Electron. Commun. Probab. **17** (2012), no. 3, 1–11.
DOI: 10.1214/ECP.v17-1704
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Almost sure asymptotics for the number of types for simple Ξ-coalescents

Fabian Freund*

Abstract

Let K_n be the number of types in the sample $\{1,\dots,n\}$ of a Ξ-coalescent $\Pi=(\Pi_t)_{t\geq 0}$ with mutation and mutation rate $r>0$. Let $\Pi^{(n)}$ be the restriction of Π to the sample. It is shown that M_n/n, the fraction of external branches of $\Pi^{(n)}$ which are affected by at least one mutation, converges almost surely and in L^p ($p\geq 1$) to $M:=\int_0^\infty re^{-rt}S_t dt$, where S_t is the fraction of singleton blocks of Π_t. Since for coalescents without proper frequencies, the effects of mutations on non-external branches is neglectible for the asymptotics of K_n/n, it is shown that $K_n/n\to M$ for $n\to\infty$ in L^p ($p\geq 1$). For simple coalescents, this convergence is shown to hold almost surely. The almost sure results are based on a combination of the Kingman correspondence for random partitions and strong laws of large numbers for weighted i.i.d. or exchangeable random variables.

Keywords: almost sure convergence; coalescent; external branches; mutation.
AMS MSC 2010: Primary 60F15; 05C05, Secondary 60F25; 92D15.
Submitted to ECP on November 10, 2009, final version accepted on October 30, 2011.

1 Introduction

Let $\mathcal{E}$ be the set of partitions of $\mathbb{N}:=\{1,2,\dots\}$, where every $e\in\mathcal{E}$ is represented by the collection of its blocks ordered by their least elements. A coalescent process $(\Pi_t)_{t\geq 0}$ (or simply coalescent) is a *càdlàg* stochastic process with state space $\mathcal{E}$. Its distinguishing feature is the following block-merging mechanism: For $n\in\mathbb{N}$ let $(\Pi_t^{(n)})_{t\geq 0}:=(\varrho_n\circ\Pi_t)_{t\geq 0}$ be the restriction of $(\Pi_t)_{t\geq 0}$ to the set $\mathcal{E}_n$ of partitions of $\{1,\dots,n\}$ (called n-coalescent). Then $(\Pi_t^{(n)})_{t\geq 0}$ is Markovian and all transitions are done by merging blocks of the current state which is a partition of $\{1,\dots,n\}$. The rate of such a transition is determined by the number of blocks present before the merger and by the numbers of blocks that are merged together to form each new block (independent of the sizes of the blocks and of n). Each forming of a new block during such a transition is called a collision.

The first and most important example is the Kingman coalescent, which allows only for one binary merger at each transition. Sagitov [19, p. 1117] and Pitman [18, Theorem 1] independently characterise the distribution of a coalescent process allowing only for one multiple collision at a transition by a finite measure Λ on the interval $[0,1]$. The rates are given by

$$\lambda_{b,k}:=\int_{[0,1]}x^{k-2}(1-x)^{b-k}\Lambda(dx),$$

*University of Hohenheim, Germany. E-mail: ffreund@uni-hohenheim.de

when k blocks of the b present blocks are merged during a transition of $(\Pi_t^{(n)})_{t\geq 0}$. These coalescents are called coalescents with multiple collisions or Λ-coalescents. There is an analog characterisation for the distribution of any coalescent (allowing for simultaneous multiple collisions) by Schweinsberg [21, Theorem 2] which uses a finite measure Ξ on the infinite simplex $\Delta := \{(x_1, x_2, \ldots) | x_i \in \mathbb{R},\ x_1 \geq x_2 \geq \cdots \geq 0,\ \sum_{i\in\mathbb{N}} x_i \leq 1\}$. Throughout this work, for convenience define $(x,x) := \sum_{i\in\mathbb{N}} x_i^2$, $|x| := \sum_{i\in\mathbb{N}} x_i$ and $a := \Xi(\{0\})$. Using the appropriate characterizing measure Ξ, the rates of a coalescent are then given by

$$\lambda(b; k_1, \ldots, k_r) := \int_{\Delta\setminus\{0\}} \Big(\sum_{l=0}^{s} \sum_{i_1\neq\cdots\neq i_{r+l}} \binom{s}{l} \prod_{j=1}^{r} x_{i_j}^{k_j} \prod_{m=1}^{l} x_{i_{r+m}} (1-|x|)^{s-l}\Big) \frac{\Xi(dx)}{(x,x)} + a 1_{\{r=1,k_1=2\}}$$

for each transition in which $r \geq 1$ sets of $k_1 \geq \ldots \geq k_r \geq 2$ of b present blocks are merged into one block each and $s \geq 0$ blocks remain unmerged. Since every coalescent is characterised by such a measure Ξ, coalescents are also called Ξ-coalescents. Note that in this paper the case of Ξ being the zero measure is excluded, which corresponds to the case of a coalescent having no collisions at all.

Coalescents can be divided into two different classes. The first class is the class of coalescents with proper frequencies. A coalescent has proper frequencies if and only if the fraction S_t of singleton blocks (i.e. blocks with only one element) of all individuals in the coalescent fulfills $S_t = 0$ almost surely for all $t > 0$ (see [21, p. 37]). This class includes important coalescents such as the Kingman coalescent (where Λ is the Dirac measure in 0) and the Bolthausen-Sznitman coalescent (where Λ is the uniform distribution on $[0,1]$). All other coalescents are called coalescents without proper frequencies. A coalescent has no proper frequencies (see [21, Proposition 30]) if and only if the characterising measure Ξ satisfies

$$\Xi(\{0\}) = 0 \text{ and } \mu_{-1} := \int_{\Delta\setminus\{0\}} |x| \frac{\Xi(dx)}{(x,x)} < \infty. \tag{1.1}$$

For Λ-coalescents this is equivalent to (see [18, Theorem 8])

$$\Lambda(\{0\}) = 0 \text{ and } \mu_{-1} := \int_{(0,1]} x^{-1} \Lambda(dx) < \infty. \tag{1.2}$$

In this paper we focus on the class of coalescents without proper frequencies, especially the subclass of simple coalescents in the spirit of Bertoin and LeGall (see [6, p. 275]), which are those Ξ-coalescents satisfying

$$\Xi(\{0\}) = 0 \text{ and } \mu_{-2} := \int_{\Delta\setminus\{0\}} \frac{\Xi(dx)}{(x,x)} < \infty. \tag{1.3}$$

For Λ-coalescents, this can also be expressed by

$$\Lambda(\{0\}) = 0 \text{ and } \mu_{-2} := \int_{(0,1]} x^{-2} \Lambda(dx) < \infty. \tag{1.4}$$

This class includes the Dirac coalescents with Ξ being the Dirac measure in a point $x \in \Delta \setminus \{0\}$ and the Poisson-Dirichlet coalescents. See [16, Section 4.1.2], [17, Section 6] and [20, Section 3] for some properties of Poisson-Dirichlet coalescents.

In population genetics, the restricted coalescent $(\Pi_t^{(n)})_{t\geq 0}$ is used as a model for the genealogical tree of a sample of n individuals in a large (infinite) haploid population.

It is also possible to introduce a mutation mechanism to this model in the following

way: For every $n \in \mathbb{N}$, let a homogeneous Poisson process $\Psi^{(n)}$ with rate $r > 0$ generate points along the branches of the genealogical tree of $(\Pi_t^{(n)})_{t\geq 0}$ and independently of that tree. This can and will be done in a pathwise consistent manner, meaning that for $m, n \in \mathbb{N}$ with $m < n$ the mutation structure on the genealogical tree of $\Pi_t^{(m)}$ given by $\Psi^{(m)}$ is pathwise the same as the mutation structure on the genealogical tree of the individuals $\{1, \ldots, m\}$ if their genealogy and mutations are tracked in the genealogical tree of $(\Pi_t^{(n)})_{t\geq 0}$ with mutations $\Psi^{(n)}$. In this paper, mutations are neutral with respect to reproduction and will behave according to the infinitely many alleles model. The infinitely many alleles model states that each individual $i \in \mathbb{N}$ inherits a common type (of the common ancestor), but this type gets changed into a completely new type not yet present in the sample/population by every mutation that occurs on the tree while following the branch leading to the external node of the tree which symbolizes i. $((\Pi_t)_{t\geq 0}, \Psi)$ is called a coalescent with mutation, where $\Psi := (\Psi^{(n)})_{n\in\mathbb{N}}$, and $r > 0$ is called its mutation rate.
An interesting quantity of coalescents with mutation is K_n, the number of different types present in the sample $\{1, \ldots, n\}$. Stated more precisely, K_n is the number of different types among the external nodes $\{1, \ldots, n\}$ of the tree generated by $((\Pi_t^{(n)})_{t\geq 0}, \Psi)$ when mutations are interpreted in the infinitely many alleles model. K_n has been analyzed for many coalescent processes with mutation, for example for the Kingman coalescent (closely linked to the celebrated Ewens' sampling formula) [10], the Bolthausen-Sznitman coalescent [2, p. 6], some beta coalescents [4, Theorem 1.9], [5, Theorem 9], [3, Theorem 3, Theorem 4] (if Λ is a $\beta(2-\alpha, \alpha)$-distribution with $1 < \alpha < 2$ or a similar distribution) and the class of coalescents without proper frequencies [11, Theorem 1.1, Theorem 1.2]. The present paper focuses on the class of coalescents without proper frequencies and especially on simple coalescents. Roughly speaking, this class of coalescents (more precisely, the restricted coalescents) can be seen as models for the genealogical trees of populations (more precisely, for a sample of that population) where one individual can have a huge number of offsprings, for example a fraction of the whole population. This class appears in the work [8] of Eldon and Wakeley, see also [1], [12] and [22] for some other models where coalescents without proper frequencies/simple coalescents arise as genealogical trees. In this paper, an alternative proof of a L^p-version ($p \geq 1$) of the convergence of K_n/n proven in [11, Theorem 1.2] is given. For the class of simple coalescents, this convergence is shown to also hold almost surely.

2 Number of mutated external branches

Let $((\Pi_t)_{t\geq 0}, \Psi)$ be a coalescent with mutation. For $n \in \mathbb{N}$ let $(\Pi_t^{(n)})_{t\geq 0}$ be the restriction of $(\Pi_t)_{t\geq 0}$ on $\mathcal{E}_n$. Recall that if the (n-)coalescent is seen as a tree, the ith external branch is the edge connecting the external node which represents the ith individual with the rest of the tree. For $i \in \mathbb{N}$ let E_i be the time until the ith individual collides for the first time. Analogously, let $E_i^{(n)}$ be the waiting time for the first collision of individual i in $(\Pi_t^{(n)})_{t\geq 0}$. These waiting times for individual i are the length of the ith external branch (either in the restricted or the unrestricted coalescent). Note that

$$E_i^{(i)} \geq E_i^{(n)} \geq E_i \text{ for all } i, n \in \mathbb{N},\ i \leq n. \tag{2.1}$$

For every $t \geq 0$, Π_t is an exchangeable partition of $\mathbb{N}$. Recall that due to Kingman's representation theorem (see [15, Theorem 2, p. 240]), the frequency

$$f_{i,t} = \lim_{n\to\infty} n^{-1} |A_i \cap \{1, \ldots, n\}|$$

of the ith biggest block A_i of Π_t exists almost surely ($A_i := \emptyset$ if there are fewer than i blocks) and an individual $i \in \mathbb{N}$ is a singleton almost surely if and only if it is not part of a block with positive frequency. This means that for $t \geq 0$

$$S_t^{(n)} := \frac{1}{n} \sum_{i=1}^{n} 1_{\{E_i^{(n)} \geq t\}} \longrightarrow S_t \text{ almost surely as } n \to \infty, \tag{2.2}$$

where $S_t = 1 - \sum_{i \in \mathbb{N}} f_{i,t}$ is the fraction of singletons of Π_t. During the waiting time for the first collision (the length of the external branch) each individual can be affected by one or several mutations according to independent Poisson processes on these branches. So every time there is a Poisson point on the ith external branch, the individual i is mutated. First, the asymptotic behaviour of M_n, the number of mutated external branches, i.e. how many external branches of $(\Pi_t^{(n)})_{t \geq 0}$ are affected by at least one mutation, is analyzed. This will be a first step towards the analysis of the asymptotic behaviour of $(K_n)_{n \in \mathbb{N}}$ for coalescents without proper frequencies, since for these coalescents it will be shown that asymptotically for $n \to \infty$, $(K_n/n)_{n \in \mathbb{N}}$ behaves like $(M_n/n)_{n \in \mathbb{N}}$. In order to analyze the asymptotics of $(M_n/n)_{n \in \mathbb{N}}$, a strong law of large numbers for weighted sums of i.i.d. random variables is used (see, for example, [7, Theorem 1.1]).

Theorem 2.1 (Strong law for weighted sums)**.** *Let $(A_{in})_{1 \leq i \leq n, n \in \mathbb{N}}$ be an array of random variables with $\sup_{1 \leq i \leq n, n \in \mathbb{N}} |A_{in}| < \infty$ almost surely and $\sum_{i=1}^{n} A_{in}/n \to S$ almost surely as $n \to \infty$. Let $X_1, X_2, \ldots$ be i.i.d. integrable random variables independent of $(A_{in})_{1 \leq i \leq n, n \in \mathbb{N}}$. Then*

$$\frac{1}{n} \sum_{i=1}^{n} A_{in} X_i \longrightarrow S \cdot E(X_1) \quad \textit{almost surely as } n \to \infty.$$

Proof. Theorem 1.1 and Remark (v) of [7] yield $\frac{1}{n} \sum_{i=1}^{n} A_{in}(X_i - E(X_i)) \to 0$ almost surely as $n \to \infty$. The result follows, since

$$\frac{1}{n} \sum_{i=1}^{n} A_{in} E(X_i) = E(X_1) \frac{1}{n} \sum_{i=1}^{n} A_{in} \to E(X_1) S \quad \text{almost surely as } n \to \infty.$$

□

This is the main tool to prove the following result.

Theorem 2.2. *Let $(\Pi_t)_{t \geq 0}$ be a Ξ-coalescent with mutation rate $r > 0$. Then,*

$$\frac{M_n}{n} \longrightarrow \int_0^\infty r e^{-rt} S_t dt \quad \textit{almost surely and in } L^p \ (p \geq 1) \textit{ as } n \to \infty. \tag{2.3}$$

Proof. The proof is divided into two parts: First, a result similar to (2.3) is proven for $M_n^{(t)}$, the number of external branches of $(\Pi_t^{(n)})_{t \geq 0}$ that are affected by at least one mutation until time $t > 0$ (t excluded). This is done by establishing upper and lower bounds for $M_n^{(t)}$ for all $n \in \mathbb{N}$ and showing that these bounds converge for $n \to \infty$. By successive refinement of the bounds for $n \in \mathbb{N}$, it will be shown that the limits of these bounds coincide eventually, thus showing that $M_n^{(t)}$ converges to the same limit. Afterwards, the convergence result for $M_n^{(t)}$ will be extended to a convergence result for M_n.
Before the proof is started, it is helpful to impose an i.i.d. structure on the model of the coalescent with mutation in the following way. Recall first that the mutations on the external branches of the n-coalescent are modelled as points of a Poisson point process

independent of the n-coalescent. Now, for every $i \in \mathbb{N}$, take a copy of $[0,\infty)$. For $i > 1$, regard all Poisson points on the ith external branch of $(\Pi_t^{(i)})_{t\geq 0}$ as points on the interval $[0, E_i^{(i)})$ of the ith copy of $[0,\infty)$ where 0 represents the leaf. For $i = 1$, regard all Poisson points on the first external branch of $(\Pi_t^{(2)})_{t\geq 0}$ as points on the interval $[0, E_1^{(2)})$ of the first copy of $[0,\infty)$ instead. Note that, by construction, the points on different copies of $[0,\infty)$ are independent. Now, take a Poisson process P_i' from an i.i.d. collection of homogeneous Poisson processes $(P_i')_{i\in\mathbb{N}}$ on $[0,\infty)$ with intensity rate r independent of the coalescent with mutation and shift all points by adding $E_i^{(i)}$ (or $E_1^{(2)}$ for $i = 1$). For every $i \in \mathbb{N}$, regard these shifted points also as points on the ith copy of $[0,\infty)$. Define P_i as the set consisting of all Poisson points on the ith external branch and all shifted points of P_i' (both regarded as points on the ith copy of $[0,\infty)$). By independence of P_i' and the Poisson points on the coalescent branches, the resulting concatenated random set P_i of points on the ith copy of $[0,\infty)$ is again a homogeneous Poisson process with rate r on $[0,\infty)$. Its restriction to $[0, E_i^{(n)})$ for $n \in \mathbb{N}$, $i \leq n$, shows the position of each mutation on the ith external branch of $(\Pi_t^{(n)})_{t\geq 0}$ starting at the leaf. This follows from (2.1) and the pathwise consistency of the mutation process $(\Psi^{(n)})_{n\in\mathbb{N}}$. Note that due to independence of both the mutations on different branches and the Poisson processes $(P_i')_{i\in\mathbb{N}}$, $(P_i)_{i\in\mathbb{N}}$ is an i.i.d. collection of Poisson processes on $[0,\infty)$.
The first step of the proof is establishing a convergence result similar to (2.3) for $M_n^{(t)}$, the number of external branches that are affected by at least one mutation until time $t > 0$ (t excluded). Define $t_j := jt/k$ for $0 \leq j \leq k$,

$$A_{inj} := 1_{\left\{t_{j-1}\leq E_i^{(n)}<t_j\right\}} \text{ for } n \in \mathbb{N}, 1 \leq i \leq n, 1 \leq j < k \text{ and}$$

$$A_{ink} := 1_{\left\{t_{k-1}\leq E_i^{(n)}\right\}} \text{ for } n \in \mathbb{N}, 1 \leq i \leq n.$$

Let $Y_i^{(j)} := 1_{\{|P_i\cap[0,t_j)|>0\}}$ for $i \in \mathbb{N}, 0 \leq j \leq k$ indicate whether there is a point (mutation) of P_i on $[0,t_j)$. Note that, for each fixed $j \in \{1,\dots,k\}$, $(Y_i^{(j)})_{i\in\mathbb{N}}$ is i.i.d. with $E(Y_i^{(j)}) = P(|P_i \cap [0,t_j)| > 0) = 1 - e^{-rt_j}$. Now construct upper and lower bounds for $M_n^{(t)}$. For each external branch i of length s with $t_{j-1} \leq s < t_j$ for some $1 \leq j \leq k$, $Y_i^{(j)}$ is an upper bound and $Y_i^{(j-1)}$ is a lower bound for $1_{\{|P_i\cap[0,s)|>0\}}$. For an external branch i with length $s \geq t$, there is only a contribution to $M_n^{(t)}$ if there is a mutation of P_i on $[0,t)$. Since $Y_i^{(k)}$ is an upper bound and $Y_i^{(k-1)}$ is a lower bound for $1_{\{|P_i\cap[0,t)|>0\}}$, these variables are also bounds for the contribution to $M_n^{(t)}$ if the external branch has length $s \geq t$. In order to get an upper/lower bound for $M_n^{(t)}$, one has to sum the appropriate bounds for the contributions to $M_n^{(t)}$ of each external branch. Thus,

$$\sum_{j=1}^{k}\sum_{i=1}^{n}\frac{A_{inj}Y_i^{(j)}}{n} \geq \frac{M_n^{(t)}}{n} \geq \sum_{j=1}^{k}\sum_{i=1}^{n}\frac{A_{inj}Y_i^{(j-1)}}{n}, \tag{2.4}$$

where the variables $(A_{inj})_{1\leq i\leq n, 1\leq j\leq k}$ decide which bound is used depending on the length of the external branch whose contribution is estimated. For $n \to \infty$, Theorem 2.1 yields for $j \in \{1,\dots,k-1\}$

$$\begin{aligned}&\sum_{i=1}^{n}\frac{A_{inj}Y_i^{(j)}}{n} \overset{a.s.}{\to} E(Y_1^{(j)})(S_{t_{j-1}} - S_{t_j})\\ =\ & (1-e^{-rt_j})(S_{t_{j-1}} - S_{t_j}) = \int_0^{t_j} re^{-ru}(S_{t_{j-1}} - S_{t_j})du \text{ and}\end{aligned}$$

$$\sum_{i=1}^{n} \frac{A_{inj} Y_i^{(j-1)}}{n} \overset{a.s.}{\to} E(Y_1^{(j-1)})(S_{t_{j-1}} - S_{t_j})$$

$$= \quad (1 - e^{rt_{j-1}})(S_{t_{j-1}} - S_{t_j}) = \int_0^{t_{j-1}} re^{-ru}(S_{t_{j-1}} - S_{t_j})du,$$

since $S_{t_{j-1}} - S_{t_j}$ is almost surely the fraction of the external branches $(E_i)_{i\in\mathbb{N}}$ of the coalescent process that are at least of length t_{j-1} but shorter than t_j. For $j = k$, the same argument shows

$$\sum_{i=1}^{n} \frac{A_{ink} Y_i^{(k)}}{n} \overset{a.s.}{\to} \int_0^{t_k} re^{-ru} S_{t_{k-1}} du \text{ and}$$

$$\sum_{i=1}^{n} \frac{A_{ink} Y_i^{(k-1)}}{n} \overset{a.s.}{\to} \int_0^{t_{k-1}} re^{-ru} S_{t_{k-1}} du$$

for $n \to \infty$.
Thus, as $n \to \infty$, (2.4) becomes (after sorting with respect to S_{t_j})

$$\int_0^t re^{-ru} \sum_{j=1}^{k} S_{t_{j-1}} 1_{\{t_{j-1} \le u < t_j\}} du \ge \limsup_{n\to\infty} \frac{M_n^{(t)}}{n}$$

$$\ge \quad \liminf_{n\to\infty} \frac{M_n^{(t)}}{n} \ge \int_0^{t_{k-1}} re^{-ru} \sum_{j=1}^{k} S_{t_j} 1_{\{t_{j-1} \le u < t_j\}} du \text{ almost surely.} \tag{2.5}$$

Since $(S_t)_{t\ge 0}$ is non-increasing in t and bounded by zero and one, hence has only countable many jump points, one has

$$\sum_{j=1}^{k} S_{t_{j-1}} 1_{\{t_{j-1} \le u < t_j\}}, \sum_{j=1}^{k} S_{t_j} 1_{\{t_{j-1} \le u < t_j\}} \to S_u$$

for λ-almost all $u \in [0, t)$ as $k \to \infty$. Hence for $k \to \infty$, bounded convergence turns (2.5) into

$$\int_0^t re^{-ru} S_u du \ge \limsup_{n\to\infty} \frac{M_n^{(t)}}{n} \ge \liminf_{n\to\infty} \frac{M_n^{(t)}}{n} \ge \int_0^t re^{-ru} S_u du \text{ a.s.} \tag{2.6}$$

Note that (2.6) holds almost surely, since outside the union of the (countably many) exception sets for the inequalities (2.5) for different values of k, the inequalities (2.5) are all true, so taking limits does not change that. So the almost sure convergence is shown for the truncated $M_n^{(t)}/n$. For every $t > 0$, decompose

$$\frac{M_n}{n} = \frac{M_n^{(t)}}{n} + \frac{M_n - M_n^{(t)}}{n}. \tag{2.7}$$

$M_n - M_n^{(t)}$ is the number of mutated external branches in $(\Pi_t^{(n)})_{t\ge 0}$ that are not affected by a mutation until t, thus $0 \le (M_n - M_n^{(t)})/n \le S_t^{(n)}$, where $S_t^{(n)}$ is the fraction of singletons of all n individuals in $\Pi_t^{(n)}$. From (2.2), for every $t \ge 0$, $\lim_{n\to\infty} S_t^{(n)} = S_t$ holds almost surely. Note that

$$E(S_t) = P(\{1\} \text{ is a block of } \Pi_t) = e^{-\mu_{-1} t} \to 0$$

for $t \to \infty$, where μ_{-1} is defined as in (1.1) (see [21, Proposition 30]). Thus $S_t \to 0$ in L^1. Since $(S_t)_{t\ge 0}$ is non-increasing, this convergence also holds almost surely. This shows

$$0 \le \liminf_{n\to\infty} \frac{M_n - M_n^{(t)}}{n} \le \limsup_{n\to\infty} \frac{M_n - M_n^{(t)}}{n} \to 0 \text{ almost surely}$$

for $t \to \infty$. The desired almost sure convergence of M_n/n for $n \to \infty$ follows by letting $n, t \to \infty$ in (2.7), as (2.6) shows, for every $t > 0$, $M_n^{(t)}/n \to \int_0^t re^{-ru} S_u du$ almost surely for $n \to \infty$. Since $0 \leq M_n/n \leq 1$ is bounded, hence uniformly integrable, this convergence also holds in L^p $(p \geq 1)$. □

Remark 2.3. *For Ξ-coalescents with proper frequencies, i.e. for coalescents satisfying (1.1), one has $S_t = 0$ almost surely for all $t > 0$. In this case, the right hand side in (2.3) is equal to 0 almost surely.*

3 Asymptotics for the number of types for coalescents without proper frequencies

Now, focus on K_n, the number of different types in the sample $\{1, \ldots, n\}$. For coalescent processes without proper frequencies there are some known results for its asymptotic behaviour. Theorem 1.2 of [11] states that K_n/n converges weakly to $\int_0^\infty re^{-rt} S_t dt$. Theorem 2.2 shows that the latter object is also the almost sure and L^p-limit of M_n/n for $n \to \infty$ $(p \geq 1)$. Note that $M_n \leq K_n$ for $n \in \mathbb{N}$, since any mutated external branch will lead to a type that does not appear anywhere else. Recall that in [11, Corollary 4.2] it is also shown that $(K_n - M_n)/n \to 0$ in L^1 as $n \to \infty$. Since $(K_n - M_n)/n \leq 1$ for $n \in \mathbb{N}$, this also implies convergence in L^p for all $p \geq 1$. Together with Theorem 2.2, this yields

Theorem 3.1. *Let K_n be the number of types among the first $n \in \mathbb{N}$ individuals in a Ξ-coalescent without proper frequencies. Then*

$$\frac{K_n}{n} \longrightarrow \int_0^\infty re^{-rt} S_t dt \quad \text{in } L^p \ (p \geq 1) \text{ as } n \to \infty,$$

where, for $t > 0$, S_t is the fraction of singletons of Π_t.

Remark 3.2. *Theorem 3.1 slightly improves the convergence result in [11].*

4 Almost sure/L^p asymptotics for the number of types for simple Ξ-coalescents

In order to get almost sure convergence in Theorem 3.1, one would need to show $(K_n - M_n)/n \to 0$ almost surely as $n \to \infty$. From [11, p. 13] it is known that for every coalescent process with mutation

$$0 \leq K_n - M_n \leq C_n, \ n \in \mathbb{N}, \tag{4.1}$$

where C_n is the number of collisions of $(\Pi_t^{(n)})_{t \geq 0}$. So it suffices to show that $C_n/n \to 0$ almost surely as $n \to \infty$. In this section, this is shown for the class of simple Ξ-coalescents, i.e. for coalescents satisfying (1.3).

Recall the Poisson construction for a simple Ξ-coalescent $(\Pi_t)_{t \geq 0}$ according to [21, Section 3]: First take a Poisson process Z on $[0, \infty) \times \Delta$ with intensity measure $\mu := dt \otimes (\Xi(dx)/(x, x))$. For a simple Ξ-coalescent μ is finite on $[0, t] \times \Delta$ for every $t \in [0, \infty)$, thus σ-finite. Order the points of Z increasingly in the t-coordinate. To construct $(\Pi_t)_{t \geq 0}$, start with Π_0 as the partition of $\mathbb{N}$ into singletons. At each successive point (t, x) of Z divide $[0, 1)$ into intervals $y_0, y_1, \ldots$ defined by $y_k := [1 - \sum_{i=k}^\infty x_i, 1 - \sum_{i=k+1}^\infty x_i)$ $(k \in \mathbb{N}_0)$ where $x = (x_1, x_2, \ldots)$ is the x-coordinate (simplex-valued coordinate) of the Poisson point and $x_0 := 1 - |x|$. This division is called Kingman's paintbox. Now for each block present in Π_{t-} throw a ball randomly (i.e. with uniform distribution) onto $[0, 1)$ divided as described above. Each ball is thrown independently of all other balls.

Then merge all blocks whose balls have fallen into the same compartment y_j ($j \in \mathbb{N}$) of $[0,1)$. Do not merge any block whose ball has fallen into compartment y_0. The new blocks resulting from all of these mergers and the unmerged blocks form Π_t.
Note that in order to construct $(\Pi_t^{(n)})_{t\geq 0}$, start with the partition of $\{1,\dots,n\}$ into singletons, at every Poisson point (t,x) of Z use the balls described above and merge all blocks of $\Pi_{t-}^{(n)}$ whose balls have landed in the same compartment $y_1, y_2, \dots$. Also note that in the case of simple Ξ-coalescents without restriction it can always be assumed that $(\Pi_t)_{t\geq 0}$ is pathwise constructed via Poisson construction. There are many classical results for the balls-in-boxes problem that occurs in this construction. Here, the following result is of most importance.

Lemma 4.1 (occupancy scheme)**.** *Let $X = (X_1, X_2, \dots)$ be a Δ-valued random variable and let $U_1, U_2, \dots$ be i.i.d. random variables with uniform distribution on $[0,1)$ independent of X. Define $X_0 := 1 - |X|$. Divide $[0,1)$ into the compartments $Y_k := [1-\sum_{i=k}^{\infty} X_i, 1-\sum_{i=k+1}^{\infty} X_i)$ ($k \in \mathbb{N}_0$) and let V_n be the number of compartments that are occupied by at least one value of $U_1, \dots, U_n$. Then $P(\lim_{n\to\infty} V_n/n = 0) = 1$.*

Proof. Define $y_k := [1-\sum_{i=k}^{\infty} x_i, 1-\sum_{i=k+1}^{\infty} x_i)$ for $x = (x_1, x_2, \dots) \in \Delta$, $x_0 := 1 - |x|$, $k \in \mathbb{N}_0$ and let $f_n(x,u) = \sum_{m=1}^{n} \sum_{i=0}^{\infty} 1_{\{u_m \in y_i, u_1, \dots, u_{m-1} \notin y_i\}}$ for $u = (u_1, u_2, \dots) \in [0,1]^{\mathbb{N}}$ be the number of compartments $y_0, y_1, \dots$ occupied by $u_1, \dots, u_n$. Note that the expected number of occupied boxes after n thrown balls $U_1, \dots, U_n$ in such a set of fixed boxes with lengths $x_0, x_1, \dots$ is

$$\mu_n(x) := E(f_n(x, (U_m)_{m\in\mathbb{N}})) = \sum_{i\in\mathbb{N}_0} (1 - (1-x_i)^n), \tag{4.2}$$

since $\sum_{m=1}^{n} 1_{\{U_m \in y_i, U_1, \dots, U_{m-1} \notin y_i\}}$ is 0 if and only if all n balls have not fallen into compartment y_i which has probability $(1-x_i)^n$. The Bernoulli inequality yields $0 \leq 1-(1-x_i)^n \leq nx_i$ for every $i \in \mathbb{N}_0$. If now the sum in (4.2) is seen as an integral of the counting measure $\mu_{\mathbb{N}_0}$ on $\mathbb{N}_0$, it follows by bounded convergence with dominating series $(x_i)_{i\in\mathbb{N}_0}$

$$\lim_{n\to\infty} \frac{\mu_n(x)}{n} = \int \lim_{n\to\infty} \frac{1-(1-x_i)^n}{n} \mu_{\mathbb{N}_0}(di) = 0. \tag{4.3}$$

[13, Theorem 8] shows that $\lim_{n\to\infty} f_n(x,\cdot)/\mu_n(x) = 1$ $P_{(U_m)_{m\in\mathbb{N}}}$-almost surely. Thus, together with (4.3),

$$P(\lim_{n\to\infty} f_n(x, (U_m)_{m\in\mathbb{N}})/n = 0) = 1 \text{ for any } x \in \Delta. \tag{4.4}$$

If you now throw balls into boxes of random lengths $X_0, X_1, \dots$, using Fubini and (4.4) leads to

$$P(\lim_{n\to\infty} \frac{V_n}{n} = 0) = P(\lim_{n\to\infty} \frac{f_n(X, (U_1, U_2, \dots))}{n} = 0) = 1.$$

□

The Poisson construction of coalescents mentioned above is the key to prove the following result.

Lemma 4.2. *Let $(\Pi_t)_{t\geq 0}$ be a simple Ξ-coalescent. Let C_n be the number of collisions of $(\Pi_t^{(n)})_{t\geq 0}$. Then $C_n/n \to 0$ almost surely and in L^p for any $p \geq 1$ as $n \to \infty$.*

Proof. Without loss of generality, assume that $(\Pi_t)_{t\geq 0}$ is pathwise constructed via the Poisson construction. Observe that collisions are always due to a Poisson point of Z. For every $j \in \mathbb{N}$, split C_n into the number of collisions due to the first j points (ordered in the first coordinate) $(T_1, X_1), \dots, (T_j, X_j)$ of Z and into the number of all other collisions.

Let $C_n^{(i)}$ be the number of collisions due to the ith Poisson point, so the number of collisions due to the first j Poisson points is $\sum_{i=1}^{j} C_n^{(i)}$. The number of collisions after the first j Poisson points can be at most the number of blocks of $\Pi_{T_j}^{(n)}$ minus 1, since $l \in \mathbb{N}$ blocks can have at most $l-1$ collisions. To analyze the number of blocks of $\Pi_{T_j}^{(n)}$, define the following random variables for $k, j \in \mathbb{N}$:

$$B_k^{(j)} := 1\Big\{ \text{ball } k \text{ has fallen into } y_0^{(1)}, \dots, y_0^{(j)} \text{ for the first } j \text{ Poisson points}\Big\},$$

where $y_0^{(i)}$ is the y_0-compartment of the paintbox generated by the i-th Poisson point. After j Poisson points, $\Pi_{T_j}^{(n)}$ has at most $V_n^{(1)} + \cdots + V_n^{(j)} + \sum_{k=1}^{n} B_k^{(j)}$ blocks, where $V_n^{(i)}$ is the number of occupied boxes in the construction step of $(\Pi_t^{(n)})_{t\geq 0}$ belonging to the ith Poisson point. This can be seen from the fact that every block of $\Pi_{T_j}^{(n)}$ comes either from a collision due to the first j Poisson points or is a singleton block. If it is a singleton block, every ball which was thrown for this block for one of the first j Poisson points has either fallen into a non-y_0-compartment where there were no other balls in this compartment or has fallen into the compartment y_0.
Thus, C_n is bounded by

$$\begin{aligned} C_n &\leq C_n^{(1)} + \cdots + C_n^{(j)} + V_n^{(1)} + \cdots + V_n^{(j)} + \sum_{k=1}^{n} B_k^{(j)} - 1 \\ &\leq 2(V_n^{(1)} + \cdots + V_n^{(j)}) + \sum_{k=1}^{n} B_k^{(j)} - 1, \end{aligned}$$

for all $j \in \mathbb{N}$, since, by construction, $0 \leq C_n^{(i)} \leq V_n^{(i)}$ for $n \in \mathbb{N}$. The next step is to analyze the asymptotics of this bound for $n \to \infty$. Lemma 4.1 shows $\lim_{n\to\infty}(V_n^{(i)}/n) = 0$ almost surely for every $i \in \mathbb{N}$, since at most n independent balls are thrown for every Poisson point. To analyze $(B_k^{(j)})_{k\in\mathbb{N}}$ it is helpful to look at the x-coordinates (the simplex coordinates) of the points of Z. Recall that these coordinates govern the distribution of the length of the compartments of the paintboxes used in the Poisson construction. In the case of a simple measure Ξ satisfying (1.3), the x-coordinates $(X_i)_{i\in\mathbb{N}}$ of the points of Z (ordered in time) are i.i.d. random variables with $X_1 \stackrel{d}{=} \nu/\nu(\Delta)$, where $\nu(dx) := (x,x)^{-1}\Xi(dx)$. This can be read from the construction of the Poisson point process in [14, p. 23] and the product structure of the intensity measure μ. It is convenient to introduce $Y_i := |X_i|$ for $i \in \mathbb{N}$, which gives the total length of all compartments other than $y_0^{(i)}$ for the corresponding Poisson point. Y_i is also the probability that a ball does not hit compartment $y_0^{(i)}$. Since it is only interesting whether a ball hits compartment $y_0^{(i)}$ or another compartment of $[0,1]$, $(B_k^{(j)})_{k\in\mathbb{N}}$ is conditional i.i.d. with $B_1^{(j)}$ Bernoulli-distributed with parameter $\prod_{i=1}^{j}(1-Y_i)$ for every j conditioned on $Y_1, \dots, Y_j$. Thus, $(B_k^{(j)})_{k\in\mathbb{N}}$ is exchangeable. By the strong law of large numbers for exchangeable random variables (see for example [9, Remark 3]),

$$\frac{\sum_{k=1}^{n} B_k^{(j)}}{n} \to \prod_{i=1}^{j}(1-Y_i) \text{ almost surely as } n \to \infty. \tag{4.5}$$

Thus,

$$0 \leq \liminf_{n\to\infty} \frac{C_n}{n} \leq \limsup_{n\to\infty} \frac{C_n}{n} \leq \prod_{i=1}^{j}(1-Y_i)$$

almost surely for all $j \in \mathbb{N}$. Since $(Y_i)_{i\in\mathbb{N}}$ is i.i.d. with $P(Y_i > 0) = 1$ (Ξ is simple, hence ν has no mass in $0 \in \Delta$), the Borel-Cantelli lemma shows that $P(\limsup_{i\to\infty}\{Y_i \geq \epsilon\}) = 1$

for some $\epsilon > 0$. This implies $\prod_{i=1}^{\infty}(1 - Y_i) = 0$ almost surely and yields the almost sure convergence. Since $0 \leq C_n/n \leq 1$ for all $n \in \mathbb{N}$, the convergence also holds in L^p for $p \geq 1$. □

This leads to the main result.

Theorem 4.3. *Let K_n be the number of types among the first $n \in \mathbb{N}$ individuals in a simple Ξ-coalescent. Then*

$$\frac{K_n}{n} \to \int_0^{\infty} re^{-rt} S_t dt \quad \textit{almost surely and in } L^p \ (p \geq 1) \textit{ as } n \to \infty \ ,$$

where, for $t > 0$, S_t is the fraction of singletons of Π_t.

Proof. By Theorem 2.2, $M_n/n \to \int_0^{\infty} re^{-rt} S_t dt$ almost surely and in L^p. With (4.1) and Lemma 4.2, $0 \leq (K_n - M_n)/n \leq C_n/n \to 0$ almost surely and in L^p as $n \to \infty$. □

References

[1] Barton, N.H., Etheridge, A.M. and Véber, A.: A new model for evolution in a spatial continuum. *Electron. J. Probab.* **15**, (2010), 162–216. MR-2594876

[2] Basdevant, A.-L. and Goldschmidt, C.: Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **13**, (2008), 486–512. MR-2386740

[3] Berestycki, J., Berestycki, N. and Limic, V.: Asymptotic sampling formulae and particle system representations for Λ-coalescents. arXiv:1101.1875

[4] Berestycki, J., Berestycki, N. and Schweinsberg, J.: Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35**, (2007), 1835–1887. MR-2349577

[5] Berestycki, J., Berestycki, N. and Schweinsberg, J.: Small-time behavior of beta-coalescents. *Ann. Inst. H. Poincaré Probab. Statist.* **44**, (2008), 214–238. MR-2446321

[6] Bertoin, J. and LeGall, J.-F.: Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields* **126**, (2003), 261–288. MR-1990057

[7] Cuzick, J.: A strong law for weighted sums of i.i.d. random variables. *J. Theor. Probab.* **8**, (1995), 625–640. MR-1340830

[8] Eldon, B. and Wakeley, J.: Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **192**, (2006), 2621–2633.

[9] Etemadi, N. and Kaminski, M.: Strong law of large numbers for 2-exchangeable random variables. *Stat. Probab. Letters* **28**, (1996), 245–250. MR-1406997

[10] Ewens, W.J.: The sampling theory of selectively neutral alleles. *Theoret. Popul. Biol.* **3**, (1972), 87–112. MR-0325177

[11] Freund, F. and Möhle, M.: On the number of allelic types for samples taken from exchangeable coalescents with mutation. *Adv. Appl. Probab.* **41**, (2009), 1082–1101. MR-2663237

[12] Huillet, T. and Möhle, M.: Population genetics models with skewed fertilities: a backward and forward analysis. *Stoch. Models.* **27**, (2011), 521–554.

[13] Karlin, S.: Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17**, (1967), 373–401. MR-0216548

[14] Kingman, J.F.C.: Poisson Processes. Oxford Studies in Probability, 3. Oxford Science Publications. *The Clarendon Press, Oxford University Press*, New York, 1993. viii+104 pp. MR-1207584

[15] Kingman, J.F.C.: The coalescent. *Stoch. Proc. Appl.* **13**, (1982), 235–248. MR-0671034

[16] Marynych, A.: On the asymptotics of moments of linear random recurrences. *Theory Stoch. Proc.* **16**, (2010), 106–119. MR-2779988

[17] Möhle, M.: Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent. *Stoch. Process. Appl.* **120**, (2010), 2159–2173. MR-2684740

[18] Pitman, J.: Coalescents with multiple collisions. *Ann. Probab.* **27**, (1999), 1870-1902. MR-1742892

[19] Sagitov, S.: The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, (1999), 1116-1125. MR-1742154

[20] Sagitov, S.: Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40**, (2003), 839–854. MR-2012671

[21] Schweinsberg, J.: Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, (2000), 1-50. MR-1781024

[22] Taylor, J. and Véber, A.: Coalescent processes in subdivided populations subject to recurrent mass extinctions. *Electron. J. Probab.* **14**, (2009), 242–288. MR-2471665

Acknowledgments. The author thanks Philip Herriger and Martin Möhle for a fruitful discussion concerning Lemma 4.1.