi'} := \begin{cases} \mu_\Xi\left(\mathscr{P}_\infty(\tilde{\pi})\right) & \text{if } \pi' = \mathrm{Coag}(\pi, \tilde{\pi}), \\ -\sum_{\tilde{\pi}\in\mathscr{P}_n} \mu_\Xi\left(\mathscr{P}_\infty(\tilde{\pi})\right) & \text{if } \pi' = \pi, \\ 0 & \text{otherwise}, \end{cases}$$

where $\mu_\Xi(\cdot) = \int_{\mathscr{P}_{[0,1]}} \varrho_\rho(\cdot)\Xi(d\rho)$*. Also let* $\mathbf{P}^{(N)}$ *be the transition matrix for the genealogy of* n *randomly chosen individuals given by*

$$\mathbf{P}^{(N)}_{\pi,\pi'} := \begin{cases} \mathbb{P}\left(\tilde{\Pi}_1^{(N)} = \tilde{\pi}\right) & \text{if } \pi' = \mathrm{Coag}\,(\pi,\tilde{\pi}) \\ 0 & \text{otherwise,} \end{cases}$$

where we have made a slight abuse of notation by writing $\tilde{\Pi}^{(N)} = \tilde{\pi}$ *instead of* $\tilde{\Pi}^{(N)}\Big|_n = \tilde{\pi}$*.*

Assume that

$$\mathbf{P}^{(N)} = \mathbf{I} + c_N\mathbf{Q} + o\,(c_N)\,. \tag{2.1}$$

I. *If* $c_N \to c > 0$ *then* $\left(\mathit{\Pi}_t^{(N,n)}\right)_{t\geq 0}$ *converges weakly in the product topology for* $\mathscr{P}_n^{\mathbb{N}}$ *to a Markov chain with initial state* $\mathbf{0}$ *and transition matrix* $\mathbf{I} + c\mathbf{Q}$*.*

II. *If* $c_N \to 0$ *then* $\left(\mathit{\Pi}_{[t/c_N]}^{(N,n)}\right)_{t\geq 0}$ *converges weakly in the Skorohod space* $C([0,\infty),\mathscr{P}_n)$ *to the* Ξ*-coalescent with initial state* $\mathbf{0}_n$*.*

Proof. **I.** Equation (2.1) plus the condition $c_N \to c > 0$ imply $\lim_{N\to\infty}\mathbf{P}^{(N)} = \mathbf{I} + c\mathbf{Q}$ which in turn gives convergence of the finite dimensional distributions; the latter is equivalent to weak convergence on $\mathscr{P}_n^{\mathbb{N}}$ (see Example 2.6 in Billingsley 1999).

II. Given that both $\left(\mathit{\Pi}_{[t/c_N]}^{(N,n)}\right)_{t\geq 0}$ and the Ξ-coalescent are Feller processes we need only prove convergence of their semigroups (Theorem 2.5 in Ethier and Kurtz 1986). By means of the equality

$$\left(\mathbf{P}^{(N)}\right)^k - (\mathbf{I} + c_N\mathbf{Q})^k = \sum_{i=1}^{k}\left(\mathbf{P}^{(N)}\right)^{i-1}\left(\mathbf{P}^{(N)} - (\mathbf{I} + c_n\mathbf{Q})\right)(\mathbf{I} + c_N\mathbf{Q})^{k-i}$$

(which follows by telescoping the sum on the right hand side), we compute, for every $t > 0$

$$\begin{aligned}\left\|\left(\mathbf{P}^{(N)}\right)^{\left[\frac{t}{c_N}\right]} - (\mathbf{I} + c_N\mathbf{Q})^{\left[\frac{t}{c_N}\right]}\right\| &\leq \left[\frac{t}{c_N}\right]\left\|\mathbf{P}^{(N)} - (\mathbf{I} + c_N\mathbf{Q})\right\|\left(1 + \left\|\mathbf{P}^{(N)} - (\mathbf{I} + c_N\mathbf{Q})\right\|\right)^{\left[\frac{t}{c_N}\right]} \\ &\leq \left[\frac{t}{c_N}\right]\left\|\mathbf{P}^{(N)} - (\mathbf{I} + c_N\mathbf{Q})\right\|\exp\left\{\left[\frac{t}{c_N}\right]\left(o\,(c_N) + o\left(c_N^2\right)\right)\right\} \\ &\overset{N\to\infty}{\longrightarrow} 0;\end{aligned}$$

where $\|\cdot\|$ refers to the usual operator norm, and we have used (2.1). Therefore

$$\lim_{N\to\infty}\left(\mathbf{P}^{(N)}\right)^{\left[\frac{t}{c_N}\right]} = \lim_{N\to\infty}(\mathbf{I} + c_N\mathbf{Q})^{\left[\frac{t}{c_N}\right]} = e^{t\mathbf{Q}}.$$

□

Möhle and Sagitov 2001 also give conditions under which (2.1) holds.

Theorem 2.1.2 (Theorem 2.1 in Möhle and Sagitov 2001)**.** *Let $\mathbf{P}^{(N)}$ and $\mathbf{Q}$ be as in the previous Theorem 2.1.1. Then the equality* (2.1) *holds for some μ_Ξ (corresponding to a Ξ-coalescent) if and only if the limits*

$$\phi_j(b_1,\dots,b_j) \coloneqq \lim_{N\to\infty} \frac{\mathbb{E}\left[\left(\xi_{1,N}\right)_{b_1}\cdots\left(\xi_{j,N}\right)_{b_j}\right]}{N^{b_1+\dots+b_j-j}c_N} \tag{2.2}$$

exist for $j\in\mathbb{N}$ and $b_1\geq \dots b_j\geq 2$.

Proof. Let $\tilde\pi$ be a coagulation increment different from $\mathbf{0}_{\#\tilde\pi}$, and denote by $\mathbf{b}=(b_1,\dots,b_{\#\tilde\pi})$ its block sizes ordered decreasingly; in particular $b_1+\dots+b_{\#\tilde\pi}=|\pi|$. We first compute

$$\begin{aligned}\frac{1}{c_N}\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde\pi)} &= \frac{1}{c_N}\frac{1}{(N)_{|\pi|}}\sum_{\substack{i_1,\dots,i_{\#\tilde\pi}\\ \text{all distinct}}}^{N}\mathbb{E}\left[\left(\xi_{i_1,N}\right)_{b_1}\cdots\left(\xi_{i_{\#\tilde\pi},N}\right)_{b_{\#\tilde\pi}}\right]\\ &= \frac{1}{c_N}\frac{(N)_{\#\tilde\pi}}{(N)_{|\pi|}}\mathbb{E}\left[\left(\xi_{1,N}\right)_{b_1}\cdots\left(\xi_{\#\tilde\pi,N}\right)_{b_{\#\tilde\pi}}\right]\\ &= \frac{\mathbb{E}\left[\left(\xi_{1,N}\right)_{b_1}\cdots\left(\xi_{\#\tilde\pi,N}\right)_{b_{\#\tilde\pi}}\right]}{N^{b_1+\dots+b_{\#\tilde\pi}-\#\tilde\pi}c_N}\left(1+\mathcal{O}\left(N^{-1}\right)\right)\end{aligned}$$

where we have used Stirling's approximation $\Gamma(m+c)/\Gamma(m+d)=m^{c-d}\left(1+\mathcal{O}\left(1/m\right)\right)$ as $m\to\infty$. Note that the last line above does not depend on π and that it is invariant under permutations of the partition $\tilde\pi$, i.e. for any finite permutation σ we have

$$\frac{1}{c_N}\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde\pi)} = \frac{1}{c_N}\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\sigma(\tilde\pi))}.$$

Thus (2.1) holds if and only if the limits in (2.2) exist for any $j\in\mathbb{N}$, $b_1\geq 2$, and $b_2,\dots,b_j\in\{1,2,\dots\}$; in this case, by Theorem 1.2.8, the quantities $\phi_j(b_1,\dots,b_j)$ define a coagulation measure through the equality

$$\phi_j(b_1,\dots,b_j)=\mu_\Xi\left(\mathscr{P}_\infty\left(\pi_{\mathbf{b}}\right)\right),$$

where

$$\pi_{\mathbf{b}}\coloneqq\{\{1,\dots,b_1\},\{b_1+1,\dots,b_1+b_2\},\dots,\{b_1+\dots+b_{j-1}+1,\dots,n\}\}.$$

Thus, it only remains to note that in fact it is enough to verify (2.2) for $b_1\geq b_2\geq\dots\geq b_j\geq 2$. Indeed, the result follows easily from the additivity of μ_Ξ which gives the recursion

$$\mu_\Xi\left(\mathscr{P}_\infty\left(\pi_{\mathbf{b}}\right)\right)=\mu_\Xi\left(\mathscr{P}_\infty\left(\pi_{\mathbf{b}}\cup\{\{n+1\}\}\right)\right)+\sum_{k=1}^{j}\mu_\Xi\left(\mathscr{P}_\infty\left(\pi_{\mathbf{b},k}\right)\right)$$

where $\pi_{\mathbf{b},k}$ is constructed from $\pi_{\mathbf{b}}$ by adding the element $n+1$ to the kth block of $\pi_{\mathbf{b}}$ (i.e. to the set $\{b_1+\dots+b_{k-1}+1,\dots,b_1+\dots+b_{k-1}+b_k\}$). This entails that if (2.2) is verified for $b_1\geq b_2\geq\dots\geq b_j\geq 2$, $j\in\mathbb{N}$, then it is also verified for $b_1\geq b_2\geq\dots\geq b_j>1=b_{j+1}$; and, recursively, that it is also verified for $b_1\geq b_2\geq\dots\geq b_j>1=b_{j+1}=\dots=b_{j+k}$. □

2.1.1 Genealogies of Supercritical Galton-Watson Processes

In this section we briefly describe (without proof) an application of Theorem 2.1.1 for the genealogy of supercritical Galton-Watson processes, a result obtained in Schweinsberg 2003.

In Schweinsberg's model, each individual has an independent and identically distributed number of children, say $X_i, 1 \leq i \leq N$, so that the random variable $X_1 + \cdots + X_N$ gives the total size of the produced offspring. In order to fix the population size at N, first the condition $\mathbb{E}[X_1] > 1$ is imposed which ensures that $X_1 + \cdots + X_N \geq N$ with sufficiently high probability, and, second, the surplus of $X_1 + \cdots + X_N - N$ children are uniformly chosen among all the offspring produced and are killed so that the definitive offspring sizes $(\xi_{1,N}, \ldots, \xi_{N,N})$ of each parent are set. On the event that $X_1 + \cdots + X_N < N$, the condition $\mathbb{E}[X_1] > 1$ allows us to set $(\xi_{1,N}, \ldots, \xi_{N,N})$ to an arbitrary value without affecting the limit behaviour of the genealogy as $N \to \infty$. Schweinsberg 2003 demonstrates the following description of the genealogy depending on the tail of the distribution of X_1 expressed in the condition

$$\lim_{k\to\infty} k^{\alpha}\mathbb{P}(X_1 \geq k) > 0. \tag{2.3}$$

Theorem 2.1.3 (Schweinsberg 2003). *As $N \to \infty$:*

I. *If $\mathbb{E}[X_1^2] < \infty$ or (2.3) holds with $\alpha = 2$, then $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ converges weakly to Kingman's coalescent in the Skorohod space $D([0,\infty), \mathscr{P}_n)$.*

II. *If (2.3) holds with $1 \leq \alpha < 2$, then $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ converges weakly to the Beta coalescent of parameter α in the Skorohod space $D([0,\infty), \mathscr{P}_n)$.*

III. *If (2.3) holds with $0 \leq \alpha < 1$, then $\left(\Pi^{(N,n)}_{t}\right)_{t\geq 0}$ converges weakly to a discrete-time coalescent process. The coagulation increments $\left(\tilde{\Pi}_h\right)_{h\geq 0}$ of the limit process are i.i.d. and their asymptotic frequencies in Kingman's representation (as in Section 1.1.6) are Poisson-Dirichlet(α,0) distributed.*

2.2 Multinomial Populations

In this section we come back to the general scenario of a population with constant size N evolving in discrete time, and let $\xi_N^{(t)} := \left(\xi_{1,N}^{(t)}, \cdots, \xi_{N,N}^{(t)}\right)$ be the offspring sizes of generation t. This time, however, instead of Cannings' conditions on $\left(\xi_N^{(t)}\right)_{t\geq 0}$ (Section 2.1), we impose the conditions

1. **Static environment:** the sequence of vectors of offspring sizes $\left(\xi_N^{(t)}\right)_{t\geq 0}$ is i.i.d.,
2. **multinomial Offspring Law:** The distribution of the vector $\xi_N^{(1)}$ is given by

$$\begin{aligned} \xi_N^{(1)}\Big|\eta^{(N)} &\stackrel{d}{=} \text{Multinomial}(\eta^{(N)}) \text{ and} \\ \mathbb{P}\left(\eta^{(N)} \in \cdot\right) &= \nu^{(N)}(\cdot), \end{aligned}$$

where of course $\nu^{(N)}$ is a probability measure supported on the set

$$\left\{\eta \in [0,1]^N : \sum_{i=1}^{N} \eta_i = 1\right\}.$$

That is, we have replaced Cannings' **neutrality** condition by the **multinomial offspring law** condition (see Cortines and Mallein 2017; Huillet and Möhle 2021).

Multinomial models are surely reminiscent of Cannings' models; however, an important distinction is that the random vector of probabilities $\eta^{(N)}$ needs not be exchangeable, and thus neither does $\xi_N^{(1)}$; in particular the offspring sizes $(\xi_{1,N}, \ldots, \xi_{N,N})$ need not be identically distributed. In general, every Cannings' model has a multinomial approximation given by setting $\eta^{(N)} = \frac{1}{\sum_{i=1}^N \xi_{i,N}} (\xi_{1,N}, \ldots, \xi_{N,N})$, in the same spirit as Huillet and Möhle 2021 can be regarded as the multinomial homologue of Schweinsberg 2003. Nonetheless, as explained in Section 2.2.1, Cannings' and multinomial models do have different modelling assumptions, such as the distinction between sampling with replacement and sampling without replacement. On the other hand, multinomial models do not have in general a Cannings analogue, unless $\eta^{(N)}$ is itself exchangeable, or is made exchangeable by permuting its entries uniformly at random, thus losing any information on the possible asymmetries of the offspring size distribution. In fact, as seen further below in Sections 2.3 and 2.2.3, multinomial models naturally arise in the context of asymmetric offspring distributions, such as those expected in populations undergoing natural selection (although note that, since children choose parents in an i.i.d manner, multinomial models cannot incorporate the inheritance of fitness traits). Thus, multinomial models are more versatile for modelling applications than Cannings' models which only cover symmetric offspring distributions.

At the same time, however, the exchangeability of Kingman's paint-box procedure makes the genealogy $\mathit{\Pi}^{(N,n)}$ of n randomly chosen individuals in the multinomial model with random probabilities $\eta^{(N)}$, identical in distribution to the genealogy of the corresponding Cannings' model where $\eta^{(N)}$ is made exchangeable by permuting its entries uniformly at random. Thus, even when multinomial models may deal with asymmetric populations not covered by Cannings' models, at the level of their genealogies, however, multinomial models can in fact be regarded as special cases of Cannings' models, inheriting their ease of management. Indeed, in the same way as in Section 2.1, the **static environment** condition allows us to study the genealogy of the multinomial population arbitrarily backwards in time through the coalescent process

$$\mathit{\Pi}_t^{(N)} := \begin{cases} \mathbf{0}_N, & t = 0, \\ \mathrm{CO}_{h=1}^{t}\, \tilde{\Pi}_h^{(N)}, & t > 0, \end{cases} \tag{2.4}$$

where the coagulation increments $\tilde{\Pi}_h^{(N)}$ are constructed through Kingman's representation (Theorem 1.1.6) from a sequence of i.i.d. random mass-partitions $\left(\tilde{\rho}_t^{(N)}\right)_{t\geq 0}$ with common distribution coinciding with that of $\left(\eta^{(N)}\right)^{\downarrow}$ the vector $\eta^{(N)}$ arranged in decreasing order. It is easily seen that the parental relations between two consecutive generations, say t and $t+1$, can be constructed by allowing children in generation $t+1$ to independently choose parents according to an independent realization of $\tilde{\rho}_t^{(N)}$. Grouping children with the same parent gives in fact the paint-box construction of $\tilde{\Pi}_t^{(N)}$ from $\tilde{\rho}_t^{(N)}$.

We are again interested in the weak limit as $N \to \infty$ of $\left(\Pi_t^{(N,n)}\right)_{t\geq 0}$ under a suitable time scale. Again the correct time scale is given by the probability c_N which, in this case, can be computed as

$$c_N = \mathbb{E}\left[\sum_{k=1}^{\infty}\left(\eta_k^{(N)}\right)^2\right]. \tag{2.5}$$

The same proof as for Theorem 2.1.1 allows us to restate it in the terminology of this section without proof, giving, however, a new general criterion for the weak convergence of the genealogy of multinomial models. The transition matrix $\mathbf{P}^{(N)}$ is now given by

$$\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde{\pi})} = \int_{\mathscr{P}_{[0,1]}} \varrho_\rho(\mathscr{P}_\infty(\tilde{\pi}))\nu^{(N)}(d\rho).$$

Condition (2.1) of Theorem 2.1.1, on the other hand, translates into

$$\mu_{\nu^{(N)}}(\mathscr{P}_\infty(\tilde{\pi})) = \begin{cases} c_N\mu_\Xi(\mathscr{P}_\infty(\tilde{\pi})) + o\left(c_N\right) & \text{if } \tilde{\pi} \in \mathscr{P}_n \setminus \{\mathbf{0}_n\} \\ 1 - c_N\sum_{\tilde{\pi}\in\mathscr{P}_n}\mu_\Xi(\mathscr{P}_\infty(\tilde{\pi})) + o\left(c_N\right) & \text{if } \tilde{\pi} = \mathbf{0}_n. \end{cases} \tag{2.6}$$

Theorem 2.2.1. *Let $\mathbf{P}^{(N)}$ and $\mathbf{Q}$ be as in Theorem 2.1.1. Assume that* (2.6) *holds.*

I. *If $c_N \to c > 0$ then, as $N \to \infty$, $\left(\Pi_t^{(N,n)}\right)_{t\geq 0}$ converges weakly in the product topology for $\mathscr{P}_n^{\mathbb{N}}$ to a Markov chain with initial state $\mathbf{0}_n$ and transition matrix $\mathbf{I} + c\mathbf{Q}$.*

II. *If $c_N \to 0$ then, as $N \to \infty$, $\left(\Pi_{[t/c_N]}^{(N,n)}\right)_{t\geq 0}$ converges weakly in the Skorohod space $C([0,\infty), \mathscr{P}_n)$ to the Ξ-coalescent with initial state $\mathbf{0}_n$.*

Corollary 2.2.2. *If $\tilde{\rho}^{(N)}$ converges weakly on $\mathscr{P}_{[0,1]}$ to $\tilde{\rho}^{(\infty)}$ as $N \to \infty$, then $\left(\Pi_t^{(N,n)}\right)_{t\geq 0}$ converges weakly in the product topology for $(\mathscr{P}_n)^{\mathbb{N}}$ to a Markov chain with initial state $\mathbf{0}_n$ and transition matrix given by*

$$\mathbf{P}^{(\infty)}_{\pi,\mathrm{Coag}(\pi,\tilde{\pi})} = \mathbb{E}\left[\varrho_{\tilde{\rho}^{(\infty)}}(\mathscr{P}_\infty(\tilde{\pi}))\right].$$

Proof. Proposition 1.1.7 readily gives $\mathbb{E}\left[\varrho_{\tilde{\rho}^{(N)}}(\mathscr{P}_\infty(\tilde{\pi}))\right] = \mathbb{E}\left[\varrho_{\tilde{\rho}^{(\infty)}}(\mathscr{P}_\infty(\tilde{\pi}))\right] + o\left(1\right)$, which is the condition of Theorem 2.2.1 for the case $\lim_{N\to\infty} c_N > 0$ if we replace c_N and c by 1. □

Corollary 2.2.3. *Assume that*

$$\mathbb{E}\left[\left(\eta_1^{(N)}\right)^3 + \cdots + \left(\eta_N^{(N)}\right)^3\right] = o\left(c_N\right). \tag{2.7}$$

Then, as $N \to \infty$, $\left(\Pi_{[t/c_N]}^{(N,n)}\right)_{t\geq 0}$ converges weakly in the Skorohod space $C([0,\infty), \mathscr{P}_n)$ to Kingman's coalescent with initial state $\mathbf{0}_n$.

Proof. Observe that for any $N \geq j$ and $b_1 \geq \cdots \geq b_j \geq 1$ we have

$$\begin{aligned}
\sum_{\substack{i_1,i_2,\ldots,i_j \\ \text{all distinct}}}^{N} \left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_j}^{(N)}\right)^{b_j} &\leq \sum_{\substack{i_1,i_2,\ldots,i_{j-1} \\ \text{all distinct}}}^{N} \left\{ \left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_{j-1}}^{(N)}\right)^{b_{j-1}} \sum_{i_j \notin \{i_1,\ldots,i_{j-1}\}} \left(\eta_{i_j}^{(N)}\right)^{b_j} \right\} \\
&\leq \sum_{\substack{i_1,i_2\ldots,i_{j-1} \\ \text{all distinct}}}^{N} \left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_{j-1}}^{(N)}\right)^{b_{j-1}} \times 1 \\
&\;\;\vdots \\
&\leq \sum_{k=1}^{N} \left(\eta_k^{(N)}\right)^{b_1}. \qquad (2.8)
\end{aligned}$$

This upper bound together with

$$\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde{\pi})} = \sum_{\substack{i_1,\ldots,i_{\#\tilde{\pi}} \\ \text{all distinct}}}^{N} \mathbb{E}\left[\left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_{|\tilde{\pi}|}}^{(N)}\right)^{b_{|\tilde{\pi}|}}\right]$$

and (2.7) give, for any $\pi \in \mathscr{P}_n$ and coagulation increment $\tilde{\pi} \in \mathscr{P}_{|\pi|}$ with ordered block sizes $b_1 \geq \cdots \geq b_j$, $j = |\tilde{\pi}|$,

$$\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde{\pi})} = \begin{cases} o\left(c_N\right) & \text{if } b_1 \geq 3 \text{ or } b_2 > 1, \\ c_N & \text{otherwise,} \end{cases}$$

so that the proof is finished by an application of Theorem 2.2.1. □

Corollary 2.2.4. *Assume that the conditions*

$$\mathbb{E}\left[\left(\eta_2^{(N)}\right)^3 + \cdots + \left(\eta_N^{(N)}\right)^3\right] = o\left(c_N\right),$$

and

$$\exists \beta > 0: \quad \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{\beta}\right] = o\left(c_N\right)$$

both hold, then $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly in the Skorohod topology for* $D([0,\infty), \mathscr{P}_n)$ *to Kingman's coalescent.*

Proof. By Corollary 2.2.3 and the first hypothesis it is enough to prove $\mathbb{E}\left[\left(\eta_1^{(N)}\right)^3\right] = o\left(c_N\right)$. The result is trivial if $\beta \leq 3$, we thus assume $\beta > 3$. Observe that for $\lambda \in (0,2)$ we have, using Hölder's inequality,

$$\begin{aligned}
\mathbb{E}\left[\left(\eta_1^{(N)}\right)^3\right] &= \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{\lambda}\left(\eta_1^{(N)}\right)^{3-\lambda}\right] \\
&\leq \mathbb{E}\left[\left(\eta_1^{(N)}\right)^2\right]^{\lambda/2} \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{\frac{2(3-\lambda)}{2-\lambda}}\right]^{1-\lambda/2}
\end{aligned}$$

so that, choosing $\lambda \in (0,2)$ to ensure $\frac{2(3-\lambda)}{2-\lambda} > \beta$ and using the second hypothesis, we obtain

$$\begin{aligned}\frac{\mathbb{E}\left[\left(\eta_1^{(N)}\right)^3\right]}{c_N} &\leq \frac{\mathbb{E}\left[\left(\eta_1^{(N)}\right)^{\frac{2(3-\lambda)}{2-\lambda}}\right]^{1-\lambda/2}}{c_N^{1-\lambda/2}} \\ &\leq \left(c_N^{-1}\mathbb{E}\left[\left(\eta_1^{(N)}\right)^{\beta}\right]\right)^{1-\lambda/2} \\ &\overset{N\to\infty}{\to} 0.\end{aligned}$$

□

By essentially the same proof as in Theorem 2.1.2 we obtain the following general necessary and sufficient condition for equation (2.6) to hold.

Theorem 2.2.5. *Let $\mathbf{P}^{(N)}$ and $\mathbf{Q}$ be as in the previous Theorem 2.2.1. Then the equality* (2.6) *holds for some Ξ-coalescent if and only if the limits*

$$\phi_j(b_1,\ldots,b_j) := \lim_{N\to\infty} \frac{\mathbb{E}\left[\left(\prod_{i=1}^j \left(s_i^{(N)}\right)^{b_j-1}\right)\prod_{i=1}^j\left(1-\left(s_1^{(N)}+\cdots+s_i^{(N)}\right)\right)\right]}{c_N} \tag{2.9}$$

exist for $j \in \mathbb{N}$ and $b_1 \geq \ldots b_j \geq 2$, where $\left(s_1^{(N)}, s_2^{(N)},\ldots\right)$ is a size-biased reordering of $\eta^{(N)}$.

Proof. We compute, recalling the notation $\mathbf{b} = (b_1,\ldots,b_{\#\tilde{\pi}})$ for the block sizes of the coagulation increment $\tilde{\pi} \neq \mathbf{0}_{\#\pi}$ in $\left(\Pi_t^{(N,n)}\right)_{t\geq 0}$,

$$\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde{\pi})} = \sum_{\substack{i_1,\ldots,i_{\#\tilde{\pi}}\\ \text{all distinct}}}^{N} \mathbb{E}\left[\left(\eta_{i_1}^{(N)}\right)^{b_1}\cdots\left(\eta_{i_{\#\tilde{\pi}}}^{(N)}\right)^{b_{\#\tilde{\pi}}}\right]$$

Now observe that if $\eta_i^{(N)} = 0$ then the corresponding terms in the right hand side above are zero, and also $\mathbb{P}\left(\eta_i^{(N)} \in \left(s_1^{(N)},\ldots,s_N^{(N)}\right)\right) = 0$. Thus we may assume $\eta_i^{(N)} > 0$ for all $i \in [N]$. Observe that for distinct indexes $i_1,\ldots,i_b$ we have

$$\begin{aligned}\mathbb{P}\left(s_k^{(N)} = \eta_{i_k}^{(N)}, 1\leq k\leq b|\eta^{(N)}\right) &= \mathbb{P}\left(s_1^{(N)} = \eta_{i_1}^{(N)}|\eta^{(N)}\right)\cdots\mathbb{P}\left(s_b^{(N)} = \eta_{i_b}^{(N)}|\eta^{(N)}, s_1^{(N)},\cdots,s_{b-1}^{(N)}\right) \\ &= \prod_{k=1}^b \frac{\eta_{i_k}^{(N)}}{1-\sum_{j=1}^{k-1}\eta_{i_j}^{(N)}}\end{aligned}$$

so that

$$\begin{aligned}
&\sum_{\substack{i_1,\dots,i_{\#\tilde\pi}\\ \text{all distinct}}}^{N} \mathbb{E}\left[\left(\eta_{i_1}^{(N)}\right)^{b_1}\cdots\left(\eta_{i_{\#\tilde\pi}}^{(N)}\right)^{b_{\#\tilde\pi}}\right]\\
&=\sum_{\substack{i_1,\dots,i_{\#\tilde\pi}\\ \text{all distinct}}}^{N} \mathbb{E}\left[\prod_{k=1}^{\#\tilde\pi}\frac{\eta_{i_k}^{(N)}}{1-\sum_{j=1}^{k-1}\eta_{i_j}^{(N)}}(1-\sum_{j=1}^{k-1}\eta_{i_j}^{(N)})\left(\eta_{i_k}^{(N)}\right)^{b_k-1}\right]\\
&=\mathbb{E}\left[\sum_{\substack{i_1,\dots,i_{\#\tilde\pi-1}\\ \text{all distinct}}}^{N}\mathbb{P}\left(s_k^{(N)}=\eta_{i_k}^{(N)},1\le k\le\#\tilde\pi|\eta^{(N)}\right)\mathbb{E}\left[\prod_{j=1}^{\#\tilde\pi}(1-\sum_{j=1}^{k-1}s_{i_j}^{(N)})\left(s_{i_k}^{(N)}\right)^{b_k-1}|\eta^{(N)},s_1^{(N)},\dots,s_k^{(N)}\right]\right]\\
&=\mathbb{E}\left[\prod_{j=1}^{\#\tilde\pi}(1-\sum_{j=1}^{k-1}s_{i_j}^{(N)})\left(s_{i_k}^{(N)}\right)^{b_k-1}\right]
\end{aligned}$$

The proof is finished by following the same argument as in Theorem 2.1.2.

□

2.2.1 A Class of Generalized Wright-Fisher Models

In this section we briefly present the model introduced in Huillet and Möhle 2021. This model can be seen as an adaptation of condition (2.3) and Theorem 2.1.3 to the multinomial model described in this section. In this adaptation, the random probabilities $\eta^{(N)}$ of the multinomial model are constructed by normalizing a sequence $(X_i)_{1\le i\le N}$ of i.i.d positive random variables which are assumed to satisfy

$$\mathbb{P}\left(X>x\right)\sim x^{-\alpha}\ell(x),\quad x\to\infty, \tag{2.10}$$

for some constant $\alpha\ge 0$, and a function $\ell\colon(0,\infty)\to(0,\infty)$ slowly varying at infinity (c.f. condition (2.3)). This models differs from that in Schweinsberg 2003 in that X need not be integer-valued; and also in that it does not require the "supercriticality" condition $\mathbb{E}[X]>1$, although this condition can be imposed by considering aX which would yield the same random vector of probabilities

$$\eta^{(N)}=\left(\frac{X_1}{\sum_{i=1}^N X_i},\cdots,\frac{X_N}{\sum_{i=1}^N X_i}\right)$$

. Also, the new generation is constructed through a sample with replacement, as opposed to the sample without replacement done in Schweinsberg 2003. In Huillet and Möhle 2021 the asymptotic genealogy of this process is derived.

Theorem 2.2.6 (Theorem 1 in Huillet and Möhle 2021)**.** *As $N\to\infty$,*

I. *If $\mathbb{E}[X^2]<\infty$ (in particular if (2.10) holds with $\alpha>2$), then $\left(\Pi_{[t/c_N]}^{(N,n)}\right)_{t\ge0}$ converges weakly to Kingman's coalescent in the Skorohod space $D([0,\infty),\mathscr{P}_n)$, and c_N satisfies $c_N\sim\frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2}N^{-1}$.*

II. *If* (2.10) *holds with* $\alpha = 2$, *then* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly to Kingman's coalescent in the Skorohod space* $D([0,\infty), \mathscr{P}_n)$, *and* c_N *satisfies* $c_N \sim \frac{2}{\mathbb{E}[X]^2}\ell^*(N)N^{-1}$, *where* $\ell^*(x) := \int_1^x \frac{\ell(s)}{s} ds$.

III. *If* (2.10) *holds with* $\alpha \in (1,2)$, *then* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly to the Beta*$(2-\alpha,\alpha)$ *coalescent in the Skorohod space* $D([0,\infty), \mathscr{P}_n)$, *and* c_N *satisfies* $c_N \sim \alpha \mathsf{B}(2-\alpha,\alpha)\mathbb{E}\left[X\right]^{-\alpha}\ell(N)N^{1-\alpha}$.

IV. *If* (2.10) *holds with* $\alpha = 1$ *then* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly to the Bolthausen-Sznitman coalescent in the Skorohod space* $D([0,\infty), \mathscr{P}_n)$, *and if* $(a_N)_{N\in\mathbb{N}}$ *satisfies* $\ell^*(a_N) \sim a_N/N$, *then* c_N *satisfies* $c_N \sim \ell(a_N)/\ell^*(a_N) = N\ell(a_N)/a_N$.

V. *If* (2.10) *holds with* $\alpha \in (0,1)$, *then* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly to the discrete-time Poisson-Dirichlet*$(\alpha, 0)$ *coalescent, and* $c_N \to 1-\alpha$.

VI. *If* (2.10) *holds with* $\alpha = 0$, *then* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly to the star-shaped coalescent and* $c_N \to 1$.

2.2.2 Single-Parent Offspring Bursts

In this section we provide a simpler sufficient condition for (2.6) to hold which gives a Λ-coalescent as the limit genealogy. There are two simple heuristics from distinct population dynamics which motivate this section: 1) all the individuals in the population have the same fitness but occasionally, by mere chance, there occur large reproductive events where a single individual takes up a large fraction of the offspring in the next generation, and 2) the population is under the effect of strong selection so that the family frequencies given by the mass-partition $\tilde{\rho}^{(N)}$ are dominated by the family frequency $\tilde{\rho}^{(N)}_1$ of the fittest individual. In both of these scenarios we expect that all the coagulations that remain observable in the limit correspond to very large reproduction events of single individuals along the generations (see Cortines and Mallein 2017; Huillet and Möhle 2021; Schweinsberg 2003); these coagulations must then be driven by $\tilde{\rho}^{(N)}_1$, the largest value of $\tilde{\rho}^{(N)}$. In practice we would rather allow more flexibility and consider any random (possibly unordered) vector of probabilities $\eta^{(N)} \overset{a.s.}{\in} \left\{\eta \in [0,1]^N \colon \sum_{i=1}^N \eta_i = 1\right\}$ and prove that all the coagulations that "survive" in the limit are driven by the frequency $\eta^{(N)}_1$. This entails that $\eta^{(N)}_1$ should approximate $\tilde{\rho}^{(N)}_1$ as $N\to\infty$, but working with $\eta^{(N)}_1$ instead of $\tilde{\rho}^{(N)}_1$ may ease the study of the underlying genealogy of the population (also note that our setup does include the case where $\eta^{(N)}_i = \tilde{\rho}^{(N)}_i$ for all $1 \leq i \leq N$). Of particular interest is the case where $\left(\eta^{(N)}_1, \dots, \eta^{(N)}_N\right)$ is a size-biased pick from $\tilde{\rho}^{(N)}$ for which the distribution is explicitly known depending on the distribution of $\tilde{\rho}^{(N)}$.

The first result, an easy consequence of Theorem 2.2.1, reads as follows.

Proposition 2.2.7. *Let $\lambda_{b,b} = \int_0^1 p^{b-2}\Lambda(dp)$ for a coagulation measure Λ, and assume that*

$$\mathbb{E}\left[\left(\eta_1^{(N)}\right)^b\right] \sim \frac{c_N \lambda_{b,b}}{\lambda_{2,2}}, \quad \forall b \geq 2.$$

Then $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ converges weakly in the Skorohod space $C([0,\infty),\mathscr{P}_n)$ to the Λ-coalescent.

Remark. *Given that $\eta^{(N)}$ need not have any particular order, the main hypothesis of the above proposition can be replaced by the seemingly weaker condition*

$$\exists k \in \mathbb{N}\colon \mathbb{E}\left[\left(\eta_k^{(N)}\right)^b\right] \sim \frac{c_N \lambda_{b,b}}{\lambda_{2,2}}, \quad \forall b \geq 2.$$

To simplify the writting we stick to the index 1.

Proof of Proposition 2.2.7. The equality (2.5) plus the hypothesis give

$$\mathbb{E}\left[\left(\eta_2^{(N)}\right)^2 + \cdots + \left(\eta_N^{(N)}\right)^2\right] = o\left(c_N\right).$$

Also, by the same type of estimate as in (2.8) we have, for $b_1 \geq 2$,

$$\sum_{\substack{i_1\geq 2, i_2,\ldots,i_j \\ \text{all distinct}}}^{N} \left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_j}^{(N)}\right)^{b_j} \leq \sum_{k=2}^{N} \left(\eta_k^{(N)}\right)^2. \tag{2.11}$$

Combining this we have, for any $\pi \in \mathscr{P}_n$ and coagulation increment $\tilde{\pi} \in \mathscr{P}_{\#\pi} \setminus \mathbf{0}_{\#\pi}$ with ordered block sizes $b_1 \geq \cdots \geq b_j$, $j = \#\tilde{\pi}$,

$$\begin{aligned}
\mathbf{P}^{(N)}_{\pi,\mathrm{Coag}(\pi,\tilde{\pi})} &= \mathbb{E}\left[\sum_{\substack{i_1=1, i_2,\ldots,i_j \\ \text{all distinct}}}^{N} \left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_j}^{(N)}\right)^{b_j}\right] + o\left(c_N\right) \\
&= \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{b_1} \sum_{\substack{i_2\geq 2, i_2,\ldots,i_j \\ \text{all distinct}}}^{N} \left(\eta_{i_2}^{(N)}\right)^{b_2} \cdots \left(\eta_{i_j}^{(N)}\right)^{b_j}\right] + o\left(c_N\right).
\end{aligned}$$

Again using (2.11) we have

$$\mathbb{E}\left[\sum_{\substack{i_1=1, i_2,\ldots,i_j \\ \text{all distinct}}}^{N} \left(\eta_{i_1}^{(N)}\right)^{b_1} \cdots \left(\eta_{i_j}^{(N)}\right)^{b_j}\right] = \begin{cases} o\left(c_N^{-1}\right) & \text{if } b_2 \geq 2, \\ \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{b_1} (1-\eta_1^{(N)})^{\#\pi - b_1}\right] & \text{otherwise.} \end{cases}$$

It thus remains to prove that $\mathbb{E}\left[\left(\eta_1^{(N)}\right)^{b} (1-\eta_1^{(N)})^{n-b}\right] \sim \frac{c_N^{-1}}{\lambda_{2,2}}\lambda_{b,n}$ for all $n \geq b \geq 2$. Given the recursion formula $\lambda_{b,n} = \lambda_{b,n+1} + \lambda_{b+1,n+1}$ and the equation

$$\mathbb{E}\left[\left(\eta_1^{(N)}\right)^{b_1} (1-\eta_1^{(N)})^{\#\pi-b_1}\right] = \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{b_1} (1-\eta_1^{(N)})^{\#\pi+1-b_1}\right] + \mathbb{E}\left[\left(\eta_1^{(N)}\right)^{b_1+1} (1-\eta_1^{(N)})^{\#\pi+1-b_1-1}\right]$$

we may assume $n = b$. The required asymptotic for this case is given by the hypothesis. ☐

The remainder of this section is inspired and extends some of the results in Cortines and Mallein 2017, with the purpose of easing their applicability. Formally, assume that there exists a sequence L_N and a function $f\colon (0,1) \to \mathbb{R}^+$ such that

$$i) \lim_{N\to\infty} L_N = \infty, \quad ii) \lim_{N\to\infty} L_N \mathbb{P}\left(\eta_1^{(N)} > x\right) = f(x), \text{ and} \tag{2.12}$$
$$iii) \lim_{N\to\infty} L_N \mathbb{E}\left[\left(\eta_2^{(N)}\right)^2 + \cdots + \left(\eta_N^{(N)}\right)^2\right] = 0.$$

In light of the inequality $\left(\eta_2^{(N)}\right)^2 + \cdots + \left(\eta_N^{(N)}\right)^2 \leq \left(\max_{2\leq i\leq N} \eta_i^{(N)}\right)\left(\eta_2^{(N)} + \cdots + \eta_N^{(N)}\right) \leq \left(\max_{2\leq i\leq N} \eta_i^{(N)}\right)$, the third condition above may be replaced by the stronger condition

$$iii') \lim_{N\to\infty} L_N \mathbb{E}\left[\max_{2\leq i\leq N} \eta_i^{(N)}\right] = 0.$$

Theorem 2.2.8. *Assume that all the conditions in* (2.12) *hold and that*

$$iv) \quad \int_0^1 x \left(\sup_{N\in\mathbb{N}} L_N \mathbb{P}\left(\eta_1^{(N)} > x\right)\right) dx < \infty$$

also holds. Then, as $N \to \infty$, $c_N \sim L_N^{-1} \int_0^1 2xf(x)dx$ *and the time-scaled coalescent processes* $\left(\Pi^{(N,n)}_{[t/L_N]}\right)_{t\geq 0}$ *converges weakly in the Skorohod space* $C([0,\infty), \mathscr{P}_n)$ *to the* Λ*-coalescent with* Λ *given by* $\int_y^1 \frac{\Lambda(dx)}{x^2} = f(y)$.

Proof. For any $b \geq 2$, we have

$$\mathbb{E}\left[\left(\eta_1^{(N)}\right)^b\right] = L_N^{-1} \int_0^1 bx^{b-1} L_N \mathbb{P}\left(\eta_1^{(N)} > x\right) dx.$$

Next, the integrability hypothesis on $x \sup_{N\in\mathbb{N}} L_N \mathbb{P}\left(\eta_1^{(N)} > x\right)$ allows us to apply dominated convergence in the integral of the right hand side, obtaining

$$\mathbb{E}\left[\left(\eta_1^{(N)}\right)^b\right] \sim L_N^{-1} \int_0^1 bx^{b-1} f(x)dx = L_N^{-1} \int_0^1 x^{b-2}\Lambda(dx). \tag{2.13}$$

Substituting $b = 2$ and using $iii)$ gives the stated asymptotic for c_N. The latter, together with (2.13) give the necessary hypothesis of Proposition 2.2.7 for the convergence to the Λ-coalescent

□

It is easy to see from the proof of Theorem 2.2.8 that the conditions $ii)$ and $iv)$ can be replaced by the weaker condition

$$ii') \lim_{N\to\infty} L_N \mathbb{E}\left[\left(\eta_1^{(N)}\right)^b\right] = \int_0^1 x^{b-2}\Lambda(dx), \forall b \geq 2.$$

In fact we have the following.

Lemma 2.2.9. *Under condition iv), conditions ii) and ii′) are equivalent.*

Proof. **Proof of ii) ⇒ ii′):** The proof of this implication is contained in the proof of Theorem 2.2.8.
Proof of ii′) ⇒ ii): Let $h^+(x) = \limsup_{N\to\infty} L_N \mathbb{P}(\eta_1^{(N)} > x)$ which is decreasing on $(0,1]$, and note that by *iv*) the functions $xh^+(x)$ and $xf(x)$ are both integrable. Again using *iv*) and the Dominated Convergence Theorem we write, for any $b \geq 2$,

$$\limsup_{N\to\infty} L_N \mathbb{E}\left[\left(\eta_1^{(N)}\right)^b\right] = \limsup_{N\to\infty} L_N \int_0^1 bx^{b-1}\mathbb{P}(\eta_1^{(N)} > x)dx = \int_0^1 bx^{b-1}h^+(x)dx.$$

This, together with *ii′*), imply that for any polynomial P we have

$$\int_0^1 P(x)xf(x)dx = \int_0^1 P(x)xh^+(x)dx.$$

Let $t > 0$. By the Stone-Weierstrass theorem we can approximate uniformly the function $x \to e^{-tx}$ by polynomials, obtaining

$$\int_0^1 e^{-tx}xf(x)dx = \int_0^1 e^{-tx}xh^+(x)dx.$$

Since this holds for every $t > 0$, this implies

$$f(x) \overset{a.e.}{=} h^+(x).$$

Repeating the same argument but now with $h^-(x) = \liminf_{N\to\infty} L_N \mathbb{P}(\eta_1^{(N)} > x)$ we obtain

$$h^+(x) \overset{a.e.}{=} f(x) \overset{a.e.}{=} h^-(x).$$

□

We now state and prove the conditions for the convergence of the genealogy to Kingman's coalescent.

Theorem 2.2.10. *Assume that all the conditions in* (2.12) *hold and that the conditions*

$$v) \int_0^1 xf(x)dx = \infty \text{ and } vi) \exists M \geq 2\colon \int_0^1 x^M \left(\sup_{N\in\mathbb{N}} L_N \mathbb{P}\left(\eta_1^{(N)} > x\right)\right) dx < \infty$$

hold. Then,

$$\lim_{N\to\infty} c_N L_N = \infty \tag{2.14}$$

and $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly in the Skorohod topology for* $D([0,\infty),\mathscr{P}_n)$ *to Kingman's coalescent.*

Proof. On the one hand observe that

$$L_N c_N \geq L_N \mathbb{E}\left[\left(\eta_1^{(N)}\right)^2\right] = \int_0^1 2x L_N \mathbb{P}\left(\eta_1^{(N)} > x\right) dx$$

so that by Fatou's lemma we obtain $\liminf_{N\to\infty} L_N c_N = \infty$, thus proving (2.14). For the stated convergence to Kingman's coalescent, observe that (2.14) implies $\lim_{N\to\infty} \frac{o\left(L_N^{-1}\right)}{c_N} = \lim_{N\to\infty} \frac{1}{c_N L_N} L_N o\left(L_N^{-1}\right) = 0$. This, together with iii), give the first condition of Corollary 2.2.4. The second condition of Corollary 2.2.4 corresponds to iv). □

Again, it is easy to see from the above proof that the convergence to Kingman's coalescent still holds if we replace v) and vi) by the weaker conditions

$$v')\ \lim_{N\to\infty} L_N c_N > 0$$

and

$$vi')\exists\beta > 0\colon\ \mathbb{E}\left[(\eta_1^{(N)})^\beta\right] = o\left(c_N\right) \text{ as } N \to \infty.$$

2.2.3 Example: Poisson-Dirichlet Size-Biased Pick

In this section we study the case in which the random mass partition $\tilde{\rho}^{(N)}$ can be constructed from a size-biased pick $(\tilde{V}_1, \tilde{V}_2, \dots)$ of a Poisson-Dirichlet (PD) random mass partition of parameters (α, θ), $0 < \alpha < 1, \theta > -\alpha$ (Pitman and Yor 1997), by setting

$$\eta_i^{(N)} = \frac{\tilde{V}_i^\kappa}{\sum_{j=1}^N \tilde{V}_j^\kappa}, \quad 1 \leq i \leq N, \tag{2.15}$$

where $\kappa \in \mathbb{R}$ is a parameter of the model.

The case $\kappa = \alpha$ of the following theorem was proved in Cortines and Mallein 2017. Here we extend their results by following their main heuristics but using different technical arguments such as applying Theorems 2.2.8 and Corollary 2.2.3 instead of their Lemmas 3.1 and 3.2, and also by working directly with the expectations appearing in conditions ii') and iii).

Theorem 2.2.11. *Assume $\alpha/2 < \kappa \leq \alpha$. Then there exists a constant $B \equiv B_{\alpha,\theta,\kappa}$ such that*

$$B\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{-2} \leq c_N = \mathbb{E}\left[\left(\eta_1^{(N)}\right)^2\right] + \mathcal{O}\left(\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{-2}\right). \tag{2.16}$$

Furthermore, setting $L_N = \ell_{\alpha,\theta,\kappa}\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}}$ where

$$\ell_{\alpha,\theta,\kappa}^{-1} = \frac{\alpha}{\kappa} \frac{\Gamma(1-\alpha)^{\frac{\theta}{\alpha}}}{\Gamma(1+\kappa-\alpha)^{1+\frac{\theta}{\alpha}}} \frac{\Gamma\left(\frac{\alpha+\theta}{\alpha}(1-\frac{\kappa}{\alpha})+1\right)}{\Gamma\left((\alpha+\theta)\left(1-\frac{\kappa}{\alpha}\right)+1\right)} \frac{\Gamma(1+\theta)\Gamma\left(1-\frac{\theta}{\alpha}\right)}{\mathsf{B}\left(1-\frac{\theta}{\alpha}, 1+\frac{\theta}{\alpha}\right)}$$

we have, as $N \to \infty$,

I. *if* $\theta \in (-\alpha, \alpha)$ *then* $c_N \sim \left(1 - \frac{\theta}{\alpha}\right)/L_N$ *and* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly in the Skorohod topology for* $D([0,\infty), \mathscr{P}_n)$ *to the Beta*$\left(1-\frac{\theta}{\alpha}, 1+\frac{\theta}{\alpha}\right)$*-coalescent;*

II. *otherwise,* $\lim_{N\to\infty} c_N \log(N)^{1+\frac{\theta}{\alpha}} > 0$ *and* $\left(\Pi^{(N,n)}_{[t/c_N]}\right)_{t\geq 0}$ *converges weakly to Kingman's coalescent.*

Remark. *Observe that the choice of* κ *does not affect the shape of the limit genealogy, but only the correct time scale* c_n*.*

Before proving this theorem we first develop some general results on the PD(α,θ) distribution. We first provide an understanding of the asymptotic behavior of the normalizing sum appearing in (2.15), namely

$$\zeta_{N,\kappa} := \sum_{i=1}^{N} \tilde{V}_i^{\kappa}.$$

For this we first recall the well-known stick-breaking construction of Poisson-Dirichlet(α,θ) size-biased picks. Let $(\tilde{V}_1, \tilde{V}_2, \dots)$ be a size-biased pick from a random partition with Poisson-Dirichlet(α,θ) distribution. The sequence $(\tilde{V}_1, \tilde{V}_2, \dots)$ may be constructed from a collection $(Y_1, Y_2, \dots)$ of independent random variables where Y_i is Beta$(1-\alpha, \theta + i\alpha)$ distributed, by setting

$$\tilde{V}_1 = Y_1, \text{ and } \tilde{V}_i = (1-Y_1)\dots(1-Y_{i-1})Y_i$$

(see Proposition 2 in Pitman and Yor 1997).

We now write $\zeta_{N,\kappa}$ in terms of the sequence $(Y_1, Y_2, \dots)$. For this observe that

$$\prod_{i=1}^{N} (1-Y_i)^{\kappa} = e^{-\kappa\mu_N} e^{\kappa S_N} \tag{2.17}$$

where

$$S_N := \mu_N + \sum_{i=1}^{N} \log(1-Y_i), \quad \mu_N := \sum_{i=1}^{N} -\mathbb{E}\left[\log(1-Y_i)\right] \tag{2.18}$$

define a martingale. We also define the martingale

$$M_{N,\kappa} := \sum_{i=1}^{N} \left(i^{\kappa-1} Y_i^{\kappa} - \mathbb{E}\left[i^{\kappa-1} Y_i^{\kappa}\right]\right).$$

With this, we have

$$\begin{aligned}
\zeta_{N,\kappa} &= \sum_{i=1}^{N} Y_i^{\kappa} e^{-\kappa\mu_{i-1}} e^{\kappa S_{i-1}} \\
&= \sum_{i=1}^{N} \left(M_{i,\kappa} - M_{i-1,\kappa} + \mathbb{E}\left[i^{\kappa-1} Y_i^{\kappa}\right]\right) i^{1-\kappa} e^{-\kappa\mu_{i-1}} e^{\kappa S_{i-1}} \\
&= \overline{M}_{N,\kappa} + \Sigma_{N,\kappa} \qquad (2.19)
\end{aligned}$$

where $\overline{M}_{N,\kappa}$ is the martingale

$$\overline{M}_{N,\kappa} := \sum_{i=1}^{N} (M_{i,\kappa} - M_{i-1,\kappa}) i^{1-\kappa} e^{-\kappa\mu_{i-1}} e^{\kappa S_{i-1}}$$

and $\Sigma_{N,\kappa}$ is the sum

$$\Sigma_{N,\kappa} := \sum_{i=1}^{N} \mathbb{E}\left[Y_i^{\kappa}\right] e^{-\kappa\mu_{i-1}} e^{\kappa S_{i-1}} = \sum_{i=1}^{N} \frac{\mathsf{B}(1+\kappa-\alpha, \theta+i\alpha)}{\mathsf{B}(1-\alpha, \theta+i\alpha)} e^{-\kappa\mu_{i-1}} e^{\kappa S_{i-1}}.$$

We now study the product $\prod_{i=1}^{N}(1-Y_i)^{\kappa}$ with the help of the martingale S_N and equation (2.17). First we provide a technical lemma concerning beta random variables. In the following we let $\psi = \frac{d}{dz}\log\Gamma(z)$ be the digamma function, and ψ_1 the trigamma function $\frac{d^2}{dt^2}\log\Gamma(t)$.

Lemma 2.2.12. *Let X be Beta(a,b) distributed. Then*

$$\mathbb{E}\left[\log(X)\right] = \psi(a) - \psi(a+b) \tag{2.20}$$

and

$$\mathbb{V}\mathrm{ar}\left(\log(X)\right) = \psi_1(b) - \psi_1(a+b). \tag{2.21}$$

Proof. We only compute (2.20), and refer the reader to Aryal and Nadarajah 2004 for the computation of $\mathbb{E}\left[\log(X)^2\right]$, from which (2.21) easily follows. We have

$$\mathbb{E}\left[\log(X)\right] = \int_0^1 \log(x) x^{a-1}(1-x)^{b-1} dx = \frac{1}{\mathsf{B}(a,b)} \int_0^1 \lim_{h\to 0} \frac{x^h - 1}{h} x^{a-1}(1-x)^{b-1} dx,$$

where $\left|x^h - 1\right| \le h\log(x)$ for all $x \in (0,1)$. Let $0 < \epsilon < a$, then $\int_0^1 \log(x) x^{a-1}(1-x)^{b-1} dx \le \|x^{\epsilon}\log(x)\|_{\mathsf{L}^\infty([0,1])} \int_0^1 x^{a-\epsilon-1}(1-x)^{b-1} dx = \frac{\mathsf{B}(a-\epsilon,b)}{\mathsf{B}(a,b)} < \infty$. Thus, Dominated Convergence yields

$$\begin{aligned}\mathbb{E}\left[\log(X)\right] &= \frac{1}{\mathsf{B}(a,b)} \int_0^1 \frac{\partial}{\partial a} x^{a-1}(1-x)^{b-1} dx = \frac{1}{\mathsf{B}(a,b)} \frac{\partial}{\partial a}\mathsf{B}(a,b) = \frac{\partial}{\partial a}\log\mathsf{B}(a,b) \\ &= \psi(a) - \psi(a+b).\end{aligned}$$

□

Lemma 2.2.13. *For $a > -1$ let Υ_a be the constant such that $-\log(N) + \sum_{i=1}^{N}(a+i)^{-1} = \Upsilon_a + o(1)$, e.g. $\Upsilon \equiv \Upsilon_0$ is the Euler-Mascheroni constant. Assume $\kappa > -(\theta+\alpha)$. Then there exists a random variable $S_\infty \in \mathsf{L}^2$ such that*

$$\lim_{N\to\infty} \left(N^{\frac{1-\alpha}{\alpha}} \prod_{i=1}^{N} (1-Y_i) \right)^{\kappa} = \alpha^{-\kappa} K_{\alpha,\theta}^{\kappa} e^{\kappa S_\infty}$$

a.s. and in L^1, where $K_{\alpha,\theta} = \exp\{\psi(\theta+1) - \frac{1}{\alpha}\Upsilon_{\theta/\alpha}\}$. Furthermore,

$$\mathbb{E}\left[e^{\kappa S_\infty}\right] = K_{\alpha,\theta}^{\kappa} \frac{\Gamma(\theta+1)}{\Gamma(\theta+\kappa+1)} \frac{\Gamma\left(\frac{\theta+\kappa}{\alpha}+1\right)}{\Gamma\left(\frac{\theta}{\alpha}+1\right)}.$$

Moreover, if $0 < \kappa < \theta + \alpha$*, then for some constant* K'

$$\mathbb{P}\left(\inf_{N \geq 0} N^{\frac{1-\alpha}{\alpha}} \prod_{i=1}^{N}(1 - Y_i) \leq y\right) \leq K' y^{\kappa} \tag{2.22}$$

for all $N \geq 1$ *and* $y \geq 0$*.*

Proof. Observe that the product $e^{\kappa \mu_N} \prod_{i=1}^{N}(1 - Y_i)^{\kappa}$ converges a.s. to a strictly positive random variable if and only if the martingale S_N converges a.s. Recall that ψ_1 satisfies $\psi_1(t+1) = \psi_1(t) - t^{-2}$ and $\lim_{t\to\infty} \psi_1(t) = 0$, then, since $1 - Y_i$ is Beta$(\theta + i\alpha, 1 - \alpha)$ distributed, we have, using Lemma 2.2.12 in the first line,

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &= \sum_{i=1}^{N} \mathbb{V}\mathrm{ar}\left(\log\left(1 - Y_i\right)\right) = \sum_{i=1}^{N} \psi_1(\theta + i\alpha) - \psi_1(\theta + (i-1)\alpha + 1) \\
&= \sum_{i=1}^{N} \psi_1(\theta + i\alpha) - \psi_1(\theta + (i-1)\alpha) + \frac{1}{(\theta + (i-1)\alpha)^2} \\
&= \psi_1(\theta + N\alpha) - \psi_1(\theta) + \sum_{i=1}^{N-1} \frac{1}{(\theta + (i-1)\alpha)^2} \\
&\overset{N\to\infty}{\to} -\psi_1(\theta) + \sum_{i=1}^{\infty} \frac{1}{(\theta + (i-1)\alpha)^2} < \infty.
\end{aligned}$$

It follows that the martingale S_N is bounded in L^2 and converges a.s. and in L^2 to a random variable S_∞, so that $e^{\kappa S_N} \overset{a.s.}{\to} e^{\kappa S_\infty}$ for all $\kappa \in \mathbb{R}$. We now prove that $e^{\kappa S_\infty} \in \mathsf{L}^1$ whenever $\kappa > -(\theta + \alpha)$; then, the function $e^{\kappa x}$ being convex and the martingale $(S_N)_{N\in\mathbb{N}}$ being closable at infinity, would imply that the submartingale $\left(e^{\kappa S_N}\right)_{N\in\mathbb{N}}$ converges a.s. and in L^1 to $e^{\kappa S_\infty}$. Observe that for all $\kappa > -(\theta + \alpha)$,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{N}(1 - Y_i)^{\kappa}\right] &= \prod_{i=1}^{N} \frac{\mathsf{B}\left(\theta + i\alpha + \kappa, 1 - \alpha\right)}{\mathsf{B}\left(\theta + i\alpha, 1 - \alpha\right)} \\
&= \frac{\Gamma(\theta + \alpha + \kappa)}{\Gamma(\theta + \kappa + 1)} \frac{\Gamma(\theta + 1)}{\Gamma(\theta + \alpha)} \times \\
&\quad \prod_{i=1}^{N-1} \left(\frac{\theta + i\alpha}{\theta + i\alpha + \kappa}\right) \left(\frac{\Gamma(\theta + (i+1)\alpha + \kappa)}{\Gamma(\theta + i\alpha + \kappa)}\right) \left(\frac{\Gamma(\theta + i\alpha)}{\Gamma(\theta + (i+1)\alpha)}\right) \\
&= \frac{\Gamma(\theta + 1)}{\Gamma(\theta + \kappa + 1)} \frac{\Gamma(\theta + N\alpha + \kappa)}{\Gamma(\theta + N\alpha)} \frac{\Gamma\left(\frac{\theta}{\alpha} + N\right)}{\Gamma\left(\frac{\theta}{\alpha} + 1\right)} \frac{\Gamma\left(\frac{\theta+\kappa}{\alpha} + 1\right)}{\Gamma\left(\frac{\theta+\kappa}{\alpha} + N\right)}
\end{aligned}$$

and, by Stirling's approximation to the Gamma ratios above,

$$\mathbb{E}\left[\prod_{i=1}^{N}(1 - Y_i)^{\kappa}\right] \sim \alpha^{\kappa} \frac{\Gamma(\theta + 1)}{\Gamma(\theta + \kappa + 1)} \frac{\Gamma\left(\frac{\theta+\kappa}{\alpha} + 1\right)}{\Gamma\left(\frac{\theta}{\alpha} + 1\right)} N^{\kappa(\alpha-1)/\alpha}, \quad \kappa > -(\theta + \alpha).$$

On the other hand, we have $-\mathbb{E}\left[\log(1 - Y_i)\right] = \psi(\theta + (i-1)\alpha + 1) - \psi(\theta + i\alpha)$. Using the well-known identity $\psi(t+1) = \psi(t) + t^{-1}$, the bounds $\log(t) - t^{-1} \leq \psi(t) \leq \log(t+1) - t^{-1}$

for $t > 0$ (see Corollary 2.3 in Muqattash and Yahdi 2006), and the definition of $\Upsilon_{\theta/\alpha}$, we obtain, as $N \to \infty$,

$$\begin{aligned} \mu_N &= \psi(\theta+1) - \psi(\theta+N\alpha) + \sum_{i=1}^{N-1} \frac{1}{\theta+i\alpha} \\ &= \psi(\theta+1) - \log(N) - \log(\alpha) + \frac{1}{\alpha}\log(N) - \frac{1}{\alpha}\Upsilon_{\theta/\alpha} + o\left(1\right). \end{aligned} \tag{2.23}$$

Thus, we arrive at

$$e^{\kappa\mu_N} \sim \alpha^{-\kappa} K^{\kappa}_{\alpha,\theta} N^{\kappa(1-\alpha)/\alpha}, \quad \kappa \in \mathbb{R}. \tag{2.24}$$

Fatou's lemma then yields

$$\mathbb{E}\left[e^{\kappa S_\infty}\right] \leq \liminf_{N\to\infty} e^{\kappa\mu_N} \mathbb{E}\left[\prod_{i=1}^{N}(1-Y_i)^{\kappa}\right] < \infty,$$

and the ensuing L^1 convergence gives

$$\mathbb{E}\left[e^{\kappa S_\infty}\right] = K^{\kappa}_{\alpha,\theta} \frac{\Gamma(\theta+1)}{\Gamma(\theta+\kappa+1)} \frac{\Gamma\left(\frac{\theta+\kappa}{\alpha}+1\right)}{\Gamma\left(\frac{\theta}{\alpha}+1\right)}.$$

Finally, to prove (2.22), observe that from Doob's maximal inequality applied to the closable submartingale $\left(e^{-\kappa S_N}\right)_{N\in\mathbb{N}}$ plus (2.24), we have

$$\begin{aligned} \exists K' > 0: \quad \mathbb{P}\left(\inf_{N\geq 0} N^{\frac{1-\alpha}{\alpha}} \prod_{i=1}^{N}(1-Y_i)^{\kappa} \leq y\right) &= \mathbb{P}\left(\sup_{N\geq 0} N^{\frac{1-\alpha}{\alpha}} e^{-\kappa S_N+\mu_N} > y^{-\kappa}\right) \\ &\leq \left(\sup_{N\geq 0} N^{\frac{1-\alpha}{\alpha}} e^{\mu_N}\right) \mathbb{E}\left[e^{-\kappa S_\infty}\right] y^{\kappa} \\ &\leq K' y^{\kappa}. \end{aligned}$$

□

We now study the asymptotic behaviour of the two terms, $\overline{M}_{N,\kappa}$ and $\Sigma_{N,\kappa}$, that compose $\zeta_{N,\kappa}$ in (2.19).

Lemma 2.2.14. *Assume $\kappa > \alpha/2$, then there exists a r.v. $\overline{M}_{\infty,\kappa}$ such that*

$$\lim_{N\to\infty} \overline{M}_{N,\kappa} = \overline{M}_{\infty,\kappa}, \text{ a.s. and in } \mathsf{L}^2;$$

in particular, if $\kappa \leq \alpha$, then

$$\lim_{N\to\infty} \frac{1}{\sum_{i=1}^{N} i^{-\frac{\kappa}{\alpha}}} \overline{M}_{N,\kappa} = 0, \text{ a.s. and in } \mathsf{L}^2.$$

Proof. We compute, for every $\kappa > 0$,

$$\mathbb{E}\left[(M_{i,\kappa} - M_{i-1,\kappa})^2\right] = \mathbb{V}\text{ar}\left(i^{\kappa-1}Y_i^{\kappa}\right) = \frac{i^{-2}}{\alpha^{2\kappa}}\left(\frac{\Gamma(1+2\kappa-\alpha)}{\Gamma(1-\alpha)} - \frac{\Gamma(1+\kappa-\alpha)^2}{\Gamma(1-\alpha)^2} + \mathcal{O}\left(i^{-1}\right)\right)$$

as $i \to \infty$ so that, using (2.24), Lemma 2.2.13, and the condition $\kappa > \frac{\alpha}{2}$

$$\begin{aligned}
&\sum_{i=1}^{\infty} \mathbb{V}\text{ar}\left(\overline{M}_{i+1,\kappa} - \overline{M}_{i,\kappa}\right) \\
&= \sum_{i=1}^{\infty} i^{-2\frac{\kappa}{\alpha}} \mathbb{E}\left[e^{2\kappa S_i}\right] \alpha^{-2\kappa}\left(\frac{\Gamma(1+2\kappa-\alpha)}{\Gamma(1-\alpha)} - \frac{\Gamma(1+\kappa-\alpha)^2}{\Gamma(1-\alpha)^2}\right)\left(1+\mathcal{O}\left(i^{-1}\right)\right) \\
&\leq C\mathbb{E}\left[e^{2\kappa S_\infty}\right]\sum_{i=1}^{\infty} i^{-2\kappa/\alpha} < \infty
\end{aligned}$$

for some $C > 0$. This implies the a.s. and L^2 convergence of $\overline{M}_{N,\kappa}$ to some r.v. $\overline{M}_{\infty,\kappa}$. □

Lemma 2.2.15. *We have, for $\kappa \leq \alpha$,*

$$\lim_{N\to\infty} \frac{1}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}} \Sigma_{N,\kappa} = \frac{\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} K_{\alpha,\theta}^{-\kappa} e^{\kappa S_\infty} \tag{2.25}$$

almost surely and in L^1. On the other hand, if $\kappa > \alpha$, then

$$\lim_{N\to\infty} \Sigma_{N,\kappa} = \Sigma_{\infty,\kappa}$$

almost surely and in L^p, $p \geq 1$, for some r.v. $\Sigma_{\infty,\kappa}$.

Proof. Observe that, as $N \to \infty$

$$\frac{\mathsf{B}(1+\kappa-\alpha, \theta+N\alpha)}{\mathsf{B}(1-\alpha, \theta+N\alpha)} = \frac{\alpha^{-\kappa}\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} N^{-\kappa}\left(1+\mathcal{O}\left(N^{-1}\right)\right). \tag{2.26}$$

Then equation (2.24) together with Lemma 2.2.13 and the Stolz-Césaro theorem yield

$$\begin{aligned}
\lim_{N\to\infty} \frac{\Sigma_{N,\kappa}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}} &\overset{a.s.}{=} \frac{\alpha^{-\kappa}\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} \lim_{N\to\infty} \frac{N^{-\kappa} e^{-\kappa\mu_{N-1}} e^{\kappa S_{N-1}}}{N^{-\frac{\kappa}{\alpha}}} \\
&\overset{a.s.}{=} \frac{\alpha^{-\kappa}\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} \alpha^{\kappa} K_{\alpha,\theta}^{-\kappa} e^{\kappa S_\infty}.
\end{aligned}$$

Hence the desired a.s. convergence follows. Similarly, by the same equation (2.24) and Lemma 2.2.13,

$$\lim_{N\to\infty} \frac{\mathbb{E}\left[\Sigma_{N,\kappa}\right]}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}} = \frac{\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} K_{\alpha,\theta}^{-\kappa}\mathbb{E}\left[e^{\kappa S_\infty}\right]$$

so that Scheffe's lemma gives the corresponding L^1 convergence.

On the other hand, when $\kappa > \alpha$ we have, using equation (2.24) plus the estimate for the Beta ratio above, and Lemma 2.2.13,

$$\exists C > 0 \text{ s.t. } \forall p \geq 1 : \quad \|\Sigma_{N,\kappa}\|_{\mathsf{L}^p}^p \leq \left(\sum_{i=1}^{\infty} C i^{-\kappa/\alpha} \left\| e^{\kappa S_\infty} \right\|_{\mathsf{L}^p} (1 + o(1)) \right)^p < \infty.$$

In particular the increasing sequence $\left(\Sigma_{N,\kappa}^p\right)_{N \geq 1}$ is a.s. bounded and thus convergent. The convergence also holds in L^p by dominated convergence which yields $\|\Sigma_{N,\kappa}\|_{\mathsf{L}^p} \to \|\Sigma_{\infty,\kappa}\|_{\mathsf{L}^p}$. □

Proposition 2.2.16. *If $\alpha/2 < \kappa \leq \alpha$, then*

$$\lim_{N \to \infty} \frac{\zeta_{N,\kappa}}{\sum_{i=1}^N i^{-\kappa/\alpha}} = \frac{\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} K_{\alpha,\theta}^{-\kappa} e^{\kappa S_\infty}$$

almost surely and in L^1. If, furthermore $\eta\kappa < \theta + \alpha$, then

$$\lim_{N \to \infty} \left(\frac{\zeta_{N,\kappa}}{\sum_{i=1}^N i^{-\kappa/\alpha}} \right)^{-\eta} = \left(\frac{\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} K_{\alpha,\theta}^{-\kappa} e^{\kappa S_\infty} \right)^{-\eta}$$

almost surely and in L^1. On the other hand, if $\kappa > \alpha$, then there exists a r.v. $\zeta_{\infty,\kappa}$ such that

$$\lim_{N \to \infty} \zeta_{N,\kappa} = \zeta_{\infty,\kappa}$$

almost surely and in L^2.

Proof. The first a.s. and L^1 convergences follow directly from (2.19) and Lemmas 2.2.14 and 2.2.15. The second convergences will follow once we prove

$$\limsup_{N \geq 0} \mathbb{E}\left[\left(\frac{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N \tilde{V}_i^\kappa} \right)^\eta \right] < \infty \tag{2.27}$$

together with an application of dominated convergence and Scheffe's lemma. To upper-bound the expectations in (2.27) observe that, if $\mathbb{E}_\Omega$ is the expectation operator on the probability space $\left(\Omega = \{f(1), \dots, f(N)\}, (\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}})^{-1} \sum_{i=1}^N \delta_{f(i)} i^{-\frac{\kappa}{\alpha}}\right)$, $f(i) := Y_i^\kappa e^{-\kappa \mu_{i-1}} e^{\kappa S_{i-1}} i^{\frac{\kappa}{\alpha}}$, then, by Jensen's inequality we have,

$$\begin{aligned}
\frac{1}{\left(\sum_{i=1}^N \tilde{V}_i^\kappa\right)^\eta} &= \exp\left\{ -\eta \log\left((\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}})(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}})^{-1} \sum_{i=1}^N Y_i^\kappa e^{-\kappa\mu_{i-1}} e^{\kappa S_{i-1}} i^{\frac{\kappa}{\alpha}} i^{-\frac{\kappa}{\alpha}} \right) \right\} \\
&= e^{-\eta \log\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)} \exp\left\{ -\eta \log\left(\mathbb{E}_\Omega [x] \right) \right\} \\
&\leq e^{-\eta \log\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)} \exp\left\{ -\eta \mathbb{E}_\Omega [\log(x)] \right\} \\
&= \left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}} \right)^{-\eta} \exp\left\{ -\eta\kappa \frac{1}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}} \sum_{i=1}^N \frac{\log(Y_i) - \mu_{i-1} + S_{i-1} + \frac{1}{\alpha}\log(i)}{i^{\frac{\kappa}{\alpha}}} \right\}.
\end{aligned}$$

Furthermore, by Lemma 2.2.13 and the condition $\eta\kappa < \alpha+\theta$ we have, for $\epsilon > 0$ small enough and using that $\left(e^{-\eta\kappa(1+\epsilon)S_N}\right)_N$ is a submartingale,

$$\exp\left\{-\eta\kappa(1+\epsilon)\frac{1}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\sum_{i=1}^N S_{i-1} i^{-\frac{\kappa}{\alpha}}\right\} \le \sup_{N\in\mathbb{N}} e^{-\eta\kappa(1+\epsilon)S_N} \in \mathsf{L}^1.$$

Thus, plugging these estimates and using Hölders inequality ($p = 1+\epsilon$),

$$\mathbb{E}\left[\left(\frac{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N \tilde{V}_i^\kappa}\right)^\eta\right] \le \left\|\sup_{N\in\mathbb{N}} e^{-\eta\kappa(1+\epsilon)S_N}\right\|_{\mathsf{L}^1} \times$$

$$\mathbb{E}\left[\exp\left\{-\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{1}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\sum_{i=1}^N \frac{\log(Y_i)-\mu_{i-1}+\frac{1}{\alpha}\log(i)}{i^{\frac{\kappa}{\alpha}}}\right\}\right]^{\frac{\epsilon}{1+\epsilon}}. \tag{2.28}$$

We now compute, for N large enough to ensure $\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{1}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}} < 1-\alpha$, and using an easy consequence of the Mean Value Theorem on the function $z \to \log\Gamma(z)$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{-\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{1}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\log(Y_i) i^{-\frac{\kappa}{\alpha}}\right\}\right] = \mathbb{E}\left[Y_i^{-\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}}\right] \\
&= \exp\left\{\log\Gamma\left(1-\alpha-\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\right) - \log\Gamma\left(1-\alpha\right)\right\} \\
&\quad \exp\left\{\log\Gamma(1+\theta+(i-1)\alpha) - \log\Gamma\left(1+\theta+(i-1)\alpha-\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\right)\right\} \\
&\le \exp\left\{\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\psi\left(1+\theta+(i-1)\alpha\right)\right\}.
\end{aligned} \tag{2.29}$$

Equation (2.23) and the identity $\psi(1+\theta+(i-1)\alpha) = \psi(\theta+(i-1)\alpha) + \frac{1}{\theta+(i-1)\alpha}$ yield, as $i \to \infty$,

$$\psi\left(1+\theta+(i-1)\alpha\right) - \mu_{i-1} + \frac{1}{\alpha}\log(i) = \frac{1}{\theta+(i-1)\alpha} - \psi(\theta+1) + \frac{1}{\alpha}\Upsilon_{\theta/\alpha} + o\left(1\right).$$

Thus we obtain, for all $i \le N$,

$$\exists C' > 0\colon \quad \mathbb{E}\left[\exp\left\{-\eta\kappa\frac{1+\epsilon}{\epsilon}\frac{i^{-\frac{\kappa}{\alpha}}(\log(Y_i)-\mu_{i-1}+\frac{1}{\alpha}\log(i))}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\right\}\right]^{\frac{\epsilon}{1+\epsilon}} \le \exp\left\{\eta\kappa\frac{i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}C'\right\}.$$

Taking the product over i, and plugging in (2.28), we conclude, for all $N \in \mathbb{N}$,

$$\mathbb{E}\left[\left(\frac{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}{\sum_{i=1}^N \tilde{V}_i^\kappa}\right)^\eta\right] < \exp\{\eta\kappa C'\};$$

this entails (2.27). Finally, the stated convergence in the case $\kappa > \alpha$ follows directly from (2.19) and Lemmas 2.2.14 and 2.2.15. □

We will now gather the results so far in order to prove Theorem 2.2.11. First, however, we recall yet another well-known fact on the Poisson-Dirichlet(α,θ) distribution, mainly that if we remove the element $\tilde V_1$ from $(\tilde V_1,\tilde V_2,\dots)$ and renormalize the resulting sequence then, in distribution, we had simply changed the value of the parameter θ by adding the value of α; formally

Lemma 2.2.17 (Poisson-Dirichlet change of parameter)**.** *Let $(\tilde V_1,\tilde V_2,\dots)$ be a sized-biased pick from a Poisson-Dirichlet(α,θ) mass partition. Then the sequence $\left(\frac{\tilde V_2}{1-\tilde V_1},\frac{\tilde V_3}{1-\tilde V_1},\dots\right)$ is distributed as a size-biased pick from a PD$(\alpha,\theta+\alpha)$ random mass partition, and is independent from $\tilde V_1$.*

Proof of Theorem 2.2.11. We first prove (2.16). On the one hand, by the reverse Hölder's inequality we have, for any $p>1$,

$$c_N=\mathbb{E}_{\alpha,\theta}\left[\frac{\zeta_{N,2\kappa}}{\zeta_{N,\kappa}^2}\right]\geq \mathbb{E}_{\alpha,\theta}\left[\frac{\zeta_{N,2\alpha}}{\zeta_{N,\kappa}^2}\right]\geq \|\zeta_{N,2\alpha}\|_{\mathsf{L}^{1/p}}\,\mathbb{E}_{\alpha,\theta}\left[\zeta_{N,\kappa}^{\frac{2}{p-1}}\right]^{p-1},$$

where $\mathbb{E}_{\alpha,\theta}$ refers to expectation with respect to the PD(α,θ) distribution. Proposition 2.2.16 yields, on the one hand,

$$\zeta_{N,2\alpha}^{1/p}\overset{a.s.,\mathsf{L}^1}{\to}\zeta_{\infty,2\alpha}^{1/p}\overset{a.s.}{>}0$$

so that $0<\liminf_{N\geq 1}\|\zeta_{N,2\alpha}\|_{\mathsf{L}^{1/p}}<\infty$. On the other hand, since $\frac{2}{1-p}\kappa<0<\alpha+\theta$, the same proposition yields

$$0<\liminf_{N\geq 1}\mathbb{E}_{\alpha,\theta}\left[\left(\frac{\zeta_{N,\kappa}}{\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}}\right)^{-\frac{2}{1-p}}\right]^{p-1}<\infty.$$

Thus, putting all together,

$$c_N\geq B\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{-2}$$

for some $B>0$. On the other hand, by Lemma 2.2.17 we have

$$\begin{aligned}
\mathbb{E}_{\alpha,\theta}\left[\sum_{i=2}^N\left(\eta_i^{(N)}\right)^2\right]&=\mathbb{E}_{\alpha,\theta}\left[\frac{\sum_{i=2}^N\left(\frac{\tilde V_i}{1-\tilde V_1}\right)^{2\kappa}(1-\tilde V_1)^{2\kappa}}{\left(\tilde V_1^\kappa+(1-\tilde V_1)^\kappa\sum_{j=2}^N\left(\frac{\tilde V_j}{1-\tilde V_1}\right)^\kappa\right)^2}\right]\\
&=\mathbb{E}_{\alpha,\theta+\alpha}\left[\int_0^1\frac{(1-y)^{2\kappa}\zeta_{N-1,2\kappa}}{(y^\kappa+(1-y)^\kappa\zeta_{N-1,\kappa})^2}\frac{y^{-\alpha}(1-y)^{\theta+\alpha-1}}{\mathsf{B}(1-y,\theta+\alpha)}dy\right]\\
&=\mathbb{E}_{\alpha,\theta+\alpha}\left[\frac{\zeta_{N-1,2\kappa}}{\zeta_{N-1,\kappa}^2}\int_0^1\frac{(1-y)^{2\kappa}}{(y^\kappa/\zeta_{N-1,\kappa}+(1-y)^\kappa)^2}\frac{y^{-\alpha}(1-y)^{\theta+\alpha-1}}{\mathsf{B}(1-y,\theta+\alpha)}dy\right]\\
&\leq\mathbb{E}_{\alpha,\theta+\alpha}\left[\frac{\zeta_{N-1,2\kappa}}{\zeta_{N-1,\kappa}^2}\right].
\end{aligned}$$

The condition $\kappa > \alpha/2$ and Proposition 2.2.16 give $\zeta_{N-1,2\kappa} \in \mathsf{L}^p, p \geq 1$, so that we may use Hölders inequality with $p = (1+\epsilon_1)(1+\epsilon_2), q = \frac{1+\epsilon_1}{\epsilon_1}(1+\epsilon_2)$, and $r = \frac{(1+\epsilon_1)(1+\epsilon_2)}{\epsilon_2(1+\epsilon_1)}$, for $\frac{\alpha-\theta}{2(\alpha+\theta)} < \epsilon_1 < \frac{\alpha}{\alpha+\theta}$ and $\epsilon_2 > 0$ small enough, to obtain

(2.30)

$$\mathbb{E}_{\alpha,\theta+\alpha}\left[\frac{\zeta_{N-1,2\kappa}}{\zeta^2_{N-1,\kappa}}\right] \leq \left\| \frac{\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}}}{\zeta_{N-1,\kappa}^{1+\frac{\theta}{\alpha}}} \right\|_{\mathsf{L}^p} \left\| \frac{\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1-\frac{\theta}{\alpha}}}{\zeta_{N-1,\kappa}^{1-\frac{\theta}{\alpha}}} \right\|_{\mathsf{L}^q} \|\zeta_{N-1,2\kappa}\|_{\mathsf{L}^r} \left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{-2}$$
$$= \mathcal{O}\left(\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{-2}\right).$$

We now prove the stated convergence of the genealogy.

Proof of I: Given equation (2.16), the condition $\theta \in (-\alpha, \alpha)$, and Theorem 2.2.8, it remains to prove condition $ii')$ of the same theorem, i.e. that

$$L_N \mathbb{E}_{\alpha,\theta}\left[\left(\eta_1^{(N)}\right)^b\right] \sim c_{\alpha,\theta,\kappa} \int_0^1 p^{b-2}\Lambda(dp) \tag{2.31}$$

where $\Lambda(dp)$ is the coagulation measure of the Beta$(1+\frac{\theta}{\alpha}, 1-\frac{\theta}{\alpha})$-coalescent. By Lemma 2.2.17 we have

$$\mathbb{E}_{\alpha,\theta}\left[\left(\eta_1^{(N)}\right)^b\right] = \mathbb{E}_{\alpha,\theta+\alpha}\left[\int_0^1 \frac{y^{\kappa b}}{(y^\kappa + (1-y)^\kappa \zeta_{N-1,\kappa})^b} y^{-\alpha}(1-y)^{\theta+\alpha-1}\frac{dy}{\mathsf{B}(1-\alpha,\theta+\alpha)}\right],$$

where, making the change of variable $u = (1-y)^\kappa \zeta_{N-1,\kappa}$, we obtain

$$\mathbb{E}_{\alpha,\theta}\left[\left(\eta_1^{(N)}\right)^b\right] = \frac{\kappa^{-1}}{\mathsf{B}(1-\alpha,\theta+\alpha)} \times$$
$$\mathbb{E}_{\alpha,\theta+\alpha}\left[\frac{1}{\zeta_{N-1,\kappa}^{1+\theta/\alpha}} \int_0^{\zeta_{N-1,\kappa}} \frac{\left(1-\frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}}\right)^{\kappa b}}{\left(\left(1-\frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}}\right)^\kappa + u\right)^b} \left(1-\frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}}\right)^{-\alpha} u^{\theta/\alpha} du\right]. \tag{2.32}$$

The integral inside the expectation above can be bounded on the set $\left\{\frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}} \leq \epsilon\right\}$ by

$$\int_0^\infty \frac{1}{(1-\epsilon^\kappa + u)^b} u^{\theta/\alpha} du < \infty,$$

yielding, by Proposition 2.2.16 and the conditions $\theta \in (-\alpha, \alpha)$ and $\frac{\alpha}{2} < \kappa \leq \alpha$,

$$\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}} \mathbb{E}_{\alpha,\theta}\left[\left(\eta_1^{(N)}\right)^b; \frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}} \leq \epsilon\right] = \mathcal{O}(1)$$

as $N \to \infty$. On the other hand, on the set $\left\{ \frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}} > \epsilon \right\}$ and using the condition $\theta < \alpha$, the integral can be bounded by

$$\mathbb{1}_{\epsilon<1}(1-\epsilon)^{2\kappa-\alpha} \int_{\epsilon^\kappa \zeta_{N-1,\kappa}}^{\zeta_{N-1,\kappa}} u^{\theta/\alpha-2} du \leq C \zeta_{N-1,\kappa}^{\frac{\theta}{\alpha}-1}$$

for some $C > 0$. This yields, together with Hölder's inequality with $p = 1+\epsilon$ for $\frac{\alpha-\theta}{2(\alpha+\theta)} < \epsilon < \frac{\alpha}{\alpha+\theta}$,

$$\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}} \mathbb{E}_{\alpha,\theta}\left[\left(\eta_1^{(N)}\right)^b ; \frac{u^{1/\kappa}}{\zeta_{N-1,\kappa}^{1/\kappa}} > \epsilon\right] \leq \mathbb{E}_{\alpha,\theta+\alpha}\left[\frac{\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}}}{\zeta_{N-1,\kappa}^{1+\frac{\theta}{\alpha}+1-\frac{\theta}{\alpha}}}\right]$$
$$\leq \left\| \frac{\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}}}{\zeta_{N-1,\kappa}^{1+\frac{\theta}{\alpha}}} \right\|_{\mathsf{L}^p} \left\| \frac{\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1-\frac{\theta}{\alpha}}}{\zeta_{N-1,\kappa}^{1-\frac{\theta}{\alpha}}} \right\|_{\mathsf{L}^q} \left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{\frac{\theta}{\alpha}-1}.$$

By a second application of Proposition 2.2.16 and the conditions $\theta \in (-\alpha, \alpha)$ and $\kappa \leq \alpha$, the last line above converges to 0 as $N \to \infty$. Thus, we may apply dominated convergence in (2.32) which, together with Proposition 2.2.16, yields

$$\begin{aligned}\lim_{N\to\infty} &\left(\sum_{i=1}^N i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta}{\alpha}} \mathbb{E}_{\alpha,\theta}\left[\left(\eta_1^{(N)}\right)^b\right] \\ &= \mathbb{E}_{\alpha,\alpha+\theta}\left[\left(\frac{\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} K_{\alpha,\theta}^{-\kappa} e^{\kappa S_\infty}\right)^{-1-\frac{\theta}{\alpha}}\right] \frac{\kappa^{-1}}{\mathsf{B}(1-\alpha, \alpha+\theta)} \times \\ &\int_0^\infty \left(\frac{1}{1+u}\right)^b u^{\theta/\alpha} du.\end{aligned} \tag{2.33}$$

By means of the change of variable $p = (1+u)^{-1}$ we may rewrite the integral above as

$$\int_0^\infty \left(\frac{1}{1+u}\right)^b u^{\theta/\alpha} du = \int_0^1 p^{b-2} p^{1-\theta/\alpha-1} (1-p)^{1+\theta/\alpha-1} dp,$$

whereas by Lemma 2.2.13

$$\begin{aligned}&\mathbb{E}_{\alpha,\alpha+\theta}\left[\left(\frac{\Gamma(1+\kappa-\alpha)}{\Gamma(1-\alpha)} K_{\alpha,\theta}^{-\kappa} e^{\kappa S_\infty}\right)^{-1-\frac{\theta}{\alpha}}\right] \frac{\kappa^{-1}}{\mathsf{B}(1-\alpha, \alpha+\theta)} \\ &= \kappa^{-1} \left(\frac{\Gamma(1-\alpha)}{\Gamma(1+\kappa-\alpha)}\right)^{1+\frac{\theta}{\alpha}} \frac{\Gamma(\alpha+\theta+1)}{\Gamma\left(\alpha+\theta-\frac{\kappa}{\alpha}(\alpha+\theta)+1\right)} \frac{\Gamma\left(\frac{1}{\alpha}(\alpha+\theta-\frac{\kappa}{\alpha}(\alpha+\theta))+1\right)}{\Gamma\left(\frac{\alpha+\theta}{\alpha}+1\right)} \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)\Gamma(\alpha+\theta)} \\ &= \frac{\alpha}{\kappa} \frac{\Gamma(1-\alpha)^{\frac{\theta}{\alpha}}}{\Gamma(1+\kappa-\alpha)^{1+\frac{\theta}{\alpha}}} \frac{\Gamma\left(\frac{\alpha+\theta}{\alpha}(1-\frac{\kappa}{\alpha})+1\right)}{\Gamma\left((\alpha+\theta)\left(1-\frac{\kappa}{\alpha}\right)+1\right)} \frac{\Gamma(1+\theta)\Gamma\left(1-\frac{\theta}{\alpha}\right)}{\mathsf{B}\left(1-\frac{\theta}{\alpha}, 1+\frac{\theta}{\alpha}\right)}.\end{aligned}$$

Substituting in (2.33) we obtain (2.31).

Proof of II: By a similar computation as in (2.30) we have, using this time $\alpha/2 < \kappa \le \alpha$ and $\theta \ge \alpha$, and setting $p = (1+\epsilon_1)(1+\epsilon_2), q = \frac{1+\epsilon_1}{\epsilon_1}(1+\epsilon_2)$, and $r = \frac{(1+\epsilon_1)(1+\epsilon_2)}{\epsilon_2(1+\epsilon_1)}$, for $\theta - 2\alpha < \eta < \theta$, $\frac{\alpha-\theta+\eta}{2(\alpha+\theta)} < \epsilon_1 < \frac{\alpha}{\alpha+\theta}$, and $\epsilon_2 > 0$ small enough; by Proposition 2.2.16,

$$\begin{aligned}
\mathbb{E}_{\alpha,\theta}\left[\sum_{i=1}^{N}\left(\eta_i^{(N)}\right)^3\right] &= \mathbb{E}_{\alpha,\theta}\left[\frac{\zeta_{N,3\kappa}}{\zeta_{N,\kappa}^3}\right] \\
&\le \left\|\frac{\left(\sum_{i=1}^{N} i^{-\frac{\kappa}{\alpha}}\right)^{1+\frac{\theta-\eta}{\alpha}}}{\zeta_{N-1,\kappa}^{1+\frac{\theta-\eta}{\alpha}}}\right\|_{\mathsf{L}^p}\left\|\frac{\left(\sum_{i=1}^{N} i^{-\frac{\kappa}{\alpha}}\right)^{2-\frac{\theta-\eta}{\alpha}}}{\zeta_{N-1,\kappa}^{2-\frac{\theta-\eta}{\alpha}}}\right\|_{\mathsf{L}^q}\left\|\zeta_{N-1,3\kappa}\right\|_{\mathsf{L}^r}\left(\sum_{i=1}^{N} i^{-\frac{\kappa}{\alpha}}\right)^{-3} \\
&= \mathcal{O}\left(\left(\sum_{i=1}^{N} i^{-\frac{\kappa}{\alpha}}\right)^{-3}\right).
\end{aligned}$$

This, together with (2.16) yield $\mathbb{E}\left[\sum_{i=1}^{N}\left(\eta_i^{(N)}\right)^3\right] = o\left(c_N\right)$ and the proof is finished by an application of Corollary 2.2.3. □

2.3 Exponential models

We study a population evolution model of N particles positioned on the real line that evolve through discrete generations $t \in \mathbb{N}$. Every generation is of size N and is constructed from the previous generation through **branching** and **selection** steps. The positions of the particles, say $(X_1^N(t), \dots, X_N^N(t))$, give the fitness levels of the individuals; individuals with a higher position have a greater probability of having more descendants in the next generation. During the branching step an individual at position x will be replaced, independently from the other particles and from the previous generations, by a countable number of children whose positions are given by a Poisson point process of intensity $e^{-(s-x)}\,ds$. For the selection step we let $1 \le \gamma \le \infty$ and $\beta > 0$ be two parameters of the model and work conditionally on the positions of all the newly produced particles. Selection occurs in two substeps, first the fittest $\lceil N^\gamma \rceil$ children are selected and the rest are discarded; second, N of the surviving children are sampled without replacement, with the probability of picking a child at position x being proportional to $e^{\beta x}$. Observe that if $\gamma = \infty$ then the first selection step is innocuous. Also, the second selection step is well defined only when $\beta > 1$; indeed, since for any $x \in \mathbb{R}$ the integral

$$\int_{-\infty}^{\infty}\left(e^{\beta s} \wedge 1\right) e^{-(s-x)} ds$$

is finite only when $\beta > 1$. This entails that the sampling weight of all the descendants of a particle at position x will be finite only in this case, so that the sampling probabilities are also well defined only in this case. Hence the model is well defined whenever $(0 < \beta \le 1, \gamma < \infty)$ or $(\beta > 1, \gamma \le \infty)$.

We are primarily interested in the limit as $N \to \infty$ of the genealogy $\left(\Pi_t^{(N,n)}\right)_{t\geq 0}$ of n randomly chosen individuals; as well as in computing the speed of selection (defined in Section 2.3.2 below), for different regimes of the parameters γ and β. The case $\gamma = 1$, in which the strength of the first selection step is maximum and the N fittest children are always chosen for the next generation, corresponds to the original exponential model of Brunet and Derrida 1997, 2012. The case ($\gamma = \infty, \beta > 1$), in which the first selection step is weakest, but the second selection step is strongest, corresponds to the model studied by Cortines and Mallein 2017. These two cases fall in the strong-selection regime of the exponential models whose dynamics are mainly governed by the fittest individuals, in this regime the genealogy $\left(\Pi_t^{(N,n)}\right)_{t\geq 0}$ converges to the Bolthausen-Sznitman coalescent (see Theorem 2.3.2), and the speed of selection is of order $\log\log(N)$ (see Theorem 2.3.12). In Theorems 2.3.2 and 2.3.12 we also present our joint (yet unpublised) work together with Emmanuel Schertzer where we study the case ($1 \leq \gamma < \infty, 0 < \beta < 1$) and find that, for large γ and small β, both selection steps become too weak in comparison to the force of mutation. In this setting the overall population falls into a weak-selection regime in which the limit genealogy is a discrete-time coalescent process, and the cloud of particles primarily explores the low-fitness landscape. On the other hand, for small γ and large β, the population once again falls into the strong-selection regime as before. Additionally, in this manuscript we provide a mild extensions on the latter results by including the case ($\beta = 1, 1 \leq \gamma < \infty$), and also provide a complimentary proof of the results first described in the physics literature (Brunet and Derrida 1997, 2012) for regime $\gamma = 1$.

Figure 2.1: Schematic representation of the passage from generation t to generation $t+1$ with $N = 5$. The selection step has been divided into two selections steps, the first one corresponds to the filter keeping the $\lceil N^\gamma \rceil$ rightmost particles, and the second corresponds to the sampling without replacement procedure described in the main text. The figure is a realization of the branching + selection steps of a single generation in the $\alpha > 0$ regime of Theorems 2.3.2 and 2.3.12 below.

Finally, before stating and proving our main results, it is also worth mentioning another interesting modification of the original exponential model in which $\gamma = 1$ (the N fittest individuals are always chosen) but, in the branching step, the offspring produced by a parent at position x are centerted around $ax, 0 < a < 1$ (instead of x), leading to Beta$(2-a^{-1}, a^{-1})$-coalescents in the limit Cortines and Mallein 2018.

2.3.1 Limit Genealogy

We first rewrite the exponential model in terms of the multinomial model of Section 2.2. For this let $(\mathcal{E}_j)_{j\geq 0}$ be i.i.d. standard exponential random variables so that

$$x_k = (\mathcal{E}_1 + \cdots + \mathcal{E}_k)^{-1}, \quad k \geq 1, \tag{2.34}$$

are the points, in decreasing order, of a Poisson point process Ξ on $(0,\infty)$ with intensity measure $z^{-2}dz$ (see the Mapping Theorem in Kingman 1992); also let $I^N := \left(I_1^N, \dots, I_N^N\right)$ be a random sample without replacement from the set of indexes $[N^\gamma]$ where index k is chosen with initial probabilities

$$\mathbb{P}\left(I_1^N = k\right) = \frac{x_k^\beta}{\sum_{j=1}^{\lceil N^\gamma \rceil} x_j^\beta}, \quad 1 \leq k \leq \lceil N^\gamma \rceil. \tag{2.35}$$

Theorem 2.3.1. *The genealogy of the exponential model of parameters (N,γ,β) is identical in distribution to the genealogy of the multinomial model whose family frequencies are given by*

$$\eta^{(N)} \overset{d}{=} \left(\frac{x_{I_k^N}}{\sum_{j=1}^N x_{I_j^N}}\right)_{1\leq k\leq N}. \tag{2.36}$$

Proof. By superimposing independent Poisson point processes (Section 2.2 in Kingman 1992) we see that an equivalent branching step to the one described above is to instead produce all the offspring of all the individuals in generation t through a single Poisson point process of intensity $e^{-(s-X_{eq}^N(t))}\,ds$ where $X_{eq}^N(t) := \log\left(\sum_{j=1}^N e^{X_j^N(t)}\right)$ (see Proposition 1.3 in Cortines and Mallein 2017). One can also see that, under this setting, a child at position x is a descendant of the kth individual in generation t with probability

$$\eta_k^{(N)} := \frac{e^{X_k^N(t)}}{\sum_{j=1}^N e^{X_j^N(t)}}; \tag{2.37}$$

i.e. children choose parents independently of their own positions with probabilities given by (2.37). Now, going one further generation backwards in time, and using the same superposition of Poisson point processes as before, we see that the offspring produced in the branching step of generation $t-1$ is equal in distribution to the set of points $(X_{eq}(t-1)+x_k)_{k\in\mathbb{N}}$, where the x_k's are as in (2.34). Indeed, by the Mapping Theorem (Kingman 1992) the points of a Poisson point processes with intensity $e^{-(s-X_{eq}^N(t-1))}\,ds$ are identical in distribution to $(X_{eq}(t-1)+x_k)_{k\in\mathbb{N}}$. Performing the selection step on the latter we see that the positions $\left(X_1^N(t),\dots,X_N^N(t)\right)$ are identical in distribution to $(X_{eq}(t-1)+\log(x_{I_1^N}),\dots,X_{eq}(t-1)+\log(x_{I_N^N}))$ and thus, plugging in (2.37) and considering the dependence on t,

$$\eta^{(N)}(t) \overset{d}{=} \left(\frac{x_{I_k^N}}{\sum_{j=1}^N x_{I_j^N}}\right)_{1\leq k\leq N}.$$

It remains to observe that the distribution of the right hand side does not depend on t and that, the branching steps being independent, the sequence $\left(\eta^{(N)}(t)\right)_{t\geq 0}$ is i.i.d. □

To ease notation we will omit the N superscript and write I instead of I^N from now on. In order to characterize the limit genealogy we start by studying the multinomial weight $e^{-X_1} \overset{d}{=} x_{I_1}$ of the first chosen individual when $0 < \beta < 1$. In particular we are interested in its asymptotic behaviour as $N \to \infty$; letting a_N be a sequence of (normalizing) positive numbers converging to infinity we compute, for any $z > 0$,

$$\begin{aligned} \mathbb{P}(a_N e^{-X_1} > z) &= \mathbb{E}\left[\mathbb{P}\left(a_N x_{I_1} > z | x_1, x_2, \ldots\right)\right] \\ &= \mathbb{E}\left[\left(\frac{1}{\sum_{k=1}^{\lceil N^\gamma \rceil} x_k^\beta}\right) \sum_{k \le \lceil N^\gamma \rceil :\, x_k^{-1} < a_N/z} x_k^\beta\right] \\ &\approx \frac{a_N^{1-\beta}}{N^{\gamma(1-\beta)}} z^{\beta-1}; \end{aligned} \tag{2.38}$$

where we have used the law of large numbers in the form $x_k \approx k^{-1}$ and the Riemann-integral approximation

$$\sum_{k=r}^{R} k^{-\beta} \sim \frac{1}{1-\beta} R^{1-\beta}$$

as $R \to \infty$. This suggests that the correct normalizing sequence is $a_N = N^\alpha$ where

$$\alpha \coloneqq \gamma - \frac{1}{1-\beta}. \tag{2.39}$$

We obtain the following theorem.

Theorem 2.3.2. *Let $\left(\Pi_t^{(N,n)}\right)_{t \ge 0}$ be the genealogy of n randomly chosen individuals in the exponential model with population size N. Then as $N \to \infty$:*
I) Weak Selection Regime: *If $(\beta < 1, \alpha > 0)$, the process $\left(\Pi_t^{(N,n)}\right)_{t \ge 0}$ converges in the product topology for $(\mathscr{P}_n)^{\mathbb{N}}$, to the discrete-time Poisson-Dirichlet $(1-\beta, 0)$ coalescent process started with n individuals.*
II) Strong Selection Regime: *If $(\beta < 1, \alpha < 0), (\beta = 1, \gamma < \infty)$, or $(\beta > 1, \gamma = \infty)$, letting*

$$\chi \coloneqq \frac{1 - \gamma(1 - \beta \wedge 1)}{\beta \wedge 1}, \tag{2.40}$$

then, as $N \to \infty$, $c_N \sim (\chi \log N)^{-1}$, and the time-scaled process $\left(\Pi_{\lfloor t/c_N \rfloor}^{(N,n)}\right)_{t \ge 0}$ converges weakly in the Skorohod space $D([0,\infty), \mathscr{P}_n)$ to the Bolthausen-Sznitman coalescent.

Regime $\gamma = 1$ (equivalently $(\gamma = \infty, \beta = \infty)$ or, trivially, $(\gamma = 1, \beta = 1)$) of the above theorem corresponds to the original exponential model studied in Brunet and Derrida 1997, 2012; Brunet, Derrida, et al. 2007; while regime $(\beta > 1, \gamma = \infty)$ was proved in Cortines and Mallein 2017. Here we describe our article together with Emmanuel Schertzer in which we

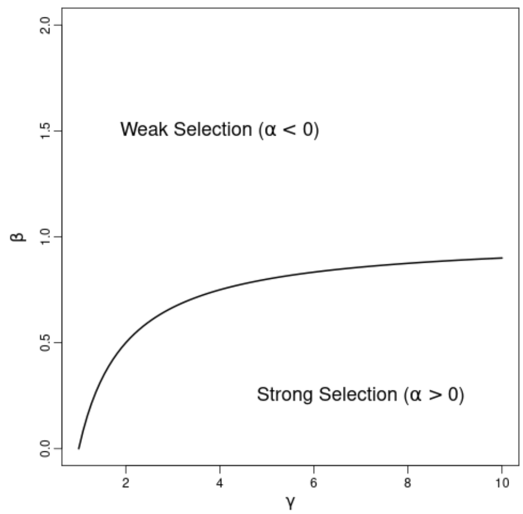

Figure 2.2: Phase diagram for the expnential models.

deal with regimes ($\beta < 1, \alpha \neq 0$), and additionally complement the results by proving regime ($\beta = 1, 1 \leq \gamma < \infty$) using the same techniques.

There are two missing regimes in the above theorem. First, in regime ($\beta > 1, \gamma < \infty$), the force of selection is stronger than in the regime ($\beta > 1, \gamma = \infty$) but weaker than in the regime ($\gamma = 1$); for this reason we conjecture that the asymptotics in this case should equal those described in Cortines and Mallein 2017 and Brunet and Derrida 1997, 2012; Brunet, Derrida, et al. 2007, and may even follow using similar techniques. The second missing regime, in which the parameters $1 < \gamma < \infty$ and $\beta \in (0, 1)$ satisfy $\gamma = \frac{1}{1-\beta}$ (i.e $\alpha = 0$), still remains to be studied.

The techniques used in the proofs of the different regimes in Theorem 2.3.2 are different enough to be put in different subsections. We first proof **I)**, and then split the proof of **II)** into the case ($0 < \beta \leq 1, 1 \leq \gamma < \infty, \alpha < 0$), and the case ($\beta > 1, \gamma = \infty$).

Proof of Theorem 2.3.2.I)

For the case $\alpha > 0$ it will be easier to work with the general setting of a deterministic positive and decreasing sequence $(a_k)_{1\leq k\leq\infty}$ satisfying $ka_k \to 1$ as $k \to \infty$ which, by the law of large numbers, the r.vs. $(x_k)_{1\leq k<\infty}$ satisfy almost surely. We wish to prove that if I is a sample without replacement of size N from $\{1, \ldots, \lceil N^\gamma \rceil\}$ with initial probabilities

$$\mathbb{P}(I_1 = k) = \left(\frac{a_k^\beta}{\sum_{j=1}^{\lceil N^\gamma \rceil} a_j^\beta} \right), \tag{2.41}$$

then

$$\left(\frac{a_{\hat{I}_k}}{\sum_{j=1}^{N} a_{\hat{I}_j}} \right)_{1\leq k\leq N} \Rightarrow \frac{W_{\infty,\infty}^Z}{\left\| W_{\infty,\infty}^Z \right\|_{1_1}},$$

in the $\mathtt{l}_1$ topology where $(\hat{I}_k)_{1\le k\le N}$ is the set I ordered increasingly (so that $\left(a_{\hat{I}_k}\right)_{1\le k\le N}$ is ordered decreasingly), and

$$W^Z_{N,\ell} := (z_1, \dots, z_{N\wedge\ell}, 0, \dots)$$

are the ranked atoms of an independent Poisson random measure Z on $(0,\infty)$ with intensity measure $(1-\beta)x^{\beta-2}dx$ (the reason for using a double subindex in the notation $W^Z_{N,\ell}$ will be made clear shortly). Recall from Pitman and Yor 1997 that the normalized sequence $\frac{W^Z_{\infty,\infty}}{\left\|W^Z_{\infty,\infty}\right\|_{\mathtt{l}_1}}$ has Poisson-Dirichlet $(1-\beta,0)$ distribution.

Independent Sampling in the Selection Step: Let (I,J) be the random coupling where J is an i.i.d sample of size N from $[N^\gamma]$ with probabilities given by (2.41), and I is constructed from J by replacing all its repeated coordinates $R = \{j\colon J_i = J_j \text{ for some } 1 \le i < j\}$ (so that only the first copy remains) with a new sample without replacement from $[N^\gamma]\setminus J$ with initial probabilities

$$\frac{a_k^\beta}{\sum_{i\in[N^\gamma]\setminus J} a_i^\beta}, \quad i \in [N^\gamma]\setminus J;$$

thus ensuring that the resulting set I is a sample without replacement with the original probabilities (2.41). Recall $\hat{I}$ (resp. $\hat{J}$) for the arrangement of I in non-decreasing order. Our first objective is to prove the following results which ensure that the sampling without replacement of the selection step can be replaced, in the limit as $N\to\infty$, by sampling with replacement. We first introduce some notation. Let $W^I_{N,\ell}$ (resp. $W^J_{N,\ell}$) be defined as

$$W^I_{N,\ell} := \left(a_{\hat{I}_1}, \dots, a_{\hat{I}_{N\wedge\ell}}, 0, \dots\right).$$

Lemma 2.3.3. *As $N\to\infty$,*

$$\left\|W^I_{N,\infty} - W^J_{N,\infty}\right\|_{\mathtt{l}_1} \xrightarrow{P} 0. \tag{2.42}$$

Before proving Lemma 2.3.3 we will first motivate ourselves by characterizing the weak limit of $W^J_{N,\infty}$ in $\mathscr{P}_{[0,1]}$ as $N\to\infty$.

Proposition 2.3.4. *As $N\to\infty$,*

$$W^J_{N,\infty} \Rightarrow W^Z_{\infty,\infty}$$

in the $\mathtt{l}_1$ topology.

Proof. We follow the steps in the proofs of Lemmas 20 and 21 in Schweinsberg 2003. We first prove the convergence of the finite dimensional distributions, i.e. for fixed $\ell\in\mathbb{N}$,

$$W^J_{N,\ell} \Rightarrow W^Z_{\infty,\ell} \tag{2.43}$$

in the $\mathtt{l}_1$ topology. Indeed, following Schweinsberg 2003, for any collection of positive real numbers $\infty = y_0 > y_1 \ge y_2 \ge \cdots \ge y_\ell$ we define the random variables $\bar{L}_i := \#\{k\in J : y_i \le N^\alpha a_k\}$ where $1\le i\le \ell$, and note that

$$\mathbb{P}\left(N^\alpha a_{\hat{J}_1} \ge y_1, \dots, N^\alpha a_{\hat{J}_\ell} \ge y_\ell\right) = \mathbb{P}\left(\bar{L}_i \ge i \text{ for } 1\le i\le \ell\right)$$

so that it is enough to prove

$$(L_1,\dots,L_\ell)\Rightarrow\left(Z\left([y_1,y_0)\right),\dots,Z\left([y_\ell,y_{\ell-1})\right)\right)$$

where $L_i \coloneqq \#\{k\in J : y_i\le N^\alpha a_k<y_{i-1}\}$. We compute, for any collection of non-negative integers $n_1,\dots,n_\ell$,

$$\begin{aligned}&\mathbb{P}\left(L_1=n_1,\dots,L_\ell=n_\ell\right)\\&=\frac{(N)_{n_1+\cdots+n_\ell}}{n_1!\dots n_\ell!}\frac{\prod_{i=1}^{\ell}\left(\sum_{k:\,N^\alpha/y_{i-1}\le a_k\le N^\alpha/y_i}^{\lceil N^\gamma\rceil}a_k^\beta\right)^{n_i}}{\left(\sum_{k=1}^{\lceil N^\gamma\rceil}a_k^\beta\right)^{n_1+\cdots+n_\ell}}\left(1-\frac{\sum_{k:\,a_k\le N^\alpha/y_\ell}^{\lceil N^\gamma\rceil}a_k^\beta}{\sum_{k=1}^{\lceil N^\gamma\rceil}a_k^\beta}\right)^{N-n_1-\cdots-n_\ell}\end{aligned}$$

which, by means of the following asymptotics:

$$(N)_{n_1+\cdots+n_\ell}\sim N^{n_1+\cdots+n_\ell},$$

$$\left(\sum_{k:\,N^\alpha/y_{i-1}\le a_k\le N^\alpha/y_i}^{\lceil N^\gamma\rceil}a_k^\beta\right)^{n_i}\sim\left(\frac{(N^\alpha/y_i)^{(1-\beta)}}{1-\beta}-\frac{(N^\alpha/y_{i-1})^{(1-\beta)}}{1-\beta}\right)^{n_i},$$

$$\left(\sum_{k=1}^{\lceil N^\gamma\rceil}a_k^\beta\right)^{n_1+\cdots+n_\ell}\sim\frac{N^{\gamma(1-\beta)(n_1+\cdots+n_\ell)}}{(1-\beta)^{n_1+\cdots+n_\ell}},$$

and

$$\begin{aligned}\lim_{N\to\infty}\left(1-\frac{\sum_{k:\,a_k\le N^\alpha/y_\ell}^{\lceil N^\gamma\rceil}a_k^\beta}{\sum_{k=1}^{\lceil N^\gamma\rceil}a_k^\beta}\right)^{N-n_1-\cdots-n_\ell}&=\exp\left\{-\lim_{N\to\infty}(N-n_1-\cdots-n_\ell)\frac{\sum_{k=1}^{N^\alpha/y_\ell}a_k^\beta}{\sum_{k=1}^{\lceil N^\gamma\rceil}a_k^\beta}\right\}\\&=\exp\left\{-\lim_{N\to\infty}N\frac{N^{\alpha(1-\beta)}y_\ell^{\beta-1}}{N^{\gamma(1-\beta)}}\right\}\\&=e^{-y_\ell^{\beta-1}},\end{aligned}$$

becomes

$$\begin{aligned}&\mathbb{P}\left(L_1=n_1,\dots,L_\ell=n_\ell\right)\\&\sim\frac{N^{n_1+\cdots+n_\ell}}{n_1!\dots n_\ell!}\frac{\prod_{i=1}^{\ell}\left((N^\alpha/y_i)^{1-\beta}-(N^\alpha/y_{i-1})^{1-\beta}\right)^{n_i}}{N^{\gamma(1-\beta)(n_1+\cdots+n_\ell)}}e^{-y_\ell^{\beta-1}}\\&=\frac{N^{n_1+\cdots+n_\ell}}{n_1!\dots n_\ell!}\frac{\prod_{i=1}^{\ell}\left(y_i^{\beta-1}-(y_{i-1})^{\beta-1}\right)^{n_i}}{N^{n_1+\cdots+n_\ell}}e^{-y_\ell^{\beta-1}}\\&=\prod_{i=1}^{\ell}\frac{e^{-(y_i^{\beta-1}-y_{i-1}^{\beta-1})}\left(y_i^{\beta-1}-(y_{i-1})^{\beta-1}\right)^{n_i}}{n_i!}\\&=\mathbb{P}\left(Z([y_i,y_{i-1}))=n_i\text{ for }1\le i\le\ell\right).\end{aligned}$$

This proves (2.43). Given (2.43) and by Theorem 3.2 in Billingsley 1999, it only remains to prove

$$\lim_{\ell\to\infty}\limsup_{N\to\infty}\mathbb{E}\left[\left\|W^J_{N,\ell}-W^J_{N,\infty}\right\|_{1_1}\wedge 1\right]=0.$$

Let $0<\epsilon<1$ and choose ℓ large enough to ensure $\mathbb{E}\left[\sum_{k=\ell}^{\infty} z_k\right]<\epsilon/2$. Also (2.43) allows us to choose N_1 large enough to ensure $\mathbb{P}\left(N^\alpha a_{\hat{J}_\ell}>\epsilon\right)<\epsilon$ for every $N\geq N_1$; thus, using the fact that $a_{\hat{J}_k}$ is decreasing in k,

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{k=\ell}^{N}N^\alpha a_{\hat{J}_k}\right)\wedge 1\right]&=\mathbb{P}\left(N^\alpha a_{\hat{J}_\ell}>\epsilon\right)+\mathbb{E}\left[\left(\sum_{k=\ell}^{N}N^\alpha a_{\hat{J}_k}\right)\wedge 1; N^\alpha a_{\hat{J}_\ell}\leq\epsilon\right]\\&\leq\epsilon+\mathbb{E}\left[\sum_{k=N_1}^{N}N^\alpha a_{\hat{J}_k}\mathbb{1}_{N^\alpha a_{\hat{J}_k}\leq\epsilon}\right].\end{aligned}$$

Finally observe that $\mathbb{P}\left(J_i=k|J_1,\ldots,J_{i-1}\right)=\frac{a_k^\beta}{\sum_{k=1}^{\lceil N^\gamma\rceil}a_k^\beta-a_{J_1}-\cdots-a_{J_{k-1}}}\leq\frac{a_k^\beta}{\sum_{k=N}^{\lceil N^\gamma\rceil}a_k^\beta}$, $1\leq i\leq N$, so that

$$\begin{aligned}\mathbb{E}\left[\sum_{k=M}^{N}N^\alpha a_{\hat{J}_k}\mathbb{1}_{N^\alpha a_{\hat{J}_k}\leq\epsilon}\right]&=N^\alpha\sum_{k\in[N^\gamma]:\ a_k<\epsilon N^{-\alpha}}a_k\mathbb{E}\left[\sum_{i=1}^{N}\mathbb{1}_{J_i=k}\right]\\&\leq N^{1+\alpha}\left(\frac{\sum_{k:\ a_k^{-1}>\epsilon^{-1}N^\alpha}^{\lceil N^\gamma\rceil}a_k^{1+\beta}}{\sum_{k=N}^{\lceil N^\gamma\rceil}a_k^\beta}\right)\\&\sim N^{1+\alpha}\left(\frac{(\epsilon^{-1}N^\alpha)^{-\beta}}{\beta}\right)\frac{1-\beta}{N^{\gamma(1-\beta)}}\\&=\epsilon^\beta\left(\frac{1-\beta}{\beta}\right).\end{aligned}\tag{2.44}$$

□

Proof of Lemma 2.3.3. Let $\delta>0$ and observe that, by a similar computation as in equation (2.44) (replacing ϵ by $N^{-\delta}$ and using that $a_k\sim k^{-1}$), both

$$\sum_{N^{-\alpha-\delta}<k\leq\lceil N^\gamma\rceil}N^\alpha\mathbb{E}\left[a_k\mathbb{1}_{k\in I}\right]\ \text{ and }\sum_{N^{-\alpha-\delta}<k\leq\lceil N^\gamma\rceil}N^\alpha\mathbb{E}\left[a_k\mathbb{1}_{k\in J}\right]$$

are of order $\mathcal{O}\left(N^{-\beta\delta}\right)$, and thus

$$\mathbb{E}\left[\left\|W^I_{N,\infty}-W^J_{N,\infty}\right\|_{1_1}\wedge 1\right]\leq\mathbb{E}\left[\left(a_1\#[A^I_\delta\Delta A^J_\delta]\right)\wedge 1\right]+\mathcal{O}\left(N^{-\beta\delta}\right)\tag{2.45}$$

where $\#[A^I_\delta\Delta A^J_\delta]$ is the number of indices k, counting repetitions, such that $k<N^{-\alpha-\delta}$ and $k\in I\Delta J$. Formally, if A^J_δ (resp. A^I_δ) is the multiset

$$A^J_\delta:=\left[J_k:\ J_k\leq N^{-\alpha-\delta}\right],$$

then $\#[A_\delta^I \Delta A_\delta^J]$ is the size of the multiset symmetrical difference $[A_\delta^I \Delta A_\delta^J]$. We now prove that $\#[A_\delta^I \Delta A_\delta^J] \xrightarrow{P} 0$ as $N \to \infty$ for δ small enough, which, when plugged in (2.45), gives (2.42). By construction of (I, J) and using that $\mathbb{P}\left(I_i = k | J\right) \leq \frac{a_k^\beta}{\sum_{j=N}^{\lceil N^\gamma \rceil} a_j^\beta}$, we obtain, on the one hand, for the indices $k < N^{-\alpha-\delta}$ such that $k \in \left[A^I \setminus A^J\right]$,

$$\begin{aligned}
\mathbb{E}\left[\#\left[A_\delta^I \setminus A_\delta^J\right]\right] &\leq \mathbb{E}\left[\#\left\{k \in I \setminus J \colon k \leq N^{\alpha+\delta}\right\}\right] \\
&\leq \sum_{k=1}^{\lfloor N^{\alpha+\delta} \rfloor} \mathbb{P}\left(k \in I \setminus J\right) \\
&= \sum_{k=1}^{\lfloor N^{\alpha+\delta} \rfloor} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{E}\left[\mathbb{1}_{J_i = J_j} \mathbb{P}\left(\mathbb{1}_{I_i = k} | J\right)\right] \\
&\leq N^2 \left(\frac{\sum_{k=1}^{\lceil N^\gamma \rceil} a_k^{-2\beta}}{\left(\sum_{k=1}^{\lceil N^\gamma \rceil} a_k^\beta\right)^2} \right) \sum_{k=1}^{\lfloor N^{\alpha+\delta} \rfloor} \left(\frac{a_k^\beta}{\sum_{j=N}^{\lceil N^\gamma \rceil} a_j^\beta} \right) \\
&= \begin{cases} \mathcal{O}\left(N^{1-2\gamma(1-\beta)+\delta(1-\beta)}\right) & \text{if } \beta > 1/2, \\ \mathcal{O}\left(N^{1-2\gamma(1-\beta)+\delta(1-\beta)} \log N\right) & \text{if } \beta = 1/2, \\ \mathcal{O}\left(N^{1-\gamma(1-\beta)+\delta(1-\beta)}\right) & \text{if } \beta < 1/2. \end{cases}
\end{aligned}$$

The latter converge to zero as $N \to \infty$ whenever $\delta < \frac{\gamma - 1}{1-\beta}$. On the other hand, by a similar computation,

$$\begin{aligned}
\mathbb{E}\left[\#\left[A_\delta^J \setminus A_\delta^I\right]\right] &\leq N^2 \frac{\sum_{k=1}^{\lfloor N^{\alpha+\delta} \rfloor} a_k^{-2\beta}}{\left(\sum_{k=1}^{\lceil N^\gamma \rceil} a_k^\beta\right)^2} \\
&= \mathcal{O}\left(N^2 \frac{\sum_{k=1}^{\lfloor N^{\alpha+\delta} \rfloor} a_k^{-2\beta}}{N^{2\gamma(1-\beta)}} \right)
\end{aligned}$$

which converges to zero whenever either $\beta > 1/2$ or $\delta < \frac{\gamma(1-\beta)-1+\beta\alpha}{1-2\beta}$. □

We are now in position to gather results and prove **I)** of Theorem 2.3.2.

Proof of Theorem 2.3.2 I. By Theorem 2.3.1 and Corollary 2.2.2 it is sufficient to prove that

$$\left(\frac{x_{I_k}}{\sum_{j=1}^N x_{I_j^N}} \right)_{1 \leq k \leq N} \Rightarrow \frac{W_{\infty,\infty}^Z}{\left\| W_{\infty,\infty}^Z \right\|_{\mathtt{l}_1}}$$

in $\mathtt{l}_1$. Proposition 2.3.4 and Lemma 2.3.3, plus Theorem 3.1 in Billingsley 1999, give

$$W_{N,\infty}^I \Rightarrow W_{\infty,\infty}^Z. \tag{2.46}$$

The convergence of the normalized sequences follows from the Continuous Mapping Theorem (Billingsley 1999) and the fact that $\mathbb{P}\left(\left\| W_{\infty,\infty}^Z \right\|_{\mathtt{l}_1} = 0\right) = 0$. □

Proof of Theorem 2.3.2.II), cases $(0 < \beta \leq 1, 1 \leq \gamma < \infty, \alpha < 0)$

In this section we assume that $0 < \beta \leq 1$ and $1 \leq \gamma < \infty$ satisfy $\alpha < 0$. We begin by proving that the high fitness children positioned above level $\approx N^{-\chi}$ are all chosen during the selection step with probability converging to 1 as $N \to \infty$ (Lemma 2.3.8); whereas the total fitness weight of the individuals in the new generation that are below level $\approx N^{-\chi}$ becomes negligible as $N \to \infty$ (Lemma 2.3.9). We then use these two heuristics in order to "reduce" the reproduction+selection steps of our model to those of the original exponential model of Brunet and Derrida 1997, 2012 with population size equal to N^{χ}, for which conditions *ii*) and *iii*) of Theorem 2.2.8 with $L_N = \chi \log(N)$ can be proved by direct computations.

First we provide two large deviation estimates which will simplify the computations further ahead. Recall the Poisson point process Ξ with non-increasing atoms (2.34); the following two lemmas roughly say that if we wish to sum the terms x_k^{β} over k then we may approximate x_k by k^{-1}. The first lemma gives high-probability lower bounds for the sum, while the second gives high-probability upper bounds.

Lemma 2.3.5. If $\beta < 1$: *Let $c > 1 - \beta$ and $\delta \in [0, 1)$. Define*

$$E_N = \left\{ \sum_{i=N^{\delta}}^{N} x_i^{\beta} \leq c^{-1} N^{1-\beta} \right\}.$$

Then for every $\eta > 0$, we have $\lim_{N\to\infty} N^{\eta} \mathbb{P}(E_N) = 0$.
If $\beta = 1$: *Let $c > 1$ and $\delta \in [0, 1)$. Define*

$$E_N = \left\{ \sum_{i=N^{\delta}}^{N} x_i \leq c^{-1}(1-\delta)\log(N) \right\}.$$

Then for $\eta < 1 - \delta$, we have $\lim_{N\to\infty} N^{\eta} \mathbb{P}(E_N) = 0$.

Proof. **Case $\beta < 1$** : By equation (2.34) and the Cramer large deviation estimate for the sum of standard exponential random variables we have $\mathbb{P}\left(x_N \geq \frac{N^{-1}}{1-\epsilon}\right) \leq e^{-N(-\epsilon - \log(\epsilon))}$ and $\mathbb{P}\left(x_{N^{\delta}} \leq \frac{N^{-\delta}}{1+\epsilon}\right) \leq e^{-N(\epsilon - \log(\epsilon))}$; thus,

$$\begin{aligned}\mathbb{P}(E_N) &= \mathbb{P}\left(E_N; x_N < \frac{N^{-1}}{1-\epsilon}, x_{N^{\delta}} > \frac{N^{-\delta}}{1+\epsilon}\right) + e^{-N(-\epsilon-\log(\epsilon))}\\ &\leq \mathbb{P}\left(\int_{\frac{N^{-1}}{1-\epsilon}}^{\frac{N^{-\delta}}{1+\epsilon}} x^{\beta}\Xi(dx) \leq \frac{N^{1-\beta}}{c}\right) + e^{-N(-\epsilon-\log(\epsilon))}\end{aligned}$$

so that it only remains to prove that for some appropriate choice of $\epsilon > 0$ we have

$$\mathbb{P}\left(\int_{\frac{N^{-1}}{1-\epsilon}}^{\frac{N^{-\delta}}{1+\epsilon}} x^{\beta}\Xi(dx) \leq \frac{1}{c(1-\epsilon)^{1-\beta}}\left(N(1-\epsilon)\right)^{1-\beta}\right) = \mathcal{O}\left(e^{-\eta N^{1-\beta}}\right), \quad \text{for some } \eta > 0. \tag{2.47}$$

Using Markov's inequality and Campbell's formula, for any $a_1 < a_2$ and $b > 0$ we have

$$\mathbb{P}\left(\int_{a_1}^{a_2} x^\beta \Xi(dx) \leq b^{-1} a_1^{\beta-1}\right) \leq e^{b^{-1}a_1^{\beta-1}} \mathbb{E}\left[\exp\left\{\int_{a_1}^{a_2} x^\beta \Xi(dx)\right\}\right]$$
$$= \exp\left\{b^{-1}a_1^{\beta-1} - \int_{a_1}^{a_2} \frac{1-e^{-x^\beta}}{x^2} dx\right\}.$$

Next, $1 - e^{-x^\beta} = x^\beta e^\theta$ for $x > 0$ and some $\theta \in [-x^\beta, 0]$ and thus

$$1 - e^{-x^\beta} \geq e^{-a_1^\beta} x^\beta$$

for all $x > a_1$; thus

$$\log\left(\mathbb{P}\left(\int_{a_1}^{a_2} x^\beta \Xi(dx) < b a_1^{\beta-1}\right)\right) \leq b^{-1}a_1^{\beta-1} - e^{-a_1^\beta} \int_{a_1}^{a_2} x^{\beta-2} dx$$
$$= a_1^{\beta-1} b^{-1}\left(1 - \frac{b\left(1-(a_2/a_1)^{\beta-1}\right)}{1-\beta} e^{-a^\beta}\right). \tag{2.48}$$

Letting $0 < \epsilon < 1$ be small enough to ensure that $c(1-\epsilon)^{1-\beta} > 1$ and setting $a_1 \equiv a_1^N = \frac{N^{-1}}{1-\epsilon}$, $a_2 \equiv a_2^N = \frac{N^{-\delta}}{1+\epsilon}$ and $b \equiv b^N = c(1-\epsilon)^{1-\beta}$, we obtain $\frac{b^N}{1-\beta} > 1$ and $\left(1 - \frac{b^N\left(1-(a_2^N/a_1^N)^{\beta-1}\right)}{1-\beta} e^{-a_N^\beta}\right) < 0$ for large enough N, so that substituting in (2.48) we get (2.47).

Case $\beta = 1$: We may assume $\delta > 0$ since

$$\left\{\sum_{i=1}^N x_i \leq c^{-1}(1-\delta)\log(N)\right\} \subset \left\{\sum_{i=N^\delta}^N x_i \leq c^{-1}(1-\delta)\log(N)\right\}.$$

By the same reasoning as in the previous case, we need to prove

$$\mathbb{P}\left(\int_{\frac{N^{-1}}{1-\epsilon}}^{\frac{N^{-\delta}}{1+\epsilon}} x^\beta \Xi(dx) \leq c^{-1}(1-\delta)\log(N)\right) = o\left(N^{-\eta}\right) \tag{2.49}$$

for $\eta < 1-\delta$. Consider any $a_1 < a_2 < 1$ and $b > 0$ then, again by the same computations as in the previous case, we have

$$\log\left(\mathbb{P}\left(\int_{a_1}^{a_2} x^\beta \Xi(dx) < b\log(a_2/a_1)\right)\right) \leq b\log(a_2/a_1) - e^{-a_1}\int_{a_1}^{a_2} \frac{1}{x} dx$$
$$= \log(a_2/a_1)(b - e^{-a_1}). \tag{2.50}$$

Setting $b^N = 1/c$, $a_1 \equiv a_1^N = \frac{N^{-1}}{1-\epsilon}$, and $a_2 \equiv a_2^N = \frac{N^{-\delta}}{1+\epsilon}$, we obtain $b^N - e^{-a_1^N} < 0$ for large enough N thus proving (2.49). □

Lemma 2.3.6. If $\beta < 1$: *Let $2c \in (0, 1-\beta)$ and*

$$E_N = \left\{\sum_{i=1}^N x_i^\beta \geq c^{-1} N^{(1-\beta)}\right\}.$$

Then for some $\eta > 0$*, we have* $\lim_{N\to\infty} N^{\eta}\mathbb{P}(E_N) = 0$*.*
If $\beta = 1$**:** *Let* $\epsilon > 0$ *and*

$$E_N = \left\{ \sum_{i=1}^{N} x_i \geq N^{\epsilon} \right\},$$

then $\mathbb{P}(E_N) = \mathcal{O}\left(N^{-\epsilon}\log N\right)$*.*

Proof. **Case** $\beta < 1$**:** As in the previous Lemma, we have $\mathbb{P}\left(x_N < \frac{N^{-1}}{1+\epsilon}\right) \leq \exp\left\{-N(\epsilon - \log(1-\epsilon))\right\}$ which for any choice of $\epsilon > 0$ is of order $o\left(N^{-\eta}\right)$ for every $\eta > 0$; thus for every $\eta > 0$

$$\mathbb{P}\left(E_N\right) \leq \mathbb{P}\left(\int_{\frac{N^{-1}}{1+\epsilon}}^{\infty} x^{-\beta}\Xi(dx) \geq \frac{N^{(1-\beta)}}{c}\right) + o\left(N^{-\eta}\right).$$

It remains to prove that with an appropriate choice of $\epsilon > 0$ the first term in the right hand side is of order $o\left(N^{-\eta}\right)$ for some $\eta > 0$. We compute, for any $a < b < \infty$ and $c > 0$,

$$\begin{aligned}\mathbb{P}\left(\int_a^b x^{\beta}\Xi(dx) \geq \frac{a^{\beta-1}}{c}\right) &\leq \mathbb{E}\left[\exp\left\{\int_a^b x^{\beta}\Xi(dx)\right\}\right] e^{-\frac{a^{\beta-1}}{c}} \\ &= \exp\left\{-\int_a^b \frac{1-e^{x^{\beta}}}{x^2}dx - \frac{a^{\beta-1}}{c}\right\}.\end{aligned}$$

We have $1 - e^{x^{\beta}} \geq -x^{\beta}e^{b^{\beta}}$ for all $x < b$, then we have

$$\begin{aligned}\log\left(\mathbb{P}\left(\int_a^b x^{-\beta}\Xi(dx) \geq \frac{a^{\beta-1}}{c}\right)\right) &\leq e^{b^{\beta}}\int_a^b x^{\beta-2}dx - \frac{a^{\beta-1}}{c} \\ &\leq \frac{a^{\beta-1}}{1-\beta}\left(e^{b^{\beta}} - \frac{1-\beta}{c}\right).\end{aligned}$$

Now let $0 < \delta < 1$ and write $a_N = \frac{N^{-1}}{1+\epsilon}$ and $b_N = a_N^{1-\delta}$, then using the above estimate and Markov's inequality

$$\begin{aligned}\mathbb{P}\left(\int_{\frac{N^{-1}}{1+\epsilon}}^{\infty} x^{-\beta}\Xi(dx) \geq \frac{N^{(1-\beta)}}{c}\right) &\leq \mathbb{P}\left(\int_{b_N}^{\infty} x^{\beta}\Xi(dx) \geq \frac{N^{(1-\beta)}}{2c}\right) + \\ &\quad \mathbb{P}\left(\int_{a_N}^{b_N} x^{\beta}\Xi(dx) \geq \frac{a_N^{\beta-1}}{2c(1+\epsilon)^{1-\beta}}\right) \\ &\leq 2ca_N^{1-\beta}\mathbb{E}\left[\int_{b_N}^{\infty} x^{\beta}\Xi(dx)\right] + \\ &\quad \exp\left\{\frac{a_N^{\beta-1}}{1-\beta}\left(e^{a_N^{\beta(1-\delta)}} - \frac{1-\beta}{2c(1+\epsilon)^{1-\beta}}\right)\right\}.\end{aligned}$$

Since $\mathbb{E}\left[\int_{b_N}^{\infty} x^{\beta}\Xi(dx)\right] = \int_{b_N}^{\infty} x^{\beta-2}dx = a_N^{(\beta-1)(1-\delta)}$, the first term above is of order $\mathcal{O}\left(a_N^{(1-\beta)\delta}\right)$, whereas the second term vanishes exponentially fast whenever $\epsilon > 0$ is chosen to ensure $\frac{1-\beta}{2c(1+\epsilon)^{1-\beta}} > 1$.

Case $\beta = 1$: Following the proof of the previous case we have

$$\mathbb{P}\left(E_N\right) \le \mathbb{P}\left(\int_{\frac{N^{-1}}{1+\epsilon}}^{\infty} x^{-\beta}\Xi(dx) \ge N^{\epsilon}\right) + o\left(N^{-\eta}\right).$$

where now

$$\begin{aligned}\mathbb{P}\left(\int_{\frac{N^{-1}}{1+\epsilon}}^{\infty} x\Xi(dx) \ge N^{\epsilon}\right) &\le \mathbb{P}\left(x_1 \ge N^{\epsilon}\right) + \mathbb{P}\left(\int_{\frac{N^{-1}}{1+\epsilon}}^{x_1} x\Xi(dx) \ge N^{\epsilon}\right)\\ &\le N^{-\epsilon} + N^{-\epsilon}\mathbb{E}\left[\int_{\frac{N^{-1}}{1+\epsilon}}^{x_1} \frac{1}{x}dx; x_1 > \frac{N^{-1}}{1+\epsilon}\right]\\ &= N^{-\epsilon} + N^{-\epsilon}\mathcal{O}\left(\log(N)\right).\end{aligned}$$

□

In the following, $\mathbb{P}^{\Xi}$ will denote the law of the N-sampled points conditional on a realization of the process Ξ.

Lemma 2.3.7. *For every $i \in [N^{\gamma}]$*

$$\mathbb{P}^{\Xi}(i \notin I) \le \exp\left\{-\frac{Nx_i^{\beta}}{\sum_{j=1}^{\lceil N^{\gamma}\rceil} x_j^{\beta}}\right\}. \tag{2.51}$$

Proof. Using that

$$\begin{aligned}\mathbb{P}^{\Xi}\left[I_{j+1} \neq i | I_1, \ldots, I_j\right] &= \left(1 - \frac{x_i^{\beta}}{\sum_{[N^{\gamma}]} x_i^{\beta} - \sum_{k=1}^{j} x_{I_j}^{\beta}}\right)\mathbb{1}_{i \notin \{I_1,\ldots,I_j\}} + \mathbb{1}_{i \in \{I_1,\ldots,I_j\}}\\ &\le \left(1 - \frac{x_i^{\beta}}{\sum_{[N^{\gamma}]} x_i^{\beta}}\right)\mathbb{1}_{i \notin \{I_1,\ldots,I_j\}} + \mathbb{1}_{i \in \{I_1,\ldots,I_j\}},\end{aligned}$$

and

$$\mathbb{P}^{\Xi}\left[I_{j+1} \neq i, \cap_{k=1}^{j} I_k \neq i \middle| I_1, \ldots, I_j\right] \le \left(1 - \frac{x_i^{\beta}}{\sum_{[N^{\gamma}]} x_i^{\beta}}\right)\mathbb{1}_{i \notin \{I_1,\ldots,I_j\}},$$

plus an inductive argument, we obtain

$$\mathbb{P}^{\Xi}(i \notin I) \le \exp\left\{N \ln\left(1 - \frac{x_i^{\beta}}{\sum_{j=1}^{\lceil N^{\gamma}\rceil} x_j^{\beta}}\right)\right\}.$$

The bound in (2.51) follows from plugging the inequality $\log(1-x) \le -x$ for $x \in (0,1)$. □

As promised, we now prove that the $\lceil N^{\chi-\epsilon}\rceil$ fittest individuals are always chosen in the limit, formally,

Lemma 2.3.8. *Let $0 < \epsilon < \chi$, $0 < \beta \le 1$, and define the event*

$$A_{N,\epsilon} \;=\; \left\{\{1, \ldots, \lceil N^{\chi-\epsilon}\rceil\} \subseteq I\right\},$$

then there exists $\eta_\epsilon > 0$ such that $\mathbb{P}\left(A_N^c\right) = \mathcal{O}\left(N^{-\eta_\epsilon}\right)$; in particular

$$\lim_{N\to\infty} \log(N)\mathbb{P}\left(A_N^c\right) = 0.$$

Proof. By (2.34) plus Cramer's estimate, and Lemma 2.3.6, we may work on the set

$$E_N = \{x_1 \le N\} \cap \left\{x_{\lceil N^{\chi-\epsilon}\rceil} > N^{-\chi+\frac{\epsilon}{2}}\right\} \cap \left\{\sum_{j=1}^{\lceil N^\gamma\rceil} x_j^\beta \le N^{\gamma(1-\beta)+\beta\epsilon/4}\right\},$$

so that, summing over i in (2.51) we obtain, for some $\eta_\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(A_{N,\epsilon}^c\right) &\le \mathbb{P}\left(E_N^c\right) + \mathbb{P}\left(A_{N,\epsilon}; E_n\right) \\
&\le N^{-\eta_\epsilon} + \mathbb{E}\left[\sum_{i=1}^{\lceil N^{\chi-\epsilon}\rceil} \exp\left\{-N^{1-\gamma(1-\beta)-\beta\epsilon/4} x_i^\beta\right\}; E_N\right] \\
&\le N^{-\eta_\epsilon} + \mathbb{E}\left[\int_{N^{-\chi+\epsilon/2}}^N \exp\left\{-N^{1-\gamma(1-\beta)-\beta\epsilon/4} x^\beta\right\} \Xi(dx)\right]
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{E}\left[\int_{N^{-\chi+\epsilon/2}}^N \exp\left\{-N^{1-\gamma(1-\beta)-\beta\epsilon/4} x^\beta\right\} \Xi(dx)\right] \\
&= \int_{N^{-\chi+\epsilon/2}}^N \exp\left\{-N^{1-\gamma(1-\beta)-\beta\epsilon/4} x^\beta\right\} \frac{dx}{x^2} \\
&\le \exp\left\{-N^{1-\gamma(1-\beta)-\beta\epsilon/4-\beta(\chi-\epsilon/2)}\right\} \int_{N^{-\chi+\epsilon/2}}^N x^{-2}dx \\
&= \exp\left\{-N^{\epsilon\beta/2}\right\} \mathcal{O}\left(N^{\chi-\epsilon/2}\right).
\end{aligned}$$

□

Now we prove that the total fitness weight of the chosen individuals below level $N^{-\chi-\epsilon}$ is negligible in the limit.

Lemma 2.3.9. *For every $\epsilon > 0$,*

$$\mathbb{E}\left[\sum_{k=1}^N x_{I_k} \mathbb{1}_{x_{I_k} \le N^{-(\chi+\epsilon)}}\right] \le \mathcal{O}\left(N^{-\epsilon\beta}\right).$$

In particular, letting

$$B_{N,\epsilon} = \left\{\sum_{k=1}^N x_{I_k} \mathbb{1}_{x_{I_k} < N^{-(\chi+\epsilon)}} < N^{-\epsilon\beta/2}\right\},$$

there exists $\eta_\epsilon > 0$ such that

$$\lim_{N\to\infty} N^{\eta_\epsilon} \mathbb{P}\left(B_{N,\epsilon}^c\right) = 0. \tag{2.52}$$

Proof. When $\beta = 1$ we trivially have

$$\sum_{k=1}^{N} x_{I_k} \mathbb{1}_{x_{I_k} \le N^{-(\chi+\epsilon)}} \le N^{1-1-\epsilon}$$

with probability one. For the case $0 < \beta < 1$ let $c > 1 - \beta$ and let E_N be the set

$$E_N = \left\{ \sum_{k=N}^{\lceil N^\gamma \rceil} x^\beta \ge c^{-1} N^{\gamma(1-\beta)} \right\};$$

then, using Lemma 2.3.5 and the fact that $\mathbb{P}^{\Xi}(k \in I) \le N \left(\frac{x_k^\beta}{\sum_{k=N}^{\lceil N^\gamma \rceil} x_k^\beta} \right)$ for every k,

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{N} x_{I_k} \mathbb{1}_{x_{I_k} \le N^{-(\chi+\epsilon)}}\right] &= \mathbb{E}\left[\sum_{k:\, x_k \le N^{-(\chi+\epsilon)}}^{\lceil N^\gamma \rceil} x_k \mathbb{P}^{\Xi}(k \in I)\right] \\
&\le \mathbb{E}\left[\sum_{k:\, x_k \le N^{-(\chi+\epsilon)}}^{\lceil N^\gamma \rceil} x_k; E_N^c\right] + N\mathbb{E}\left[\frac{\sum_{k:\, x_k < N^{-\chi-\epsilon}}^{\lceil N^\gamma \rceil} x_k^{1+\beta}}{\sum_{k=N}^{\lceil N^\gamma \rceil} x^\beta}; E_N\right] \\
&\le N^{1-\chi-\epsilon}\mathbb{P}\left(E_N^c\right) + \mathcal{O}\left(N^{1-\gamma(1-\beta)}\right) \int_0^{N^{-\chi-\epsilon}} x^{\beta-1} dx \\
&= o\left(N^{-\beta\epsilon}\right) + \mathcal{O}\left(N^{1-\gamma(1-\beta)-\beta(\chi+\epsilon)}\right) \\
&= \mathcal{O}\left(N^{-\beta\epsilon}\right).
\end{aligned}$$

Finally, Markov's inequality easily yields (2.52). □

Proof of Theorem 2.3.2.II), case $(0 < \beta \le 1, 1 \le \gamma < \infty)$*.* The proof consists of checking the conditions of Theorem 2.2.8 with $L_N = \chi \log N$.

Proof of condition iii): Let $\epsilon > 0$ and recall the set $A_{N,\epsilon}$ of Lemma 2.3.8; observe that under this set we have $\sum_{k=1}^{N} x_{\hat{I}_k} \ge \sum_{k=1}^{\lceil N^{\chi-\epsilon} \rceil} x_k$ and $\tilde{\rho}_1^{(N)} = \frac{x_1}{\sum_{k=1}^{N} x_{\hat{I}_k}}$ so that $\tilde{\rho}_i^{(N)} \le \frac{x_i}{\sum_{k=1}^{\lceil N^{\chi-\epsilon} \rceil} x_k}$ for all $2 \le i \le N$. Let $c > 1$ and let E_N be the set

$$E_N = \left\{ \sum_{k=1}^{\lceil N^{\chi-\epsilon} \rceil} x_k \ge c^{-1} \log(N) \right\},$$

then, using Lemmas 2.3.5 and 2.3.8,

$$\begin{aligned}
\mathbb{E}\left[\left(\tilde{\rho}_2^{(N)}\right)^2 + \cdots + \left(\tilde{\rho}_N^{(N)}\right)^2\right] &\le \mathbb{E}\left[\frac{x_2^2 + \cdots + x_N^2}{\left(\sum_{k=1}^{\lceil N^{\chi-\epsilon} \rceil} x_k\right)^2}; E_N\right] + \mathbb{P}\left(A_{N,\epsilon}^c\right) + \mathbb{P}\left(E_N^c\right) \\
&\le c^{5/4}\mathbb{E}\left[\left(\frac{x_2}{\log N}\right)^{5/4}\right] + \mathbb{E}\left[\frac{x_3^2 + \cdots + x_N^2}{\log^2(N)}\right] + o\left(\log(N)^{-1}\right) \\
&= o\left(\log(N)^{-1}\right)
\end{aligned}$$

where for the second inequality we have also used that $x^2 < x^{5/4}$ whenever $0 < x < 1$ and, for the latter equality, we have used that $\mathbb{E}\left[x_2^{5/4}\right] = \int_0^\infty x^{-1-1/4} x e^{-x} dx < \infty$ and $\sum_{k=3}^\infty \mathbb{E}\left[x_k^2\right] = \sum_{k=3}^\infty \frac{\Gamma(k-2)}{\Gamma(k)} = \sum_{k=3}^\infty \frac{1}{(k-1)(k-2)} < \infty$. □

Proof of conditions iv) and ii): Working on the sets $A_{N,\epsilon}$ and $B_{N,\epsilon}$ of Lemmas 2.3.8 and 2.3.9 respectively, we have

$$\frac{x_1}{\sum_{k=1}^{\lceil N^{\chi+\epsilon}\rceil} x_k + N^{-\epsilon\beta}} \leq \tilde{\rho}_1^{(N)} \leq \frac{x_1}{\sum_{k=1}^{\lceil N^{\chi+\epsilon}\rceil} x_k}. \tag{2.53}$$

Then, letting E_N be as above,

$$\begin{aligned}\mathbb{P}\left(\tilde{\rho}_1^{(N)} > x\right) &\leq \mathbb{P}\left(x_1 > x \sum_{k=1}^{\lceil N^{\chi+\epsilon}\rceil} x_k\right) \\ &\leq \mathbb{P}\left(x_1 > xc^{-1}\log(N)\right) + \mathbb{P}\left(E_N^c\right) \\ &= 1 - \exp\left\{\frac{c}{x\log(N)}\right\} + o\left(\log(N)^{-1}\right) \\ &= o\left(\log(N)^{-1}\right)\end{aligned}$$

which proves *iv*). In view of *iv*) we may prove *ii'*) instead of *ii*) by Lemma 2.2.9. In turn, by *iii*) and (2.53), *ii'*) follows directly from Lemma 2.3.10 below. □

Lemma 2.3.10 (Equation (29) in Brunet and Derrida 2012)**.** *Let a_N be any sequence of positive numbers such that $a_N \to 0$. Then*

$$\lim_{N\to\infty} \log(N)\mathbb{E}\left[\left(\frac{x_1}{\sum_{k=1}^N x_k + a_N}\right)^b\right] = \frac{1}{b-1} \equiv \int_0^1 bx^{b-1}\frac{1-x}{x}dx.$$

Proof. We first provide an outline of the proof for the case $a_N \equiv 0$, i.e. that

$$\log(N)\ \mathbb{E}\left[\sum_{j=1}^N \left(\frac{x_j}{\sum_{k=1}^N x_k}\right)^b\right] = \frac{1}{b-1} + \mathcal{O}\left(\frac{1}{\log N}\right), \tag{2.54}$$

as $N \to \infty$, by repeating the steps in Section III of Brunet, Derrida, et al. 2007. We give yet another construction of the random (this time unordered) set of points $\{x_1, \ldots, x_N\}$. By the Mapping Theorem (Kingman 1992) the random vector $(x_1, \ldots, x_N)$ is equal in distribution to $\left(e^{\xi_1}, \ldots, e^{\xi_N}\right)$ where $(\xi_k)_{k\geq 1}$ are the points of a Poisson point process with intensity measure $e^{-s}ds$ arranged in decreasing order. In turn, it can be seen using the properties of Poisson point processes that, conditional on the value of ξ_{N+1}, the unordered collection of points $\left\{e^{\xi_1}, \ldots, e^{\xi_N}\right\}$ is equal in distribution to a set $\{\xi_1, \ldots, \xi_N\}$ of i.i.d. r.vs. with density

$$\mathbb{1}_{s>\xi_{N+1}} \frac{e^{-s}ds}{\int_{\xi_{N+1}}^\infty e^{-s}ds} = \mathbb{1}_{s>\xi_{N+1}} e^{-(s-\xi_{N+1})}ds.$$

The latter are in turn equal in distribution to $\{\xi_{N+1}+\mathcal{E}_1,\cdots,\xi_{N+1}+\mathcal{E}_N\}$ where the $\mathcal{E}_k$'s are i.i.d. standard exponential random variables independent of ξ_{N+1}. Since $e^{\xi_{N+1}}\overset{d}{=}x_{N+1}$ we obtain, for every N,

$$\{x_k\}_{k=1}^N \overset{d}{=} \left\{e^{\xi_{\sigma(k)}}\right\}_{k=1}^N \overset{d}{=} \left\{x_{N+1}e^{\mathcal{E}_k}\right\}_{k=1}^N. \tag{2.55}$$

where $x_{N+1},\mathcal{E}_1,\ldots,\mathcal{E}_N$ are all independent. Thus the expectation in (2.54) can be expressed as

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^N\left(\frac{x_j}{\sum_{k=1}^N x_k}\right)^b\right] &= \mathbb{E}\left[\sum_{j=1}^N\left(\frac{e^{\mathcal{E}_j}}{\sum_{k=1}^N e^{\mathcal{E}_k}}\right)^b\right]\\ &= N\mathbb{E}\left[\left(\frac{e^{\mathcal{E}_1}}{\sum_{k=1}^N e^{\mathcal{E}_k}}\right)^b\right]\\ &= N\int_0^\infty dy_1\cdots\int_0^\infty dy_N e^{-y_1-\cdots-y_N}\frac{e^{by_1}}{\left(\sum_{k=1}^N e^{y_k}\right)^b}. \qquad (2.56)\end{aligned}$$

Using the identity

$$(K)^{-b}=\frac{1}{\Gamma(b)}\int_0^\infty ds \quad s^{b-1}e^{-sK},\quad K>0,\quad b>0,$$

with $K=\sum_{k=1}^N e^{y_k}$, and Tonelli's theorem, the integrals in (2.56) become

$$\begin{aligned}&\frac{N}{\Gamma(b)}\int_0^\infty ds\int_0^\infty dy_1\cdots\int_0^\infty dy_N e^{-y_1-\cdots-y_N}e^{by_1}s^{b-1}\exp\left\{-\sum_{k=1} se^{y_k}\right\}\\ &=\frac{N}{\Gamma(b)}\int_0^\infty ds \quad s^{b-1}I_b(s)\left(I_0(s)\right)^{N-1}, \qquad (2.57)\end{aligned}$$

where, for $p\geq 0$, the function $I_p(s)$ is defined as

$$I_p(s):=\int_0^\infty dy \quad e^{(p-1)y-e^{-sy}}=s^{1-p}\int_s^\infty du \quad u^{p-2}e^{-u}.$$

Observe that, for any p, the function $I_p(s)$ is decreasing on s and, furthermore, $\lim_{s\to\infty}s^\beta I_p(s)=0,\forall\beta\in\mathbb{R}$. Now, for large N, the integral in (2.57) is dominated by values of s near 0, expanding I_p correspondingly (see eq. (21) in Brunet and Derrida 2012) we have, as $s\to 0$,

$$I_0(s)=1+s(\log(s)+\Upsilon-1)+\mathcal{O}\left(s^2\right),\text{ and } I_{p\geq 2}(s)=\frac{(p-2)!}{s^{p-1}}+\mathcal{O}\left(s^{2-p}\right),$$

where we recall that Υ is the Euler-Mascheroni constant. It follows that for values of s that are of order $1/N\log(N)$, and as $N\to\infty$, making the change of variable $s=\mu/N\log(N)$, we obtain

$$I_0(s)^N=e^{-\mu}\left(1+\mu\frac{\log(\mu)-\log\log N+\Upsilon-1}{\log N}+\mathcal{O}\left(\frac{\mu\log\mu}{\log N}\right)^2\right).$$

Thus making $s = \mu/N\log(N)$ in 2.57 and plugging the above estimates in (2.56), we obtain

$$\mathbb{E}\left[\sum_{j=1}^{N}\left(\frac{x_j}{\sum_{k=1}^{N} x_k}\right)^b\right] = \frac{N}{\Gamma(b)}\int_0^\infty ds \quad s^{b-1} I_b(s)\left(I_0(s)\right)^{N-1}$$
$$= \frac{N}{N\log(N)}\frac{(b-2)!}{\Gamma(b)} + \mathcal{O}\left(\frac{N}{N\log^2(N)}\right) = \frac{1}{\log(N)(b-1)} + \mathcal{O}\left(\frac{1}{\log^2(N)}\right),$$

thus proving (2.54).

Finally, for the extended case $a_N > 0$, observe that the condition $a_N > 0$ and (2.54) easily give

$$\lim_{N\to\infty}\log(N)\mathbb{E}\left[\sum_{j=1}^{N}\left(\frac{x_j}{a_N + \sum_{k=1}^{N} x_k}\right)^b\right] \leq \frac{1}{b-1}.$$

Now, for the reverse inequality, we first write

$$\mathbb{E}\left[\sum_{j=1}^{N}\left(\frac{x_j}{a_N + \sum_{k=1}^{N} x_k}\right)^b\right] = \mathbb{E}\left[\left(\frac{1}{\frac{a_N}{\sum_{k=1}^{N} x_k} + 1}\right)^b \sum_{j=1}^{N}\left(\frac{x_j}{\sum_{k=1}^{N} x_k}\right)^b\right].$$

Second, by Markov's inequality followed by Jensen's inequality,

$$\mathbb{P}\left(\frac{a_N}{\sum_{k=1}^{N} x_k} > \epsilon\right) \leq \frac{1}{\epsilon}a_N\mathbb{E}\left[\frac{1}{\sum_{k=1}^{N} x_k}\right] = \frac{1}{\epsilon}\frac{a_N}{\sum_{k=1}^{N} k^{-1}}\mathbb{E}\left[\frac{\sum_{k=1}^{N} k^{-1}}{\sum_{k=1}^{N} x_k k k^{-1}}\right]$$
$$\leq o\left(\log(N)^{-1}\right)\mathbb{E}\left[\frac{\sum_{k=1}^{N} x_k^{-1} k^{-2}}{\sum_{k=1}^{N} k^{-1}}\right] = o\left(\log(N)^{-1}\right)\frac{\sum_{k=1}^{N} k k^{-2}}{\sum_{k=1}^{N} k^{-1}},$$

where we have used (2.34) to compute $\mathbb{E}\left[x_k^{-1}\right] = k$. Combining the two expressions together with 2.54 we obtain, for every $\epsilon > 0$,

$$\lim_{N\to\infty}\log(N)\mathbb{E}\left[\left(\frac{1}{\frac{a_N}{\sum_{k=1}^{N} x_k} + 1}\right)^b \sum_{j=1}^{N}\left(\frac{x_j}{\sum_{k=1}^{N} x_k}\right)^b\right]$$
$$\geq \lim_{N\to\infty}\log(N)\mathbb{E}\left[\frac{1}{\epsilon+1}\sum_{j=1}^{N}\left(\frac{x_j}{\sum_{k=1}^{N} x_k}\right)^b\right] + \lim_{N\to\infty}\log(N)\mathbb{P}\left(\frac{a_N}{\sum_{k=1}^{N} x_k} > \epsilon\right)$$
$$\geq \frac{1}{\epsilon+1}\frac{1}{b-1}.$$

The proof is finished by taking $\epsilon \to 0$. □

Proof of Theorem 2.3.2 II, case $(\beta > 1, \gamma = \infty)$

By equation (2.34) and Proposition 10 in Pitman and Yor 1997 the random mass partition $\left(\frac{x_1^\beta}{\sum_{k=1}^{\infty} x_k^\beta}\right)$ has Poisson-Dirichlet $(\frac{1}{\beta}, 0)$ distribution. Thus,

$$\left(\frac{x_{I_1}^\beta}{\sum_{k=1}^{\infty} x_k^\beta}, \ldots, \frac{x_{I_N}^\beta}{\sum_{k=1}^{\infty} x_k^\beta}\right) \stackrel{d}{=} (s_1, \ldots, s_N)$$

where $(s_1,\ldots,s_N)$ is a Poisson-Dirichlet$(\frac{1}{\beta},0)$ size-biased pick of size N. The result then follows by an application of Theorem 2.2.11.

2.3.2 Speed of Selection

Recall from the proof of Theorem 2.3.1 that

$$\left(X_1^N(t),\ldots,X_N^N(t)\right) \stackrel{d}{=} (X_{eq}(t-1)+\log(x_{I_1}),\ldots,X_{eq}(t-1)+\log(x_{I_N})) \tag{2.58}$$

where $X_{eq}^N(t)=\log\left(\sum_{i=1}^N e^{X_j^N(t)}\right)$. This observation allows for an easy definition, and computation, of the speed of selection $\nu_N \equiv \nu_{N,\beta,\gamma}$ as

$$\nu_N := \lim_{t\to\infty}\max_{1\le j\le N}\frac{X_j^N(t)}{t} = \lim_{t\to\infty}\min_{1\le j\le N}\frac{X_j^N(t)}{t}. \tag{2.59}$$

Indeed, we have the following.

Lemma 2.3.11 (Lemma 1.5 in Cortines and Mallein 2017)**.** *With probability one,*

$$\lim_{t\to\infty}\max_{1\le j\le N}\frac{X_j^N(t)}{t} = \mathbb{E}\left[\log\left(\sum_{i=1}^N x_{I_i}\right)\right] = \lim_{t\to\infty}\min_{1\le j\le N}\frac{X_j^N(t)}{t}.$$

Thus ν_N is well defined and, furthermore, $\nu_N=\mathbb{E}\left[\log\left(\sum_{i=1}^N x_{I_i}\right)\right]$.

Proof. First, from (2.58) and the definition of $X_{eq}^N(t)$ we obtain $X_{eq}^N(t)-X_{eq}^N(t-1) \stackrel{d}{=} \log\left(\sum_{i=1}^N x_{I_i}\right)$. Thus, the branching steps being i.i.d., and by the law of large numbers, we have

$$\lim_{t\to\infty}\frac{X_{eq}^N(t)}{t} \stackrel{a.s.}{=} \mathbb{E}\left[\log\left(\sum_{i=1}^N x_{I_i}\right)\right]. \tag{2.60}$$

Second, note that $\left|\max_{1\le j\le N}X_j^N(t)-X_{eq}^N(t-1)\right|$ is upperbounded by $|\log(x_1)|+\left|\log(x_{\lceil N^\gamma\rceil})\right|$ which has finite expectation. Thus, by dominated convergence,

$$\mathbb{E}\left[\limsup_{t\to\infty}\frac{\left|\max_{1\le j\le N}X_j^N(t)-X_{eq}^N(t-1)\right|}{t}\right] = \limsup_{t\to\infty}\frac{\mathbb{E}\left[\left|\max_{1\le j\le N}X_j^N(t)-X_{eq}^N(t-1)\right|\right]}{t} = 0.$$

This, together with (2.60), imply

$$\begin{aligned}\lim_{t\to\infty}\frac{\max_{1\le j\le N}X_j^N(t)}{t} &= \lim_{t\to\infty}\frac{X_{eq}^N(t-1)}{t}+\frac{\max_{1\le j\le N}X_j^N(t)-X_{eq}^N(t-1)}{t}\\ &= \mathbb{E}\left[\log\left(\sum_{i=1}^N x_{I_i}\right)\right].\end{aligned}$$

Repeating the argument but this time replacing $\max_{1\le j\le N}X_j^N(t)$ with $\min_{1\le j\le N}X_j^N(t)$, we conclude the proof of the lemma. □

Computing $\nu_N = \mathbb{E}\left[\log\left(\sum_{i=1}^N x_{I_i}\right)\right]$ in the above Lemma 2.3.11 we obtain the following theorem.

Theorem 2.3.12. *For ν_N as above, and as $N \to \infty$,*
I) Weak Selection Regime: *If $(\beta < 1, \alpha > 0)$, then*

$$\nu_N = -\alpha \log(N) + \mathbb{E}\left[\log(Y_\beta)\right] + o(1) \tag{2.61}$$

where Y_β is the positive $(1-\beta)$-stable law with Laplace transform $\mathbb{E}\left[e^{-\theta Y_\beta}\right] = \exp\{-\Gamma(\beta)\theta^{1-\beta}\}$.
II) Strong Selection Regime: *If $(\beta < 1, \alpha < 0), (\beta = 1, \gamma < \infty)$, or $(\beta > 1, \gamma \leq \infty)$, then*

$$\nu_N = \log\left(\chi \log N\right) + o(1). \tag{2.62}$$

Proof of Theorem 2.3.12.I)

In this section we prove

$$\nu_N = -\alpha \log(N) + \mathbb{E}\left[\log\left(Y_\beta\right)\right] + o(1) \tag{2.63}$$

as $N \to \infty$ where, recalling the Poisson random measure Z with intensity measure $(1-\beta)z^{\beta-2}dz$ of Section 2.3.1, the r.v.

$$Y_\beta := \int_0^\infty zZ(dz)$$

is the positive $(1-\beta)$-stable r.v. with Laplace transform $\mathbb{E}\left[e^{-\theta Y_\beta}\right] = \exp\left\{-\Gamma(\beta)\theta^{1-\beta}\right\}$ (see (12) in Pitman and Yor 1997).

In the upcoming proof of Theorem 2.3.12.I) we will simplify the computations with the help of the following lemma, whose proof we postpone until the end of the section.

Lemma 2.3.13. *Let $(E_N)_{N\geq 1}$ be a sequence of events such that, for some $\delta > 0$,*

$$\lim_{N\to\infty} N^\delta \mathbb{P}(E_N) \;=\; 0.$$

Then, for every $\eta > 0$,

$$\lim_{N\to\infty} \mathbb{E}\left[\left|\log\left(\sum_{k=1}^N x_{I_k}\right)\right|^\eta ; E_N\right] \;=\; 0.$$

Assuming the above Lemma 2.3.13, we first give the main proof of this section.

Proof of Theorem 2.3.12.I). We first write

$$\mathbb{E}\left[\log\left(\sum_{k=1}^N x_{I_k}\right)\right] = -\alpha\log(N) + \mathbb{E}\left[\log\left(\sum_{k=1}^N N^\alpha x_{I_k^N}\right)\right],$$

then from equation (2.46) and the Continuous Mapping Theorem we have

$$\log\left(\sum_{k=1}^N N^\alpha x_{I_k^N}\right) \Rightarrow \log\left(Y_\beta\right),$$

so it only remains to prove that the collection of r.vs. $(\log\left(\sum_{k=1}^N N^\alpha x_{I_k^N}\right))_{N\geq 1}$ is uniformly integrable, which would imply the corresponding convergence in expectation (Theorem 3.5 in Billingsley 1999). The latter boils down to proving that, for some $\eta > 1$, we have

$$\limsup_{N\to\infty} \mathbb{E}\left[\left|\log\left(N^\alpha \sum_{k=1}^N x_{I_k^N}\right)\right|^\eta; N^\alpha \sum_{k=1}^N x_{I_k^N} < 1\right] < \infty \tag{2.64}$$

and

$$\limsup_{N\to\infty} \mathbb{E}\left[\left|\log\left(N^\alpha \sum_{k=1}^N x_{I_k^N}\right)\right|^\eta; N^\alpha \sum_{k=1}^N x_{I_k^N} > 1\right] < \infty. \tag{2.65}$$

See e.g. (3.18) in Billingsley 1999.

Proof of (2.64)**:** We begin by computing, for an arbitrary subset E_N to be specified further ahead,

$$\begin{aligned}
&\mathbb{E}\left(\left|\log\left(N^\alpha \sum_{k=1}^N x_{I_k}\right)\right|^\eta \mathbb{1}_{\sum_{k=1}^N N^\alpha x_{I_k} \leq 1}; E_N^c\right)\\
&\leq \mathbb{E}\left(\left(-\log\left(\max_{1\leq k\leq N} N^\alpha x_{I_k}\right)\right)^\eta, \max_{1\leq k\leq N} N^\alpha x_{I_k} \leq 1, E_N^c\right)\\
&= \int_0^\infty \mathbb{P}\left(-\log\left(\max_{1\leq k\leq N} N^\alpha x_{I_k}\right) \geq u^{\frac{1}{\eta}}, \max_{1\leq k\leq N} N^\alpha x_{I_k} \leq 1, E_N^c\right) du\\
&\leq \int_0^\infty \mathbb{P}\left(\max_{1\leq k\leq N} N^\alpha x_{I_k} \leq \exp(-u^{\frac{1}{\eta}}), E_N^c\right) du\\
&\leq \int_0^\infty \mathbb{P}\left(\max_{1\leq k\leq N} N^\alpha x_{J_k} \leq \exp(-u^{\frac{1}{\eta}}), E_N^c\right) du\\
&= \int_0^\infty \mathbb{E}\left[\left(1 - \frac{\sum_{i: x_i \geq \frac{1}{N^\alpha}\exp(-u^{\frac{1}{\eta}})} x_i^\beta}{\sum_{i=1}^{N^\gamma} x_i^\beta}\right)^N, E_N^c\right] du\\
&\leq \int_0^\infty \mathbb{E}\left[\exp\left(-N\frac{\sum_{i: x_i \geq \frac{1}{N^\alpha}\exp(-u^{\frac{1}{\eta}})} x_i^\beta}{\sum_{i=1}^{N^\gamma} x_i^\beta}\right), E_N^c\right] du
\end{aligned} \tag{2.66}$$

where for the second inequality we have used the coupling (I, J) for the selection step (conditional on the x_k's) which satisfies

$$\max_{1\leq k\leq N} N^\alpha x_{J_k} \leq \max_{1\leq k\leq N} N^\alpha x_{I_k}.$$

Now to deal with the expectation appearing in (2.66) let $2c \in (0, 1-\beta)$ and set

$$E_N = \left\{\sum_{i=1}^{\lceil N^\gamma \rceil} x_i^\beta \leq c^{-1} N^{\gamma(1-\beta)}\right\}.$$

By Lemma 2.3.6 together with Lemma 2.3.13 we have

$$\begin{aligned}&\limsup_{N\to\infty}\ \mathbb{E}\left[\left|\log\left(N^\alpha\sum_{k=1}^N x_{I_k^N}\right)\right|^\eta; N^\alpha\sum_{k=1}^N x_{I_k^N}<1\right]\\&=\limsup_{N\to\infty}\ \mathbb{E}\left[\left|\log\left(N^\alpha\sum_{k=1}^N x_{I_k^N}\right)\right|^\eta; N^\alpha\sum_{k=1}^N x_{I_k^N}<1; E_N\right].\end{aligned}$$

Further, using (2.66), we have

$$\begin{aligned}&\mathbb{E}\left[\left|\log\left(N^\alpha\sum_{k=1}^N x_{I_k}\right)\right|^\eta; \sum_{k=1}^N\ N^\alpha x_{I_k}\leq 1; E_n\right]\\&\leq\int_0^\infty \mathbb{E}\left[\exp(-cN\frac{\sum_{i:x_i\geq\frac{1}{N^\alpha}\exp(-u^{\frac{1}{\eta}})}x_i^\beta}{N^{\gamma(1-\beta)}})\right]du\\&=\int_0^\infty \mathbb{E}\left[\exp\left(-cN^{\chi\beta}\sum_{i:x_i\geq\frac{1}{N^\alpha}\exp(-u^{\frac{1}{\eta}})}x_i^\beta\right)\right]du.\end{aligned}$$

Define $\bar{c}=\min_{\mathbb{R}+}\frac{1}{x^\beta}(1-\exp(-cx^\beta))$. By Campbell's formula, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(-cN^{\chi\beta}\sum_{i:x_i\geq\frac{1}{N^\alpha}\exp(-u^{\frac{1}{\eta}})}x_i^\beta\right)\right]&=\exp\left(\int_{\frac{\exp\left(-u^{\frac{1}{\eta}}\right)}{N^\alpha}}^\infty e^{-c(N^\chi x)^\beta}-1\frac{dx}{x^2}\right)\\&=\exp\left(N^\chi\int_{\frac{N^\chi}{N^\alpha}\exp(-u^{\frac{1}{\eta}})}^\infty e^{-cx^\beta}-1\frac{dx}{x^2}\right)\\&\leq\exp\left(-\bar{c}N^\chi\int_{\frac{N^\chi}{N^\alpha}\exp(-u^{\frac{1}{\eta}})}^\infty\frac{dx}{x^{2-\beta}}\right)\\&=\exp\left(-\frac{\bar{c}}{1-\beta}\exp\left((1-\beta)u^{\frac{1}{\eta}}\right)\right).\end{aligned}$$

Finally, since

$$\int_0^\infty\exp\left\{-\frac{\bar{c}}{1-\beta}\exp\left((1-\beta)u^{\frac{1}{\eta}}\right)\right\}du<\infty,$$

it follows that

$$\limsup_{N\to\infty}\ \mathbb{E}\left(\left|\log\left(N^\alpha\sum_{k=1}^N x_{I_k}\right)\right|^\eta, \sum_{k=1}^N\ N^\alpha x_{I_k}\leq 1; E_n\right)\ <\ \infty.$$

Proof of (2.65)**:** Let $\delta\in(0,1-\beta)$ and pick $c>0$ to ensure that $\log(b)^\eta<cb^\delta$ for every $b>1$. A direct application of the Mean Value Theorem and our choice of c shows the existence of $\tilde{c}>0$ such that

$$\forall x>0, b\geq 1,\quad \log^\eta(b+x)\leq cb^\delta+\tilde{c}x.$$

Let $K > 1-\beta$ and let E_N be the set

$$E_N := \left\{ \sum_{k=N}^{\lceil N^\gamma \rceil} x_k^\beta \geq K^{-1} N^{\gamma(1-\beta)} \right\}.$$

By the large deviation Lemma 2.3.5 we may apply Lemma 2.3.13, so that it is enough to prove (2.65) restricted on the set E_N. On the set $\{N^\alpha \sum_{k=1}^N x_{I_k} > 1\}$, we have

$$\log\left(N^\alpha \sum_{k=1}^N x_{I_k}\right)^\eta = \mathbb{1}\left\{\max_{1\leq k\leq N} N^\alpha x_{I_k} > 1\right\} \log^\eta\left(N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k} > N^{-\alpha}} + N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k}\leq N^{-\alpha}}\right) +$$
$$\mathbb{1}\left\{\max_{1\leq k\leq N} N^\alpha x_{I_k} \leq 1\right\} \log^\eta\left(N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k}\leq N^{-\alpha}}\right)$$
$$\leq c\left(N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k} > N^{-\alpha}}\right)^\delta + (c+\tilde{c}) N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k}\leq N^{-\alpha}}.$$

It remains to control the expected value of the RHS of the latter inequality on the set E_N. On the one hand,

$$\begin{aligned}
\mathbb{E}\left[N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k}\leq N^{-\alpha}}; E_N\right] =& N^\alpha \sum_{k=1}^N \mathbb{E}\left[x_{I_k}; x_{I_k} \leq N^{-\alpha}; E_N\right] \\
\leq& N^\alpha N \mathbb{E}\left[\frac{\sum_{k\leq[N^\gamma]:\, x_k\leq N^{-\alpha}} x_k^{\beta+1}}{\sum_{k=N}^{\lceil N^\gamma\rceil} x_k^\beta}; E_N\right] \\
\leq& K N^{\alpha+1}\mathbb{E}\left[\frac{\sum_{k:\, x_k\leq N^{-\alpha}} x_k^{\beta+1}}{N^{\gamma(1-\beta)}}\right] \\
\leq& K N^{\alpha+1-\gamma(1-\beta)} \int_0^{\frac{1}{N^\alpha}} \frac{dx}{x^{1-\beta}} \\
=& \mathcal{O}(1).
\end{aligned}$$

On the other hand, by the $\mathtt{l}_{\delta^{-1}}$ triangle inequality (recall $\delta < 1-\beta < 1$) we have

$$\left(\sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k} > aN^{-\alpha}}\right)^\delta \leq \sum_{k=1}^N x_{I_k}^\delta \mathbb{1}_{x_{I_k} > aN^{-\alpha}}$$

so that

$$\begin{aligned}
\mathbb{E}\left[\left(N^\alpha \sum_{k=1}^N x_{I_k}\mathbb{1}_{x_{I_k} > N^{-\alpha}}\right)^\delta; E_N\right] \leq& N^{\alpha\delta} \sum_{k=1}^N \mathbb{E}\left[x_{I_k}^\delta \mathbb{1}_{x_{I_k} > aN^{-\alpha}}; E_N\right] \\
\leq& N^{\alpha\delta} \sum_{k=1}^N \mathbb{E}\left[\frac{\int_{aN^{-\alpha}}^\infty x^{\delta+\beta}\Xi(dx)}{\sum_{k=N}^{\lceil N^\gamma\rceil} x_k^\beta}; E_N\right] \\
\leq& (1-\beta) K N^{\alpha\delta} N^{1-\gamma(1-\beta)} \int_{aN^{-\alpha}}^\infty x^{\delta+\beta}\frac{dx}{x^2} \\
=& \mathcal{O}(1).
\end{aligned}$$

□

Proof of Lemma 2.3.13. Let $\eta > 0$ be fixed. Since $Nx_{\lceil N^\gamma \rceil} \le \sum_{k=1}^N x_{I_k} \le Nx_1$ we have

$$\mathbb{E}\left[\left|\log\left(\sum_{k=1}^N x_{I_k}\right)\right|^\eta ; E_N\right] \le \mathbb{E}\left[\left(\max\{\log(Nx_1), -\log(Nx_{\lceil N^\gamma\rceil}), 0\}\right)^\eta ; E_N\right]$$
$$\le \mathbb{E}\left[\log^\eta(Nx_1); Nx_1 > 1, E_N\right] + \mathbb{E}\left[\log^\eta(x_{\lceil N^\gamma\rceil}^{-1}/N); x_{\lceil N^\gamma\rceil}^{-1}/N > 1, E_N\right].$$

On the one hand, for the first term above, by Hölder's inequality and using that $\forall \epsilon > 0, \exists c_\epsilon > 0$ such that $\log^\eta(x) \le c_\epsilon x^\epsilon, x > 1$, we obtain

$$\mathbb{E}\left[\log^\eta(Nx_1); Nx_1 > 1, E_N\right] \le \mathbb{P}(E_N)^{1/2} \left\|\log^\eta(Nx_1); Nx_1 > 1\right\|_{\mathsf{L}^2}$$
$$\le \mathbb{P}(E_N)^{1/2} c_\epsilon N^\epsilon \left\|x_1^\epsilon\right\|_{\mathsf{L}^2}.$$

By the hypothesis, the latter converges to zero whenever $0 < \epsilon < (1/2) \wedge (2\delta)$ which ensures $\|x_1^\epsilon\|_{\mathsf{L}^2} < \infty$ and $\epsilon < 2\delta$. Similarly, for the second term,

$$\mathbb{E}\left[\log^\eta(x_{\lceil N^\gamma\rceil}^{-1}/N); x_{\lceil N^\gamma\rceil}^{-1}/N > 1, E_N\right] \le \mathbb{P}\left(E_N\right)^{1/2} \left\|\log^\eta(x_{\lceil N^\gamma\rceil}^{-1}/N); x_{\lceil N^\gamma\rceil}^{-1}/N > 1\right\|_{\mathsf{L}^2}$$
$$\le \mathbb{P}\left(E_N\right)^{1/2} c_\epsilon N^{-\epsilon} \left\|x_{\lceil N^\gamma\rceil}^{-\epsilon}\right\|_{\mathsf{L}^2}$$

where, by equation (2.34) plus Stirling's approximation, $\left\|x_{\lceil N^\gamma\rceil}^{-\epsilon}\right\|_{\mathsf{L}^2}^2 = \frac{\Gamma(\lceil N^\gamma\rceil + 2\epsilon)}{\Gamma(\lceil N^\gamma\rceil)} \sim N^{\gamma\epsilon}$. Thus, if $0 < \epsilon(\gamma - 1) < 2\delta$, a second application of the hypothesis yields $\lim_{N\to\infty} \mathbb{E}\left[\log^\eta(x_{\lceil N^\gamma\rceil}^{-1}/N); x_{\lceil N^\gamma\rceil}^{-1}/N > 1, E_N\right] = 0$, which finishes the proof of the lemma. □

Proof of Theorem 2.3.12.II)

Proof of Theorem 2.3.12.II). Recall the sets $A_{N,\epsilon}$ and $B_{N,\epsilon}$ of Propositions 2.3.8 and 2.3.9, so that by these propositions together with Lemma 2.3.13 we have, for every $\epsilon > 0$, $\mathbb{E}\left[\log\left(\sum_{k=1}^N x_{I_k}\right)\right] = \mathbb{E}\left[\log\left(\sum_{k=1}^N x_{I_k}\right); A_{N,\epsilon}, B_{N,\epsilon}\right] + o(1)$ as $N \to \infty$ and, hence,

(2.67)
$$\lim_{N\to\infty}\left|\log\log(N) - \mathbb{E}\left[\log\left(\sum_{k=1}^N x_{I_k}\right)\right]\right| = \lim_{N\to\infty}\left|\log\log(N) - \mathbb{E}\left[\log\left(\sum_{k=1}^N x_{I_k}\right); A_{N,\epsilon}, B_{N,\epsilon}\right]\right|$$

for all $\epsilon > 0$. Observe that on the event $A_{N,\epsilon}, B_{N,\epsilon}$ we have

$$\log\left(\sum_{k=1}^{\lceil N^{\chi-\epsilon}\rceil} x_k\right) \le \log\left(\sum_{k=1}^N x_{I_k}\right) \le \log\left(\sum_{k=1}^{\lceil N^{\chi+\epsilon}\rceil} x_k + N^{-\epsilon\beta/2}\right). \tag{2.68}$$

Also note that

$$\mathbb{E}\left[\log\left(\sum_{k=1}^N x_k\right)\right] = \log\left(\sum_{k=1}^N k^{-1}\right) + \mathbb{E}\left[\log\left(\frac{\sum_{k=1}^N x_k}{\sum_{k=1}^N k^{-1}}\right)\right] = \log\log N + o(1); \tag{2.69}$$

indeed, on the one hand, by Jensen's inequality

$$\mathbb{E}\left[\log\left(\frac{\sum_{k=1}^{N} x_k}{\sum_{k=1}^{N} k^{-1}}\right)\right] \le \log\left(\frac{\sum_{k=2}^{N} \mathbb{E}\left[x_k\right]}{\sum_{k=1}^{N} k^{-1}}\right) \overset{N\to\infty}{\to} 0,$$

where, for the last limit, we have used that

$$\mathbb{E}\left[x_k\right] = \frac{\Gamma(k-1)}{\Gamma(k)} = \frac{1}{k-1}.$$

On the other hand, by a second and third applications of Jensen's inequality,

$$\begin{aligned}-\mathbb{E}\left[\log\left(\frac{\sum_{k=1}^{N} x_k}{\sum_{k=1}^{N} k^{-1}}\right)\right] &= \mathbb{E}\left[\log\left(\frac{\sum_{k=1}^{N} k^{-1}}{\sum_{k=1}^{N} x_k}\right)\right] \le \log\left(\mathbb{E}\left[\frac{\sum_{k=1}^{N} k^{-1}}{\sum_{k=1}^{N} x_k}\right]\right) \\ &= \log\left(\mathbb{E}\left[\frac{\sum_{k=1}^{N} k^{-1}}{\sum_{k=1}^{N} x_k k k^{-1}}\right]\right) \le \log\left(\mathbb{E}\left[\frac{\sum_{k=1}^{N} x_k^{-1} k^{-2}}{\sum_{k=1}^{N} k^{-1}}\right]\right) = \log(1),\end{aligned}$$

where we have used $\mathbb{E}\left[x_k^{-1}\right] = k$ for the last equality. Thus, taking expectations in (2.68) and plugging in (2.69) we obtain, for every $\epsilon > 0$,

$$\begin{aligned}\log(\chi-\epsilon) + \log\log(N) + o(1) &\le \mathbb{E}\left[\log\left(\sum_{k=1}^{N} x_{I_k}\right); A_{N,\epsilon}, B_{N,\epsilon}\right] \\ &\le \log(\chi+\epsilon) + \log\log(N) + o(1);\end{aligned}$$

which, together with (2.67), imply

$$\log(\chi-\epsilon) \le \lim_{N\to\infty}\left|\log\log(N) - \mathbb{E}\left[\log\left(\sum_{k=1}^{N} x_{I_k}\right)\right]\right| \le \log(\chi+\epsilon).$$

The proof is finished by taking $\epsilon \to 0$. □

2.4 A Model in Continuous Time

In this section we describe a type of Moran model with mutation and selection which, in contrast to the exponential models of Section 2.3, evolves in continuous time, and the fitness of its individuals is not given by positions in the $\mathbb{R}$ continuum but are instead a function of the (discrete) number of mutations acquired by each individual. The model was rigorously studied by Schweinsberg 2017a,b for the first time, although the main heuristic arguments can be traced back to at least Desai and Fisher 2007; Desai, Walczak, and Fisher 2013. Assuming that the strength of selection is much larger than that of mutation, Schweinsberg 2017a,b shows that the genealogy of this model is once again described by the Bolthausen-Sznitman coalescent, as is the case for the strong-selection regime (Theorem 2.3.2) of the exponential models. For the speed of selection, however, the authors show that it is (approximately) of order $o(1)$ as $N \to \infty$, which contrasts with the order $\mathcal{O}(\log\log N)$ for the speed of selection in the strong-selection regime of Theorem 2.3.12 for the exponential models.

Formally, the model consists of a population of fixed size N evolving in continuous time under the effect of mutation and natural selection. All mutations are beneficial and their effect in fitness is measured by a parameter $s_N > 0$. Every individual acquires mutations according to an independent Poisson point process of intensity μ_N, and the fitness of an individual with j mutations is given by

$$\max\{0, 1 + s_N (j - M(t))\},$$

where $M(t)$ is the empirical mean number of mutations at time t. Letting $X_j(t), j \in \mathbb{N}$, be the number of individuals at time t that have exactly j mutations, we may write $M(t)$ as

$$M(t) = \frac{1}{N}\sum_{j=1}^{\infty} jX_j(t).$$

Individuals die at rate one and are replaced by a copy of one individual in the population chosen with probability proportional to its fitness.

Two important quantities for the study of this model are

$$k_N := \frac{\log N}{\log(s_N/\mu_N)}, \text{ and } a_N := \frac{\log(s_N/\mu_N)}{s_N}.$$

The quantity k_N gives the natural scale for the number of mutations; it turns out that if

$$M^*(t) := \max\{j \in \mathbb{N} \colon X_j(t) > 0\}$$

is the maximum number of mutations present in any individual at time t, then the difference

$$Q(t) := M^*(t) - M(t)$$

is typically a constant multiple of k_N (Theorem 2.4.1). On the other hand, the quantity a_N is the right time scale to study the genealogy of the process, since the time to the most recent common ancestor of two randomly chosen individuals is also typically a constant multiple of a_N. The value of a_N is also the amount of time between the first appearance of an individual with j mutations and the time when $M(t)$ equals j.

The main assumptions needed for the results of this section are

A1. $\lim_{N\to\infty} \frac{k_N}{\log(1/s_N)} = \infty$,

A2. $\lim_{N\to\infty} \frac{k_N \log(k_N)}{\log(s_N/k_N)} = 0$,

A3. $\lim_{N\to\infty} s_N k_N = 0$.

These assumptions have the following main implications concerning the asymptotic behaviour of s_N and μ_N:

(2.70) $$\lim_{N\to\infty} s_N = 0,$$

(2.71) $$\lim_{N\to\infty} k_N = \infty = \lim_{N\to\infty} a_N, \lim_{N\to\infty} \frac{k_N}{a_N} = 0, \text{ and}$$

(2.72) $$\forall a > 0, \lim_{N\to\infty} \frac{\mu_N}{s_N^a} = 0 \text{ and } \lim_{N\to\infty} \frac{\mu_N}{N^{-a}} = \infty.$$

2.4.1 The Speed of Selection and the Genealogy

In this section we describe the main results for the speed of selection and the genealogy of the process proven in Schweinsberg 2017a,b, and then give a brief summary of the heuristic arguments which guide the corresponding proofs.

This first theorem says that, if we time-scale the population by a_N, then the difference in number of mutations between the fittest individual and the average individual is typically of the order of k_N.

Theorem 2.4.1. *Assume that* **A1 - A3** *hold. There is a unique bounded function* $q\colon [0,\infty) \to [0,\infty)$ *such that*

$$q(t) = \begin{cases} e^t & \text{if } 0 \le t < 1, \\ \int_{t-1}^{t} q(s)ds & \text{if } t \ge 1. \end{cases}$$

This function satisfies

$$\lim_{t\to\infty} q(t) = 2,$$

and, for every compact subset $S \subset (0,\infty) \setminus \{1\}$,

$$\sup_{t\in S} \left| \frac{Q(a_N t)}{k_N} - q(t) \right| \xrightarrow{P} 0, \quad \text{as } N \to \infty. \tag{2.73}$$

The following theorem gives the asymptotic behaviour of the mean number of mutations under the same time-scale a_N of the population.

Theorem 2.4.2. *Let* $m\colon [0,\infty) \to \mathbb{R}$ *be the function*

$$m(t) := \begin{cases} 0 & \text{if } 0 \le t < 1, \\ 1 + \int_0^{t-1} q(s)ds & \text{if } t \ge 1, \end{cases}$$

and also define

$$m^*(t) := m(t) + q(t) = 1 + \int_0^t q(s)ds.$$

Assume that **A1-A3** *hold. Then, for any compact subset* $S \subset [0,\infty) \setminus \{1\}$,

$$\sup_{t\in S} \left| \frac{M(a_N t)}{k_N} - m(t) \right| \xrightarrow{P} 0, \quad \text{as } N \to \infty; \tag{2.74}$$

whereas for any compact $S \subset (0,\infty)$,

$$\sup_{t\in S} \left| \frac{M^*(a_N t)}{k_N} - m^*(t) \right| \xrightarrow{P} 0, \quad \text{as } N \to \infty. \tag{2.75}$$

In Schweinsberg 2017a the authors go further and give a characterization of the fitness distribution at time t; we refer the reader to this article for further details on the subject.

Note that by Theorem 2.4.1 we have

$$\lim_{t\to\infty} \frac{m(t)}{t} = 2,$$

so that, combining with Theorem 2.4.2 we obtain the approximation for the speed of selection given by, for large t,

$$\frac{M(a_N t)}{a_N t} \approx \frac{k_N m(t)}{a_N t} \approx 2\frac{k_N}{a_N} = \frac{s_N \log(N)}{\log^2(s_N/\mu_N)} \overset{N\to\infty}{\to} 0,$$

where we have used (2.71) for the convergence to zero.

On the other hand, for the description of the genealogy of the process we have the following theorem.

Theorem 2.4.3. *Assume* **A1-A3** *hold. Fix positive real numbers t_0 and T such that $t_0 > 0$ and $T + t_0 > 2$. Consider a sample of n individuals of the population at time $a_N T$ and, for $0 \leq u \leq t_0 + 1$, let $\Pi_u^{(N,n)}$ be the partition that describes their ancestry at time $a_N(T-u)$. Then*

$$\mathbb{P}\left(\Pi_1^{(N,n)} = \{\{1\},\ldots,\{n\}\}\right) = 1$$

and the finite dimensional distributions of $\left(\Pi_{1+u}^{(N,n)}\right)_{0\leq u\leq t_0}$ converge as $N \to \infty$ to those of the Bolthausen-Sznitman coalescent.

Heuristics for the Genealogy

We first describe the dynamics of the population at the initial stage when $M(t) \approx 0$, and, thus, $X_0(t) \approx N$. This means that the process can be approximated by a multitype branching process where a type i individual dies at rate $\left(1 - \frac{1+s_N(i-M(t))}{\sum_{j=1}^{\infty} X_j(t)(1+s_N(j-M(t)))}\right) \approx \left(1 - \frac{1+s_N i}{N}\right) \approx 1$, gives birth to another type i individual at rate $1 + s_N(i - M(t)) \approx 1 + s_N i$, and mutates to type $i+1$ at rate $X_{i+1}(t)\left(1 + s_N(i+1-M(t))\right) + \mu_N \approx \mu_N$. In particular, new type $i+1$ individuals are created by mutation at rate $X_i(t)\mu_N$, and, if one such mutation occurs at time u, then its type i descendency at time $t > u$ will be, on average, $e^{(1+s_N(i+1)-1-\mu_N)(t-u)} = e^{(s_N(i+1)-\mu_N)(t-u)} \overset{\mu_N << s_N}{\approx} e^{s_N j(t-u)}$. Integrating over u we obtain the approximation

$$\mathbb{E}\left[X_1(t)\right] \overset{M(t)\approx 0}{\approx} \int_0^t \mu_N \mathbb{E}\left[X_0(u)\right] e^{s_N(t-u)} du \approx \int_0^1 \mu_N N e^{s_N(t-u)} = \frac{N\mu_N(e^{s_N t}-1)}{s_N}, \tag{2.76}$$

and, recursively,

$$\mathbb{E}\left[X_{i+1}(t)\right] \overset{M(t)\approx 0}{\approx} \int_0^t \mu_N \mathbb{E}\left[X_i(u)\right] e^{s_N(t-u)} du \approx \int_0^1 \mu_N N e^{s_N(t-u)} = \frac{N\mu_N^i}{s_N^i i!}(e^{s_N t}-1)^i. \tag{2.77}$$

Since these approximations are valid only when the total number of mutations is close to zero, they will be valid for t such that $X_t(t) << N$. From (2.76) we see that this will (approximately) hold whenever $e^{s_N t} \leq \frac{s_N}{\mu_N}$ or

$$t < \frac{1}{s_N}\log(s_N/\mu_N) = a_N.$$

It turns out that $X_i(t) \approx \mathbb{E}\left[X_i(t)\right]$ for $i \leq k_N$, but for $i > k_N$ the expectation $\mathbb{E}\left[X_i(t)\right]$ is dominated by the rare events where a particle acquires i mutations in an unusually short amount of time, leading X_i to be unusually large. In this scenario, the evolution of X_i has two main stages, the first stochastic, and the second deterministic. Before before time τ_i, where τ_i is the first time when there are at least s_N/μ_N type $i-1$ individuals, X_i is approximately 0. The stochastic stage occurs between times $[\tau_i, \tau_{i+1}]$, during this stage the type i population becomes established; during the second stage, after time τ_{i+1}, the number of type i individuals evolves roughly deterministically. It can be shown that shortly after time τ_{i+1} the increase in X_i that is driven by mutations from type $i-1$ individuals becomes negligible so that X_i grows deterministically at rate $s_N(i - M(t))$, which is the selective advantage of type i over the average individual. That is, for times $t > \tau_{i+1}$ such that $X_i(t)$ is not yet near zero,

$$X_i(t) \approx \frac{s_N}{\mu_N} e^{\int_{\tau_{i+1}}^t s_N(i-M(u))du}, \text{ if } t > \tau_{i+1}, X_i(t) > 0. \tag{2.78}$$

On the other hand, between times $[\tau_i, \tau_{i+1}]$, type $i-1$ individuals will acquire a new mutation at rate μ_N, so that the total growth rate of X_i driven by mutations will be $\mu_N X_{i-1}(u) \overset{(2.78)}{\approx} \mu_N \frac{s_N}{\mu_N} e^{\int_{\tau_i}^u s_N(i-1-M(\eta))d\eta}$; and if $u - \tau_i > 0$ is small enough so that the term $i - 1 - M(\eta) \approx i - 1 - M(\tau_i) = Q(\tau_i)$ inside the integral remains approximately constant in $\eta \in [\tau_i, u]$, then

$$\mu_N X_{i-1}(u) \approx s_N e^{s_N(i-1-M(u))(u-\tau_i)} = s_N e^{s_N Q(\tau_i)(u-\tau_i)}. \tag{2.79}$$

Since a new type i individual produced by a mutation at time $\tau_i < u < t$ will have on average $e^{s_N(Q(\tau_i)+1)(t-u)}$ descendants at time t, then, putting both estimates together, for $t \in [\tau_i, \tau_{i+1})$,

$$\begin{aligned} X_i(t) &\approx \int_{\tau_i}^t e^{s_N Q(\tau_i)(u-\tau_i)} e^{s_N(Q(\tau_i)+1)(t-u)} du \\ &= s_N e^{s_N(Q(\tau_i)+1)(t-\tau_i)} \int_{\tau_i}^t e^{-s_N(u-\tau_i)} du \\ &= e^{s_N(Q(\tau_i)+1))(t-\tau_i)}(1 - e^{-t}) \\ &\approx e^{s_N(Q(\tau_i)+1)(t-\tau_i)}, \quad t >> 1/s_N. \end{aligned} \tag{2.80}$$

Equating the above approximation to $\frac{s}{\mu}$ we obtain an approximation for $\tau_{i+1} - \tau_i$, mainly

$$\tau_{i+1} - \tau_i \approx \frac{1}{s_N(Q(\tau_i) + 1)} \log\left(\frac{s_N}{\mu_N}\right) \approx \frac{a_N}{Q(\tau_i)}.$$

This gives the overall high probability behaviour of X_i. Now, to study the unusual appearance of large families, we approximate the evolving descendancy of a single mutation that occurs at time $u > \tau_i$ (producing a type i individual) by a supercritical branching process with birth rate $1 + s_N Q(\tau_i)$ and death rate 1. Such a branching process will survive with probability approximately $s_N Q(\tau_i)$ and, conditional on survival, its population size at time $t > u$ will approximate

$$\frac{E}{s_N Q(\tau_i)} e^{s_N Q(\tau_i)(t-u)},$$

where E is standard exponentially distributed. Writing $u = \tau_i - \frac{\log(s_N Q(\tau_i))}{s_N Q(\tau_i)} + \eta$, we see that the offspring size of such a mutation can be approximated by

$$E e^{-s_N Q(\tau_i)\eta} e^{s_N Q(\tau_i)(t-\tau_i)} \overset{(2.80)}{\approx} E e^{-s_N Q(\tau_i)\eta} X_i(t).$$

Thus, the probability that a mutation that occurs at time $\tau_i < u$ has an offspring of size at least $xX_i(t)$, $x \in (0,1)$, at time $t > u$, is approximately

$$s_N Q(\tau_i)\mathbb{P}\left(Ee^{-s_N Q(\tau_i)\eta} X_i(t) > xX_i(t)\right) = s_N Q(\tau_i) e^{-xe^{s_N Q(\tau_i)\eta}}.$$

Integrating over the possible mutation times which occur at rate (2.79), the rate at which a family of size at least $xX_i(t)$ appears is approximately

$$\int_{-\infty}^{\infty} \left(s_N e^{s_N Q(\tau_i)\left(-\frac{\log(s_N Q(\tau_i))}{s_N Q(\tau_i)} + \eta\right)} \right) s_N Q(\tau_i) e^{-xe^{s_N Q(\tau_i)\eta}} d\eta = \int_{-\infty}^{\infty} s_N e^{-s_N Q(\tau_i)\eta} e^{-xe^{s_N Q(\tau_i)\eta}} d\eta$$
$$= \frac{1}{Q(\tau_i)x}.$$

Here we recognize the term $x^{-1} = \int_x^1 y^{-2} dy$ as the rate at which a y-merger with $y > x$ occurs in the Bolthausen-Sznitman coalescent. Finally, the heuristic argument, and the accompanying formal proof given by Schweinsberg 2017b, end by proving that these are the only type of large reproductive events that occur in the population, i.e., if a large fraction of siblings with i mutations coalesce, then their parent will have $i-1$ mutations with probability tending to 1 as $N \to \infty$.

Chapter 3

The Site Frequency Spectrum of the Bolthausen-Sznitman Coalescent

Here we present our joint article together with Götz Kersting and Arno Siri-Jégousse (Kersting, Siri-Jégousse, and H. Wences 2021).

A measure of the genetic diversity in a present day sample of a population is often used in population genetics in order to infer its evolutionary past and the forces at play in its dynamics. The *Site Frequency Spectrum* (SFS) is a well known theoretical model of the genetic diversity present in a population, it assumes that neutral mutations arrive to the population as a Poisson Process and that each arriving mutation falls in a different site of the genome (infinite sites model), in contrast to the *Allele Frequency Spectrum* in which mutations are assumed to fall on the same site but create a new allele every time (infinite alleles model). Heuristically, the SFS is a random vector constructed from a coalescent process $(\Pi_t)_{t\geq 0}$ with values in $\mathscr{P}_n$ in the following way: first a genealogical tree is constructed in the natural way according to the evolution of the blocks and the jump times of $(\Pi_t)_{t\geq 0}$ (i.e. individuals find a common ancestor whenever blocks coalesce), and then a random set of points is thrown upon the tree according to a Poisson point process with rate θ (see Figure 3). These points are interpreted as neutral mutations that occur to the individuals in the population and that are inherited to the individuals in generation 0 according to the topology of the tree given by $(\Pi_t)_{t\geq 0}$. Each mutation is assumed to occur at a different place in the genome so that each creates a new segregating site. Finally, for each integer $1 \leq b \leq n-1$ the random variable $SFS_{n,b}$ is set to be the number of mutations (segregating sites) that are shared by a exactly b individuals in generation 0 (see Figure 3), the SFS is the random vector $SFS = (SFS_{n,1}, \dots, SFS_{n,n-1})$. Given the close relation between the SFS and the whole structure of the underlying genealogical tree (topology + branch lengths), the SFS can be used as a model selection tool for the evolutionary dynamics of a population from a dataset of present-day genetic diversity (Eldon et al. 2015; Freund and Siri-Jégousse 2021; Koskela 2018).

In this chapter we give explicit expressions of the first and second moments for the whole Site Frequency Spectrum $(SFS_{n,b})_{1\leq b<n}$ of the Bolthausen-Sznitman coalescent, which to our knowledge were only known for Kingman's coalescent until now (Fu 1995). For the

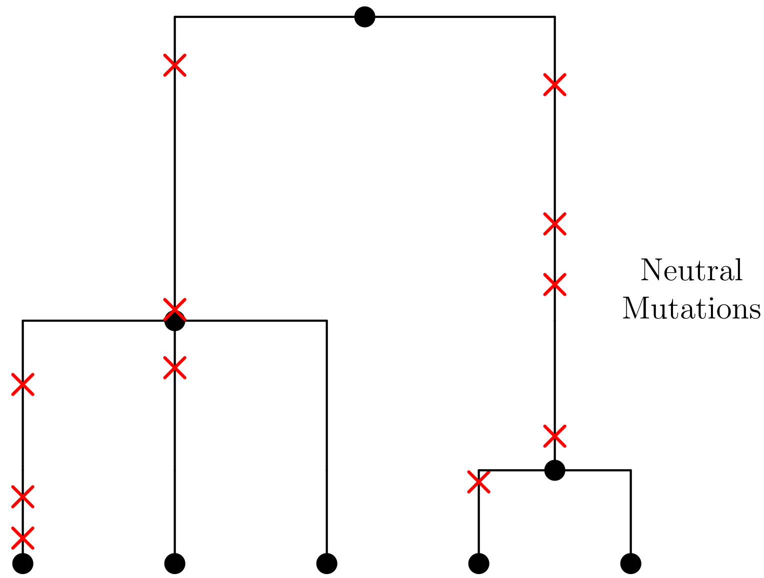

Figure 3.1: Schematic representation of a Poisson point process over a genealogical tree. In this example we have $SFS_{5,1} = 5, SFS_{5,2} = 4, SFS_{5,3} = 2$, and $SFS_{5,4} = 0$.

expectation we obtain the formula

$$\mathbb{E}\left[SFS_{n,b}\right] = \theta n \int_0^1 \frac{\Gamma(b-p)}{\Gamma(b+1)} \frac{\Gamma(n-b+p)}{\Gamma(n-b+1)} \frac{dp}{\Gamma(1-p)\Gamma(1+p)},$$

where θ denotes the mutation rate. This expression is easily evaluated using existing numerical routines that compute $\log\Gamma(a), a > 0$, and writing the gamma ratios $\frac{\Gamma(a)}{\Gamma(b)}$ as $e^{\log\Gamma(a)-\log\Gamma(b)}$.

This formula allows no insight into the shape of the expected site frequency spectrum. For this purpose, and in order to characterize the asymptotic behaviour of the SFS, approximations are helpful. A simple approximation resting on Stirling's formula reads for $2 \leq b \leq n-1$

$$\mathbb{E}[SFS_{n,b}] \approx \frac{\theta}{n-1} \frac{b-1}{b} f_1\left(\frac{b-1}{n-1}\right) \tag{3.1}$$

where f_1 is a convex, non-monotone function on $(0, 1)$ defined by

$$f_1(u) := \int_0^1 u^{-p-1}(1-u)^{p-1} \frac{\sin(\pi p)}{\pi p}\, dp\ . \tag{3.2}$$

We remark that this integral may be reduced to the (complex) exponential integral $Ei(\cdot)$. These formulas show that the shape of the Site Frequency Spectrum, restricted to the range $2 \leq b < n$, is explained essentially by one function not depending on the population size n. Also our approximations update those given in Neher and Hallatschek 2013 for the case of families with frequencies close to 0 and 1, since we have $f_1(u) {\sim} (u \log u)^{-2}$ close to 0 and $f_1(u) {\sim} ((u-1)\log(1-u))^{-1}$ close to 1, see equations (3.28) and (3.29) below. The case

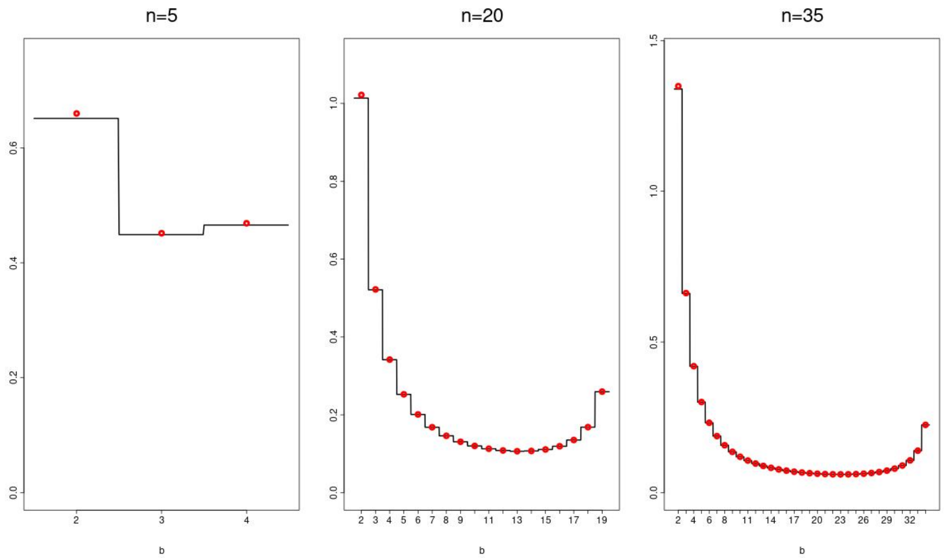

Figure 3.2: Comparison of exact and approximated values of $\mathbb{E}[SFS_{n,b}]$, red circles present the exact values for $b = 2$ to $n-1$, and the black lines their refined approximations (3.4).

$b = 1$ is not covered by (3.1), it has to be treated separately, which reflects the dominance of external branches in the Bolthausen-Sznitman coalescent. In this case we have

$$\frac{\log n}{n}\mathbb{E}[SFS_{n,1}] \approx \theta. \tag{3.3}$$

See Theorem 3.2.4 for a rigorous and complete summary of the asymptotic behaviour of $E[SFS_{n,b}]$.

The above approximation is accurate also from a numerical point of view. Only for $b = 2$ we encounter an enlarged relative error which anyhow remains less than 10 percent for $n \geq 8$. If a more precise result is desired then the following refined approximation may be applied for $2 \leq b \leq n$:

$$\mathbb{E}\left[SFS_{n,b}\right] \approx \theta n \frac{b-1}{b}\left(\frac{1}{(n-1)^2} f_1\left(\frac{b-1}{n-1}\right) - \frac{1}{(n-1)^3} g_1\left(\frac{b-1}{n-1}\right)\right), \tag{3.4}$$

with a positive function g_1 on $(0,1)$ given by

$$g_1(u) := \frac{1}{2u^2(1-u)^2} \frac{\pi^2 + \log^2\frac{1-u}{u} + \frac{2}{u}\log\frac{1-u}{u}}{\left(\pi^2 + \log^2\frac{1-u}{u}\right)^2}. \tag{3.5}$$

With this formula we have a relative error remaining below 1 percent for $b = 2$ and $n \geq 10$, below 0.5 percent for $b = 2$ and $n \geq 150$, and below 0.3 percent for $b \geq 3$ and $n \geq 10$. Thus this approximation appears well-suited for practical purpose. Figure 3.2 illustrates its precision in the cases $n = 5$, 20, 35 and $\theta = 1$.

For $b=1$ the approximation formula corresponding to (3.1) reads

$$\begin{aligned}\mathbb{E}\left[SFS_{n,1}\right] &= \theta n \int_0^1 \frac{\Gamma(n-1+p)}{\Gamma(n)} \frac{dp}{\Gamma(1+p)} \\ &\approx \theta n \int_0^1 (n-1)^{p-1} \frac{dp}{\Gamma(1+p)},\end{aligned}$$

which is an immediate consequence of Stirling's approximation. It is precise for small n and requires no further correction as in the case $b \geq 2$.

We also study the asymptotic behavior of the second moments which, together with the above asymptotics for the first moment, leads to the following L^2 convergences:

$$\frac{\log n}{n} SFS_{n,1} \to \theta,$$

and, whenever $b \geq 2$ and $b = o\left(\sqrt{n}/\log n\right)$,

$$\frac{b(b-1)\log^2(n/b)}{n} SFS_{n,b} \to \theta.$$

These generalize and strengthen the results in Diehl and Kersting 2019 for the Bolthausen-Sznitman coalescent.

Finally we provide the joint distribution function of the branch lengths of large families, i.e families of size at least half the total population size, and their marginal distribution function. These results are useful to obtain the marginal distribution function of the Site Frequency Spectrum and a sampling formula for the half of the vector corresponding to large family sizes, although we do not present such tedious computations here.

Asymptotic results for related functionals on the Bolthausen-Sznitman coalescent have been derived by studying the block count chain of the coalescent through a coupling with a random walk as in Iksanov and Möhle 2007 and Kersting, Pardo, and Siri-Jégousse 2014, where asymptotics for the total number of jumps, and the total, internal, and external branch lengths of the Bolthausen-Sznitman coalescent are described; these results give the asymptotic behaviour of the total number of mutations present in the population, the number of mutations present in a single individual, and the number of mutations present in at least 2 individuals. Also, a Markov chain approximation of the initial steps of the process was developed in Diehl and Kersting 2019 where asymptotics for the total tree length and the Site Frequency Spectrum of small families were derived for a class of Λ-coalescents containing the Bolthausen-Sznitman coalescent.

Progress has also been made for the finite coalescent even for the general $\Lambda-$ and Ξ-coalescents. The finite Bolthausen-Sznitman coalescent has been studied through the spectral decomposition of its jump rate matrix described in Kukla and Pitters 2015. This lead the authors to derive explicit expressions for the transition probabilities and the Green's matrix of this coalescent, and also of Kingman's coalescent. The spectral decomposition of the jump rate matrix of a general coalescent, including coalescents with multiple mergers, is also used in Spence, Kamm, and Song 2016 where an expression for the expected Site Frequency Spectrum is given in terms of matrix operations which in the case of the Bolthausen-Sznitman coalescent result in an algorithm requiring on the order of n^2 computations. In Hobolth, Siri-Jégousse, and Bladt 2019 another expression in terms of matrix operations is given for this and other functionals of general coalescent processes, both in expected value (and higher moments) and in distribution; these expressions however are deduced from the theory of phase-type distributions, in particular distributions of rewards constructed on top of coalescent processes, and also require vast computations for large population sizes.

Our method, mainly based on the Random Recursive Tree construction of the Bolthausen-Sznitman coalescent given in Goldschmidt and Martin 2005, gives easy-to-compute expressions for the first and second moments of the Site Frequency Spectrum of this particular coalescent. This combinatorial construction not only allows us to study the bottom, but also the top of the tree; thus providing an additional insight into the past of the population and large families, both asymptotically and for any fixed population size.

In Section 3.1 we layout the basic intuitions that compose the bulk of our method, including the Random Recursive Tree construction of the Bolthausen-Sznitman coalescent and the derivation of the first moment of the Site Frequency Spectrum for the infinite coalescent as a first application (Corollary 3.1.2). In Section 3.2 we present our results on the first and second moments of the branch lengths (Theorem 3.2.1) and of the Site Frequency Spectrum (Corollary 3.2.2) for any fixed family size and initial population. We then use these expressions to obtain asymptotic approximations of these moments as the initial population goes to infinity (Theorems 3.2.4 and 3.2.5) which lead to L^2 convergence results on the SFS (Corollary 3.2.6). In Section 3.3 we restrict ourselves to the case of large family sizes and present the joint and marginal distribution functions of their branch lengths (Theorems 3.3.1 and 3.3.3), along with a limit in law result (Corollary 3.3.2). Finally, in Sections 3.2.1 and 3.3.1 we provide detailed proofs of our results.

3.1 Random Recursive Tree Construction of the BSC

Consider the Bolthausen-Sznitman coalescent $(\Pi_t)_{t\geq 0}$ with values in $\mathscr{P}_\infty$, the space of partitions of $\mathbb{N}$, and the ranked coalescent $(|\Pi_t|^\downarrow)_{t\geq 0}$, with values in the space of mass partitions $\mathscr{P}_{[0,1]}$, made of the asymptotic frequencies of Π_t reordered in a non-increasing way. In what follows we present the Random Recursive Tree (RRT) construction of the Bolthausen-Sznitman coalescent given in Goldschmidt and Martin 2005; then we follow the argument given in the same paper to establish that

$$|\Pi_t|^\downarrow \overset{d}{=} PD(e^{-t}, 0), \tag{3.6}$$

where $PD(\alpha, \theta)$ is the $(\alpha, \theta)-$Poisson-Dirichlet distribution.

Briefly, the construction of the Bolthausen-Sznitman coalescent in terms of Random Recursive Trees proceeds as follows. We work on the set of recursive trees whose labeled nodes form a partition π of $[n] := \{1, \ldots, n\}$, where the ordering of the nodes that confers the term "recursive" is given by ordering the blocks of π according to their smallest elements. A cutting-merge procedure is defined on the set of recursive trees of this form with a marked edge, this procedure consists of cutting the marked edge and merging all the labels in the subtree below with the node above, thus creating a new recursive tree whose labels form a new (coarser) partition of $[n]$ (see Figure 3.3). With this operation in mind we consider a RRT with labels $\{1\}, \cdots, \{n\}$, say T, to which we also attach independent standard exponential

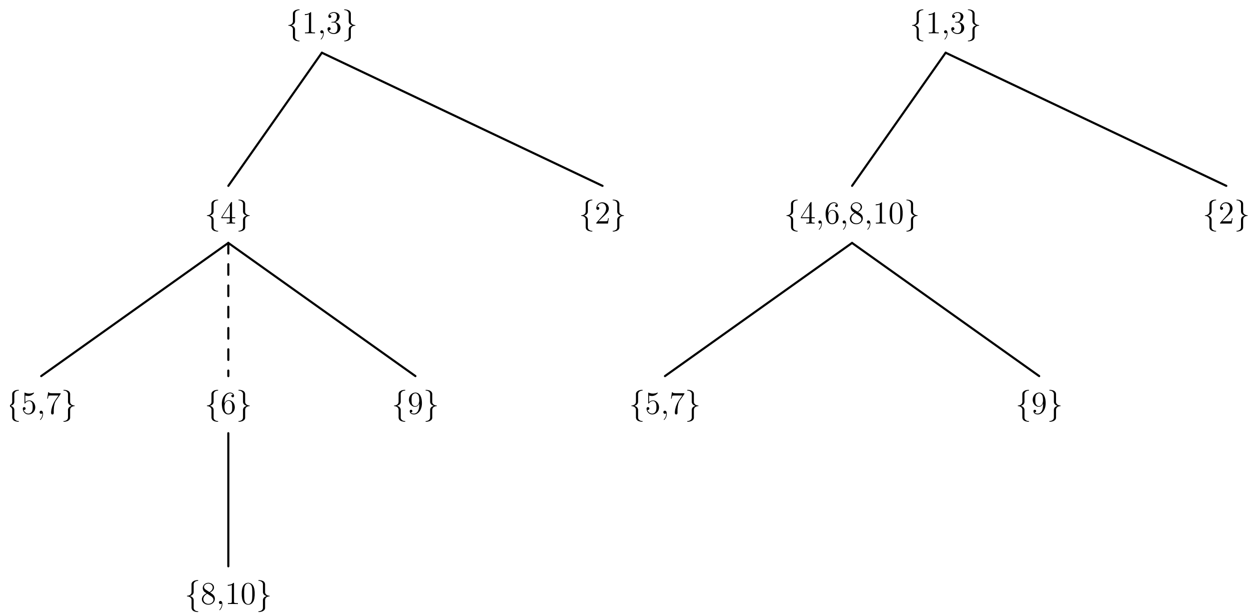

Figure 3.3: On the left, an example of a recursive tree whose labels constitute a partition of $\{1, \cdots, 10\}$. On the right, the resulting recursive tree after a cutting-merge procedure performed on the marked edge (dashed line) of the first tree.

variables to each edge. Then, for each time $t > 0$ we retrieve the partition of $[n]$ obtained by performing a cutting-merge procedure on all the edges of T whose exponential variable is less than t. This gives a stochastic process $(\Pi_t^{(n)})_{t \geq 0}$ with values on the set of partitions of $[n]$ that can be proven to be the n-Bolthausen-Sznitman coalescent (Goldschmidt and Martin 2005).

The fact that $|\Pi_t|^{\downarrow} \stackrel{d}{=} PD(e^{-t}, 0)$ now follows readily. To see this, consider the construction of T where nodes arrive sequentially and each arriving node attaches to any of the previous nodes with equal probability. Considering also their exponential edges and having in mind the cutting-merge procedure we see that for any fixed time t, and assuming that $b-1$ nodes have arrived and formed k blocks of sizes $s_1, \ldots, s_k$ in $\Pi_t^{(b-1)}$, the next arriving node, node $\{b\}$, will form a new block in $\Pi_t^{(b)}$ if and only if it attaches to any of the roots of the sub-trees of T that form the said k blocks and if, furthermore, its exponential edge is greater than t; this occurs with probability $\frac{ke^{-t}}{b-1}$. On the other hand, in order for $\{b\}$ to join the jth block of size s_j it must either attach to the root of the sub-tree of T that builds this block and its exponential edge must be less than t, which happens with probability $\frac{1-e^{-t}}{b-1}$, or it must attach to any other node of the said sub-tree, which happens with probability $\frac{s_j-1}{b-1}$; thus, the probability of attaching to the jth block is $\frac{s_j-e^{-t}}{b-1}$. We recognize in these expressions the probabilities that define the Chinese Restaurant Process with parameters $\alpha = e^{-t}$ and $\theta = 0$.

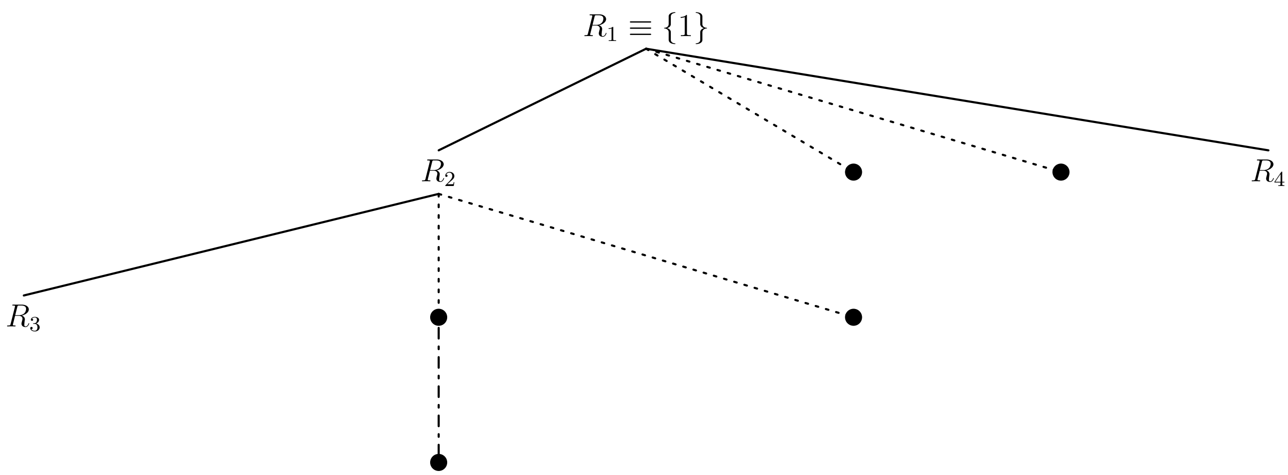

Figure 3.4: Schematic representation of passing from $\Pi_t^{(n)}$ to $\Pi_t^{(n+1)}$ for fixed t, by adding a new node (blue) to a RRT. Solid lines and dotted lines represent edges whose exponential variables are greater than t and less than or equal to t, respectively. In this case at time t there are four subtrees rooted at R_1, R_2, R_3, and R_4, which form the blocks that constitute $\Pi_t^{(n)}$; these blocks are also the tables of a Chinese Restaurant Process. In case (i) the new node will be included in the block formed by R_2 at time t, irrespective of whether its exponential edge is greater than t or not. In case (ii) the new node forms part of the block rooted at R_4 because its exponential edge is less than t. Finally, in case (iii) the new node is a new root of a subtree that will form an additional block of $\Pi_t^{(n+1)}$ (i.e. the new node opens a new table in the Chinese Restaurant Process).

We now provide two straightforward applications of the RRT construction described above which nonetheless contain the essential intuitions underlying the forthcoming proofs.

3.1.1 Site Frequency Spectrum in the infinite coalescent

For the first application consider a subset $I \subset (0,1)$ and define $(C_I(t))_{t\geq 0}$ to be the process of the number of blocks in Π_t with asymptotic frequencies in I. Then

$$\ell_I := \int_0^\infty C_I(t)\, dt \tag{3.7}$$

gives the total branch length of families with size frequencies in I in the infinite coalescent.

Our first theorem is a simple corollary of the equality in law (3.6).

Theorem 3.1.1. *For $I \subset (0,1)$, we have*

$$\mathbb{E}[\ell_I] = \int_I \int_0^1 u^{-p-1}(1-u)^{p-1} \frac{\sin(\pi p)}{\pi p}\, dp\, du.$$

In particular, note that if in the infinite sites model with mutation rate θ we define SFS_I to be the number of mutations shared by a proportion u of individuals with u ranging in I, then by conditioning on ℓ_I we get

Corollary 3.1.2. *For $I \subset (0,1)$, we have*

$$\mathbb{E}\left[SFS_I\right] = \theta \int_I \int_0^1 u^{-p-1}(1-u)^{p-1} \frac{\sin(\pi p)}{\pi p}\, dp\, du. \tag{3.8}$$

Proof of Theorem 3.1.1. Since

$$\mathbb{E}\left[\ell_I\right] = \int_0^\infty \mathbb{E}\left[C_I(t)\right]\, dt$$

it only remains to compute $\mathbb{E}\left[C_I(t)\right]$ and simplify the expressions, but this is a straightforward consequence of Equation (6) in Pitman and Yor 1997 which states that if $\varrho = (a_1, \cdots)$ is $PD(\alpha,\theta)$ distributed, and $f : \mathbb{R} \to \mathbb{R}$ is a function, then

$$\mathbb{E}\left[\sum_{i=1}^\infty f(a_i)\right] = \frac{\Gamma(\theta+1)}{\Gamma(\theta+\alpha)\Gamma(1-\alpha)} \int_0^1 f(u) \frac{(1-u)^{\alpha+\theta-1}}{u^{\alpha+1}}\, du. \tag{3.9}$$

Taking $f(u) = \mathbb{1}_I(u)$we get

$$\mathbb{E}[C_I(t)] = \frac{1}{\Gamma(e^{-t})\Gamma(1-e^{-t})} \int_0^1 \mathbb{1}_I(u) \frac{(1-u)^{e^{-t}-1}}{u^{e^{-t}+1}}\, du.$$

The proof is finished by using Euler's reflection formula, making $p = e^{-t}$ on the above expression, and integrating on $[0,\infty)$. $\square$

3.1.2 Time to the absorption

In this section we prove a useful lemma for the upcoming proofs, but a first consequence of this lemma gives the distribution function of the time to absorption, A_n, in the n-coalescent, a result already proved in Möhle and Pitters 2014.

We recall B which stands for the Beta function

$$\mathsf{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

and Ψ for the digamma function

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = -\Upsilon - \sum_{n=1}^{\infty}\left(\frac{1}{z+n-1} - \frac{1}{n}\right)$$

where Υ stands for the Euler-Mascheroni constant.

Lemma 3.1.3. *Let T be a RRT on a set of n labels and with independent exponential edges. Define the two functionals $m(T)$ and $M(T)$ that give the minimum and the maximum of the exponential edges attached to the root of T. Then*

$$\mathbb{P}(m(T) > s) = \frac{1}{(n-1)\mathsf{B}(n-1, e^{-s})}, \tag{3.10}$$

and

$$\mathbb{P}(M(T) \leq s) = \frac{1}{(n-1)\mathsf{B}(n-1, 1-e^{-s})}. \tag{3.11}$$

Also, for independent trees T_1 and T_2 of respective size n_1 and n_2, we have

$$\begin{aligned} &\mathbb{P}(m(T_2) - M(T_1) > s) \\ =&\frac{1}{(n_1-1)(n_2-1)}\int_0^1 \frac{\Psi(n_1-p) - \Psi(1-p)}{\mathsf{B}(n_2-1, e^{-s}p)\mathsf{B}(n_1-1, 1-p)}\,dp. \end{aligned} \tag{3.12}$$

The proof of (3.11) follows the same lines as in Möhle and Pitters 2014 where the law of the time to absorption of the Bolthausen-Sznitman coalescent is derived, since this time is the maximum of the exponential edges attached to the root of a RRT. That is,

$$\mathbb{P}(A_n \leq s) = \frac{1}{(n-1)\mathsf{B}(n-1, 1-e^{-s})}, \tag{3.13}$$

and, as $n \to \infty$,

$$A_n - \log\log n \xrightarrow{d} -\log E \tag{3.14}$$

where E is a standard exponential random variable. The latter convergence in distribution was elegantly proved in Goldschmidt and Martin 2005 using a construction of random recursive trees in continuous time, whereas in this case it follows from Stirling's approximation to the Gamma functions appearing in (3.13).

On the other hand, the equality (3.12) will be used in the computation of the distribution function of branch lengths with large family sizes presented in Section 3.3.

Proof of Lemma 3.1.3. Let $\mathcal{E}_2, \cdots, \mathcal{E}_n$ be the exponential edges associated to the nodes of T. For the proof of (3.10) we consider the event $\{m(T) > s\}$. This event occurs when, in the recursive construction of T along with the exponential edges, the ith node ($2 \leq i \leq n$) does not attach to $\{1\}$ whenever $\mathcal{E}_i < s$; this happens with probability $1 - \frac{1-e^{-s}}{i-1}$. Thus, considering the n nodes, we obtain

$$\begin{aligned}\mathbb{P}(m(T) > s) &= e^{-s}\left(\frac{1+e^{-s}}{2}\right)\cdots\left(\frac{n-2+e^{-s}}{n-1}\right)\\ &= \frac{1}{(n-1)\mathsf{B}(n-1, e^{-s})}.\end{aligned}$$

For (3.11) we instead build the tree such that the ith node does not attach to $\{1\}$ whenever $\mathcal{E}_i > s$; this happens with probability $1 - \frac{e^{-s}}{i-1}$. Thus we obtain

$$\begin{aligned}\mathbb{P}(M(T) \leq s) &= (1-e^{-s})\left(\frac{2-e^{-s}}{2}\right)\cdots\left(\frac{n-1-e^{-s}}{n-1}\right)\\ &= \frac{1}{(n-1)\mathsf{B}(n-1, 1-e^{-s})}.\end{aligned}$$

Finally we compute

$$\begin{aligned}&\mathbb{P}(m(T_2) - M(T_1) > s)\\ =&\frac{1}{(n_1-1)(n_2-1)}\int_0^\infty \frac{1}{\mathsf{B}(n_2-1, e^{-(s+t)})}\frac{d}{dt}\left(\frac{1}{\mathsf{B}(n_1-1, 1-e^{-t})}\right)\,dt\end{aligned}$$

and by changing the variable $p = e^{-x}$ we obtain (3.12). □

3.2 Moments of the Site Frequency Spectrum

By a simple adaptation of our previous notation for branch lengths in the infinite coalescent (C_I and ℓ_I), in the finite case we also define for $1 \leq b \leq n-1$ the process $(C_{n,b}(t))_{t\geq 0}$ and the random variables $(\ell_{n,b})$, where $C_{n,b}(t)$ is the number of blocks of size b in $\Pi_t^{(n)}$, and

$$\ell_{n,b} := \int_0^\infty C_{n,b}(t)\,dt. \tag{3.15}$$

We now provide explicit expressions for $\mathbb{E}[\ell_{n,b}]$ and $\mathbb{E}[\ell_{n,b_1}\ell_{n,b_2}]$; for this we define the functions

$$F_1(n,b) = \int_0^1 \frac{\Gamma(b-p)}{\Gamma(b+1)}\frac{\Gamma(n-b+p)}{\Gamma(n-b+1)}\frac{dp}{\Gamma(1-p)\Gamma(1+p)},$$

$$\begin{aligned}F_2(n,b_1,b_2) = &\int_0^1\int_0^{p_1}\frac{\Gamma(b_1-p_1)}{\Gamma(b_1+1)}\frac{\Gamma(b_2-b_1+p_1-p_2)}{\Gamma(b_2-b_1+1)}\\ &\times\frac{\Gamma(n-b_2+p_2)}{\Gamma(n-b_2+1)}\frac{dp_2\,dp_1}{p_1\Gamma(1-p_1)\Gamma(p_1-p_2)\Gamma(p_2+1)}\end{aligned}$$

and

$$F_3(n, b_1, b_2) = \int_0^1 \int_0^1 \frac{\Gamma(b_1 - p_1)}{\Gamma(b_1 + 1)} \frac{\Gamma(b_2 - p_2)}{\Gamma(b_2 + 1)} \times \frac{\Gamma(n - b_1 - b_2 + p_1 + p_2)}{\Gamma(n - b_1 - b_2 + 1)} \frac{dp_2\, dp_1}{\Gamma(1 - p_1)\Gamma(1 - p_2)(p_1 \vee p_2)\Gamma(p_1 + p_2)}.$$

Theorem 3.2.1. *For any pair of integers n, b such that $1 \le b \le n - 1$, we have*

$$\mathbb{E}[\ell_{n,b}] = nF_1(n, b) \tag{3.16}$$

Also, for any triple of integers n, b_1, b_2, with $1 \le b_1 \le b_2 \le n - 1$, we have

$$\mathbb{E}\left[\ell_{n,b_1} \ell_{n,b_2}\right] = nF_2(n, b_1, b_2) + nF_3(n, b_1, b_2)\mathbb{1}_{\{b_1 + b_2 \le n\}} \tag{3.17}$$

As before, we may define $SFS_{n,b}$ as the number of mutations shared by b individuals in the n-coalescent. By conditioning on the value of the associated branch lengths we get

Corollary 3.2.2. *For $1 \le b \le n - 1$,*

$$\mathbb{E}[SFS_{n,b}] = \theta n F_1(n, b)$$

and, for $1 \le b_1 \le b_2 \le n - 1$, we have,

$$\begin{aligned} \mathbb{Cov}\left(SFS_{n,b_1}, SFS_{n,b_2}\right) =& \theta^2 n F_2(n, b_1, b_2) + \theta^2 n F_3(n, b_1, b_2) \mathbb{1}_{b_1 + b_2 \le n} \\ & - \theta^2 n^2 F_1(n, b_1) F_1(n, b_2) + \theta n F_1(n, b) \mathbb{1}_{b_1 = b = b_2}. \end{aligned}$$

We also characterize the asymptotic behavior of the functions F_1, F_2 and F_3 as $n \to \infty$, which in turn give asymptotic approximations for the first and second moments of the branch lengths and of SFS. For this we recall the function f_1 defined in (3.2) and also define for $0 < u_1 < u_2 < 1$,

$$f_2(u_1, u_2) := \int_0^1 \int_0^{p_1} \frac{u_1^{-p_1 - 1}(u_2 - u_1)^{p_1 - p_2 - 1}\,(1 - u_2)^{p_2 - 1}}{p_1 \Gamma(1 - p_1)\Gamma(p_1 - p_2)\Gamma(p_2 + 1)}\, dp_2\, dp_1, \tag{3.18}$$

and, for $u_1, u_2 > 0, u_1 + u_2 < 1$,

$$f_3(u_1, u_2) := \int_0^1 \int_0^1 \frac{u_1^{-p_1 - 1} u_2^{-p_2 - 1}\,(1 - u_1 - u_2)^{p_1 + p_2 - 1}}{\Gamma(1 - p_1)\Gamma(1 - p_2)(p_1 \vee p_2)\Gamma(p_1 + p_2)}\, dp_2\, dp_1\ . \tag{3.19}$$

Lemma 3.2.3. *We have as $n \to \infty$,*

$$\max_{2 \le b \le n-1} \left| \frac{n^2 F_1(n, b)}{f_1\left(\frac{b-1}{n-1}\right)} - \frac{b - 1}{b} \right| \to 0, \tag{3.20}$$

whereas for $b = 1$,

$$\frac{n^2}{(\log n) f_1\left(\frac{1}{n-1}\right)} F_1(n, 1) \to 1. \tag{3.21}$$

Similarly

$$\max_{2\le b_1<b_2\le n-1}\left|\frac{n^3F_2(n,b_1,b_2)}{f_2\left(\frac{b_1-1}{n-1},\frac{b_2-1}{n-1}\right)}-\frac{b_1-1}{b_1}\right|\to 0, \tag{3.22}$$

and if also $b_1\vee(n-b_2)\to\infty$ *then*

$$\max_{\substack{2\le b_1\le b_2\le n-1\\ b_1+b_2<n}}\left|\frac{n^3F_3(n,b_1,b_2)}{f_3\left(\frac{b_1-1}{n-2},\frac{b_2-1}{n-2}\right)}-\left(\frac{b_1-1}{b_1}\right)\left(\frac{b_2-1}{b_2}\right)\right|\to 0. \tag{3.23}$$

Remark. *The above lemma does not cover the cases* $b_1=1$ *or* $b_1=b_2$ *for* F_2*, nor the cases* $b_1=1$*,* $b_2=1$*,* $n=b_1+b_2$ *or* $b_1\vee(n-b_2)\not\to\infty$ *for* F_3*. However, using the same techniques we also obtain asymptotics in these cases which are used in Theorem* 3.2.5 *below.*

The proof of the above lemma also gives asymptotic expressions for the functions f_1, f_2 and f_3, leading to straightforward asymptotics for the expectation and covariance of SFS. The complete picture for the first moment is given in the next result.

Theorem 3.2.4. *As n goes to infinity,*
(i) The expected number of external mutations $(b=1)$ *has the following asymptotics*

$$\frac{\log n}{n}\mathbb{E}[SFS_{n,1}]\to\theta.$$

(ii) If $b\ge 2$ *and* $\frac{b}{n}\to 0$*, then*

$$\frac{b(b-1)}{n}\log^2\left(\frac{n}{b}\right)\mathbb{E}[SFS_{n,b}]\to\theta.$$

(iii) If $\frac{b}{n}\to u\in(0,1)$*, then*

$$n\mathbb{E}[SFS_{n,b}]\to\theta f_1(u)=\theta\int_0^1 u^{-1-p}(1-u)^{p-1}\frac{\sin(\pi p)}{\pi p}\,dp.$$

(iv) If $\frac{n-b}{n}\to 0$*, then*

$$(n-b)\log\left(\frac{n}{n-b}\right)\mathbb{E}[SFS_{n,b}]\to\theta.$$

(v) Let $I=(x,y)$ *with* $0<x<y<1$ *and define*

$$SFS_{n,I}:=\sum_{b=\lceil nx\rceil}^{\lfloor ny\rfloor}SFS_{n,b}.$$

Then

$$\mathbb{E}\left[SFS_{n,I}\right]\to\mathbb{E}[SFS_I]$$

as it is defined in (3.8).

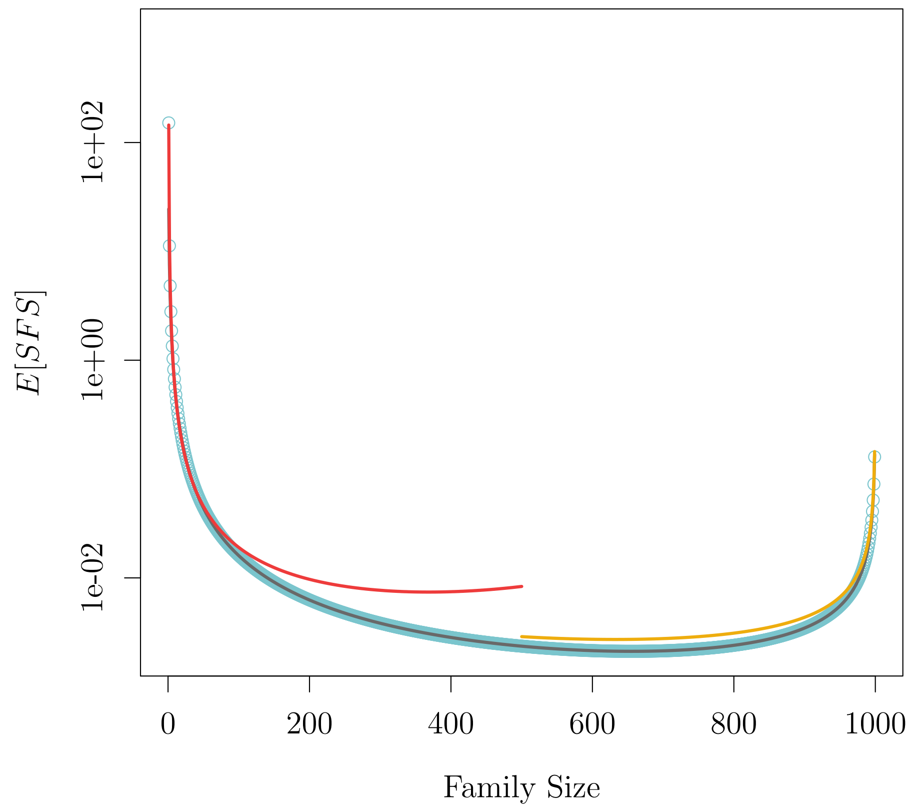

Figure 3.5: Exact and asymptotic approximations for $\mathbb{E}[SFS]$ in a population of size 1000: The blue circles give the exact value as given in Corollary 3.2.2. The gray line is the asymptotic approximation as given in Theorem 3.2.4 (iii). Red (resp. yellow) line is given by Theorem 3.2.4 (ii) (resp. (iv)).

Case (i) and case (ii) for fixed b also follow from Theorem 4 in Diehl and Kersting 2019. Cases (ii) and (iv) give an update to the approximation of the SFS for small and large families made in Neher and Hallatschek 2013.

In the same spirit and using the same techniques we now provide the complete picture for the second moments. In what follows we recall the notation $f(n) \sim g(n)$ to denote that

$$\frac{f(n)}{g(n)} \to 1$$

as $n \to \infty$.

Theorem 3.2.5. *The covariance function has the following asymptotics as n goes to infinity, in each of the following cases:*

b_1	$b_2 - b_1$	$n - b_2$	$\mathbb{Cov}(SFS_{n,b_1}, SFS_{n,b_2})$
> 1	> 0	$\sim n$	$\frac{\theta^2}{b_1(b_1-1)b_2(b_2-1)} \mathcal{O}\left(\frac{n^2}{\log^5 n}\right)$
$\sim n$	> 0	> 0	$\frac{\theta^2}{(b_2-b_1)(n-b_1)} \frac{1}{\log^2 n}$
$\sim n$	0	> 0	$\frac{\theta^2+\theta}{n-b_2} \frac{1}{\log n}$
> 1	$\sim n$	$= b_1$	$\theta^2 \mathcal{O}\left(\frac{n}{\log^4 n}\right)$
> 1	$\sim n$	$= b_1 + const^+$	$\theta^2 F_1(n - b_2, b_1) \frac{n}{\log n}$
1	0	$\sim n$	$\theta^2 \mathcal{O}\left(\frac{n^2}{\log^3 n}\right)$
1	> 0	$\sim n$	$\theta^2 \mathcal{O}\left(\frac{n^2}{\log^4 n}\right)$
1	$\sim nu$	$\sim n(1-u)$	$\theta^2 \mathcal{O}\left(\frac{1}{\log^2 n}\right)$
1	$\sim n$	> 1	$\theta^2 \mathcal{O}\left(\frac{n}{\log^3 n}\right)$
1	$\sim n$	1	$\theta^2 \mathcal{O}\left(\frac{n}{\log^3 n}\right)$
> 1	0	$\sim n$	$\theta^2 \mathcal{O}\left(\frac{n^2}{\log^5 n}\right)$
$\sim nu$	> 0	$\sim n(1-u)$	$\frac{\theta^2}{(1-u)(b_2-b_1)} \frac{1}{n \log^2 n}$
$\sim nu$	0	$\sim n(1-u)$	$\frac{\theta f_1(u)}{n}$
> 1	$\sim nu$	$\sim n(1-u)$	$\theta^2 \mathcal{O}\left(\frac{1}{\log^3 n}\right)$
$\sim nu$	$\sim n(1-u)$	> 0	$-\frac{\theta^2 f_1(u)}{n-b_2} \frac{1}{\log n}$
$\sim nu_1$	$\sim nu_2$	$\sim n(1-u_1-u_2)$	$\frac{\theta^2\left(f_2(u_1,u_1+u_2)+f_3(u_1,u_1+u_2)\mathbb{1}_{2u_1+u_2\leq 1}-f_1(u_1)f_1(u_1+u_2)\right)}{n^2}$
$\sim nu$	$\sim n(1-2u)$	$= b_1$	$\frac{\theta^2 \int_0^\infty \int_0^\infty \frac{e^{-y_1} e^{-y_2}}{y_1 \vee y_2} \, dy_1 \, dy_2}{u(1-u)} \frac{1}{n \log n}$
$\sim nu$	$\sim n(1-2u)$	$= b_1 + const^+$	$\frac{\theta^2 \int_0^\infty \int_0^\infty \frac{e^{-y_1} e^{-y_2} (y_1+y_2)}{y_1 \vee y_2} \, dy_1 \, dy_2}{u(1-u)(n-b_2-b_1)} \frac{1}{n \log^2 n}$

Also for $I, \widehat{I} \subset (0,1)$, and $SFS_{n,I}, SFS_{n,\widehat{I}}$ as defined in Theorem 3.2.4 (V), we have

$$\mathbb{Cov}\left(SFS_{n,I}, SFS_{n,\widehat{I}}\right) \to \theta^2 \int_I \int_{\widehat{I}} f_2(u_1,u_2) + f_3(u_1,u_2)\mathbb{1}_{u_1+u_2<1} - f_1(u_1)f_1(u_2)\, du_2\, du_1 + \theta \int_{I\cap\widehat{I}} f_1(u)\, du. \tag{3.24}$$

These approximations follow from the asymptotics for F_1, F_2, and F_3 substituted in the covariance formula given in Corollary 3.2.2. For the sake of simplicity we do not provide the explicit computations. We only treat the case where the expected value $\mathbb{E}[SFS_{n,b}]$ diverges, then an application of Chebyshev's inequality allows us to prove the following weak law of large numbers with L^2-convergence, which generalizes and strengthens results on the Bolthausen-Sznitman coalescent derived in Diehl and Kersting 2019.

Corollary 3.2.6. *Suppose that $b/n \to 0$ in such a way that $\mathbb{E}[SFS_{n,b}] \to \infty$, or equivalently that $b = o(\sqrt{n}/\log n)$. Then we have the following L^2-convergence*

$$\frac{SFS_{n,b}}{\mathbb{E}[SFS_{n,b}]} \xrightarrow{\mathsf{L}^2} \theta.$$

In view of Theorem 3.2.4 this means that for $b = 1$

$$\frac{\log n}{n}\; SFS_{n,1} \xrightarrow{\mathsf{L}^2} \theta,$$

and for $b \geq 2$, $b = o(\sqrt{n}/\log n)$

$$\frac{b(b-1)\log^2(n/b)}{n}\; SFS_{n,b} \xrightarrow{\mathsf{L}^2} \theta.$$

3.2.1 Proofs of Section 3.2

As in the infinite coalescent case, the proof of Theorem 3.2.1 begins with the definition (3.15) and by noting that

$$\mathbb{E}[\ell_{n,b}] = \mathbb{E}\left[\int_0^\infty C_{n.b}(t)\, dt\right] = \int_0^\infty \mathbb{E}[C_{n,b}(t)]\; dt,$$

and similarly

$$\mathbb{E}[\ell_{n,b_1}\ell_{n,b_2}] = \int_0^\infty \int_0^\infty \mathbb{E}[C_{n,b_1}(t_1)C_{n,b_2}(t_2)]\; dt_1\, dt_2,$$

so it only remains to compute $\mathbb{E}[C_{n,b}(t)]$ and $E[C_{n,b_1}(t)C_{n,b_2}(t)]$ in each case and simplify the expressions.

Proof of Theorem 3.2.1 *(first moment).* Let $\mathcal{B}$ be the collection of all possible blocks of size b in a partition of $[n]$. Then

$$\mathbb{E}[C_{n,b}(t)] = \mathbb{E}\left[\sum_{B\in\mathcal{B}} \mathbb{1}_{B\in\Pi_t^{(n)}}\right] = \sum_{B\in\mathcal{B}} \mathbb{P}\left(B \in \Pi_t^{(n)}\right),$$

and by exchangeability of $\Pi_t^{(n)}$,

$$\mathbb{E}\left[C_{n,b}(t)\right] = \binom{n}{b}\mathbb{P}\left(\{1,\cdots,b\}\in\Pi_t^{(n)}\right).$$

Thus, using (3.9), the fact that $|\Pi_t|^{\downarrow} =: (A_1, A_2, \dots) \stackrel{d}{=} PD(e^{-t}, 0)$, and writing $\Pi^{(n)}$ as $\Pi_{|n}$, we obtain

$$\begin{aligned}\mathbb{E}[C_{n,b}(t)] &= \binom{n}{b}\mathbb{E}_{(e^{-t},0)}\left[\sum_{i=1}^{\infty} A_i^b(1-A_i)^{n-b}\right]\\ &= \binom{n}{b}\int_0^1 u^{b-1}(1-u)^{n-b}\frac{u^{-e^{-t}}(1-u)^{e^{-t}-1}}{\Gamma(1-e^{-t})\Gamma(e^{-t})}\,du\\ &= \frac{n\Gamma(n)}{\Gamma(n-b+1)\Gamma(b+1)}\frac{\mathsf{B}(b-e^{-t},n-b+e^{-t})}{\Gamma(1-e^{-t})\Gamma(1+e^{-t})}.\end{aligned}$$

Finally, by changing the variable $p = e^{-t}$, we obtain (3.16). □

Now we use the random tree construction of the n-Bolthausen-Sznitman coalescent in order to compute the second moments of $\ell_{n,b}$.

Proof of Theorem 3.2.1 *(second moments).* Let $1 \le b_1 \le b_2 \le n-1$, and $\mathcal{B}_1, \mathcal{B}_2$ be the collection of all possible blocks of sizes b_1 and b_2 respectively in a partition of $[n]$. Then

$$\begin{aligned}\mathbb{E}\left[\ell_{n,b_1}\ell_{n,b_2}\right] &= \int_0^\infty\int_0^\infty \mathbb{E}\left[C_{n,b_1}(t_1)C_{n,b_2}(t_2)\right]\,dt_2\,dt_1\\ &= \int_0^\infty\int_0^\infty \sum_{B_1\in\mathcal{B}_1}\sum_{B_2\in\mathcal{B}_2}\mathbb{P}\left(B_1\in\Pi_{t_1}^{(n)}, B_2\in\Pi_{t_2}^{(n)}\right)\,dt_2\,dt_1.\end{aligned}\tag{3.25}$$

We now compute $\mathbb{P}\left(B_1\in\Pi_{t_1}^{(n)}, B_2\in\Pi_{t_2}^{(n)}\right)$ by cases.

i) Suppose that $B_1\cap B_2=\emptyset$. By exchangeability we have

$$\mathbb{P}\left(B_1\in\Pi_{t_1}^{(n)}, B_2\in\Pi_{t_2}^{(n)}\right) = \mathbb{P}(\{1,\cdots,b_1\}\in\Pi_{t_1}^{(n)}, \{b_1+1,\cdots,b_1+b_2\}\in\Pi_{t_2}^{(n)})$$

where this probability is of course 0 if $b_1+b_2>n$. Now suppose that $t_1\le t_2$. In terms of the RRT construction of the Bolthausen-Sznitman coalescent, the event

$$\{\{1,\cdots,b_1\}\in\Pi_{t_1}^{(n)}, \{b_1+1,\cdots,b_1+b_2\}\in\Pi_{t_2}^{(n)}\}$$

is characterized by a RRT with exponential edges, say $\mathcal{E}_2,\cdots,\mathcal{E}_n$, constructed as follows: for $i\in\{1,\cdots,b_1-1\}$ the node $\{i+1\}$ along with $\mathcal{E}_{i+1}$ arrive to the tree but with the imposed restriction that it may not attach to $\{1\}$ and have $\mathcal{E}_{i+1}>t_1$ at the same time, which occurs with probability e^{-t_1}/i; this ensures that $\{i+1\}$ coalesces with $\{1\}$ before time t_1 for all $i<b_1$, thus creating the block $\{1,\cdots,b_1\}$ up to time t_1. After $\{1\},\cdots,\{b_1\}$ have arrived, the node $\{b_1+1\}$ must attach to $\{1\}$ and $\mathcal{E}_{b_1+1}$ must be greater than t_2, which occurs with probability e^{-t_2}/b_1; the node $\{b_1+1\}$ will be the root of a sub-tree formed with the nodes $\{b_1+2\},\cdots,\{b_1+b_2\}$ which will build the block $\{b_1+1,\cdots,b_1+b_2\}$ at time t_2. Thus, for each

$i \in \{1, \cdots, b_2 - 1\}$ the node $\{b_1+i+1\}$ must arrive and attach to any of $\{b_1+1\}, \cdots, \{b_1+i\}$, which occurs with probability $\frac{i}{b_1+i}$, and, furthermore, conditional on this event, it may not attach to $\{b_1+1\}$ and have $\mathcal{E}_{b_1+i+1} > t_2$ at the same time, which occurs with probability $\frac{e^{-t_2}}{i}$. Finally, if $n - b_1 - b_2 > 0$, for $i \in \{0, \cdots, n-b_1-b_2-1\}$ the node $\{b_1+b_2+i+1\}$ must either attach to any of $\{b_1+b_2+j\}$, $1 \le j \le i$, or attach to $\{1\}$ or $\{b_1+1\}$ and have $\mathcal{E}_{b_1+b_2+i+1} > t_1$ or $\mathcal{E}_{b_1+b_2+i+1} > t_2$ respectively; this occurs with probability $\frac{e^{-t_1}+e^{-t_2}+i}{b_1+b_2+i}$. Putting all together we obtain

$$\begin{aligned}
&\mathbb{P}\left(B_1 \in \Pi^{(n)}_{t_1}, B_2 \in \Pi^{(n)}_{t_2}\right) \\
=&\left[\prod_{i=1}^{b_1-1}\left(1 - \frac{e^{-t_1}}{i}\right)\right]\left[\frac{e^{-t_2}}{b_1}\prod_{i=1}^{b_2-1}\left(1 - \frac{e^{-t_2}}{i}\right)\frac{i}{b_1+i}\right]\left[\prod_{i=0}^{n-b_1-b_2-1}\frac{e^{-t_1}+e^{-t_2}+i}{b_1+b_2+i}\right] \\
=&\frac{1}{(n-1)!}\frac{\Gamma(b_1-e^{-t_1})}{\Gamma(1-e^{-t_1})}e^{-t_2}\frac{\Gamma(b_2-e^{-t_2})}{\Gamma(1-e^{-t_2})}\frac{\Gamma(n-b_1-b_2+e^{-t_1}+e^{-t_2})}{\Gamma(e^{-t_1}+e^{-t_2})},
\end{aligned}$$

where the last product is set to 1 if $n - b_2 - b_1 = 0$. On the other hand, if $t_2 < t_1$, by exchangeability we may instead compute

$$\mathbb{P}(\{1, \cdots, b_2\} \in \Pi^{(n)}_{t_2}, \{b_2+1, \cdots, b_2+b_1\} \in \Pi^{(n)}_{t_1})$$

obtaining

$$\begin{aligned}
&\mathbb{P}\left(B_1 \in \Pi^{(n)}_{t_1}, B_2 \in \Pi^{(n)}_{t_2}\right) \\
=&\frac{1}{(n-1)!}\frac{\Gamma(b_2-e^{-t_2})}{\Gamma(1-e^{-t_2})}e^{-t_1}\frac{\Gamma(b_1-e^{-t_1})}{\Gamma(1-e^{-t_1})}\frac{\Gamma(n-b_2-b_1+e^{-t_2}+e^{-t_1})}{\Gamma(e^{-t_2}+e^{-t_1})}.
\end{aligned}$$

ii) Suppose that $B_1 \subset B_2$. Of course if $t_1 > t_2$ we have $\mathbb{P}\left(B_1 \in \Pi^{(n)}_{t_1}, B_2 \in \Pi^{(n)}_{t_2}\right) = 0$ whenever B_1 is strictly contained in B_2. Assuming that $t_1 \le t_2$ and using the same rationale as before we obtain

$$\begin{aligned}
&\mathbb{P}\left(B_1 \in \Pi^{(n)}_{t_1}, B_2 \in \Pi^{(n)}_{t_2}\right) \\
=&\left[\prod_{i=1}^{b_1-1}\frac{i-e^{-t_1}}{i}\right]\left[\prod_{i=0}^{b_2-b_1-1}\frac{i+e^{-t_1}-e^{-t_2}}{b_1+i}\right]\left[\prod_{i=0}^{n-b_2-1}\frac{e^{-t_2}+i}{b_2+i}\right] \\
=&\frac{1}{(n-1)!}\frac{\Gamma(b_1-e^{-t_1})}{\Gamma(1-e^{-t_1})}\frac{\Gamma(b_2-b_1+e^{-t_1}-e^{-t_2})}{\Gamma(e^{-t_1}-e^{-t_2})}\frac{\Gamma(n-b_2+e^{-t_2})}{\Gamma(e^{-t_2})},
\end{aligned}$$

where the product in the middle is set to 1 if $B_1 = B_2$.

iii) If $B_1 \cap B_2 \neq \emptyset$ and $B_1 \not\subset B_2$, we clearly have $\mathbb{P}\left(B_1 \in \Pi^{(n)}_{t_1}, B_2 \in \Pi^{(n)}_{t_2}\right) = 0$.

From the previous computations, and summing over the corresponding cases, we see that if $b_1 + b_2 \le n$ then, changing the variable $p = e^{-t}$, the integral in (3.25) is given by

$$\begin{aligned}\mathbb{E}\left[\ell_{n,b_1}\ell_{n,b_2}\right] =& \frac{n}{b_1! b_2! (n-b_1-b_2)!} \\ & \int_0^1 \int_0^1 \frac{\Gamma(b_1-p_1)}{\Gamma(1-p_1)} \frac{\Gamma(b_2-p_2)}{\Gamma(1-p_2)} \frac{\Gamma(n-b_1-b_2+p_1+p_2)}{\Gamma(p_1+p_2)} \frac{dp_1\, dp_2}{p_1 \vee p_2} \\ +& \frac{n}{b_1!(b_2-b_1)!(n-b_2)!} \\ & \int_0^1 \int_0^{p_1} \frac{\Gamma(b_1-p_1)}{\Gamma(1-p_1)} \frac{\Gamma(b_2-b_1+p_1-p_2)}{\Gamma(p_1-p_2)} \frac{\Gamma(n-b_2+p_2)}{\Gamma(p_2+1)} \frac{dp_2\, dp_1}{p_1}\end{aligned}$$

whereas if $b_1 + b_2 > n$ the first summand in the above expression is set to zero. Rearranging terms we obtain (3.17). □

Proof of Lemma 3.2.3 *(asymptotics for F_1).* Again, we have from Stirling's formula that $\Gamma(m+c)/\Gamma(m+d) = m^{c-d}(1+\mathcal{O}(1/m))$ for any real numbers c and d, where the $\mathcal{O}(1/m)$ term holds uniformly for $0 \le c, d \le 1$. Letting $m = b-1$ and $n-b$ leads to the following equality:

$$\begin{aligned}&\frac{n}{b(n-b)} \frac{\Gamma(n-b+p)}{\Gamma(n-b)} \frac{\Gamma(b-p)}{\Gamma(b)} \\ &\quad = \frac{n}{b(n-b)} (n-b)^p (b-1)^{-p} \left(1 + \mathcal{O}\left(\frac{1}{b}\right) + \mathcal{O}\left(\frac{1}{n-b}\right)\right).\end{aligned}$$

Thus, using Euler's reflection formula to write $\Gamma(1-p)\Gamma(1+p)$ as $\pi p/\sin(\pi p)$ in the definition of F_1, we get

$$\begin{aligned}F_1(n,b) &= \left(1 + \mathcal{O}\left(\frac{1}{b}\right) + \mathcal{O}\left(\frac{1}{n-b}\right)\right) \frac{1}{b(n-b)} \int_0^1 \frac{\sin(\pi p)}{\pi p} \left(\frac{n-b}{b-1}\right)^p dp \\ &= \left(1 + \mathcal{O}\left(\frac{1}{b}\right) + \mathcal{O}\left(\frac{1}{n-b}\right)\right) \frac{b-1}{b(n-1)^2} f_1\left(\frac{b-1}{n-1}\right)\end{aligned}$$

Thus, for every $\epsilon > 0$ there is a $b_0 \in \mathbb{N}$ such that for large enough $n \in \mathbb{N}$ we have

$$\max_{b_0 \le b \le n-b_0} \left| \frac{n^2 F_1(n,b)}{f_1\left(\frac{b-1}{n-1}\right)} - \frac{b-1}{b} \right| < \epsilon. \tag{3.26}$$

It remains to study the approximation as $n \to \infty$ in the cases where $n-b$ or b remain constant. In the first case, when $n-b = c$, we have $b \to \infty$ as $n \to \infty$ and, by Stirling's approximation and dominated convergence and substituting $p = y/\log b$ on the one hand

$$\begin{aligned}F_1(n,b) \sim& \int_0^1 \frac{\sin(\pi p)}{\pi p} b^{-p-1} \frac{\Gamma(c+p)}{\Gamma(c+1)}\, dp \\ =& \frac{1}{bc} \int_0^{\log b} \frac{\sin(\pi y/\log b)}{\pi y/\log b} e^{-y} \frac{\Gamma(c+y/\log b)}{\Gamma(c)} \frac{dy}{\log b} \\ \sim& \frac{1}{bc\log b} \int_0^\infty e^{-y}\, dy.\end{aligned}$$

and on the other hand because of $b \to \infty$

$$\begin{aligned}\frac{1}{n^2} f_1\left(\frac{b-1}{n-1}\right) \sim & \frac{1}{bc}\int_0^1 \frac{\sin(\pi p)}{\pi p} b^{-p} c^p \, dp \\ = & \frac{1}{bc}\int_0^{\log b} \frac{\sin(\pi y/\log b)}{\pi y/\log b} e^{-y} c^{y/\log b} \frac{dy}{\log b} \\ \sim & \frac{1}{bc\log b}\int_0^\infty e^{-y} \, dy.\end{aligned}$$

Thus $F_1(n,b) \sim n^{-2} f_1((b-1)/(n-1))$ which extends (3.26) for $b > n - b_0$.

Similarly for the second case, if $b \geq 2$ is fixed, we have $n - b \to \infty$ as $n \to \infty$. Thus, with $1 - p = y/\log n$

$$\begin{aligned}F_1(n,b) \sim & \int_0^1 \frac{\sin(\pi p)}{\pi p} \frac{\Gamma(b-p)}{\Gamma(b+1)} n^{p-1} \, dp \\ = & \frac{1}{\log^2(n)} \int_0^{\log n} \frac{\sin(\pi - \pi y/\log n)}{(1 - y/\log n)\pi y/\log n} \frac{\Gamma\left(b - 1 + \frac{y}{\log n}\right)}{\Gamma(b+1)} y e^{-y} \, dy \\ \sim & \frac{1}{b(b-1)\log^2 n} \int_0^\infty y e^{-y} \, dy\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n^2} f_1\left(\frac{b-1}{n-1}\right) \sim & \frac{1}{(b-1)^2}\int_0^1 \frac{\sin \pi p}{\pi p} (b-1)^{1-p} n^{p-1} \, dp \\ = & \frac{1}{(b-1)^2 \log^2 n} \int_0^{\log n} \frac{\sin(\pi - \pi y/\log n)}{(1 - y/\log n)\,\pi y/\log n} (b-1)^{y/\log n} y e^{-y} \, dy \\ \sim & \frac{1}{(b-1)^2 \log^2 n} \int_0^\infty y e^{-y} \, dy. \qquad (3.27)\end{aligned}$$

Thus $F_1(n,b) \sim (b-1) n^{-2} f_1((b-1)/(n-1))/b$, which extends (3.26) for $b < b_0$. This extends (3.26) for $b < b_0$. Thus we proved (3.20).

For the proof of (3.21), we substitute b by 1 and perform similar computations:

$$\begin{aligned}F_1(n,1) = & \int_0^1 \frac{\Gamma(1-p)}{\Gamma(2)} \frac{\Gamma(n-1+p)}{\Gamma(n)} \frac{dp}{\Gamma(1-p)\Gamma(1+p)} \\ \sim & \int_0^1 n^{p-1} \frac{dp}{\Gamma(1+p)} \\ = & \int_0^{\log n} e^{-y} \frac{dy}{(\log n)\Gamma(2 - y/\log n)} \\ \sim & \frac{1}{\log n} \int_0^\infty e^{-y} \, dy,\end{aligned}$$

and from (3.27) with choosing $b = 2$

$$\frac{1}{n^2} f_1\left(\frac{1}{n-1}\right) \sim \frac{1}{\log^2 n} \int_0^\infty y e^{-y} \, dy.$$

This proves (3.21). □

Proof of Lemma 3.2.3 *(asymptotics for F_2 and F_3)* . The arguments here are similar to the arguments in the proof of the asymptotics for F_1, but we avoid repeating similar and tedious computations. We only layout the first steps of the proof. By Stirling's approximation applied to the integrands appearing in F_2 and F_3, we obtain, for $b_2 - b_1 > 0$,

$$\frac{\Gamma(b_1-p_1)}{\Gamma(b_1+1)}\frac{\Gamma(b_2-b_1+p_1-p_2)}{\Gamma(b_2-b_1+1)}\frac{\Gamma(n-b_2+p_2)}{\Gamma(n-b_2+1)} =$$
$$\frac{1}{(n-1)^3}\left(\frac{b_1-1}{n-1}\right)^{-p_1-1}\left(\frac{b_2-b_1}{n-1}\right)^{p_1-p_2-1}\left(\frac{n-b_2}{n-1}\right)^{p_2-1}\times$$
$$\left(1+\mathcal{O}\left(\frac{1}{b_1}\right)+\mathcal{O}\left(\frac{1}{b_2-b_1}\right)+\mathcal{O}\left(\frac{1}{n-b_2}\right)\right),$$

and, for $n - b_2 - b_1 > 0$,

$$\frac{\Gamma(b_1-p_1)}{\Gamma(b_1+1)}\frac{\Gamma(b_2-p_2)}{\Gamma(b_2+1)}\frac{\Gamma(n-b_1-b_2+p_1+p_2)}{\Gamma(n-b_1-b_2+1)} =$$
$$\frac{1}{(n-2)^3}\left(\frac{b_1-1}{n-2}\right)^{-p_1-1}\left(\frac{b_2-1}{n-2}\right)^{-p_2-1}\left(1-\frac{b_1+b_2}{n-2}\right)^{p_1+p_2-1}\times$$
$$\left(1+\mathcal{O}\left(\frac{1}{b_1}\right)+\mathcal{O}\left(\frac{1}{b_2}\right)+\mathcal{O}\left(\frac{1}{n-b_1-b_2}\right)\right);$$

thus

$$F_2(n,b_1,b_2) = \frac{1}{(n-1)^3}f_2\left(\frac{b_1-1}{n-1},\frac{b_2-1}{n-1}\right)\left(1+\mathcal{O}\left(\frac{1}{b_1}\right)+\mathcal{O}\left(\frac{1}{b_2-b_1}\right)+\mathcal{O}\left(\frac{1}{n-b_2}\right)\right),$$

and

$$F_3(n,b_1,b_2) = \frac{1}{(n-2)^3}f_3\left(\frac{b_1-1}{n-2},\frac{b_2-1}{n-2}\right)\left(1+\mathcal{O}\left(\frac{1}{b_1}\right)+\mathcal{O}\left(\frac{1}{b_2}\right)+\mathcal{O}\left(\frac{1}{n-b_1-b_2}\right)\right).$$

Similar to the analysis in the proof of (3.20), to obtain (3.22) it remains to study the cases where at least one of $b_1, b_2 - b_1$, or $n - b_2$ remains constant, whereas for (3.23) the cases of interest are where one of b_1, b_2, or $n - b_2 - b_1$ remain constant. □

Proof of Theorem 3.2.4. We first derive the asymptotic behavior of the function f_1. We have

$$f_1(u) \sim \frac{1}{u^2\log^2 u} \quad \text{as } u \downarrow 0. \tag{3.28}$$

For the proof note that for $u < 1/2$ we have $(1-u)^{p-1} \le 2$. Therefore dominated convergence implies for $u \downarrow 0$

$$\begin{aligned}
f_1(u) &= \frac{1}{u^2}\int_0^1 u^{1-p}(1-u)^{p-1}\frac{dp}{\Gamma(1-p)\Gamma(1+p)}\\
&= \frac{1}{u^2}\int_0^1 e^{-(p-1)\log u}(1-u)^{p-1}(1-p)\frac{dp}{\Gamma(2-p)\Gamma(1+p)}\\
&= \frac{1}{u^2}\int_0^{-\log u} e^{-y}(1-u)^{y/\log\frac{1}{u}}\frac{y}{\log\frac{1}{u}}\cdot\frac{dy}{\log\frac{1}{u}\Gamma(1-\frac{y}{\log u})\Gamma(2+\frac{y}{\log u})}\\
&\sim \frac{1}{u^2\log^2 u}\int_0^\infty ye^{-y}\,dy
\end{aligned}$$

implying (3.28). Also

$$f_1(u) \sim -\frac{1}{(1-u)\log(1-u)} \qquad \text{as } u \uparrow 1, \tag{3.29}$$

which we obtain again by means of dominated convergence in the limit $u \uparrow 1$ as follows:

$$\begin{aligned} f_1(u) &= \frac{1}{u(1-u)} \int_0^1 e^{p\log(1-u)} u^{-p} \frac{dp}{\Gamma(1-p)\Gamma(1+p)} \\ &= \frac{1}{u(1-u)} \int_0^{-\log(1-u)} \frac{e^{-y} u^{y/\log(1-u)}\, dy}{(-\log(1-u))\Gamma(1+\frac{y}{\log(1-u)})\Gamma(1-\frac{y}{\log(1-u)})} \\ &\sim -\frac{1}{(1-u)\log(1-u)} \int_0^\infty e^{-y}\, dy. \end{aligned}$$

These asymptotics together with Lemma 3.2.3 imply our claims. Without loss of generality let $\theta = 1$. From (3.21) we obtain

$$\mathbb{E}\left[SFS_{n,1}\right] = nF_1(n,1) \sim \frac{1}{n} f_1\left(\frac{1}{n-1}\right) \sim \frac{\log n}{n} \frac{(n-1)^2}{\log^2(n-1)}$$

which yields claim (i).

Similary from (3.20) we get for $b \geq 2$ and $b/n \to 0$

$$\mathbb{E}\left[SFS_{n,b}\right] = nF_1(n,b) \sim \frac{b-1}{nb} f_1\left(\frac{b-1}{n-1}\right) \sim \frac{b-1}{nb} \frac{(n-1)^2}{(b-1)^2 \log^2 \frac{b-1}{n-1}}$$

which in view of $b/n \to 0$ yields assertion (ii).

Claim (iii) is an immediate consequence of formula (3.20), since here we have $(b-1)/b \to 1$.

Next under the condition $(n-b)/n \to 0$ we get from (3.20) and (3.29)

$$\mathbb{E}\left[SFS_{n,b}\right] \sim \frac{b-1}{nb} f_1\left(\frac{b-1}{n-1}\right) \sim -\frac{b-1}{nb} \frac{n-1}{(n-b)\log\frac{n-b}{n-1}} \sim \frac{1}{(n-b)\log\frac{n}{n-b}}$$

which confirms assertion (iv).

Finally, we have from (3.20)

$$\mathbb{E}\left[SFS_{n,I}\right] \sim \frac{1}{n} \sum_{\frac{b}{n} \in I} f_1\left(\frac{b}{n}\right) \sim \int_I f_1(u)\, du,$$

which is claim (v). This finishes the proof. □

Proof of Theorem 3.2.5. The approximations follow from the asymptotics for F_1, F_2, and F_3 substituted in the covariance formula given in Corollary 3.2.2. □

Proof of Corollary 3.2.6. We have to prove that

$$\mathbb{V}ar(SFS_{n,b}) = o\big(\mathbb{E}[SFS_{n,b}]^2\big).$$

From the monotonicity properties of the gamma function we have for $1 \le b \le n-1$

$$\begin{aligned}
F_2(n,b,b) &= \int_0^1 \int_0^{p_1} \frac{\Gamma(b-p_1)}{\Gamma(b+1)} \frac{\Gamma(n-b+p_2)}{\Gamma(n-b+1)} \frac{dp_2\, dp_1}{p_1 \Gamma(1-p_1)\Gamma(p_2+1)} \\
&\le \int_0^1 \frac{\Gamma(b-p_1)}{\Gamma(b+1)} \frac{\Gamma(n-b+p_1)}{\Gamma(n-b+1)} \frac{1}{\Gamma(1-p_1)p_1} \int_0^{p_1} \frac{\Gamma(1+p_1)}{\Gamma(1+p_1)\Gamma(1+p_2)}\, dp_2\, dp_1 \\
&\le \sup_{1\le x\le y\le 2} \frac{\Gamma(y)}{\Gamma(x)} \int_0^1 \frac{\Gamma(b-p_1)}{\Gamma(b+1)} \frac{\Gamma(n-b+p_1)}{\Gamma(n-b+1)} \frac{dp_1}{\Gamma(1-p_1)\Gamma(p_1+1)} \\
&= \sup_{1\le x\le y\le 2} \frac{\Gamma(y)}{\Gamma(x)} F_1(n,b). \qquad (3.30)
\end{aligned}$$

Concerning $F_3(n,b,b)$ we have for $b = o(n)$ by Stirling's approximation uniformly in $0 \le p_1, p_2 \le 1$

$$\frac{\Gamma(n-2b+p_1+p_2)}{\Gamma(n-2b+1)} \sim n \frac{\Gamma(n-b+p_1)}{\Gamma(n-b+1)} \frac{\Gamma(n-b+p_2)}{\Gamma(n-b+1)},$$

hence, with $1 < \eta < 2$

$$\begin{aligned}
&\iint\limits_{\substack{0\le p_1,p_2\le 1\\ \eta<p_1+p_2\le 2}} \frac{\Gamma(b-p_1)}{\Gamma(b+1)} \frac{\Gamma(b-p_2)}{\Gamma(b+1)} \frac{\Gamma(n-2b+p_1+p_2)}{\Gamma(n-2b+1)} \\
&\qquad \times \frac{dp_2\, dp_1}{\Gamma(1-p_1)\Gamma(1-p_2)(p_1\vee p_2)\Gamma(p_1+p_2)} \\
&\sim n \iint\limits_{\substack{0\le p_1,p_2\le 1\\ \eta<p_1+p_2\le 2}} \frac{\Gamma(b-p_1)}{\Gamma(b+1)} \frac{\Gamma(b-p_2)}{\Gamma(b+1)} \frac{\Gamma(n-b+p_1)}{\Gamma(n-b+1)} \frac{\Gamma(n-b+p_2)}{\Gamma(n-b+1)} \\
&\qquad \times \frac{dp_2\, dp_1}{\Gamma(1-p_1)\Gamma(1-p_2)(p_1\vee p_2)\Gamma(p_1+p_2)} \\
&\le \frac{n}{\eta-1} \sup_{\eta\le x\le 2} \frac{1}{\Gamma(x)} \int_0^1 \int_0^1 \frac{\Gamma(b-p_1)}{\Gamma(b+1)} \frac{\Gamma(b-p_2)}{\Gamma(b+1)} \frac{\Gamma(n-b+p_1)}{\Gamma(n-b+1)} \\
&\qquad \times \frac{\Gamma(n-b+p_2)}{\Gamma(n-b+1)} \frac{dp_2\, dp_1}{\Gamma(1-p_1)\Gamma(1-p_2)\Gamma(1+p_1)\Gamma(1+p_2)} \\
&= \frac{n}{\eta-1} \sup_{\eta\le x\le 2} \frac{1}{\Gamma(x)} F_1(n,b)^2. \qquad (3.31)
\end{aligned}$$

Also, by another application of Stirling's approximation and for $b = o(n)$

$$\iint\limits_{\substack{0\le p_1,p_2\le 1\\ 0<p_1+p_2\le\eta}} \frac{\Gamma(b-p_1)}{\Gamma(b+1)}\frac{\Gamma(b-p_2)}{\Gamma(b+1)}\frac{\Gamma(n-2b+p_1+p_2)}{\Gamma(n-2b+1)}$$

$$\times\frac{dp_2\,dp_1}{\Gamma(1-p_1)\Gamma(1-p_2)(p_1\vee p_2)\Gamma(p_1+p_2)}$$

$$= O\Big(\iint\limits_{\substack{0\le p_1,p_2\le 1\\ 0<p_1+p_2\le\eta}} b^{-p_1-p_2-2}(n-2b)^{p_1+p_2-1}$$

$$\times\frac{dp_2\,dp_1}{\Gamma(1-p_1)\Gamma(1-p_2)(p_1\vee p_2)\Gamma(p_1+p_2)}\Big)$$

$$= O\Big(b^{-\eta-2}(n-2b)^{\eta-1}\iint\limits_{0\le p_1 p_2\le 1}\frac{dp_2\,dp_1}{\Gamma(1-p_1)\Gamma(1-p_2)(p_1\vee p_2)\Gamma(p_1+p_2)}\Big)$$

$$= o\Big(\frac{n}{b^4\log^4 n}\Big) \tag{3.32}$$

Combining (3.31) and (3.32) with Theorem 3.2.4 (i) and (ii) and letting $\eta\to 2$ we obtain

$$F_3(n,b,b) = nF_1(n,b)^2(1+o(1)) + o(n^{-1}\mathbb{E}[SFS_{n,b}]^2).$$

Using this estimate together with (3.30) and with Theorem 3.2.1, Corollary 3.2.2 yields

$$\mathbb{V}ar(SFS_{n,b}) = O(\mathbb{E}[SFS_{n,b}]) + o(\mathbb{E}[SFS_{n,b}]^2)$$

Because of our assumption $\mathbb{E}[SFS_{n,b}]\to\infty$ our claim is proved. □

3.3 Distribution of the Family-Sized Branch Lengths

In this section we discuss the particular case of $\ell_{n,b}$ when $b > n/2$. In this case we are able to provide an explicit formula for the distribution function of the length of the coalescent of order b. This leads to convergence in law results, but also to the law of $SFS_{n,b}$. Observe that in this case, for all $t\ge 0$, $C_{n,b}(t)\in\{0,1\}$ and $\ell_{n,b}$ is just the time during which the block of size b survives before coalescing with other blocks (if it ever exists, otherwise obviously $\ell_{n,b}=0$). We first find an expression for the distribution function of $\ell_{n,b}$.

Theorem 3.3.1. *Suppose that $\frac{n}{2} < b < n$. For any $s\ge 0$,*

$$\mathbb{P}(\ell_{n,b} > s) = \frac{n}{(n-b)b(b-1)}\int_0^1 \frac{\Psi(b-p)-\Psi(1-p)}{\mathsf{B}(n-b,e^{-s}p)\mathsf{B}(b-1,1-p)}\,dp. \tag{3.33}$$

From the derived distribution of $\ell_{n,b}$ in Theorem 3.3.1 we obtain that, conditioned on $\ell_{n,b} > 0$, the variable $(\log n)\,\ell_{n,b}$ has a limiting distribution.

Corollary 3.3.2. *Suppose that $b/n \to u \in [1/2, 1)$ as $n \to \infty$, then letting $\alpha = \log(1-u) - \log u$, we have*

$$\frac{n}{\log n}\mathbb{P}(\ell_{n,b} > 0) \to \frac{G(\alpha)}{u(1-u)}$$

where

$$G(x) = \int_0^1 e^{px}\frac{\sin \pi p}{\pi}\, dp = \frac{1+e^x}{\pi^2+x^2}.$$

Furthermore,

$$\mathbb{P}((\log n)\, \ell_{n,b} > s | \ell_{n,b} > 0) \to \frac{G(\alpha - s)}{G(\alpha)}.$$

We now give the joint distribution of the branch lengths for large families, i.e. the joint distribution of the vector $(\ell_{n,b})_{b>n/2}$. For this we introduce the following events: for any collection of integers $\mathbf{b} = (b_1, \cdots, b_m)$ such that $n/2 < b_1 < b_2 < \cdots < b_m < n$, and any collection of nonnegative numbers $\mathbf{s} = (s_1, \cdots, s_m)$, define the event

$$\Lambda_{\mathbf{b},\mathbf{s}} := \left(\bigcap_{i=1}^m \{\ell_{b_i} > s_i\}\right) \bigcap \left(\bigcap_{\substack{b>b_1 \\ b\notin \mathbf{b}}} \{\ell_b = 0\}\right),$$

that is, the event that a block of size b_1 exists for a time larger than s_1, that this block then merges with some other blocks of total size exactly $b_2 - b_1$, that this new block exists for a time larger than s_2, and so on, until the last merge of the growing block occurs with the remaining blocks of total size exactly $n - b_m$.

Theorem 3.3.3. *For $\mathbf{b} = (b_1, \cdots, b_m)$ and $\mathbf{s} = (s_1, \cdots, s_m)$ as above, we have*

(3.34)

$$\mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}\right) = \frac{n}{b_1(b_2-b_1)\cdots(n-b_m)}\frac{\exp\{-\langle (m:1), \mathbf{s}\rangle\}}{m!}\int_0^1 p^m \frac{\Psi(b_1-p)-\Psi(1-p)}{\mathsf{B}(b_1-1, 1-p)}\, dp$$

and

(3.35)

$$\mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}, \bigcap_{n/2<b<b_1} \{\ell_{n,b} = 0\}\right)$$
$$= \frac{n}{(b_2-b_1)\cdots(n-b_m)}\frac{\exp\{-\langle (m:1), \mathbf{s}\rangle\}}{m!}\times$$
$$\left(\int_0^1 \frac{p^m}{b_1}\frac{\Psi(b_1-p)-\Psi(1-p)}{\mathsf{B}(b_1-1,1-p)} - \frac{p^{m+1}}{m+1}\sum_{n/2<b<b_1}\frac{1}{b(b_1-b)}\frac{\Psi(b-p)-\Psi(1-p)}{\mathsf{B}(b-1,1-p)}\, dp\right),$$

where

$$(m:1) := (m, m-1, \ldots, 1).$$

and $\langle \cdot, \cdot \rangle$ is the usual inner product in Euclidean space.

By conditioning on $(\ell_{n,b})_{b>n/2}$ and using equation (3.35) one can obtain a sampling formula for the vector $(SFS_{n,b})_{b>n/2}$, although the computations are rather convoluted and we do not present them here.

3.3.1 Proofs of Section 3.3

Proof of Theorem 3.3.1. Note that since $b > n/2$, and by the exchangeability of $\Pi^{(n)}$, we have:

$$\mathbb{P}(\ell_{n,b} > s) = \binom{n}{b}\mathbb{P}\left(\mathsf{Leb}(\{t : \{1, \cdots, b\} \in \Pi_t^{(n)}\}) > s\right),$$

where Leb is the Lebesgue measure, and $\mathsf{Leb}(\{t : \{1, \cdots, b\} \in \Pi_t^{(n)}\})$ gives the time that the block $\{1, \cdots, b\}$ exists in the Bolthausen-Sznitman coalescent starting with n individuals.

We now describe the event $\{\mathsf{Leb}(\{t : \{1, \cdots, b\}\} \in \Pi_t^{(n)}) > s\}$ in terms of the RRT construction of the Bolthausen-Sznitman coalescent. Let $\mathcal{G}$ be the event that the nodes $\{1\}, \{2\}, \cdots, \{b\}$ and $\{1\}, \{b+1\}, \cdots, \{n\}$ form two sub-trees, say T_1 and T_2 rooted at $\{1\}$; i.e.

$$\mathcal{G} :=\{T\colon \{j\} \text{ does not attach to } \{i\}, \text{ for all } 2 \le i \le b \text{ and } b < j \le n\}.$$

Then

$$\mathsf{Leb}(\{t : \{1, \cdots, b\}\} \in \Pi_t^{(n)}) = \begin{cases} 0 & \text{if } T \notin \mathcal{G} \\ (m\left(T_2\right) - M\left(T_1\right)) \vee 0 & \text{if } T \in \mathcal{G}. \end{cases}$$

Indeed, observe that by the cutting-merge procedure $T \notin \mathcal{G}$ if and only if any block of $\Pi^{(n)}$ that contains all of $\{1, \cdots, b\}$ also contains some $j \in \{b+1, \cdots, n\}$. On the other hand, on the event $\{T \in \mathcal{G}\}$, the random variable $M(T_1)$ is just the time at which the block $\{1, \cdots, b\}$ appears in $\Pi^{(n)}$, while $m(T_2)$ is the time at which it coalesces with some other block in T_2. Furthermore, observe that conditioned on $\{T \in \mathcal{G}\}$, T_1 and T_2 are two independent RRTs of sizes b and $n - b + 1$ respectively. Thus, by Lemma 3.1.3 we have

$$\begin{aligned}
&\mathbb{P}(\ell_{n,b} > s) \\
&= \binom{n}{b}\mathbb{P}(T \in \mathcal{G})\mathbb{P}(m(T_2) - M(T_1) > s) \\
&= \binom{n}{b}\prod_{i=0}^{n-b-1}\left(\frac{1+i}{b+i}\right)\frac{1}{(b-1)(n-b)}\int_0^1 \frac{\Psi(b-p) - \Psi(1-p)}{\mathsf{B}(n-b, e^{-s}p)\mathsf{B}(b-1, 1-p)}\, dp \\
&= \frac{n}{(n-b)b(b-1)}\int_0^1 \frac{\Psi(b-p) - \Psi(1-p)}{\mathsf{B}(n-b, e^{-s}p)\mathsf{B}(b-1, 1-p)}\, dp.
\end{aligned}$$

□

Proof of Corollary 3.3.2. Observe that, uniformly for $p \in (0, 1)$, we have

$$\Psi(b-p) - \Psi(1-p) = \sum_{k=1}^{b-1}\frac{1}{k-p} = \frac{1}{1-p} + \log b + \mathcal{O}\left(1\right),$$

thus, substituting in (3.33) and also using Stirling's approximation and Euler's reflection formula, we obtain

$$\begin{aligned}\mathbb{P}(\ell_{n,b} > 0) \sim & \frac{1}{u(1-u)n}\int_0^1 \left(\frac{1-u}{u}\right)^p \frac{\sin \pi p}{\pi}\left(\frac{1}{1-p} + \log n + \mathcal{O}(1)\right) dp \\ \sim & \frac{\log n}{u(1-u)n}\int_0^1 e^{p\alpha}\frac{\sin \pi p}{\pi}\, dp \\ = & \frac{\log n}{u(1-u)n} G(\alpha).\end{aligned}$$

On the other hand, for any $s > 0$ we have

$$\begin{aligned}\mathbb{P}\left((\log n)\,\ell_{n,b} > s\right) \sim & \frac{1}{u(1-u)n}\int_0^1 \frac{b^{-p}\,(n-b)^{pe^{-s/\log n}}}{\Gamma(1-p)\Gamma(pe^{-s/\log n})}\left(\frac{1}{1-p} + \log b + \mathcal{O}(1)\right) dp \\ \sim & \frac{\log n}{u(1-u)n}\int_0^1 e^{p\alpha}(n-b)^{p(e^{-s/\log n}-1)}\frac{1}{\Gamma(1-p)\Gamma(p)}\, dp \\ \sim & \frac{\log n}{u(1-u)n}\int_0^1 e^{p\alpha}(n-b)^{-ps/\log n}\frac{1}{\Gamma(1-p)\Gamma(p)}\, dp \\ \sim & \frac{\log n}{u(1-u)n}\int_0^1 e^{p(\alpha-s)}\frac{\sin \pi p}{\pi}\, dp \\ = & \frac{\log n}{u(1-u)n} G(\alpha - s).\end{aligned}$$

□

Proof of Theorem 3.3.3. Letting $\ell_\pi := \mathsf{Leb}\left(t : \pi \in \varPi_t^{(n)}\right)$ for any subset $\pi \subset [n]$, by exchangeability of $\varPi_t^{(n)}$ we have

$$\mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}\right) = \frac{n!}{b_1!(b_2-b_1)!\cdots(n-b_m)!}\mathbb{P}\left(\bigcap_{1\le i\le m} A_{b_i,s_i}, \bigcap_{\substack{b>b_1\\ b\notin \mathbf{b}}} \bar{A}_{b,0}\right)$$

where

$$A_{b,s} = \{\ell_{\{1,\dots,b\}} > s\}$$

and

$$\bar{A}_{b,0} = \{\ell_{\{1,\dots,b\}} = 0\}.$$

Recall that $M\left(T\big|_{b_1}\right)$ is defined as the maximum of the exponential edges associated to the root of $T\big|_{b_1}$. Letting $b_{m+1} := n$, and also letting $\mathcal{E}_b$, $1 \le b \le n$, be the exponential variable associated to b, we have

$$\begin{aligned}&\mathbb{P}\left(\bigcap_{1\le i\le m} A_{b_i,s_i}, \bigcap_{\substack{b>b_1\\ b\notin \mathbf{b}}} \bar{A}_{b,0}\right) \\ &= \left(\prod_{i=1}^{m+1} \frac{1\cdot 2\cdots(b_{i+1}-b_i)}{b_i(b_i+1)\cdots(b_{i+1}-1)}\right)\mathbb{P}\left(\mathcal{E}_{b_1+1} - M\left(T\big|_{b_1}\right) > s_1, \bigcap_{i=2}^{m} \mathcal{E}_{b_i+1} - \mathcal{E}_{b_{i-1}+1} > s_i\right),\end{aligned}$$

where the product above is the probability that T is structured in such a way that $\{b_1+1\}$ attaches to $\{1\}$ and is the root of a subtree formed with $\{b_1+1,\dots,b_2\}$, that $\{b_2+1\}$ attaches to $\{1\}$ and is the root of a subtree formed with $\{b_2+1,\dots,b_3\}$, and so forth. Using the independence of the exponential variables we obtain

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}_{b_1+1} - M\left(T\big|_{b_1}\right) > s_1, \bigcap_{i=2}^{m}\{\mathcal{E}_{b_i+1} - \mathcal{E}_{b_{i-1}+1} > s_i\}\right)\\
&= \int_0^\infty dt_1 \int_{t_1+s_1}^\infty dt_2 \cdots \int_{t_m+s_m}^\infty dt_{m+1}\left(\frac{d}{dt_1}\mathbb{P}\left(M(T\big|_{b_1}) \le t_1\right)\right) e^{-t_2}\dots e^{-t_{m+1}}\\
&= \int_0^\infty dt_1 \int_{t_1+s_1}^\infty dt_2 \cdots \int_{t_{m-1}+s_{m-1}}^\infty dt_m\left(\frac{d}{dt_1}\mathbb{P}\left(M(T\big|_{b_1}) \le t_1\right)\right) e^{-t_2}\dots e^{-2t_m}e^{-s_m}\\
&\vdots\\
&= \frac{\exp\{-\langle (m:1),\mathbf{s}\rangle\}}{m!}\int_0^\infty e^{-mt_1}\frac{d}{dt_1}\mathbb{P}\left(M(T\big|_{b_1}) \le t\right)\, dt_1.
\end{aligned}$$

From (3.11) and making $p = e^{-t}$ in the above integral, and putting all together we obtain (3.34). Finally (3.35) follows from

$$\mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}, \ell_{n,b_1-1} = 0\right) = \mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}\right) - \mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}, \ell_{n,b_1-1} > 0\right)$$

and, recursively,

$$\mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}, \bigcap_{n/2<b<b_1}\{\ell_{n,b}=0\}\right) = \mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}\right) - \sum_{n/2<b<b_1}\mathbb{P}\left(\Lambda_{\mathbf{b},\mathbf{s}}, \ell_{n,b}>0, \bigcap_{i=1}^{b_1-b-1}\{\ell_{n,b+i}=0\}\right).$$

Substituting (3.34) in the above expression, we obtain (3.35). □

3.4 Proof of the approximations

Here we derive the approximations given in the beginning of the chapter. From Stirling's approximation we have the well-known formula $\Gamma(m+c)/\Gamma(m) \approx m^c$. Its application requires some care, since we shall apply this approximation also for small values of m down to $m=1$. It is known and easily confirmed by computer that the approximation is particularly accurate within the range $0 \le c \le 1$. Thus we use for $p \in (0,1)$ and $b \ge 2$ the approximations

$$\frac{\Gamma(b-p)}{\Gamma(b+1)} = \frac{1}{b(b-1)}\frac{\Gamma(b-1+(1-p))}{\Gamma(b-1)} \approx \frac{1}{b(b-1)}(b-1)^{1-p} = \frac{(b-1)^{-p}}{b}$$

and

$$\frac{\Gamma(n-b+p)}{\Gamma(n-b+1)} = \frac{1}{n-b}\frac{\Gamma(n-b+p)}{\Gamma(n-b)} \approx (n-b)^{p-1}.$$

Also by Euler's reflection formula $\Gamma(1-p)\Gamma(1+p) = \pi p/\sin(\pi p)$. Inserting these formulas into the expression (3.16) for the expected SFS we obtain

$$\begin{aligned}\mathbb{E}[SFS_{n,b}] &\approx \theta n \frac{b-1}{b} \int_0^1 (b-1)^{-p-1}(n-b)^{p-1} \frac{\sin(\pi p)}{\pi p}\, dp \\ &= \theta \frac{n}{(n-1)^2} \frac{b-1}{b} f_1\Big(\frac{b-1}{n-1}\Big).\end{aligned}$$

It turns out that this approximation overestimates the expected SFS, which can be somewhat counterbalanced by replacing the scaling factor $n/(n-1)^2$ by $1/(n-1)$. This yields our first approximation (3.1).

For the second approximation (3.4) we apply the expansion

$$\frac{\Gamma(m+c)}{\Gamma(m)} = m^c\Big(1 - \frac{c(1-c)}{2m} + O(m^{-2})\Big),$$

see Erdélyi and Tricomi 1951. Again this approximation is particularly accurate for $0 \le c \le 1$ leading for $p \in (0,1)$ and $b \ge 2$ to

$$\begin{aligned}\frac{\Gamma(b-p)}{\Gamma(b+1)} \frac{\Gamma(n-b+p)}{\Gamma(n-b+1)} &\approx \frac{(b-1)^{-p}}{b}(n-b)^{p-1}\Big(1 - \frac{(1-p)p}{2(b-1)}\Big)\Big(1 - \frac{p(1-p)}{2(n-b)}\Big) \\ &\approx \frac{(b-1)^{-p}}{b}(n-b)^{p-1}\Big(1 - (n-1)\frac{p(1-p)}{2(b-1)(n-b)}\Big).\end{aligned}$$

Using this approximation in the expression for the expected SFS we get for $b \ge 2$

$$\begin{aligned}\mathbb{E}[SFS_{n,b}] \approx \theta n \frac{b-1}{b}\Bigg(&\frac{1}{(n-1)^2} f_1\Big(\frac{b-1}{n-1}\Big) \\ &- \frac{n-1}{2}\int_0^1 (b-1)^{-p-2}(n-b)^{p-2}\frac{\sin(\pi p)}{\pi}(1-p)\, dp\Bigg) \\ = \theta n \frac{b-1}{b}\Big(&\frac{1}{(n-1)^2} f_1\Big(\frac{b-1}{n-1}\Big) - \frac{1}{(n-1)^3} g_1\Big(\frac{b-1}{n-1}\Big)\Big)\end{aligned}$$

with the function g_1 as defined in (3.5). This integral can be evaluated by elementary means yielding formula (3.4).

Bibliography

Aryal, Gokarna and S. Nadarajah (Jan. 2004). “Information matrix for beta distributions”. In: *Serdica Mathematical Journal* 30.

Basdevant, Anne-Laure and Christina Goldschmidt (2008). “Asymptotics of the Allele Frequency Spectrum Associated with the Bolthausen-Sznitman Coalescent”. In: *Electronic Journal of Probability* 13.none, pp. 486–512. DOI: 10.1214/EJP.v13-494. URL: https://doi.org/10.1214/EJP.v13-494.

Berestycki, Julien, Nathanaël Berestycki, and Vlada Limic (2010). “The Λ-coalescent speed of coming down from infinity”. In: *The Annals of Probability* 38.1, pp. 207–233. DOI: 10.1214/09-AOP475. URL: https://doi.org/10.1214/09-AOP475.

— (2014). “Asymptotic sampling formulae for Λ-coalescents”. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 50.3, pp. 715–731. DOI: 10.1214/13-AIHP546. URL: https://doi.org/10.1214/13-AIHP546.

Berestycki, Julien, Nathanaël Berestycki, and Jason Schweinsberg (2007). “Beta-coalescents and continuous stable random trees”. In: *The Annals of Probability* 35.5, pp. 1835–1887. DOI: 10.1214/009117906000001114. URL: https://doi.org/10.1214/009117906000001114.

— (2008). “Small-time behavior of beta coalescents”. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 44.2, pp. 214–238. DOI: 10.1214/07-AIHP103. URL: https://doi.org/10.1214/07-AIHP103.

— (2013). “THE GENEALOGY OF BRANCHING BROWNIAN MOTION WITH ABSORPTION”. In: *The Annals of Probability* 41.2, pp. 527–618. ISSN: 00911798. URL: http://www.jstor.org/stable/23469426.

Bertoin, Jean (2006). *Random Fragmentation and Coagulation Processes*. Cambridge Studies in Advanced Mathematics. Cambridge University Press. DOI: 10.1017/CBO9780511617768.

— (May 2008). “Two-Parameter Poisson–Dirichlet Measures and Reversible Exchangeable Fragmentation–Coalescence Processes”. In: *Comb. Probab. Comput.* 17.3, pp. 329–337. ISSN: 0963-5483. DOI: 10.1017/S0963548307008784. URL: https://doi.org/10.1017/S0963548307008784.

Bertoin, Jean and Jean-François Le Gall (June 2000). “The Bolthausen–Sznitman coalescent and the genealogy of continuous-state branching processes”. In: *Probability Theory and Related Fields* 117, pp. 249–266. DOI: 10.1007/s004400050006.

— (June 2003). “Stochastic flows associated to coalescent processes”. In: *Probability Theory and Related Fields* 126.2, pp. 261–288. ISSN: 1432-2064. DOI: 10.1007/s00440-003-0264-4. URL: https://doi.org/10.1007/s00440-003-0264-4.

Bertoin, Jean and Jean-François Le Gall (2006). "Stochastic flows associated to coalescent processes. III. Limit theorems". In: *Illinois Journal of Mathematics* 50.1-4, pp. 147–181. DOI: 10.1215/ijm/1258059473. URL: https://doi.org/10.1215/ijm/1258059473.

Billingsley, Patrick (1999). *Convergence of probability measures.* 2nd ed. Wiley series in probability and statistics. Probability and statistics section. New York: Wiley. ISBN: 9780471197454.

Birkner, Matthias, Jochen Blath, et al. (2005). "Alpha-Stable Branching and Beta-Coalescents". In: *Electronic Journal of Probability* 10.none, pp. 303–325. DOI: 10.1214/EJP.v10-241. URL: https://doi.org/10.1214/EJP.v10-241.

Birkner, Matthias, Huili Liu, and Anja Sturm (2018). "Coalescent results for diploid exchangeable population models". In: *Electronic Journal of Probability* 23.none, pp. 1–44. DOI: 10.1214/18-EJP175. URL: https://doi.org/10.1214/18-EJP175.

Bolthausen, Eric and Alain-Sol Sznitman (Oct. 1998). "On Ruelle's Probability Cascades and an Abstract Cavity Method". In: *Communications in Mathematical Physics* 197.2, pp. 247–276. ISSN: 1432-0916. DOI: 10.1007/s002200050450. URL: https://doi.org/10.1007/s002200050450.

Brunet, Éric and Bernard Derrida (Sept. 1997). "Shift in the velocity of a front due to a cutoff". In: *Phys. Rev. E* 56 (3), pp. 2597–2604. DOI: 10.1103/PhysRevE.56.2597. URL: https://link.aps.org/doi/10.1103/PhysRevE.56.2597.

— (2012). "How genealogies are affected by the speed of evolution". In: *Philosophical Magazine* 92.1-3, pp. 255–271. DOI: 10.1080/14786435.2011.620028. URL: https://doi.org/10.1080/14786435.2011.620028.

Brunet, Éric, Bernard Derrida, et al. (Oct. 2007). "Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization". In: *Phys. Rev. E* 76 (4), p. 041104. DOI: 10.1103/PhysRevE.76.041104. URL: https://link.aps.org/doi/10.1103/PhysRevE.76.041104.

Cannings, Chris (June 1974). "The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models". In: *Advances in Applied Probability* 6.2, pp. 260–290. DOI: 10.2307/1426293. URL: https://doi.org/10.2307/1426293.

— (June 1975). "The latent roots of certain Markov chains arising in genetics: A new approach, II. Further haploid models". In: *Advances in Applied Probability* 7.2, pp. 264–282. DOI: 10.2307/1426077. URL: https://doi.org/10.2307/1426077.

Cortines, Aser and Bastien Mallein (2017). "A N-branching random walk with random selection". In: *ALEA Lat. Am. J. Probab. Math. Stat.* 14.1, pp. 117–137. DOI: 10.30757/alea.v14-07. URL: https://doi.org/10.30757/alea.v14-07.

— (2018). "The genealogy of an exactly solvable Ornstein–Uhlenbeck type branching process with selection". In: *Electronic Communications in Probability* 23.none, pp. 1–13. DOI: 10.1214/18-ECP197. URL: https://doi.org/10.1214/18-ECP197.

Delmas, Jean-François, Jean-Stéphane Dhersin, and Arno Siri-Jégousse (2008). "Asymptotic results on the length of coalescent trees". In: *The Annals of Applied Probability* 18.3, pp. 997–1025. DOI: 10.1214/07-AAP476. URL: https://doi.org/10.1214/07-AAP476.

Desai, Michael M. and Daniel S. Fisher (2007). "Beneficial Mutation–Selection Balance and the Effect of Linkage on Positive Selection". In: *Genetics* 176.3, pp. 1759–1798. ISSN: 0016-6731. DOI: 10.1534/genetics.106.067678. URL: https://www.genetics.org/content/176/3/1759.

Desai, Michael M., Aleksandra M. Walczak, and Daniel S. Fisher (2013). “Genetic Diversity and the Structure of Genealogies in Rapidly Adapting Populations”. In: *Genetics* 193.2, pp. 565–585. ISSN: 0016-6731. DOI: 10.1534/genetics.112.147157. URL: https://www.genetics.org/content/193/2/565.

Dhersin, Jean-Stéphane, Fabian Freund, et al. (2013). “On the length of an external branch in the Beta-coalescent”. In: *Stochastic Processes and their Applications* 123.5, pp. 1691–1715. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2012.12.010. URL: https://www.sciencedirect.com/science/article/pii/S0304414912002712.

Dhersin, Jean-Stéphane and Martin Möhle (2013). “On the external branches of coalescents with multiple collisions”. In: *Electronic Journal of Probability* 18.none, pp. 1–11. DOI: 10.1214/EJP.v18-2286. URL: https://doi.org/10.1214/EJP.v18-2286.

Diehl, Christina S. and Götz Kersting (2019). “Tree lengths for general Λ-coalescents and the asymptotic site frequency spectrum around the Bolthausen–Sznitman coalescent”. In: *The Annals of Applied Probability* 29.5, pp. 2700–2743. DOI: 10.1214/19-AAP1462. URL: https://doi.org/10.1214/19-AAP1462.

Drmota, Michael et al. (2007). “Asymptotic results concerning the total branch length of the Bolthausen–Sznitman coalescent”. In: *Stochastic Processes and their Applications* 117.10, pp. 1404–1421. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2007.01.011. URL: https://www.sciencedirect.com/science/article/pii/S030441490700018X.

— (2009). “A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree”. In: *Random Structures & Algorithms* 34.3, pp. 319–336. DOI: https://doi.org/10.1002/rsa.20233. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20233.

Eldon, Bjarki et al. (2015). “Can the Site-Frequency Spectrum Distinguish Exponential Population Growth from Multiple-Merger Coalescents?” In: *Genetics* 199.3, pp. 841–856. ISSN: 0016-6731. DOI: 10.1534/genetics.114.173807. URL: https://www.genetics.org/content/199/3/841.

Erdélyi, Arthur and Francesco G. Tricomi (1951). “The asymptotic expansion of a ratio of gamma functions.” In: *Pacific Journal of Mathematics* 1.1, pp. 133–142. DOI: pjm/1102613160. URL: https://doi.org/.

Ethier, Stewart N. and Thomas G. Kurtz (1986). *Markov processes : characterization and convergence*. New York: Wiley. ISBN: 978-0-470-31732-7.

Ewens, Warren J. (1972). “The sampling theory of selectively neutral alleles”. In: *Theoretical Population Biology* 3.1, pp. 87–112. ISSN: 0040-5809. DOI: https://doi.org/10.1016/0040-5809(72)90035-4. URL: https://www.sciencedirect.com/science/article/pii/0040580972900354.

Foucart, Clément, Chunhua Ma, and Bastien Mallein (2019). “Coalescences in continuous-state branching processes”. In: *Electronic Journal of Probability* 24.none, pp. 1–52. DOI: 10.1214/19-EJP358. URL: https://doi.org/10.1214/19-EJP358.

Freund, Fabian (Apr. 2020). “Cannings models, population size changes and multiple-merger coalescents”. In: *Journal of Mathematical Biology* 80.5, pp. 1497–1521. ISSN: 1432-1416. DOI: 10.1007/s00285-020-01470-5. URL: https://doi.org/10.1007/s00285-020-01470-5.

Freund, Fabian and Arno Siri-Jégousse (2021). “The impact of genetic diversity statistics on model selection between coalescents”. In: *Computational Statistics and Data Analysis*

156, p. 107055. ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2020.107055. URL: https://www.sciencedirect.com/science/article/pii/S0167947320301468.

Fu, Yun-Xin (1995). "Statistical Properties of Segregating Sites". In: *Theoretical Population Biology* 48.2, pp. 172–197. ISSN: 0040-5809. DOI: https://doi.org/10.1006/tpbi.1995.1025. URL: https://www.sciencedirect.com/science/article/pii/S0040580985710258.

Gnedin, Alexander, Alexander Iksanov, and Alexander Marynych (2011). "On Λ-coalescents with dust component". In: *Journal of Applied Probability* 48.4, pp. 1133–1151. DOI: 10.1239/jap/1324046023. URL: https://doi.org/10.1239/jap/1324046023.

Gnedin, Alexander, Alexander Iksanov, Alexander Marynych, and Martin Möhle (2014). "On asymptotics of the beta coalescents". In: *Advances in Applied Probability* 46.2, pp. 496–515. DOI: 10.1239/aap/1401369704. URL: https://doi.org/10.1239/aap/1401369704.

Gnedin, Alexander and Yuri Yakubovich (2007). "On the Number of Collisions in Λ-Coalescents". In: *Electronic Journal of Probability* 12.none, pp. 1547–1567. DOI: 10.1214/EJP.v12-464. URL: https://doi.org/10.1214/EJP.v12-464.

Goldschmidt, Christina and James Martin (2005). "Random Recursive Trees and the Bolthausen-Sznitman Coalesent". In: *Electronic Journal of Probability* 10.none, pp. 718–745. DOI: 10.1214/EJP.v10-265. URL: https://doi.org/10.1214/EJP.v10-265.

González Casanova, Adrián, Verónica Miró Pina, and Arno Siri-Jégousse (2020). "The Symmetric Coalescent and Wright-Fisher models with bottlenecks". In:

Hobolth, Asger, Arno Siri-Jégousse, and Mogens Bladt (2019). "Phase-type distributions in population genetics". In: *Theoretical Population Biology* 127, pp. 16–32. ISSN: 0040-5809. DOI: https://doi.org/10.1016/j.tpb.2019.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0040580919300140.

Huillet, Thierry and Martin Möhle (2013). "On the extended Moran model and its relation to coalescents with multiple collisions". In: *Theoretical Population Biology* 87. Coalescent Theory, pp. 5–14. ISSN: 0040-5809. DOI: https://doi.org/10.1016/j.tpb.2011.09.004. URL: https://www.sciencedirect.com/science/article/pii/S0040580911000840.

— (2021). arXiv: 2106.10939 [math.PR].

Iksanov, Alexander, Alexander Marynych, and Martin Möhle (2009). "On the number of collisions in beta $(2, b)$-coalescents". In: *Bernoulli* 15.3, pp. 829–845. DOI: 10.3150/09-BEJ192. URL: https://doi.org/10.3150/09-BEJ192.

Iksanov, Alexander and Martin Möhle (2007). "A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree". In: *Electronic Communications in Probability* 12.none, pp. 28–35. DOI: 10.1214/ECP.v12-1253. URL: https://doi.org/10.1214/ECP.v12-1253.

— (2008). "On the Number of Jumps of Random Walks with a Barrier". In: *Advances in Applied Probability* 40.1, pp. 206–228. ISSN: 00018678. URL: http://www.jstor.org/stable/20443577.

Kaj, Ingemar and Stephen M. Krone (2003). "The coalescent process in a population with stochastically varying size". In: *Journal of Applied Probability* 40.1, pp. 33–48. DOI: 10.1239/jap/1044476826.

Kersting, Götz (2012). “The asymptotic distribution of the length of Beta-coalescent trees”. In: *The Annals of Applied Probability* 22.5, pp. 2086–2107. DOI: 10.1214/11-AAP827. URL: https://doi.org/10.1214/11-AAP827.

Kersting, Götz, J. C. Pardo, and Arno Siri-Jégousse (2014). “Total internal and external lengths of the Bolthausen-Sznitman coalescent”. In: *Journal of Applied Probability* 51A, pp. 73–86. ISSN: 00219002. URL: http://www.jstor.org/stable/43284110.

Kersting, Götz, Jason Schweinsberg, and Anton Wakolbinger (2014). “The evolving beta coalescent”. In: *Electronic Journal of Probability* 19.none, pp. 1–27. DOI: 10.1214/EJP.v19-3332. URL: https://doi.org/10.1214/EJP.v19-3332.

Kersting, Götz, Arno Siri-Jégousse, and Alejandro H. Wences (2021). “Site Frequency Spectrum of the Bolthausen-Sznitman Coalescent”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 18.1, p. 1483. DOI: 10.30757/alea.v18-53. URL: https://doi.org/10.30757/alea.v18-53.

Kingman, John (1978). “The Representation of Partition Structures”. In: *Journal of the London Mathematical Society* s2-18.2, pp. 374–380. DOI: https://doi.org/10.1112/jlms/s2-18.2.374. eprint: https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/jlms/s2-18.2.374. URL: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/s2-18.2.374.

— (1982). “The coalescent”. In: *Stochastic Processes and their Applications* 13.3, pp. 235–248. ISSN: 0304-4149. DOI: https://doi.org/10.1016/0304-4149(82)90011-4. URL: https://www.sciencedirect.com/science/article/pii/0304414982900114.

— (1992). *Poisson Processes*. Oxford Studies in Probability. Clarendon Press. ISBN: 9780191591242. URL: https://books.google.com.mx/books?id=VEiM-OtwDHkC.

Koskela, Jere (2018). “Multi-locus data distinguishes between population growth and multiple merger coalescents”. In: *Statistical Applications in Genetics and Molecular Biology* 17.3, p. 20170011. DOI: doi:10.1515/sagmb-2017-0011. URL: https://doi.org/10.1515/sagmb-2017-0011.

Kukla, Jonas and Helmut Pitters (2015). “A spectral decomposition for the Bolthausen-Sznitman coalescent and the Kingman coalescent”. In: *Electronic Communications in Probability* 20.none, pp. 1–13. DOI: 10.1214/ECP.v20-4612. URL: https://doi.org/10.1214/ECP.v20-4612.

Limic, Vlada (2010). “On the Speed of Coming Down from Infinity for Ξ-Coalescent Processes”. In: *Electronic Journal of Probability* 15.none, pp. 217–240. DOI: 10.1214/EJP.v15-742. URL: https://doi.org/10.1214/EJP.v15-742.

Melissa, Matthew J. et al. (2021). “Population genetics of polymorphism and divergence in rapidly evolving populations”. In: *bioRxiv*. DOI: 10.1101/2021.06.28.450258. eprint: https://www.biorxiv.org/content/early/2021/06/30/2021.06.28.450258.full.pdf. URL: https://www.biorxiv.org/content/early/2021/06/30/2021.06.28.450258.

Möhle, Martin (2010). “Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson–Dirichlet coalescent”. English. In: *Stochastic Processes and their Applications* 120.11, pp. 2159–2173. DOI: 10.1016/j.spa.2010.07.004.

Möhle, Martin and Helmut Pitters (2014). “A spectral decomposition for the block counting process of the Bolthausen-Sznitman coalescent”. In: *Electronic Communications in Prob-*

ability 19.none, pp. 1–11. DOI: 10.1214/ECP.v19-3464. URL: https://doi.org/10.1214/ECP.v19-3464.
Möhle, Martin and Serik Sagitov (2001). "A Classification of Coalescent Processes for Haploid Exchangeable Population Models". In: *The Annals of Probability* 29.4, pp. 1547–1562. DOI: 10.1214/aop/1015345761. URL: https://doi.org/10.1214/aop/1015345761.
Muqattash, Isa and Mohammed Yahdi (2006). "Infinite family of approximations of the Digamma function". In: *Mathematical and Computer Modelling* 43.11, pp. 1329–1336. ISSN: 0895-7177. DOI: https://doi.org/10.1016/j.mcm.2005.02.010. URL: https://www.sciencedirect.com/science/article/pii/S0895717705004735.
Neher, Richard A. and Oskar Hallatschek (2013). "Genealogies of rapidly adapting populations". In: *Proceedings of the National Academy of Sciences* 110.2, pp. 437–442. ISSN: 0027-8424. DOI: 10.1073/pnas.1213113110.
Pitman, Jim (1999). "Coalescents With Multiple Collisions". In: *The Annals of Probability* 27.4, pp. 1870–1902. DOI: 10.1214/aop/1022874819. URL: https://doi.org/10.1214/aop/1022874819.
Pitman, Jim and Marc Yor (1997). "The Two-Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator". In: *The Annals of Probability* 25.2, pp. 855–900. ISSN: 00911798. URL: http://www.jstor.org/stable/2959614.
Sagitov, Serik (1999). "The general coalescent with asynchronous mergers of ancestral lines". In: *Journal of Applied Probability* 36.4, pp. 1116–1125. DOI: 10.1239/jap/1032374759.
Schweinsberg, Jason (2000a). "A Necessary and Sufficient Condition for the Λ-Coalescent to Come Down from Infinity." In: *Electronic Communications in Probability* 5.none, pp. 1–11. DOI: 10.1214/ECP.v5-1013. URL: https://doi.org/10.1214/ECP.v5-1013.
— (2000b). "Coalescents with Simultaneous Multiple Collisions". In: *Electronic Journal of Probability* 5.none, pp. 1–50. DOI: 10.1214/EJP.v5-68. URL: https://doi.org/10.1214/EJP.v5-68.
— (2003). "Coalescent processes obtained from supercritical Galton–Watson processes". In: *Stochastic Processes and their Applications* 106.1, pp. 107–139. ISSN: 0304-4149. DOI: https://doi.org/10.1016/S0304-4149(03)00028-0. URL: https://www.sciencedirect.com/science/article/pii/S0304414903000280.
— (2017a). "Rigorous results for a population model with selection I: evolution of the fitness distribution". In: *Electronic Journal of Probability* 22.none, pp. 1–94. DOI: 10.1214/17-EJP57. URL: https://doi.org/10.1214/17-EJP57.
— (2017b). "Rigorous results for a population model with selection II: genealogy of the population". In: *Electronic Journal of Probability* 22.none, pp. 1–54. DOI: 10.1214/17-EJP58. URL: https://doi.org/10.1214/17-EJP58.
Spence, Jeffrey P., John A. Kamm, and Yun S. Song (2016). "The Site Frequency Spectrum for General Coalescents". In: *Genetics* 202.4, pp. 1549–1561. ISSN: 0016-6731. DOI: 10.1534/genetics.115.184101. URL: https://www.genetics.org/content/202/4/1549.