Electronic Journal of Probability

Vol. 13 (2008), Paper no. 17, pages 486–512.

Journal URL
http://www.math.washington.edu/~ejpecp/

Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent

Anne-Laure Basdevant
Institut de Mathématiques
Université Paul Sabatier (Toulouse III).
31062 Toulouse Cedex 9, France.
abasdeva@math.ups-tlse.fr
http://www.math.univ-toulouse.fr/~abasdeva

Christina Goldschmidt
Department of Statistics
University of Oxford.
1 South Parks Road
Oxford OX1 3TG, UK.
goldschm@stats.ox.ac.uk
http://www.stats.ox.ac.uk/~goldschm

Abstract

We consider a coalescent process as a model for the genealogy of a sample from a population. The population is subject to neutral mutation at constant rate ρ per individual and every mutation gives rise to a completely new type. The allelic partition is obtained by tracing back to the most recent mutation for each individual and grouping together individuals whose most recent mutations are the same. The allele frequency spectrum is the sequence $(N_1(n), N_2(n), \ldots, N_n(n))$, where $N_k(n)$ is number of blocks of size k in the allelic partition with sample size n. In this paper, we prove law of large numbers-type results for the allele frequency spectrum when the coalescent process is taken to be the Bolthausen-Sznitman coalescent. In particular, we show that $n^{-1}(\log n)N_1(n) \xrightarrow{p} \rho$ and, for $k \geq 2$, $n^{-1}(\log n)^2 N_k(n) \xrightarrow{p} \rho/(k(k-1))$ as $n \to \infty$. Our method of proof involves tracking the formation of the allelic partition using a certain Markov process, for which we prove a fluid limit.

Key words: Allelic partition, Bolthausen-Sznitman coalescent, mutation, fluid limit.

AMS 2000 Subject Classification: Primary 60C05; Secondary: 05A18, 60F05, 60J75.

Submitted to EJP on July 16, 2007, final version accepted March 17, 2008.

1 Introduction

1.1 Exchangeable random partitions

In recent years, the topic of exchangeable random partitions has received a lot of attention (see Pitman [35] for a lucid introduction). A random partition of $\mathbb{N}$ is said to be *exchangeable* if, for any permutation $\sigma : \mathbb{N} \to \mathbb{N}$ such that $\sigma(i) = i$ for all i sufficiently large, we have that the distribution of the partition is unaffected by the application of σ. It was proved by Kingman [28; 29] that if the partition has blocks $(B_i, i \geq 1)$ listed in increasing order of least elements then the *asymptotic frequencies*,

$$f_i \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{|B_i \cap \{1, 2, \ldots, n\}|}{n}, \quad i \geq 1,$$

exist almost surely. Let $(f_i^{\downarrow})_{i\geq 1}$ be the collection of asymptotic frequencies ranked in decreasing order. Then we can view $(f_i^{\downarrow})_{i\geq 1}$ as a partition of $[0, 1]$ into intervals of decreasing length. In general, since it is possible that $\sum_{i\geq 1} f_i^{\downarrow} < 1$, there will also be a distinguished interval of length $1 - \sum_{i\geq 1} f_i^{\downarrow}$. Consider now the following *paintbox process*, which creates a random partition of $\mathbb{N}$ starting from the frequencies. Take independent uniform random variables $U_1, U_2, \ldots$ on $[0, 1]$. If U_i and U_j land in the same non-distinguished interval of the partition then assign i and j to be in the same block. If U_i lands in the distinguished interval, assign i to a singleton block. The partition we create in this way is exchangeable and has the same distribution as the partition with which we began. This procedure can also be thought of in terms of a classical balls-in-boxes problem with infinitely many unlabelled boxes, see in particular Karlin [27] and Gnedin, Hansen and Pitman [18].

There are several natural questions that we may ask about an exchangeable random partition restricted to the first n integers (or, equivalently, about the partition formed by the first n uniform random variables in the paintbox process). How many blocks does this partition have? How many blocks does it have of size exactly k, for $1 \leq k \leq n$? Even in the absence of precise distributional information for finite n, can we obtain $n \to \infty$ limits for these quantities, in an appropriate sense? These questions have been studied for various classes of exchangeable random partitions and random compositions, see in particular [1; 17; 20; 21; 23; 24].

1.2 Coalescent process and allelic partitions

In this paper, we study a particular exchangeable random partition which derives from a coalescent process. The origins of this partition lie in population genetics and we will now describe how it arises and give a brief review of the relevant literature. For large populations, genealogies are often modelled using Kingman's coalescent [30]. This is a Markov process taking values in the space of partitions of $\mathbb{N}$ (or $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$), such that the partition becomes coarser and coarser with time. Whenever the current state has b blocks, any pair of them coalesces at rate 1, independently of the other blocks and irrespective of the block sizes. We start with a sample of genetic material from n individuals. Here, n is taken to be small compared to the total underlying population size. We imagine tracing the genealogy of the sample *backwards in time* from the present. Then the blocks of the coalescent process at time t correspond to the groups of individuals having the same ancestor time t ago (where time is measured in units of the total

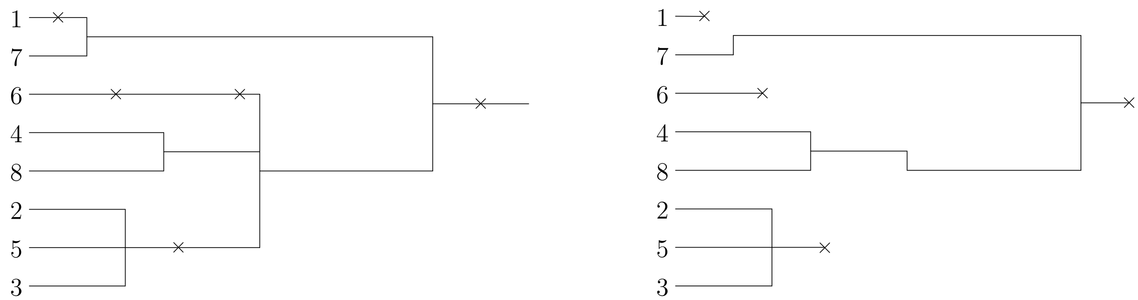

Figure 1: Left: a coalescent tree with mutations. Right: the sections of the tree relevant for the formation of the allelic partition. Note that from each individual we look back only to the last mutation, so that the second mutation on the lineage of 6 is ignored. The allelic partition here is $\{1\}, \{2, 3, 5\}, \{4, 7, 8\}, \{6\}$. If $N_k(n)$ is the number of blocks of size k when we start with n individuals, then we have $N_1(8) = 2$, $N_2(8) = 0$, $N_3(8) = 2$, $N_3(8) = N_4(8) = \cdots = N_8(8) = 0$.

underlying effective population size). See Ewens [16] or Durrett [14] for full introductions to this subject. In the population genetics setting, it is natural to introduce the concept of mutation into this model. One of the most celebrated results in this area is the *Ewens Sampling Formula*, which was proved by Ewens [15] in 1972. It concerns the *infinitely many alleles model*, in which every mutation gives rise to a completely new type. It says that if we take a sample of n genes subject to neutral mutation (that is, mutation which does not confer a selective advantage) which occurs at rate ρ for each individual, then the probability $q(m_1, m_2, \ldots)$ that there are m_j types which occur exactly j times is given by

$$q(m_1, m_2, \ldots) = \frac{n! \theta^{\sum_{i \geq 1} m_i}}{(\theta)_{n\uparrow} \prod_{j \geq 1} j^{m_j} m_j!},$$

where $\theta = 2\rho$, $(\theta)_{n\uparrow} = \theta(\theta + 1) \cdots (\theta + n - 1)$ and we must have $\sum_{j \geq 1} j m_j = n$. Another way of expressing this (due to Kingman [28]) is to picture the coalescent tree associated with Kingman's coalescent and place mutations along the length of the skeleton as a Poisson process of intensity $\rho = \theta/2$. For each individual, trace backwards in time (i.e. forwards in coalescent time) to the most recent mutation. Group together those individuals whose most recent mutations are the same; this gives the *allelic partition*. Then m_j is the number of blocks in the allelic partition containing exactly j individuals.

It is natural to extend these ideas to more general coalescent processes. See Figure 1 for an example of a general coalescent tree and its allelic partition. The Λ-coalescents are a class of Markovian coalescent processes which were introduced by Pitman [34] and Sagitov [37]. Like Kingman's coalescent, they take as their state-space the set of partitions of $[n]$ (or, indeed, of the whole set of natural numbers). Their evolution is such that only one block is formed in any coalescence event and rates of coalescence depend only on the number of blocks present and not on their sizes. Take Λ to be a finite measure on $[0, 1]$. In order to give a formal description of the Λ-coalescent, it is sufficient to give its jump rates. Whenever there are b blocks present, any particular k of them coalesce at rate

$$\lambda_{b,k} \stackrel{\text{def}}{=} \int_0^1 x^{k-2}(1 - x)^{b-k} \Lambda(dx), \quad 2 \leq k \leq b.$$

Note that, in contrast to Kingman's coalescent, here we allow *multiple collisions*; that is, we allow more than two blocks to join together. However, we do not allow more than one group of blocks to coalesce at once. Kingman's coalescent is the case $\Lambda(dx) = \delta_0(dx)$, where unit mass is placed at 0. The case $\Lambda(dx) = dx$, called the Bolthausen-Sznitman coalescent, was introduced by Bolthausen and Sznitman [7] in the context of spin glasses. It has many nice properties and appears to be more tractable than most Λ-coalescents. For example, its marginal distributions are known explicitly [34]. It has been studied in some detail: see, for example, Pitman [34], Bertoin and Le Gall [5], Basdevant [2] and Goldschmidt and Martin [25].

Another sub-class of the Λ-coalescents which has recently been particularly studied is the *Beta coalescents*, so-called because Λ here is a beta density:

$$\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}dx,$$

for some $\alpha \in (0,2)$. (The $\alpha = 1$ case is the Bolthausen-Sznitman coalescent and, in some sense, $\alpha = 2$ corresponds to Kingman's coalescent.) See Birkner et al [6] for a representation in terms of continuous-state branching processes when $\alpha \in (0,2)$.

If we suppose that, instead of Kingman's coalescent, the genealogy of the population evolves according to a general Λ-coalescent then, except in the special case of the degenerate star-shaped coalescent (where $\Lambda(dx) = \delta_1(dx)$), there is no known explicit expression for the probability $q(m_1, m_2, \ldots)$ of having m_j blocks in the allelic partition of size j. However, Möhle [31] has shown that the probabilities q must satisfy the following recursion:

$$q(m) = \frac{n\rho}{\lambda_n + n\rho} q(m - e_1) + \sum_{i=1}^{n-1} \frac{\binom{n}{i+1}\lambda_{n,i+1}}{\lambda_n + n\rho} \sum_{j=1}^{n-i} \frac{j(m_j+1)}{n-i} q(m + e_j - e_{i+j}),$$

where $\lambda_n = \sum_{k=2}^{n} \binom{n}{k} \lambda_{n,k}$, $\rho = \theta/2$, $m = (m_1, m_2, \ldots)$ and e_i is the vector with a 1 in the ith co-ordinate and 0 in all the rest. He has also shown [33] that, except in the cases of the star-shaped coalescent and Kingman's coalescent, the allelic partition is not *regenerative* in the sense of Gnedin and Pitman [19]. Dong, Gnedin and Pitman [12] have studied various properties of the allelic partition of a general Λ-coalescent. In particular, they view the allelic partition as the final partition of a coalescent process with *freeze* (see Section 2 where we use this formalism) and also give an alternative description of q as the stationary distribution of a certain discrete-time Markov chain.

Consider again the Beta coalescents. Suppose that we start the coalescent process from the partition of $[n]$ into singletons. Let $N_k(n)$ be the number of blocks of size k, for $k \geq 1$, and let $N(n) = \sum_{k=1}^{n} N_k(n)$. From the biological point of view, $N_k(n)$ is the number of types which appear exactly k times in a sample of size n and $N(n)$ is the total number of types in the sample. The complete allele frequency spectrum is the vector

$$(N_1(n), N_2(n), N_3(n), \ldots).$$

In the case of $\alpha \in (1,2)$, Berestycki, Berestycki and Schweinsberg [3; 4] have proved that

$$n^{\alpha-2}N(n) \xrightarrow{p} \frac{\rho\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha} \tag{1}$$

and, for $k \geq 1$, that

$$n^{\alpha-2} N_k(n) \xrightarrow{p} \frac{\rho\alpha(\alpha-1)^2\Gamma(k+\alpha-2)}{k!},$$

as $n \to \infty$.

The corresponding convergence results for Kingman's coalescent can be derived from the Ewens sampling formula: without rescaling, we have

$$(N_1(n), N_2(n), \ldots) \xrightarrow{d} (Z_1, Z_2, \ldots),$$

where $Z_1, Z_2, \ldots$ are independent Poisson random variables such that Z_i has mean θ/i. It follows that

$$\frac{N(n)}{\log n} \xrightarrow{\text{a.s.}} \theta,$$

as $n \to \infty$ and, moreover, that

$$\frac{N(n) - \theta \log n}{\sqrt{\theta \log n}} \xrightarrow{d} \mathrm{N}(0,1).$$

It is clear that the Beta coalescents belong to a completely different asymptotic regime.

A related problem concerns the *infinitely many sites* model. Here, as before, we put mutations on the coalescent tree, but this time we imagine that we trace the genealogy of long stretches of chromosome from each of our n individuals. Each time a mutation arrives, it affects a different site on the chromosome. The number of *segregating sites* is the number of sites at which there exists more than one allele in our sample of chromosomes. This is simply the number of mutations on the skeleton of the coalescent tree. Let $S(n)$ be the number of segregating sites when we start with a sample of n individuals. Clearly the distributions of $S(n)$ and $N(n)$ are related, in that in both cases we count mutations along the skeleton of the coalescent tree; for $N(n)$, we discard any mutation which arises on a lineage all of whose members have already mutated. For the Beta coalescents with $\alpha \in (1,2)$, the asymptotics of $S(n)$ are given in [3] and are the same as those of $N(n)$ given at (1). In [32], Möhle has studied the limiting distribution of $S(n)$ in the case where the measure $x^{-1}\Lambda(dx)$ is finite (which includes the Beta coalescents with $\alpha \in (0,1)$). He proves that

$$\frac{S(n)}{n} \xrightarrow{d} \rho \int_0^\infty \exp(-\sigma_t) dt, \tag{2}$$

where $(\sigma_t)_{t \geq 0}$ is a drift-free subordinator with Lévy measure given by the image under the transformation $x \mapsto -\log(1-x)$ of the measure $x^{-2}\Lambda(dx)$.

The number of segregating sites is, in turn, closely related to the *length* of the coalescent tree (i.e. the sum of the lengths of all of the branches) and to the total number of collisions before absorption. This has been studied for various Λ-coalescents in [3; 4; 11; 13; 22; 26].

1.3 The Bolthausen-Sznitman allelic partition

Turning now to the Bolthausen-Sznitman coalescent, Drmota, Iksanov, Möhle and Rösler [13] have proved that

$$\frac{\log n}{n} S(n) \xrightarrow{p} \rho,$$

where $S(n)$ is the number of segregating sites. They also give the fluctuations of S_n:

$$\frac{S(n) - \rho a_n}{\rho b_n} \xrightarrow{d} S,$$

where

$$a_n = \frac{n}{\log n} + \frac{n \log\log n}{(\log n)^2}, \qquad b_n = \frac{n}{(\log n)^2}$$

and S is a stable random variable with $\mathbb{E}\left[e^{itS}\right] = \exp\left(-\frac{1}{2}\pi|t| + it\log t\right)$, $t \in \mathbb{R}$.

The purpose of this paper is to prove the following theorem concerning the complete allele frequency spectrum of the Bolthausen-Sznitman coalescent.

Theorem 1.1. *For $k \geq 1$, let $N_k(n)$ be the number of blocks of the allelic partition of size k when we start with n singleton blocks. Then*

$$\frac{\log n}{n} N_1(n) \xrightarrow{p} \rho$$

and, for $k \geq 2$,

$$\frac{(\log n)^2}{n} N_k(n) \xrightarrow{p} \frac{\rho}{k(k-1)}.$$

We note that the same asymptotics hold for $S(n)$ and $N_1(n)$, which bound $N(n)$ above and below respectively. Thus, as a corollary, we also get the asymptotics for $N(n)$:

$$\frac{\log n}{n} N(n) \xrightarrow{p} \rho.$$

Suppose that we start a general Λ-coalescent $(\Pi(t))_{t\geq 0}$ from the partition of $\mathbb{N}$ into singletons. Then it has been proved by Pitman [34] that either $\Pi(t)$ has only finitely many blocks for all $t > 0$ ($(\Pi(t))_{t\geq 0}$ *comes down from infinity*) or $\Pi(t)$ has infinitely many blocks for all time ($(\Pi(t))_{t\geq 0}$ *stays infinite*). See Schweinsberg [38] for an explicit criterion for when a Λ-coalcescent comes down from infinity, in terms of the $\lambda_{b,k}$'s. The fundamental difference between the Beta coalescents for $\alpha \in (1, 2)$ and $\alpha \in (0, 1]$ (including the Bolthausen-Sznitman coalescent) is that the former coalescents come down from infinity and the latter do not. This accounts for the fact that in Berestycki, Berestycki and Schweinsberg's result, the scalings are the same for all different sizes of block as n becomes large, whereas in our theorem, the singletons must be scaled differently. Essentially, coalescence occurs rather slowly and the overwhelming first-order effect is mutation, which causes the allelic partition to consist mostly of singletons. However, at the second order (i.e. considering $(N_2(n), N_3(n), \ldots)$), we can feel the effect of the coalescence.

We do not claim that our results are of any application in population genetics: to the best of our knowledge, the Bolthausen-Sznitman coalescent has not been used to model the genealogy of any biological population. Nonetheless, our method may extend to the case of coalescents which are more biologically realistic.

Our method of proof is of some interest in itself. We track the formation of the allelic partition using a certain Markov process, for which we then prove a fluid limit (functional law of large numbers). The terminal value of our process gives the allele frequency spectrum and the fluid limit result, after a little extra work, allows us to read off the asymptotics.

Fluid limits have been widely used in the analysis of stochastic networks (see, for example, [8], [39]) and in the study of random graphs ([9], [36], [40]). In some sense, the prototypical result of the type in which we are interested is the following: suppose we take a Poisson process, $(X(t))_{t\geq 0}$ of rate 1, started from 0. Then the re-scaled process $(N^{-1}X(Nt))_{t\geq 0}$ stays close (in a rather strong sense) to the deterministic function $x(t) = t$, at least on compact time-intervals. For a general pure jump Markov process, the fluid limit is determined as the solution to a differential equation. In this article we have relied on the neat formulation in Darling and Norris [10]. However, our fluid limit is somewhat unusual. Firstly, instead of scaling time up, we actually scale it down, by a factor of $\log n$. Moreover, we have three different "space" scalings for different co-ordinates of our (multidimensional) process.

2 Fluid limit

Consider the formation of the allelic partition, starting from the partition into singletons and run until every individual has received a mutation. The easiest way to think of this is to use the terminology of Dong, Gnedin and Pitman [12] in which blocks have two possible states: *active* and *frozen*. We start with all blocks active and equal to singletons. Active blocks coalesce according to the rules of the Bolthausen-Sznitman coalescent: if there are b active blocks present then any particular k of them coalesce at rate $\lambda_{b,k} = (k-2)!(b-k)!/(b-1)!$. Moreover, every active block becomes frozen at rate ρ and stays frozen forever (this act of freezing creates a block in the allelic partition).

The data we will track are as follows. Let $X_k^n(t)$ be the number of active blocks of the *coalescent* partition at time t containing k individuals, $k \geq 1$, where we start at time 0 with n active individuals in singleton blocks. For $k \geq 1$, let $Z_k^n(t)$ be the number of blocks of the *allelic* partition of size k which have already been formed by time t (this is the number of times so far that an active block containing precisely k individuals has become frozen). For $d \geq 1$, let $Y_{d+1}^n(t) = \sum_{k=d+1}^{\infty} X_k^n(t)$, the number of active blocks containing at least $d+1$ individuals.

It is straightforward to see that, for any $d \geq 1$,

$$X^{n,d}(t) \stackrel{\text{def}}{=} (X_1^n(t), X_2^n(t), \ldots, X_d^n(t), Y_{d+1}^n(t), Z_d^n(t))_{t\geq 0}$$

is a (time-homogeneous) Markov jump process taking values in $\{0, 1, 2, \ldots, n\}^{d+2}$, with

$$X_1^n(0) = n, \quad X_k^n(0) = 0, \quad 2 \leq k \leq d, \quad Y_{d+1}^n(0) = 0, \quad Z_d^n(0) = 0.$$

Now put

$$\bar{X}_1^n(t) = \frac{1}{n} X_1^n\left(\frac{t}{\log n}\right), \qquad \bar{X}_k^n(t) = \frac{\log n}{n} X_k^n\left(\frac{t}{\log n}\right) \quad \text{for } k \geq 2,$$
$$\bar{Z}_1^n(t) = \frac{\log n}{n} Z_1^n\left(\frac{t}{\log n}\right), \quad \bar{Z}_k^n(t) = \frac{(\log n)^2}{n} Z_k^n\left(\frac{t}{\log n}\right) \quad \text{for } k \geq 2$$

and

$$\bar{Y}_{d+1}^n(t) = \frac{\log n}{n} Y_{d+1}^n\left(\frac{t}{\log n}\right) \quad \text{for } d \geq 1.$$

Fix $d \geq 1$ and write

$$\bar{X}^{n,d}(t) = (\bar{X}_1^n(t), \bar{X}_2^n(t), \dots, \bar{X}_d^n(t), \bar{Y}_{d+1}^n(t), \bar{Z}_d^n(t))$$

and define a stopping time

$$T_n = \inf\{t \geq 0 : X_1^n(t) = X_2^n(t) = \dots = X_d^n(t) = Y_{d+1}^n(t) = 0\}.$$

(Note that T_n is the same regardless of the value of d.)

For $t \geq 0$, let

$$\begin{aligned} x_1(t) &= e^{-t}, & x_k(t) &= \frac{te^{-t}}{k(k-1)}, \quad 2 \leq k \leq d, \\ z_1(t) &= \rho(1 - e^{-t}), & z_k(t) &= \frac{\rho}{k(k-1)}(1 - e^{-t} - te^{-t}), \quad 2 \leq k \leq d \end{aligned}$$

and

$$y_{d+1}(t) = \frac{te^{-t}}{d}.$$

Finally, let

$$x^{(d)}(t) = (x_1(t), x_2(t), \dots, x_d(t), y_{d+1}(t), z_d(t)).$$

We write $\|\cdot\|$ for the Euclidean norm on $\mathbb{R}^{d+2}$.

Proposition 2.1. *Fix $d \geq 1$ and let $t_0 < \infty$. Then, given $\epsilon > 0$,*

$$\mathbb{P}\left(\sup_{0 \leq t \leq t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| > \epsilon\right) \to 0$$

as $n \to \infty$.

This is the key to the following result.

Proposition 2.2. *Take $\delta > 0$. Then*

$$\mathbb{P}\left(\left|\frac{\log n}{n} Z_1^n(T_n) - \rho\right| > \delta\right) \to 0$$

and, for $k \geq 2$,

$$\mathbb{P}\left(\left|\frac{(\log n)^2}{n} Z_k^n(T_n) - \frac{\rho}{k(k-1)}\right| > \delta\right) \to 0,$$

as $n \to \infty$.

Theorem 1.1 now follows directly, since $N_k(n) = Z_k^n(T_n)$ for $k \geq 1$. Note that Proposition 2.1 tells us *how* the allele frequency spectrum is formed.

Remark. Delmas, Dhersin and Siri-Jegousse [11] have recently considered the lengths of coalescent trees associated with Beta coalescents for $\alpha \in (1, 2)$. Part (1) of their Theorem 5.1 appears to be a result analogous to our Proposition 2.1.

3 Comments

3.1 Asymptotic frequencies

It would be very interesting to have a better understanding of the distribution of the asymptotic frequency sequence of the allelic partition associated with the Bolthausen-Sznitman coalescent. In [18], Gnedin, Hansen and Pitman obtain relations between the total number of blocks $N(n)$ of an exchangeable random partition restricted to the set $\{1, \ldots, n\}$ and the asymptotic form of the sequence $(f_i^{\downarrow})_{i \geq 1}$. More precisely, they prove that, for any $\alpha \in (0, 1)$ and any function $\ell : \mathbb{R}_+ \to \mathbb{R}_+$, slowly varying at infinity, we have

$$\frac{N(n)}{\Gamma(1-\alpha)n^{\alpha}\ell(n)} \xrightarrow{\text{a.s.}} 1 \iff \frac{\#\{i \geq 1 : f_i^{\downarrow} \geq x\}}{\ell(1/x)x^{-\alpha}} \xrightarrow{\text{a.s.}} 1 \text{ as } x \to 0+ \iff \frac{f_i^{\downarrow}}{\ell^*(i)i^{-1/\alpha}} \xrightarrow{\text{a.s.}} 1,$$

where ℓ^* is also a slowly varying function which can be expressed in term of α and ℓ.

It would be nice to have a similar result for the allelic partition associated with the Bolthausen-Sznitman coalescent. There are, however, two main difficulties: first, we would need almost sure convergence of the rescaled process $N(\cdot)$, whereas here we have only established convergence in probability. Second, the Bolthausen-Sznitman coalescent corresponds to the critical case $\alpha = 1$ for which the first of the above equivalences no longer holds. In this setting, according to Proposition 18 of Gnedin, Hansen and Pitman [18], we have only the implication:

$$x(\log x)^2 \#\{i \geq 1 : f_i^{\downarrow} \geq x\} \xrightarrow{\text{a.s.}} \rho \text{ as } x \to 0+ \implies \frac{\log n}{n} N(n) \xrightarrow{\text{a.s.}} \rho$$

and, in addition, that

$$\frac{\log n}{n} N_1(n) \xrightarrow{\text{a.s.}} \rho \quad \text{and} \quad \frac{(\log n)^2}{n} N_k(n) \xrightarrow{\text{a.s.}} \frac{\rho}{k(k-1)}, k \geq 2.$$

The form of the limits is, of course, basically the same as in our Theorem 1.1 and so we might expect the following result.

Conjecture 3.1. *Let $(f_i^{\downarrow})_{i \geq 1}$ be the asymptotic frequency sequence of the allelic partition associated with the Bolthausen-Sznitman coalescent. Then*

$$f_i^{\downarrow} \sim \frac{\rho}{i(\log i)^2} \quad \text{as } i \to \infty.$$

3.2 Fluctuations

Another interesting and natural question concerns the form of the fluctuations around the deterministic limits in Theorem 1.1. The methods used in this paper do not give us access to this information. In view of the fact that $S(n)$ and $N(n)$ have the same first-order asymptotics, one might be tempted to conjecture that they should have the same second-order asymptotics as well. However, we do not have a cogent argument for why this should be the case.

3.3 Beta coalescents

The fluid limit methods used in this paper can, in principle, be extended to deal with other classes of coalescent process. For instance, the method seems to work for the Beta coalescents with parameter $\alpha \in (1,2)$. However, the calculations are more complicated than in the Bolthausen-Sznitman case. Indeed, for the Bolthausen-Sznitman coalescent, the active partition is mostly composed of singletons at any time, which essentially enables us to neglect collisions between non-singleton blocks. This approximation does not hold for the Beta coalescents with $\alpha \in (1,2)$. Since the relevant result has already been proved by Berestycki, Berestycki and Schweinsberg [3; 4] by other methods, we will not give the details.

We may also consider the Beta coalescents with parameter $\alpha \in (0,1)$. Möhle's result (2) that the total number of mutations along the coalescent tree, re-scaled by n, converges in distribution to some non-degenerate random variable suggests that here we may expect to have convergence in distribution of the allelic partition to a random vector. Clearly, the fluid limit methods used in the present paper do not adapt to this situation, but we can still use them to investigate the expected value of the number of blocks of different sizes. Indeed, the drift of the re-scaled process

$$\begin{array}{ll} \left(\frac{X_1^n(t)}{n}, \frac{Y_{d+1}^n(t)}{n^\alpha}, \frac{Z_d^n(t)}{n}\right) & \text{if } d=1 \\ \left(\frac{X_1^n(t)}{n}, \frac{X_2^n(t)}{n^\alpha}, \dots, \frac{X_d^n(t)}{n^\alpha}, \frac{Y_{d+1}^n(t)}{n^\alpha}, \frac{Z_d^n(t)}{n^\alpha}\right) & \text{if } d \geq 2 \end{array}$$

converges to an explicit function $b^{(d)}$ (but the variance $\bar{\alpha}^{n,d}$ does not tend to 0). This enables us to conjecture that

$$\frac{N_1(n)}{n} \xrightarrow{d} C_1 \quad \text{and} \quad \frac{N_k(n)}{n^\alpha} \xrightarrow{d} C_k \quad \text{for } k \geq 2,$$

where $C_1, C_2, \dots$ are strictly positive random variables. We intend to address this problem in a future paper.

4 Proofs

In this section, we prove Proposition 2.1 and deduce Proposition 2.2. In order to do so, we use the fluid limit methodology described in Darling and Norris [10]. Firstly, we need to set up some notation. Let $\beta^{n,d}(m)$ be the drift of the process $X^{n,d}$ when it is in state $m = (m_1, m_2, \dots, m_{d+2}) \in \{0, ..., n\}^{d+2}$, so that

$$\beta^{n,d}(m) = \sum_{m' \neq m} (m' - m) q^{n,d}(m, m'),$$

where $q^{n,d}(m,m')$ is the jump rate from m to m'. Let $\alpha^{n,d}(m)$ be the corresponding variance of a jump, in the sense that

$$\alpha^{n,d}(m) = \sum_{m' \neq m} \|m' - m\|^2 q^{n,d}(m, m').$$

Let us also introduce the notation

$$\alpha_k^{n,d}(m) = \sum_{m' \neq m} |m'_k - m_k|^2 q^{n,d}(m, m'),$$

for $1 \le k \le d+2$, so that we may decompose $\alpha^{n,d}(m)$ as

$$\alpha^{n,d}(m) = \sum_{k=1}^{d+2} \alpha_k^{n,d}(m).$$

Finally, let $M \stackrel{\text{def}}{=} \sum_{k=1}^{d+1} m_k$ denote the total number of active blocks in the partition. We will need to compute the drift and infinitesimal variance of the re-scaled process $\bar{X}^{n,d}$, which takes values in the set

$$\mathcal{S}^{n,d} \stackrel{\text{def}}{=} \left\{0, \frac{1}{n}, \ldots, 1\right\} \times \left\{0, \frac{\log n}{n}, 2\frac{\log n}{n}, \ldots, \log n\right\}^d \times \left\{0, \frac{(\log n)^r}{n}, 2\frac{(\log n)^r}{n}, \ldots, (\log n)^r\right\},$$

where $r = 1$ if $d = 1$ and $r = 2$ if $d \ge 2$. Denote by $\bar{\beta}^{n,d}(\xi)$ and $\bar{\alpha}^{n,d}(\xi)$ the drift and infinitesimal variance of $\bar{X}^{n,d}$ when it is in the state $\xi = (\xi_1, \xi_2, \ldots, \xi_{d+2}) \in \mathcal{S}^{n,d}$. Then, letting $m = (n\xi_1, \frac{n}{\log n}\xi_2, \ldots, \frac{n}{\log n}\xi_{d+1}, \frac{n}{(\log n)^r}\xi_{d+2})$, we have

$$\bar{\beta}_k^{n,d}(\xi) = \begin{cases} \frac{1}{n \log n}\beta_1^{n,d}(m) & k = 1 \\ \frac{1}{n}\beta_k^{n,d}(m) & 2 \le k \le d+1 \\ \frac{(\log n)^{r-1}}{n}\beta_{d+2}^{n,d}(m) & k = d+2, \end{cases} \tag{3}$$

$$\bar{\alpha}_k^{n,d}(\xi) = \begin{cases} \frac{1}{n^2 \log n}\alpha_1^{n,d}(m) & k = 1 \\ \frac{\log n}{n^2}\alpha_k^{n,d}(m) & 2 \le k \le d+1 \\ \frac{(\log n)^{2r-1}}{n^2}\alpha_{d+2}^{n,d}(m) & k = d+2 \end{cases} \tag{4}$$

and

$$\bar{\alpha}^{n,d}(\xi) = \sum_{k=1}^{d+1} \bar{\alpha}_k^{n,d}(\xi).$$

Now define $b^{(d)} : \mathbb{R}^{d+2} \to \mathbb{R}^{d+2}$ co-ordinatewise by

$$b_k^{(d)}(\xi) = \begin{cases} -\xi_1 & k = 1 \\ \frac{1}{k(k-1)}\xi_1 - \xi_k & 2 \le k \le d \\ \frac{1}{d}\xi_1 - \xi_{d+1} & k = d+1 \\ \rho\xi_d & k = d+2. \end{cases}$$

Then the vector field $b^{(d)}$ is Lipschitz in the Euclidean norm with constant $K \stackrel{\text{def}}{=} \sqrt{\rho^2 + \pi^2/3}$. The function $x^{(d)}(t)$ of the previous section is the unique solution of the differential equation

$$\frac{d}{dt}x^{(d)}(t) = b^{(d)}(x^{(d)}(t)).$$

In order to prove Proposition 2.1, we need a few lemmas. Firstly, we prove two analytic results. For $n \in \mathbb{N}$, let $h(n) = \sum_{i=1}^{n-1} \frac{1}{i}$, the $(n-1)$th harmonic number.

Lemma 4.1. *Fix $R \ge e$. Then for $x \in \frac{1}{n}\mathbb{Z} \cap [R^{-1}, 1]$,*

$$\left| \frac{h(nx)}{\log n} - 1 \right| \le \frac{\log R}{\log n}.$$

Proof. It is an elementary fact that, for $k \geq 2$,

$$\log k \leq h(k) \leq 1 + \log(k-1) \leq 1 + \log k.$$

This entails that

$$\left| \frac{h(nx)}{\log n} - 1 \right| \leq \max \left\{ -\frac{\log x}{\log n}, \frac{1 + \log x}{\log n} \right\} \leq \frac{\log R}{\log n}$$

in the specified range of x. □

Lemma 4.2. *For $0 \leq j \leq n$ and $k \geq 0$,*

$$0 \leq 1 - \frac{\binom{n}{j}}{\binom{n+k}{j}} \leq \frac{kj}{n-j+1}.$$

Proof. We have

$$\log \left(\frac{\binom{n}{j}}{\binom{n+k}{j}} \right) = -\sum_{i=0}^{j-1} (\log(n-i+k) - \log(n-i)).$$

By the mean value theorem and monotonicity of the logarithm,

$$\log(n-i+k) - \log(n-i) \leq \frac{k}{n-i}, \quad 0 \leq i \leq n-1.$$

Hence,

$$\sum_{i=0}^{j-1} (\log(n-i+k) - \log(n-i)) \leq \sum_{i=0}^{j-1} \frac{k}{n-i} \leq \frac{kj}{n-j+1}$$

and so

$$\log \left(\frac{\binom{n}{j}}{\binom{n+k}{j}} \right) \geq -\frac{kj}{n-j+1}.$$

Since

$$\exp \left(-\frac{kj}{n-j+1} \right) \geq 1 - \frac{kj}{n-j+1},$$

the result follows. □

We now have the necessary tools to begin proving the fluid limit result.
Fix $R \geq e$ and let $l(n,R,d) = R^{-1} + d/n$ and

$$\tilde{\mathcal{S}}^{n,d} = \left\{ \xi \in \mathcal{S}^{n,d} : \xi_1 \geq l(n,R,d), \quad \sum_{i=2}^{d+1} \xi_i \leq R \right\}. \tag{5}$$

Let

$$\begin{aligned} T_n^{R,d,1} &= \inf \left\{ t \geq 0 : \bar{X}_1^n(t) < l(n,R,d) \right\}, \\ T_n^{R,d,2} &= \inf \left\{ t \geq 0 : \bar{Y}_2^n(t) > R \right\} \end{aligned}$$

and set $T_n^{R,d} = T_n^{R,d,1} \wedge T_n^{R,d,2}$.

Lemma 4.3. *For $\xi \in \tilde{\mathcal{S}}^{n,d}$, there exists a constant $C(R)$, depending only on R, such that*

$$\|\bar{\beta}^{n,d}(\xi) - b^{(d)}(\xi)\| \leq \frac{C(R)}{\log n}.$$

It follows that for $t_0 < \infty$,

$$\int_0^{T_n^{R,d} \wedge t_0} \|\bar{\beta}^{n,d}(\bar{X}^{n,d}(t)) - b^{(d)}(\bar{X}^{n,d}(t))\| dt \leq \frac{C(R)t_0}{\log n}.$$

Proof. We must perform some elementary (but rather involved) calculations. From the rates of the process we will calculate the co-ordinates of $\beta^{n,d}(m)$ in turn. Recall first that the Λ-coalescents are *consistent* in the sense that, regardless of how many blocks are present in total, if we look just at a subcollection of size b of them, any k of those b blocks coalesce at rate $\lambda_{b,k}$. Furthermore, freezing occurs in a Markovian way. Hence, if M active blocks are present in the partition, the next event involves the coalescence of precisely j of them at rate $\binom{M}{j}\lambda_{M,j} = \frac{M}{j(j-1)}$ (see Theorem 3.1 of [12]). Thus, we have

$$\beta_1^{n,d}(m) = -\rho m_1 - \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_1=1}^{m_1} b_1 \frac{\binom{m_1}{b_1}\binom{M-m_1}{j-b_1}}{\binom{M}{j}}$$

Then, using the fact that $b_1\binom{m_1}{b_1} = m_1\binom{m_1-1}{b_1-1}$, we get

$$\sum_{b_1=1}^{m_1} b_1 \binom{m_1}{b_1}\binom{M-m_1}{j-b_1} = m_1 \sum_{b=0}^{m_1-1} \binom{m_1-1}{b}\binom{M-1-(m_1-1)}{j-1-b} = m_1\binom{M-1}{j-1}.$$

Thus,

$$\begin{aligned}\beta_1^{n,d}(m) &= -\rho m_1 - m_1 \sum_{j=2}^{M} \frac{1}{j-1} \\ &= -\rho m_1 - m_1 h(M).\end{aligned}$$

For $2 \leq k \leq d$,

$$\begin{aligned}\beta_k^{n,d}(m) = -\rho m_k &- \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_k=1}^{m_k} b_k \frac{\binom{m_k}{b_k}\binom{M-m_k}{j-b_k}}{\binom{M}{j}} \\ &+ \sum_{j=2}^{k} \frac{M}{j(j-1)} \sum_{\substack{0 \leq b_1, b_2, \ldots, b_{k-1} \leq j \\ \sum_{l=1}^{k-1} l b_l = k, \sum_{l=1}^{k-1} b_l = j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}} \\ = -\rho m_k - m_k h(M) &+ \sum_{j=2}^{k} \frac{M}{j(j-1)} \sum_{\substack{0 \leq b_1, b_2, \ldots, b_{k-1} \leq j \\ \sum_{l=1}^{k-1} l b_l = k, \sum_{l=1}^{k-1} b_l = j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}}.\end{aligned}$$

For the $(d+1)$th co-ordinate we have

$$\beta_{d+1}^{n,d}(m) = -\rho m_{d+1} - \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_{d+1}=1}^{m_{d+1}} (b_{d+1}-1) \frac{\binom{m_{d+1}}{b_{d+1}}\binom{M-m_{d+1}}{j-b_{d+1}}}{\binom{M}{j}}$$
$$+ \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{\substack{0\le b_1,b_2,\ldots,b_d\le j \\ \sum_{l=1}^{d} lb_l \ge d+1, \sum_{l=1}^{d} b_l = j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_d}{b_d}}{\binom{M}{j}}$$

$$= -\rho m_{d+1} - m_{d+1} h(M) + \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_{d+1}=1}^{m_{d+1}} \frac{\binom{m_{d+1}}{b_{d+1}}\binom{M-m_{d+1}}{j-b_{d+1}}}{\binom{M}{j}}$$
$$+ \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{\substack{0\le b_1,b_2,\ldots,b_d\le j \\ \sum_{l=1}^{d} lb_l \ge d+1, \sum_{l=1}^{d} b_l = j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_d}{b_d}}{\binom{M}{j}}.$$

Note that we have

$$\sum_{\substack{0\le b_1,b_2,\ldots,b_{d+1}\le j \\ \sum_{l=1}^{d+1} lb_l \ge d+1, \sum_{l=1}^{d+1} b_l = j}} \binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}$$
$$= \sum_{\substack{0\le b_1,b_2,\ldots,b_d\le j \\ \sum_{l=1}^{d} lb_l \ge d+1, \sum_{l=1}^{d} b_l = j}} \binom{m_1}{b_1}\cdots\binom{m_d}{b_d} + \sum_{b_{d+1}=1}^{m_{d+1}} \binom{m_{d+1}}{b_{d+1}}\binom{M-m_{d+1}}{j-b_{d+1}}$$

(split the sum on the left-hand side according as $b_{d+1}=0$ or $b_{d+1}\neq 0$). Thus, we get

$$\beta_{d+1}^{n,d}(m) = -\rho m_{d+1} - m_{d+1} h(M) + \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{\substack{0\le b_1,b_2,\ldots,b_{d+1}\le j \\ \sum_{l=1}^{d+1} lb_l \ge d+1, \sum_{l=1}^{d+1} b_l = j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}}{\binom{M}{j}}.$$

Finally,
$$\beta_{d+2}^{n,d}(m) = \rho m_d.$$

Using (3) and the notation $m = (m_1,\ldots,m_{d+2})$, we obtain the following expressions:

$$\bar{\beta}_1^{n,d}(\xi) = -\frac{\rho}{\log n}\xi_1 - \frac{\xi_1 h(M)}{\log n}, \tag{6}$$

for $2 \leq k \leq d$,

$$\bar{\beta}^{n,d}_k(\xi) = -\frac{\rho}{\log n}\xi_k - \frac{\xi_k h\left(M\right)}{\log n} + \frac{1}{n}\sum_{j=2}^{k}\frac{M}{j(j-1)}\sum_{\substack{0\leq b_1,b_2,\ldots,b_{k-1}\leq j\\ \sum_{l=1}^{k-1} lb_l=k,\sum_{l=1}^{k-1} b_l=j}}\frac{\binom{m_1}{b_1}\cdots\binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}}, \tag{7}$$

$$\bar{\beta}^{n,d}_{d+1}(\xi) = -\frac{\rho}{\log n}\xi_{d+1} - \frac{\xi_{d+1} h\left(M\right)}{\log n} + \frac{1}{n}\sum_{j=2}^{M}\frac{M}{j(j-1)}\sum_{\substack{0\leq b_1,b_2,\ldots,b_{k-1}\leq j\\ \sum_{l=1}^{d+1} lb_l\geq d+1,\sum_{l=1}^{d+1} b_l=j}}\frac{\binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}}{\binom{M}{j}}, \tag{8}$$

$$\bar{\beta}^{n,d}_{d+2}(\xi) = \rho\xi_d. \tag{9}$$

Bearing in mind that $M = n\left(\xi_1 + \frac{1}{\log n}\sum_{i=2}^{d+1}\xi_i\right)$ and that $\xi \in \tilde{S}^{n,d}$ (defined at (5)), we can apply Lemma 4.1, to get from (6) that

$$|\bar{\beta}^{n,d}_1(\xi) - b^{(d)}_1(\xi)| \leq \frac{(\rho + \log R)}{\log n}\xi_1.$$

Consider now the sum in the expression (7) for $\bar{\beta}^{n,d}_k(\xi)$ when $2 \leq k \leq d$. We split it into two parts, $j = k$ and $2 \leq j \leq k-1$. The $j = k$ term is

$$\frac{\xi_1 + \frac{1}{\log n}\sum_{i=2}^{d+1}\xi_i}{k(k-1)}\frac{\binom{n\xi_1}{k}}{\binom{n\xi_1 + \frac{n}{\log n}\sum_{i=2}^{d+1}\xi_i}{k}}.$$

By Lemma 4.2 we have

$$\left|\frac{\xi_1 + \frac{1}{\log n}\sum_{i=2}^{d+1}\xi_i}{k(k-1)}\frac{\binom{n\xi_1}{k}}{\binom{n\xi_1 + \frac{n}{\log n}\sum_{i=2}^{d+1}\xi_i}{k}} - \frac{1}{k(k-1)}\xi_1\right| \leq \frac{1}{\log n}\left(1 + \frac{\xi_1}{\xi_1 - d/n}\right)\sum_{i=2}^{d+1}\xi_i.$$

Turning now to the other term, if $2 \leq j \leq k-1$, we have

$$\sum_{\substack{0\leq b_1,b_2,\ldots,b_{k-1}\leq j\\ \sum_{l=1}^{k-1} lb_l=k,\sum_{l=1}^{k-1} b_l=j}}\frac{\binom{m_1}{b_1}\cdots\binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}} \leq 1 - \frac{\binom{m_1}{j}}{\binom{M}{j}} \leq \frac{j}{\log n}\frac{\sum_{i=2}^{d+1}\xi_i}{(\xi_1 - d/n)}$$

and so

$$\frac{1}{n}\sum_{j=2}^{k-1}\frac{M}{j(j-1)}\sum_{\substack{0\leq b_1,b_2,\ldots,b_{k-1}\leq j\\ \sum_{l=1}^{k-1} lb_l=k,\sum_{l=1}^{k-1} b_l=j}}\frac{\binom{m_1}{b_1}\cdots\binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}} \leq \frac{1}{\log n}\left(\xi_1 + \frac{1}{\log n}\sum_{i=2}^{d+1}\xi_i\right)\frac{\sum_{i=2}^{d+1}\xi_i}{\xi_1 - d/n}h(d).$$

With another application of Lemma 4.1, it follows that, for $\xi \in \tilde{\mathcal{S}}^{n,d}$

$$\begin{aligned}&|\bar{\beta}^{n,d}_k(\xi) - b^{(d)}_k(\xi)|\\ &\leq \frac{1}{\log n}\left((\rho + \log R)\xi_k + \left(1 + \frac{\xi_1}{\xi_1 - d/n}\right)\sum_{i=2}^{d+1}\xi_i + \left(\xi_1 + \frac{1}{\log n}\sum_{i=2}^{d+1}\xi_i\right)\frac{\sum_{i=2}^{d+1}\xi_i}{\xi_1 - d/n}h(d)\right).\end{aligned}$$

We turn finally to the expression (8) for $\bar{\beta}^n_{d+1}(\xi)$. Consider the sum which constitutes the third term. We have

$$\sum_{\substack{0\le b_1,b_2,\dots,b_{d+1}\le j\\ \sum_{l=1}^{d+1} lb_l\ge d+1,\sum_{l=1}^{d+1} b_l=j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}}{\binom{M}{j}} = 1-\sum_{\substack{0\le b_1,b_2,\dots,b_{d+1}\le j\\ \sum_{l=1}^{d+1} lb_l\le d,\sum_{l=1}^{d+1} b_l=j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}}{\binom{M}{j}}$$

$$= 1-\sum_{k=2}^{d}\sum_{\substack{0\le b_1,b_2,\dots,b_{d+1}\le j\\ \sum_{l=1}^{d+1} lb_l=k,\sum_{l=1}^{d+1} b_l=j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}}{\binom{M}{j}}.$$

But then

$$\frac{1}{n}\sum_{j=2}^{M}\frac{M}{j(j-1)}\sum_{\substack{0\le b_1,b_2,\dots,b_{d+1}\le j\\ \sum_{l=1}^{d+1} lb_l\ge d+1,\sum_{l=1}^{d+1} b_l=j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{d+1}}{b_{d+1}}}{\binom{M}{j}}$$

$$= \frac{1}{n}\left(M-1-\sum_{k=2}^{d}\frac{M}{k(k-1)}\frac{\binom{m_1}{k}}{\binom{M}{k}}-\sum_{k=2}^{d}\sum_{j=2}^{k-1}\frac{M}{j(j-1)}\sum_{\substack{0\le b_1,b_2,\dots,b_{k-1}\le j\\ \sum_{l=1}^{k-1} lb_l=k,\sum_{l=1}^{k-1} b_l=j}} \frac{\binom{m_1}{b_1}\cdots\binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}}\right)$$

and so, arguing as before, we obtain, for $\xi\in\tilde{\mathcal{S}}^{n,d}$

$$|\bar{\beta}^{n,d}_{d+1}(\xi)-b^{(d)}_{d+1}(\xi)|\le\frac{1}{n}+\frac{1}{\log n}\left((\rho+\log R)\xi_{d+1}+\left(\xi_1+\frac{1}{\log n}\sum_{i=2}^{d+1}\xi_i\right)\frac{\sum_{i=2}^{d}\xi_i}{\xi_1-d/n}dh(d)\right.$$
$$\left.+\,d\left(1+\frac{\xi_1}{\xi_1-d/n}\right)\sum_{i=2}^{d+1}\xi_i\right).$$

It is clear from (9) that

$$|\bar{\beta}^{n,d}_{d+2}(\xi)-b^{(d)}_{d+2}(\xi)|=0.$$

Putting everything together, we obtain that

$$\|\bar{\beta}^{n,d}(\xi)-b^{(d)}(\xi)\|\le\frac{C(R)}{\log n},$$

for some constant $C(R)$, whenever $\xi\in\tilde{\mathcal{S}}^{n,d}$. The final deduction follows easily. □

Lemma 4.4. *Fix $R\ge e$. Then there exists a constant $C'(R)$, depending only on R, such that for $\xi\in\tilde{\mathcal{S}}^{n,d}$,*

$$\bar{\alpha}^{n,d}(\xi)\le\frac{C'(R)}{\log n}.$$

It follows that for $t_0<\infty$,

$$\int_0^{T_n^{R,d}\wedge t_0}\bar{\alpha}^{n,d}(X_t)dt\le\frac{C'(R)t_0}{\log n}.$$

Proof. Recall that for $1 \le k \le d+2$ we have

$$\alpha_k^{n,d}(m) = \sum_{m' \neq m} |m'_k - m_k|^2 q^{n,d}(m,m'),$$

so that

$$\alpha^{n,d}(m) = \sum_{k=1}^{d+2} \alpha_k^{n,d}(m).$$

We will deal with the co-ordinates in turn.

$$\begin{aligned}\alpha_1^{n,d}(m) &= \rho m_1 + \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_1=1}^{m_1} b_1^2 \frac{\binom{m_1}{b_1}\binom{M-m_1}{j-b_1}}{\binom{M}{j}} \\ &= \rho m_1 + m_1(m_1-1) + m_1 h(M).\end{aligned}$$

Hence, by (4),

$$\bar{\alpha}_1^{n,d}(\xi) = \frac{1}{n^2 \log n} \alpha_1^{n,d}(m) \le \frac{\xi_1^2}{\log n} + \frac{C_1(R)}{n}$$

for some constant $C_1(R)$. For $2 \le k \le d$,

$$\begin{aligned}\alpha_k^{n,d}(m) = \rho m_k &+ \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_k=1}^{m_k} b_k^2 \frac{\binom{m_k}{b_k}\binom{M-m_k}{j-b_k}}{\binom{M}{j}} \\ &+ \sum_{j=2}^{k} \frac{M}{j(j-1)} \sum_{\substack{0 \le b_1, b_2, \ldots, b_{k-1} \le j \\ \sum_{l=1}^{k-1} l b_l = k, \sum_{l=1}^{k-1} b_l = j}} \frac{\binom{m_1}{b_1} \cdots \binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}} \\ = \rho m_k + m_k(m_k-1) + m_k h(M) &+ \sum_{j=2}^{k} \frac{M}{j(j-1)} \sum_{\substack{0 \le b_1, b_2, \ldots, b_{k-1} \le j \\ \sum_{l=1}^{k-1} l b_l = k, \sum_{l=1}^{k-1} b_l = j}} \frac{\binom{m_1}{b_1} \cdots \binom{m_{k-1}}{b_{k-1}}}{\binom{M}{j}}.\end{aligned}$$

By (4) we obtain

$$\bar{\alpha}_k^{n,d}(\xi) = \frac{\log n}{n^2} \alpha_k^{n,d}(m) \le \frac{\xi_k^2}{\log n} + \frac{C_k(R) \log n}{n},$$

for some constant $C_k(R)$. Furthermore,

$$\begin{aligned}\alpha_{d+1}^{n,d}(m) &= \rho m_{d+1} + \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{b_{d+1}=1}^{m_{d+1}} (b_{d+1}-1)^2 \frac{\binom{m_{d+1}}{b_{d+1}}\binom{M-m_{d+1}}{j-b_{d+1}}}{\binom{M}{j}} \\ &\quad + \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{\substack{0 \le b_1, b_2, \ldots, b_d \le j \\ \sum_{l=1}^{d} l b_l \ge d+1, \sum_{l=1}^{d} b_l = j}} \frac{\binom{m_1}{b_1} \cdots \binom{m_d}{b_d}}{\binom{M}{j}} \\ &\le \rho m_{d+1} + m_{d+1}(m_{d+1}-1) + m_{d+1} h(M) \\ &\quad + \sum_{j=2}^{M} \frac{M}{j(j-1)} \sum_{\substack{0 \le b_1, b_2, \ldots, b_d \le j \\ \sum_{l=1}^{d} l b_l \ge d+1, \sum_{l=1}^{d} b_l = j}} \frac{\binom{m_1}{b_1} \cdots \binom{m_d}{b_d}}{\binom{M}{j}}.\end{aligned}$$

So by (4) we also get

$$\bar{\alpha}_{d+1}^{n,d}(\xi) = \frac{\log n}{n^2}\alpha_{d+1}^{n,d}(m) \leq \frac{\xi_{d+1}^2}{\log n} + \frac{C_{d+1}(R)\log n}{n},$$

for some constant $C_{d+1}(R)$. Finally, we have

$$\alpha_{d+2}^{n,d}(m) = \rho m_d.$$

So

$$\bar{\alpha}_{d+2}^{n,d}(\xi) = \frac{\rho\xi_d \log n}{n} \text{ if } d = 1 \text{ and } \bar{\alpha}_{d+2}^{n,d}(\xi) = \frac{\rho\xi_d(\log n)^2}{n} \text{ if } d \geq 2.$$

Hence,

$$\bar{\alpha}^{n,d}(\xi) \leq \frac{\|\xi\|^2}{\log n} + \frac{D(R)(\log n)^2}{n},$$

for some constant $D(R)$. Since $\|\xi\|^2 \leq (1+R)^2$ in $\tilde{\mathcal{S}}^{n,d}$, we obtain

$$\bar{\alpha}^{n,d}(\xi) \leq \frac{C'(R)}{\log n},$$

for some constant $C'(R)$ depending on R. □

Proof of Proposition 2.1

We will, in fact, prove the stronger result that, for any $d \geq 1$ and any $0 < \delta < 1$,

$$\mathbb{P}\left(\sup_{0\leq t\leq t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| \geq (\log n)^{\frac{\delta-1}{2}}\right) \to 0 \tag{10}$$

as $n \to \infty$. Fix $d \geq 1$. We follow the method used in Theorem 3.1 of Darling and Norris [10] and start by noting that $\bar{X}^{n,d}(t)$ has the following standard decomposition

$$\bar{X}^{n,d}(t) = \bar{X}^{n,d}(0) + M^{n,d}(t) + \int_0^t \bar{\beta}^{n,d}(\bar{X}^{n,d}(s))ds, \tag{11}$$

where $(M^{n,d}(t))_{t\geq 0}$ is a zero-mean martingale in the natural filtration of $\bar{X}^{n,d}$. Since

$$x^{(d)}(t) = x^{(d)}(0) + \int_0^t b^{(d)}(x^{(d)}(s))ds$$

and $\bar{X}^{n,d}(0) = x^{(d)}(0)$ for all $n \in \mathbb{N}$, we have

$$\begin{aligned}
&\sup_{0\leq s\leq t} \|\bar{X}^{n,d}(s) - x^{(d)}(s)\| \\
&\quad \leq \sup_{0\leq s\leq t} \|M^{n,d}(s)\| + \int_0^t \|\bar{\beta}^{n,d}(\bar{X}^{n,d}(s)) - b^{(d)}(\bar{X}^{n,d}(s))\|ds \\
&\qquad + \int_0^t \|b^{(d)}(\bar{X}^{n,d}(s)) - b^{(d)}(x^{(d)}(s))\|ds.
\end{aligned} \tag{12}$$

Recall that K is the Lipschitz constant of $b^{(d)}$. Fix $R \geq e$ and let

$$\Omega_{n,d,1} = \left\{ \int_0^{T_n^{R,d} \wedge t_0} \|\bar{\beta}^{n,d}(\bar{X}^{n,d}(t)) - b^{(d)}(\bar{X}^{n,d}(t))\| dt \leq \frac{1}{2} (\log n)^{\frac{\delta-1}{2}} e^{-Kt_0} \right\},$$

$$\Omega_{n,d,2} = \left\{ \sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|M^{n,d}(t)\| \leq \frac{1}{2} (\log n)^{\frac{\delta-1}{2}} e^{-Kt_0} \right\}$$

and

$$\Omega_{n,d,3} = \left\{ \int_0^{T_n^{R,d} \wedge t_0} \bar{\alpha}^{n,d}(\bar{X}^{n,d}(t)) dt \leq \frac{C'(R) t_0}{\log n} \right\}.$$

From (12) we obtain that for $t < T_n^{R,d} \wedge t_0$ and on the event $\Omega_{n,d,1} \cap \Omega_{n,d,2}$,

$$\sup_{0 \leq s \leq t} \|\bar{X}^{n,d}(s) - x^{(d)}(s)\| \leq (\log n)^{\frac{\delta-1}{2}} e^{-Kt_0} + K \int_0^t \sup_{0 \leq r \leq s} \|\bar{X}^{n,d}(r) - x^{(d)}(r)\| ds.$$

Hence, by Gronwall's lemma,

$$\sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| \leq (\log n)^{\frac{\delta-1}{2}}.$$

Now, by Doob's L^2-inequality,

$$\mathbb{E}\left[\sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|M^{n,d}(t)\|^2 \right] \leq 4\mathbb{E}\left[\|M^{n,d}(T_n^{R,d} \wedge t_0)\|^2 \right] \leq 4\mathbb{E}\left[\int_0^{T_n^{R,d} \wedge t_0} \bar{\alpha}^{n,d}(\bar{X}^{n,d}(s)) ds \right].$$

Combined with Chebyshev's inequality, this tells us that

$$\mathbb{P}\left(\sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|M^{n,d}(t)\| \geq \frac{1}{2} (\log n)^{\frac{\delta-1}{2}} e^{-Kt_0}, \Omega_{n,d,3} \right) \leq \frac{16 C'(R) t_0 e^{2Kt_0}}{(\log n)^{\delta}}.$$

Hence, $\mathbb{P}(\Omega_{n,d,2} \setminus \Omega_{n,d,3}) \to 0$. By Lemmas 4.3 and 4.4, we have $\mathbb{P}(\Omega_{n,d,1}) \to 1$ and $\mathbb{P}(\Omega_{n,d,3}) \to 1$ as $n \to \infty$. But

$$\mathbb{P}\left(\sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| > (\log n)^{\frac{\delta-1}{2}} \right) \leq \frac{16 C'(R) t_0 e^{2Kt_0}}{(\log n)^{\delta}} + \mathbb{P}\left(\Omega_{n,d,1}^{c} \cup \Omega_{n,d,3}^{c} \right),$$

which clearly tends to 0 as $n \to \infty$.

In fact, we wish to prove this result for t_0 rather than $T_n^{R,d} \wedge t_0$. Set

$$\Omega_{n,R,d} = \left\{ \sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| \leq (\log n)^{\frac{\delta-1}{2}} \right\}.$$

Since $T_n^{R,d} = T_n^{R,d,1} \wedge T_n^{R,d,2}$, it will suffice for us to show that $T_n^{R,d,1} > t_0$ and $T_n^{R,d,2} > t_0$ on $\Omega_{n,R,d}$ for all large enough n and R.

Note firstly that $x_1(t) > 0$ for all $t \geq 0$ and that $x_1(t)$ decreases to 0 as $t \to \infty$. Take n and R large enough that $x_1(t_0) > (\log n)^{\frac{\delta-1}{2}} + l(n,R,d)$. Then on $\Omega_{n,R,d}$,

$$\begin{aligned}\inf_{0 \leq t \leq T_n^{R,d,1} \wedge t_0} |\bar{X}_1^n(t)| &\geq \inf_{0 \leq t \leq T_n^{R,d,1} \wedge t_0} \Big| x_1(t) - |\bar{X}_1^n(t) - x_1(t)| \Big| \\ &\geq x_1(t_0) - \sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| \\ &> l(n,R,d).\end{aligned}$$

Note now that $0 \leq y_2(t) \leq e^{-1} < R$ for all $t \geq 0$. Take n to be sufficiently big that $(\log n)^{\frac{\delta-1}{2}} + e^{-1} < R$. Then on $\Omega_{n,R,d}$, we have

$$\sup_{0 \leq t \leq T_n^{R,d,2} \wedge t_0} |\bar{Y}_2^n(t)| \leq \sup_{0 \leq t \leq T_n^{R,d} \wedge t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| + \sup_{0 \leq t \leq t_0} |y_2(t)| < R.$$

The desired result (10) follows.

Proof of Proposition 2.2

For convenience we will write

$$z_1(\infty) \stackrel{\text{def}}{=} \lim_{t \to \infty} z_1(t) = \rho \quad \text{and, for } k \geq 2, \quad z_k(\infty) \stackrel{\text{def}}{=} \lim_{t \to \infty} z_k(t) = \frac{\rho}{k(k-1)}.$$

Since we can take t_0 arbitrarily large, we can make $z_d(t_0)$ arbitrarily close to $z_d(\infty)$. With high probability, on the time interval $[0,t_0]$, $\frac{\log n}{n} Z_1^n(\frac{t}{\log n})$ stays close to $z_1(t)$ and, likewise, for $d \geq 2$, $\frac{(\log n)^2}{n} Z_d^n(\frac{t}{\log n})$ stays close to $z_d(t)$. So the work of this proof will be to demonstrate that $Z_d^n(t)$ does not do anything "nasty" between times $t_0/(\log n)$ and T_n. (Note that this interval is potentially quite long: absorption for the coalescent takes place at about time $\log\log n$; see Proposition 3.4 of [25].) We will have to split the interval $[t_0/(\log n), T_n]$ into two parts and deal with the process separately on each.

The statement of Proposition 2.2 fixes some $\delta > 0$. Take also $\eta > 0$ and fix t_0 such that

$$2x_1(t_0) + y_2(t_0) < \frac{\delta\eta}{24\rho}$$
$$\text{and, for } k \geq 1, \quad |z_k(t_0) - z_k(\infty)| < \frac{\delta}{3}.$$

(We can do this uniformly in k because of the special form of the functions $x_1(t)$, $y_2(t)$ and $z_k(t), k \geq 1$.) Take $0 < \epsilon < \frac{\delta}{3} \wedge \frac{\delta\eta}{72\rho} \wedge x_1(t_0)$. Let

$$\Omega_{n,d} = \left\{ \sup_{0 \leq t \leq t_0} \|\bar{X}^{n,d}(t) - x^{(d)}(t)\| < \epsilon \right\}.$$

Since $(\log n) T_n^{R,d,1} \leq T_n$ and $\epsilon < x_1(t_0)$, we know by the argument in the proof of Proposition 2.1 that $T_n > t_0/(\log n)$ on $\Omega_{n,1}$ for all sufficiently large n and R. Let

$$\tau_n = \inf\left\{ t \geq \frac{t_0}{\log n} : X_1^n(t) < \frac{n}{(\log n)^3} \text{ and } Y_2^n(t) < \frac{n}{(\log n)^3} \right\}.$$

We will first deal with the time interval $[t_0/(\log n), \tau_n)$.

Lemma 4.5. *For sufficiently large n, we have*

$$\frac{\log n}{n}\mathbb{E}\left[\left(Z_1^n(\tau_n) - Z_1^n\left(\frac{t_0}{\log n}\right)\right)\mathbb{1}_{\Omega_{n,1}}\right] \leq \frac{\delta\eta}{6} \tag{13}$$

and, for $k \geq 2$,

$$\frac{(\log n)^2}{n}\mathbb{E}\left[\left(Z_k^n(\tau_n) - Z_k^n\left(\frac{t_0}{\log n}\right)\right)\mathbb{1}_{\Omega_{n,1}}\right] \leq \frac{\delta\eta}{6}. \tag{14}$$

Proof. Consider a new process $(\tilde{X}_1^n(t), \tilde{Y}_2^n(t))_{t\geq 0}$ which starts from $(\bar{X}_1^n(t_0), \bar{Y}_2^n(t_0))$ and has the same dynamics as $(\bar{X}_1^n(t), \bar{Y}_2^n(t))_{t\geq 0}$, except with a stochastic time-change which means that time is now run at instantaneous rate $(\tilde{X}_1^n(t) + \tilde{Y}_2^n(t))^{-1}$. In other words, if

$$U^n(s) = \int_0^s \left[\frac{1}{n}X_1^n\left(\frac{t_0+u}{\log n}\right) + \frac{\log n}{n}Y_2^n\left(\frac{t_0+u}{\log n}\right)\right] du$$

and

$$V^n(t) = \inf\{s \geq 0 : U^n(s) \geq t\}$$

then

$$\tilde{X}_1^n(t) = \frac{1}{n}X_1^n\left(\frac{t_0+V^n(t)}{\log n}\right), \quad \tilde{Y}_2^n(t) = \frac{\log n}{n}Y_2^n\left(\frac{t_0+V^n(t)}{\log n}\right).$$

Let

$$\tilde{\tau}_n = U^n((\log n)\tau_n - t_0) = \inf\left\{t \geq 0 : \tilde{X}_1^n(t) < \frac{1}{(\log n)^3} \text{ and } \tilde{Y}_2^n(t) < \frac{1}{(\log n)^2}\right\}.$$

Then we have $\tilde{X}_1^n(\tilde{\tau}_n) = \frac{1}{n}X_1^n(\tau_n)$ and $\tilde{Y_2^n}(\tilde{\tau}_n) = \frac{\log n}{n}Y_2^n(\tau_n)$.
The process $(\tilde{X}_1^n(t), \tilde{Y}_2^n(t))_{t\geq 0}$ has drift vector $\tilde{\beta}^n(\xi)$ in state ξ, where

$$\begin{aligned}
\tilde{\beta}_1^n(\xi) &= -\frac{\rho\xi_1}{(\xi_1+\xi_2)\log n} - \frac{\xi_1 h(n\xi_1 + \frac{n}{\log n}\xi_2)}{(\xi_1+\xi_2)\log n} \\
\tilde{\beta}_2^n(\xi) &= -\frac{\rho\xi_2}{(\xi_1+\xi_2)\log n} - \frac{\xi_2 h(n\xi_1 + \frac{n}{\log n}\xi_2)}{(\xi_1+\xi_2)\log n} + \frac{\xi_1 + \frac{1}{\log n}\xi_2 - \frac{1}{n}}{(\xi_1+\xi_2)}.
\end{aligned}$$

Let

$$A^n(t) = 2\tilde{X}_1^n(t) + \tilde{Y}_2^n(t) + t, \quad t \geq 0.$$

Then $A^n(t)$ has drift

$$\begin{aligned}
&2\tilde{\beta}_1^n(\xi) + \tilde{\beta}_2^n(\xi) + 1 \\
&\quad = -\frac{2\xi_1+\xi_2}{\xi_1+\xi_2}\left(\frac{h(n\xi_1 + \frac{n}{\log n}\xi_2)}{\log n} - 1\right) + \frac{1}{\log n}\frac{\xi_2}{\xi_1+\xi_2} - \frac{\rho(2\xi_1+\xi_2)}{(\xi_1+\xi_2)\log n} - \frac{1}{n(\xi_1+\xi_2)},
\end{aligned}$$

in state ξ. Intuitively, this is small for large n and so $(A^n(t))_{t\geq 0}$ is almost a martingale. More rigorously, we have

$$2\tilde{\beta}_1^n(\xi) + \tilde{\beta}_2^n(\xi) + 1 \leq \frac{2\xi_1+\xi_2}{\xi_1+\xi_2}\left(1 - \frac{h(n\xi_1 + \frac{n}{\log n}\xi_2)}{\log n}\right) + \frac{1}{\log n}\frac{\xi_2}{\xi_1+\xi_2}.$$

Lemma 4.1 remains true if we replace R by $(\log n)^3$. So, since $\frac{\xi_1}{\xi_1+\xi_2}, \frac{\xi_2}{\xi_1+\xi_2} \leq 1$ in $\mathcal{S}^{n,d}$, we obtain

$$2\tilde{\beta}_1^n(\xi) + \tilde{\beta}_2^n(\xi) + 1 \leq \frac{6 \log\log n + 1}{\log n},$$

whenever $\xi \in \mathcal{S}^{n,d}$ and $\xi_1 + \frac{1}{\log n}\xi_2 \in \frac{1}{n}\mathbb{Z} \cap \left[1/(\log n)^3, 1\right]$.

By the same standard decomposition as in (11), there exists a zero-mean martingale $(\tilde{M}^n(t))_{t\geq 0}$ in the filtration $\mathcal{F}_t = \sigma(\tilde{X}^n(s), s \leq t)$ such that

$$A^n(t) = A^n(0) + \tilde{M}^n(t) + \int_0^t (2\tilde{\beta}_1^n(\tilde{X}^n(s)) + \tilde{\beta}_2^n(\tilde{X}^n(s)) + 1)ds.$$

Using the Markov property, we also note that, conditionally on $\mathcal{F}_0$, $\mathbb{1}_{\Omega_{n,1}}$ and $\tilde{M}_n(t)$ are independent random variables. Hence, for any bounded stopping time τ, we have

$$\mathbb{E}\left[\tilde{M}_n(\tau)\mathbb{1}_{\Omega_{n,1}}\right] = \mathbb{E}\left[\mathbb{E}\left[\tilde{M}_n(\tau)|\mathcal{F}_0\right]\mathbb{E}\left[\mathbb{1}_{\Omega_{n,1}}|\mathcal{F}_0\right]\right] = \mathbb{E}\left[\tilde{M}_n(0)\mathbb{E}\left[\mathbb{1}_{\Omega_{n,1}}|\mathcal{F}_0\right]\right] = 0.$$

Fix $t_1 > 0$. For any particular n, $A^n(t)$ and $\int_0^t (2\tilde{\beta}_1^n(\tilde{X}^n(s)) + \tilde{\beta}_2^n(\tilde{X}^n(s)) + 1)ds$ are bounded on the time interval $[0, t_1]$ and so we may apply the Optional Stopping Theorem to obtain that

$$\begin{aligned}
&\mathbb{E}\left[\left(2\tilde{X}_1^n(\tilde{\tau}_n \wedge t_1) + \tilde{Y}_2^n(\tilde{\tau}_n \wedge t_1) + (\tilde{\tau}_n \wedge t_1)\right)\mathbb{1}_{\Omega_{n,1}}\right] \\
&\quad = \mathbb{E}\left[A^n(\tilde{\tau}_n \wedge t_1)\mathbb{1}_{\Omega_{n,1}}\right] \\
&\quad = \mathbb{E}\left[A^n(0)\mathbb{1}_{\Omega_{n,1}}\right] + \mathbb{E}\left[\int_0^{\tilde{\tau}_n \wedge t_1} (2\tilde{\beta}_1^n(\tilde{X}^n(t)) + \tilde{\beta}_2^n(\tilde{X}^n(t)) + 1)dt\mathbb{1}_{\Omega_{n,1}}\right] \\
&\quad = \mathbb{E}\left[\left(2\bar{X}_1^n(t_0) + \bar{Y}_2^n(t_0)\right)\mathbb{1}_{\Omega_{n,1}}\right] + \mathbb{E}\left[\int_0^{\tilde{\tau}_n \wedge t_1} (2\tilde{\beta}_1^n(\tilde{X}^n(t)) + \tilde{\beta}_2^n(\tilde{X}^n(t)) + 1)dt\mathbb{1}_{\Omega_{n,1}}\right].
\end{aligned}$$

We have that $\tilde{X}_1^n(\tilde{\tau}_n \wedge t_1)$ and $\tilde{Y}_2^n(\tilde{\tau}_n \wedge t_1)$ are both non-negative and so

$$\begin{aligned}
\mathbb{E}\left[(\tilde{\tau}_n \wedge t_1)\mathbb{1}_{\Omega_{n,1}}\right] &\leq \left(1 - \frac{6\log\log n + 1}{\log n}\right)^{-1} \mathbb{E}\left[\left(2\bar{X}_1^n(t_0) + \bar{Y}_2^n(t_0)\right)\mathbb{1}_{\Omega_{n,1}}\right] \\
&\leq 2(2x_1(t_0) + y_2(t_0) + 3\epsilon) \\
&< \frac{\delta\eta}{6\rho},
\end{aligned}$$

since, for large enough n, $\left(1 - \frac{6\log\log n+1}{\log n}\right)^{-1}$ is bounded above by 2 and we have assumed that $2x_1(t_0) + y_2(t_0) < \delta\eta/(24\rho)$ and $\epsilon < \delta\eta/(72\rho)$. Letting $t_1 \uparrow \infty$, we obtain by monotone convergence that

$$\mathbb{E}\left[\tilde{\tau}_n\mathbb{1}_{\Omega_{n,1}}\right] < \frac{\delta\eta}{6\rho}.$$

Now, by a further application of the Optional Stopping Theorem and monotone convergence,

$$\begin{aligned}
\frac{\log n}{n}\mathbb{E}\left[\left(Z_1^n(\tau_n) - Z_1^n\left(\frac{t_0}{\log n}\right)\right)\mathbb{1}_{\Omega_{n,1}}\right] &= \frac{\log n}{n}\mathbb{E}\left[\int_{\frac{t_0}{\log n}}^{\tau_n} \rho X_1^n(t)dt\mathbb{1}_{\Omega_{n,1}}\right] \\
&= \mathbb{E}\left[\int_0^{\tilde{\tau}_n} \frac{\rho\tilde{X}_1^n(s)}{\tilde{X}_1^n(s) + \tilde{Y}_2^n(s)}ds\mathbb{1}_{\Omega_{n,1}}\right] \\
&\leq \rho\mathbb{E}\left[\tilde{\tau}_n\mathbb{1}_{\Omega_{n,1}}\right],
\end{aligned}$$

by changing variable in the integral. Similarly, for $k \geq 2$,

$$\begin{aligned}
\frac{(\log n)^2}{n}\mathbb{E}\left[\left(Z_k^n(\tau_n) - Z_k^n\left(\frac{t_0}{\log n}\right)\right)\mathbb{1}_{\Omega_{n,1}}\right] &= \frac{(\log n)^2}{n}\mathbb{E}\left[\int_{\frac{t_0}{\log n}}^{\tau_n} \rho X_k^n(t)dt\mathbb{1}_{\Omega_{n,1}}\right] \\
&\leq \frac{(\log n)^2}{n}\mathbb{E}\left[\int_{\frac{t_0}{\log n}}^{\tau_n} \rho Y_2^n(s)ds\mathbb{1}_{\Omega_{n,1}}\right] \\
&= \mathbb{E}\left[\int_0^{\tilde{\tau}_n} \frac{\rho Y_2^{\tilde{n}}(s)}{\tilde{X}_1^n(s) + \tilde{Y}_2^n(s)}ds\mathbb{1}_{\Omega_{n,1}}\right] \\
&\leq \rho\mathbb{E}\left[\tilde{\tau}_n\mathbb{1}_{\Omega_{n,1}}\right].
\end{aligned}$$

The result follows. □

From (13) and (14) and Markov's inequality,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{\log n}{n}\left(Z_1^n(\tau_n) - Z_1^n\left(\frac{t_0}{\log n}\right)\right)\right| > \frac{\delta}{3}, \Omega_{n,1}\right) \\
&\qquad \leq \frac{3\log n}{n\delta}\mathbb{E}\left[\left(Z_1^n(\tau_n) - Z_1^n\left(\frac{t_0}{\log n}\right)\right)\mathbb{1}_{\Omega_{n,1}}\right] \leq \frac{\eta}{2}
\end{aligned}$$

and, for $k \geq 2$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{(\log n)^2}{n}\left(Z_k^n(\tau_n) - Z_k^n\left(\frac{t_0}{\log n}\right)\right)\right| > \frac{\delta}{3}, \Omega_{n,1}\right) \\
&\qquad \leq \frac{3(\log n)^2}{n\delta}\mathbb{E}\left[\left(Z_k^n(\tau_n) - Z_k^n\left(\frac{t_0}{\log n}\right)\right)\mathbb{1}_{\Omega_{n,1}}\right] \leq \frac{\eta}{2}.
\end{aligned}$$

Note that we necessarily have $\tau_n \leq T_n$. Since $Z_k^n(t)$ is increasing for all $k \geq 1$ and $Z_k^n(T_n) - Z_k^n(\tau_n) \leq X_1^n(\tau_n) + Y_2^n(\tau_n) < 2n/(\log n)^3$ for all $k \geq 1$, we have that

$$\frac{\log n}{n}\left(Z_1^n(T_n) - Z_1^n(\tau_n)\right) < \frac{2}{(\log n)^2}$$

and, for $k \geq 2$,

$$\frac{(\log n)^2}{n}\left(Z_k^n(T_n) - Z_k^n(\tau_n)\right) < \frac{2}{\log n}.$$

For $n > \exp(6/\delta)$ these quantities are both less than $\delta/3$. On $\Omega_{n,1}$ we have

$$\left|\frac{\log n}{n}Z_1^n\left(\frac{t_0}{\log n}\right) - z_1(t_0)\right| \leq \frac{\delta}{3}.$$

By taking n sufficiently large, we have by Proposition 2.1 that $\mathbb{P}\left(\Omega_{n,1}^{\mathrm{c}}\right) < \eta/2$ and so we conclude that

$$\mathbb{P}\left(\left|\frac{\log n}{n}Z_1^n(T_n) - z_1(\infty)\right| > \delta\right) < \eta.$$

Now consider the case $d \geq 2$. On $\Omega_{n,d}$ we have

$$\left| \frac{(\log n)^2}{n} Z_d^n \left(\frac{t_0}{\log n} \right) - z_d(t_0) \right| \leq \frac{\delta}{3}$$

and, by taking n sufficiently large, we have by Proposition 2.1 that $\mathbb{P}\left(\Omega_{n,1}^{\mathrm{c}}\right) + \mathbb{P}\left(\Omega_{n,d}^{\mathrm{c}}\right) < \eta/2$. Hence,

$$\mathbb{P}\left(\left| \frac{(\log n)^2}{n} Z_d^n(T_n) - z_d(\infty) \right| > \delta \right) < \eta.$$

But η was arbitrary and so this completes the proof of Proposition 2.2.

Acknowledgments

A.-L. B. would like to thank the Statistical Laboratory at the University of Cambridge for its kind invitation, which was the starting point of this paper. The early part of the work was done while C. G. held the Stokes Research Fellowship at Pembroke College, Cambridge. Pembroke's support is most gratefully acknowledged. For the later part, C. G. was funded by EPSRC Postdoctoral Fellowship EP/D065755/1. We would like to thank James Norris for several extremely helpful discussions.

References

[1] A. D. Barbour and A. V. Gnedin. Regenerative compositions in the case of slow variation. *Stochastic Process. Appl.*, 116(7):1012–1047, 2006. MR2238612

[2] A.-L. Basdevant. Ruelle's probability cascades seen as a fragmentation process. *Markov Process. Related Fields*, 12(3):447–474, 2006. MR2246260

[3] J. Berestycki, N. Berestycki, and J. Schweinsberg. Beta-coalescents and continuous stable random trees. *Ann. Probab.*, 35(5):1835–1887, 2007. MR2349577

[4] J. Berestycki, N. Berestycki, and J. Schweinsberg. Small-time behavior of Beta-coalescents. `arXiv:math/0601032`. To appear in *Ann. Inst. H. Poincaré Probab. Statist.*, 2007.

[5] J. Bertoin and J.-F. Le Gall. The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields*, 117(2):249–266, 2000. MR1771663

[6] M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger. Alpha-stable branching and beta-coalescents. *Electron. J. Probab.*, 10:303–325 (electronic, paper no. 9), 2005. MR2120246

[7] E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998. MR1652734

[8] H. Chen and D. D. Yao. *Fundamentals of Queueing Networks. Performance, Asymptotics and Optimization*, volume 46 of *Applications of Mathematics*. Springer-Verlag, New York, 2001. Stochastic Modelling and Applied Probability. MR1835969

[9] R. W. R. Darling and J. R. Norris. Structure of large random hypergraphs. *Ann. Appl. Probab.*, 15(1A):125–152, 2005. MR2115039

[10] R. W. R. Darling and J. R. Norris. Differential equation approximations for Markov chains. `arXiv:0710.3269`, 2007.

[11] J.-F. Delmas, J.-S. Dhersin, and A. Siri-Jegousse. Asymptotic results on the length of coalescent trees. `arXiv:0706.0204`. To appear in *Ann. Appl. Probab.*, 2007.

[12] R. Dong, A. Gnedin, and J. Pitman. Exchangeable partitions derived from Markovian coalescents. *Ann. Appl. Probab.*, 17(4):1172–1201, 2007. MR2344303

[13] M. Drmota, A. Iksanov, M. Moehle, and U. Roesler. Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stochastic Process. Appl.*, 117(10):1404–1421, 2007. MR2353033

[14] R. Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer-Verlag, New York, 2002. MR1903526

[15] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Popul. Biol.*, 3:87–112, 1972. MR0325177

[16] W. J. Ewens. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. Theoretical introduction. MR2026891

[17] A. Gnedin. Regenerative composition structures: characterisation and asymptotics of block counts. In *Mathematics and computer science. III*, Trends Math., pages 441–443. Birkhäuser, Basel, 2004. Based on joint work with Jim Pitman and Marc Yor. MR2090532

[18] A. Gnedin, B. Hansen, and J. Pitman. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.*, 4:146–171 (electronic), 2007. MR2318403

[19] A. Gnedin and J. Pitman. Regenerative partition structures. *Electron. J. Combin.*, 11(2):Paper no. 12, 21 pp. (electronic), 2004/06. MR2120107

[20] A. Gnedin, J. Pitman, and M. Yor. Asymptotic laws for compositions derived from transformed subordinators. *Ann. Probab.*, 34(2):468–492, 2006. MR2223948

[21] A. Gnedin, J. Pitman, and M. Yor. Asymptotic laws for regenerative compositions: gamma subordinators and the like. *Probab. Theory Related Fields*, 135(4):576–602, 2006. MR2240701

[22] A. Gnedin and Y. Yakubovich. On the number of collisions in Λ-coalescents. *Electron. J. Probab.*, 12:1547–1567 (electronic, paper no. 56), 2007. MR2365877

[23] A. V. Gnedin. The Bernoulli sieve. *Bernoulli*, 10(1):79–96, 2004. MR2044594

[24] A. V. Gnedin and Y. Yakubovich. Recursive partition structures. *Ann. Probab.*, 34(6):2203–2218, 2006. MR2294980

[25] C. Goldschmidt and J. B. Martin. Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.*, 10:718–745 (electronic, paper no. 21), 2005. MR2164028

[26] A. Iksanov and M. Möhle. A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.*, 12:28–35 (electronic), 2007.

[27] S. Karlin. Central limit theorems for certain infinite urn schemes. *J. Math. Mech.*, 17:373–401, 1967. MR0216548

[28] J. F. C. Kingman. Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A*, 361(1704):1–20, 1978. MR0526801

[29] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc. (2)*, 18(2):374–380, 1978. MR0509954

[30] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3):235–248, 1982. MR0671034

[31] M. Möhle. On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli*, 12(1):35–53, 2006. MR2202319

[32] M. Möhle. On the number of segregating sites for populations with large family sizes. *Adv. in Appl. Probab.*, 38(3):750–767, 2006. MR2256876

[33] M. Möhle. On a class of non-regenerative sampling distributions. *Combin. Probab. Comput.*, 16:435–444, 2007. MR2312437

[34] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 1999. MR1742892

[35] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. With a foreword by Jean Picard. MR2245368

[36] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giant k-core in a random graph. *J. Combin. Theory Ser. B*, 67(1):111–151, 1996. MR1385386

[37] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):1116–1125, 1999. MR1742154

[38] J. Schweinsberg. A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Comm. Probab.*, 5:1–11 (electronic), 2000. MR1736720

[39] W. Whitt. *Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and Their Application to Queues.* Springer Series in Operations Research. Springer-Verlag, New York, 2002. MR1876437

[40] N. C. Wormald. Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, 5(4):1217–1235, 1995. MR1384372