Asymptotics of the frequency spectrum for general Dirichlet Ξ-coalescents[*]

Adrián González Casanova[†]   Verónica Miró Pina[‡]
Emmanuel Schertzer[§]   Arno Siri-Jégousse[¶]

Abstract

In this work, we study general Dirichlet coalescents, which are a family of Ξ-coalecents constructed from i.i.d mass partitions, and are an extension of the symmetric coalescent. This class of models is motivated by population models with recurrent demographic bottlenecks. We study the short time behavior of the multidimensional block counting process whose ith component counts the number of blocks of size i. Compared to standard coalescent models (such as the class of Λ-coalescents coming down from infinity), our process has no deterministic speed of coming down from infinity. In particular, we prove that, under appropriate re-scaling, it converges to a stochastic process which is the unique solution of a martingale problem. We show that the multivariate Lamperti transform of this limiting process is a Markov Additive Process (MAP). This allows us to provide some asymptotics for the n-Site Frequency Spectrum, which is a statistic widely used in population genetics. In particular, the rescaled number of mutations converges to the exponential functional of a subordinator.

Keywords: coalescent processes; partitions; coming down form infinity; subordinator; Markov Additive Processes.
MSC2020 subject classifications: 60J90.
Submitted to EJP on September 27, 2022, final version accepted on December 4, 2023.

[†]Instituto de Matemáticas de la Universidad Nacional Autónoma de México, CDMX, México.
[‡]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology and Universitat Pompeu Fabra (UPF), Barcelona, Spain. E-mail: veronica.miro@crg.eu
[§]Faculty of Mathematics, University of Vienna, Austria.
[¶]Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, CDMX, México.

1 Introduction

1.1 General Dirichlet Ξ-coalescents

Coalescents with simultaneous and multiple collisions (Ξ-coalescents, [41, 38, 8]) are exchangeable coagulating continuous-time Markov chains with values in the set of partitions of $\mathbb{N}$. They can be built via a paintbox construction (see [6], Section 4.2.3). Consider a finite measure Ξ on $\Delta = \{\vec{p} = (p_1, p_2, \dots) : \sum p_j \leq 1\}$, the simplex on $(0, 1)$, and a Poisson point process (Ppp) on $\mathbb{R}_+ \times \Delta$, of intensity $dt \otimes \Xi(d\vec{p})/\sum p_j^2$, giving the coagulation times and rules. At every jump time, (1) consider a family of disjoint subintervals $(I_1, I_2, \dots)$ of $(0, 1)$, where I_j has length p_j, (2) throw every block present at this time uniformly at random and independently in $(0, 1)$, and (3) blocks falling into the same subinterval among $(I_1, I_2, \dots)$ merge into one.

In this work, we consider a particular subclass of Ξ-coalescents where the interval partition of the paintbox has a generalized version of a Dirichlet distribution with a random number of components. More precisely, consider a sequence of non-negative numbers $(R(k); k \in \mathbb{N})$ and m a probability measure on $(0, \infty)$. Then, generate at rate $R(k)$ a partition $(p_1^{(k)}, \dots, p_k^{(k)})$ where

$$\forall j \in [k], \quad p_j^{(k)} := \frac{w_j}{s_k} \tag{1.1}$$

where $(w_1, \dots, w_k)$ are i.i.d. random variables with law m on $(0, \infty)$, $s_k := \sum_{i=1}^{k} w_i$ and $[k] := \{1, \dots, k\}$. As in the previous paintbox construction, blocks are assigned a uniform random variable, and we merge all the blocks falling in the same interval. This corresponds to a Ξ-coalescent where the characteristic finite measure Ξ on the infinite simplex Δ is described as follows. For every $k \in \mathbb{N}$, let us define ν_k, a probability measure on the infinite simplex Δ, s.t.

$$\nu_k = \mathcal{L}\big((w_1/s_k, w_2, s_k, \dots, w_k/s_k, 0, 0, \dots)\big),$$

where $\mathcal{L}(X)$ denotes the law of the random variable X. Then, for every measurable $B \subset \Delta$,

$$\Xi(B) = \sum_{k=1}^{\infty} \int_B R(k) \left(\sum_{i=1}^{\infty} x_i^2 \right) \nu_k\big(d(x_1, x_2, \dots)\big).$$

The case where the w_i's are Gamma distributed corresponds to the standard Dirichlet mass-partition. In particular, if the w_i's are exponentially distributed, it corresponds to a symmetric Dirichlet distribution. We refer to this model as the general Dirichlet coalescent. The name Dirichlet coalescent was coined in [23]. Therein the authors consider paintbox construction according to a Dirichlet distribution with a fixed number of components.

Another example of such a process is the symmetric coalescent defined in [24], which corresponds to the case where $w_i = 1$ a.s. In that case, in order to correspond to the paintbox construction described above, the sequence R must satisfy that $\sum R(k)/k < \infty$, see [24]. Here, we assume a finite second moment for m and that the rate of k-events has a heavy tail, in the following sense.

Assumption 1.1.   *1. There exist $\alpha \in (0, 1)$ and $\rho > 0$ such that $R(k)k^\alpha \to \rho$ as $k \to \infty$.*
  2. $\int_{\mathbb{R}_+} x^2 m(dx) < \infty$.

In fact, a Ξ-coalescent is well defined if the rate at which two blocks merge into one is finite [41]. It is easy to see from the paintbox construction that, since the vector $(p_1^{(k)}, \dots, p_k^{(k)})$, is exchangeable, this rate is given by

$$\sum_{k \geq 1} R(k) k \mathbb{E}\big((p_1^{(k)})^2\big),$$

which is finite under Assumptions 1.1 (see Proposition A.1 in Appendix A).

Suppose that the general Dirichlet coalescent starts with n singletons. Denote by $\hat{\mu}_t^n = (\hat{\mu}_t^n(1), \ldots, \hat{\mu}_t^n(n))$ the vector such that $\hat{\mu}_t^n(i)$ is the number of blocks containing i elements at time t, and denote by $|\hat{\mu}_t^n|$ the total number of blocks. Define the rescaled vector

$$\mu_t^n = \frac{1}{n}\hat{\mu}_{tn^{\alpha-1}}^n. \tag{1.2}$$

In this paper, we aim at studying the limiting behavior of the Markov process $(\mu_t^n; t \geq 0)$ as $n \to \infty$. We prove (in Theorem 2.2) that it converges towards a stochastic process $(\mu_t; t \geq 0)$, defined as the unique solution to a martingale problem associated to a continuous coagulation operator (see Theorem 2.1).

Intuitively, the result can be understood as follows. In the paintbox construction, when there are n lineages, a k-merging event corresponds to throwing n balls into k boxes (with probabilities $(p_1^{(k)}, \ldots, p_k^{(k)})$), and merging the balls that land in the same box. For $k \gg n$, the chance that non-trivial merging occurs is negligible, whereas for $k \ll n$, all lineages will be merged into a few lineages (which disappear when rescaling the number of blocks by n). The total rate of k-events with $\epsilon n \leq k \leq Mn$ for some small $\epsilon > 0$ and large $M < \infty$ can by approximated by $\int_{\epsilon n}^{Mn} \rho y^{-\alpha} dy = C n^{1-\alpha}$ with some constant C, which explains why time is slowed down by this factor. The heuristics behind the form of the coagulation operator that is the central part of the generator of the limit process are explained in Section 2.2.

We also show that this limit process is self-similar with negative index $\beta := \alpha - 1$ (see Theorem 2.3). In particular, the limit of the rescaled block counting process $(|\mu_t|; t \geq 0)$ is the exponential of a time-changed subordinator. As a direct corollary, if we define

$$A_t := \inf\left\{ s > 0 : \int_0^s \frac{1}{|\mu_r|^\beta} dr > t \right\}, \quad \text{and} \quad \xi_t := -\log(|\mu_{A_t}|), \tag{1.3}$$

then $(\xi_t; t \geq 0)$ is a subordinator. The law of the subordinator can be identified as a direct consequence of Theorem 7.2 (see Section 7). This shows that the short time behavior of the block counting process remains stochastic. This is in sharp contrast with previous studies where it is shown that classical models (such as Λ-coalescents) exhibit a deterministic speed of coming down from infinity, see Section 1.2 for a more detailed discussion. Our result can be interpreted as a stochastic speed of coming down from infinity.

1.2 Speed of coming down from infinity

We say that a Ξ-coalescent comes down from infinity if there are finitely many blocks at any time $t > 0$ almost surely, even if the coalescent is started with infinitely many blocks. In his original work, Schweinsberg already established a criterion for coming down from infinity [41]. In the case that the characteristic measure of the coalescent is supported on the set of finite mass partitions $\Delta^* = \{\vec{p} = (p_1, \ldots, p_k) : \sum_{j=1}^k p_j = 1, \text{ for some } k\}$ (which is our case of interest), the process comes down from infinity if and only if

$$\int_\Delta \frac{\Xi(d\vec{p})}{\sum_j p_j^2} = \infty.$$

Otherwise, the number of blocks stays infinite for a finite amount of time. As another example, coalescents whose characteristic measures are supported only on infinite mass partitions, i.e. for which $\Xi(\Delta^*) = 0$, either come down from infinity or always stay infinite. We consider coalescents supported on Δ^* and that come down from infinity. Limic [34] studied the small time behavior of Ξ-coalescents under what she called a

regularity assumption

$$\int_\Delta \frac{(\sum_j p_j)^2}{\sum_j p_j^2} \Xi(d\vec{p}) < \infty.$$

In this setting, and starting with infinitely many lineages, there exists a speed of coming down from infinity, i.e., a deterministic function $\nu_\Xi(t)$, which is finite for all $t > 0$ such that, if $|\hat{\mu}_t|$ is the number of blocks at time t in a coalescent starting with infinitely many lineages,

$$\lim_{t \to 0^+} \frac{|\hat{\mu}_t|}{\nu_\Xi(t)} = 1, \text{ almost surely.}$$

This mirrors the behavior of the class of Λ-coalescents coming down from infinity [5, 4]. To summarize, most of the previous studies have shown that the block counting process of a large class of exchangeable coalescents exhibits a deterministic behavior at small time scale.

In the present work, we consider Ξ-coalescents belonging to the first family (for which Ξ is supported on Δ^*), and which come down from infinity (see [41], Section 5.5). We take a different approach, since we study the rescaled number of blocks, starting from n lineages, as $n \to \infty$. In our case, when time is re-scaled by $n^{\alpha-1}$, the block counting process converges to a *stochastic* self-similar process, so there is no deterministic speed of coming down from infinity.

Our results have similar flavor to those of Haas and Miermont [25] for Λ coalescents with dust, and of Möhle and co-authors [23, 37] for a class of Ξ-coalescents with dust. In the first work, a self similar behavior of the rescaled number of blocks is obtained in the limit. In the second work, they prove that the frequency of singletons, as well as the number of blocks rescaled by n, converges to the exponential of a subordinator (without any time-rescaling). A natural prospect of research would be to identify conditions that would partition Ξ-coalescents (coming down from infinity) into two main classes: a first class with a deterministic limiting behavior, and a second one with a stochastic descent from infinity.

1.3 Perspectives on coming down from infinity

Our results deal with processes valued on the partitions of n when n goes to infinity. Although this is heuristically related to the case $n = \infty$, which corresponds to working with partitions of $\mathbb{N}$, we expect that there are important technical challenges when studying the process starting with infinitely many blocks. To be precise, the latter would require an entrance law at infinity for the limit of the multidimensional block counting process. In our approach, we avoid this problem by rescaling the block counting process by n so that $|\mu_0^n| = 1$ and there is no need for entrance laws for the limit process $(\mu_t; t \geq 0)$.

The study of entrance laws of self-similar Markov processes has recently been an active area of research. In the one dimensional case there is an extensive literature (see for example [7, 10, 19] and the references therein). Recent results in the finite dimensional case can be found in [31]. This is also a classic problem for Markov additive processes [16]. We believe that our results can motivate the study of entrance laws for infinite dimensional self-similar processes.

1.4 Biological motivation

The symmetric coalescent [24] can be obtained as the limiting genealogy of a Wright-Fisher population that undergoes rare recurrent bottlenecks reducing the population size to a random number k of individuals for only one generation. In this case, the second point of Assumptions 1.1 always holds and the first point is fulfilled if the measure

characterizing the size of the bottlenecks has power tails of order α. General Dirichlet coalescents naturally arise in an extension of this model, that can be seen as multinomial non-exchangeable reproductive events [44].

The analysis of the asymptotics of the multidimensional block counting process allows us to characterize the limiting behavior of the Site Frequency Spectrum (SFS) of our family of Ξ-coalescents as some functional of the limit process $(\mu_t; t \geq 0)$. The SFS is one of the most widespread statistics in population genetics. It consists in a vector of size $n-1$ whose ith component counts the number of mutations that are shared by i individuals in a sample of size n. We suppose that mutations occur at a constant rate over the coalescent tree started with n individuals, so that the SFS is closely related to its branch lengths. In general, this is a complex combinatorial problem and most of the previous works have relied on some approximations of the short time behavior of the block counting process to derive asymptotics for the lower part of the SFS (i.e., number of singletons, pairs etc.). Some examples are [18] for the case of the Kingman coalescent, and [5, 13] for coalescents with multiple collisions, such as Beta-coalescents, or [2, 28, 20, 29] for the special case of the Bolthausen-Sznitman coalescent. For fixed n, some studies on the law of the SFS can be found in [22, 26, 29].

There are few results available regarding the SFS of Ξ-coalescents. Works like [14, 42] present computational algorithms based on recursions to derive the expected SFS for finite n. Asymptotic properties of Ξ-coalescents started with n lineages were studied previously, in particular regarding the number of blocks [35]. Theorem 2.4 describes the asymptotics of the SFS for general Dirichlet coalescents.

2 Main results

2.1 Notation

Let us start this section with some notations. We denote by $\mathbb{N}$ the positive integers and by $\mathbb{N}_0$ the non-negative integers. Let $\ell^1(\mathbb{R}_+)$ be the set of all sequences with positive coordinates and with finite sum. For every $\mathbf{z} = (z(1), z(2), \dots) \in \ell^1(\mathbb{R}_+)$, we denote the sum of all its elements by $|\mathbf{z}| = \sum_{i=1}^{\infty} z(i)$. We also denote by $\ell^1(\mathbb{N}_0)$ the set of sequences with coefficients valued in $\mathbb{N}_0$ and finite sum. Define

$$\mathcal{Z} := \left\{ \mathbf{z} \in \ell^1(\mathbb{R}^+) : \sum_{i=1}^{\infty} iz(i) = 1 \right\}, \quad \mathcal{Z}_n := \left\{ \mathbf{z} \in \mathcal{Z} : n\mathbf{z} \in \ell^1(\mathbb{N}_0) \right\}.$$

The space $\mathcal{Z}$ will be equipped with the $\ell^1(\mathbb{R}_+)$ norm. The latter definitions are motivated by partitions of $n \in \mathbb{N}$. Recall that a partition of $n \in \mathbb{N}$ denotes an unordered sequence of integers $\{m(1), \dots, m(k)\}$ such that $\sum_{i=1}^{k} m(i) = n$. For every $i \in \mathbb{N}$, define $z(i) = \frac{1}{n}\#\{k : m(k) = i\}$, the "frequency" of i in the partition of n. Then $\mathbf{z} = (z(1), z(2), \dots)$ is an element of $\mathcal{Z}_n$. Starting with n singletons, we study the multidimensional block counting process of the general Dirichlet coalescent as a Markov process valued in $\mathcal{Z}_n$ (as already outlined in the introduction). Its first component denotes the frequency of singletons, its second component is the frequency of pairs, etc.

For every $c = (c(1), c(2), \dots) \in \ell^1(\mathbb{N}_0)$, we define $\varphi(c) = \sum_{i=1}^{\infty} ic(i)$ so that for every $\ell \in \mathbb{N}$,

$$\varphi^{-1}(\ell) := \left\{ c \in \ell^1(\mathbb{N}_0) : \sum_{i=1}^{\infty} ic(i) = \ell \right\}.$$

Observe that $\varphi^{-1}(\ell)$ is a way of encoding partitions of ℓ. For any partition $c \in \varphi^{-1}(\ell)$, we also define $c! = \prod_{i=1}^{\ell} c(i)!$.

For any random variable X valued in $\ell^1(\mathbb{R}_+)$, we denote by $\mathbb{E}(X)$ the deterministic vector obtained by taking the expected value coordinatewise. For $T > 0$, we denote by

$D([0,T], \mathcal{Z})$ the space of càdlàg functions on $(\mathcal{Z}, \ell^1(\mathbb{R}_+))$ equipped with the Skorokhod M_1 topology [43, 45].

For any $\lambda \in [0,1]$, we define

$$\forall \mathbf{z} \in \mathcal{Z}, \ \psi_\lambda(\mathbf{z}) := \sum_{i=1}^\infty z(i)\lambda^i,$$

and for every $\vec{\lambda} \in [0,1]^K, K \in \mathbb{N}$, we define

$$\forall \mathbf{z} \in \mathcal{Z}, \ \ \psi_{\vec{\lambda}}(\mathbf{z}) := \big(\psi_{\lambda_1}(\mathbf{z}), \ldots, \psi_{\lambda_K}(\mathbf{z})\big).$$

We define the following set of test functions

$$\mathcal{T} := \big\{ f : \mathcal{Z} \to \mathbb{R} : \exists \vec{\lambda} = (\lambda_1, \ldots, \lambda_K) \in [0,1]^K, F \text{ is Lipshitz continuous on } [0,1]^K,$$
$$\text{s.t. } f(\mathbf{z}) = F\big(\psi_{\lambda_1}(\mathbf{z}), \ldots, \psi_{\lambda_K}(\mathbf{z})\big) = F \circ \psi_{\vec{\lambda}}(\mathbf{z}) \big\}.$$

2.2 Convergence of the rescaled partition process

We are now ready to enunciate and comment on our main results. We start by describing the random coagulation corresponding to the jump events of the general Dirichlet Ξ-coalescent. Set $\mathbf{z}_n \in \mathcal{Z}_n$ where $nz_n(i)$ is the number of balls of size i (the size of a ball refers to the number of samples/lineages it represents). Then throw $nz_n(i)$ balls of size i, $i \geq 1$, at random into k boxes in such a way that the probability of falling into box $j \in [k]$ is $p_j^{(k)}$, as defined in (1.1). Now, define $\Lambda^{k,n}(\mathbf{z}_n) \in \mathcal{Z}_n$ as

$$\forall \ell \in \mathbb{N}, \ \Lambda^{k,n}(\mathbf{z}_n)(\ell) = \frac{1}{n}\#\{j \leq k : \text{sum of the sizes of the balls falling in box } j \text{ is } \ell\}. \tag{2.1}$$

Note that, for $\ell \geq n, \Lambda^{k,n}(\mathbf{z}_n)(\ell) = 0$. By a slight abuse of notation, we define the (random) operator $\Lambda^{k,n}$ acting on $\mathcal{Z}_n$ such that, for every function g defined on $\mathcal{Z}_n$,

$$\Lambda^{k,n}g(\mathbf{z}_n) = g\big(\Lambda^{k,n}(\mathbf{z}_n)\big).$$

Thanks to these notations we can define the infinitesimal generator of the $\mathcal{Z}_n$-valued process $(\mu_t^n; t \geq 0)$ defined in (1.2) as

$$\mathcal{A}_n f(\mathbf{z}_n) = \frac{1}{n^{1-\alpha}}\sum_{k \geq 1} R(k)\big(\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big) \tag{2.2}$$

for every measurable and bounded $f : \mathcal{Z}_n \to \mathbb{R}$ and $\mathbf{z}_n \in \mathcal{Z}_n$.

Before diving into technicalities, let us first motivate the coming results. Assume that $k, n \to \infty$ with $k/n \sim x \in (0, \infty)$, i.e., a large number of balls (n) and boxes (k), of the same order. Under this restriction, if an event involving k boxes occurs, the number of balls of size i falling in box 1 is well approximated by a Poisson random variable with parameter

$$p_1^{(k)} n z_n(i) \approx \frac{\Gamma}{x} z_n(i),$$

where

$$\Gamma := \frac{w_1}{\mathbb{E}(w_1)}.$$

Further, since the number of balls/boxes is large, the total number of boxes with r balls of size i should be well approximated by

$$k \times \left(\frac{\Gamma}{x} z_n(i)\right)^r \frac{e^{-\frac{\Gamma}{x}z_n(i)}}{r!}.$$

By a similar heuristic, if $k, n \to \infty$ with $k/n \sim x \in (0, \infty)$, we expect

$$\Lambda^{k,n}(\mathbf{z}_n)(\ell) \approx \mathbb{E}\left(x \exp(-|\mathbf{z}_n|\Gamma/x) \sum_{c \in \varphi^{-1}(\ell)} \prod_{i=1}^{\ell} \frac{(z_n(i)\Gamma/x)^{c(i)}}{c(i)!} \right),$$

where the expectation is taken with respect to the random variable Γ. This justifies the limit operator introduced later on in (2.3).

2.3 A martingale problem

We now define the martingale problem, associated to a continuous coagulation operator, described as follows. Let $x > 0$ and define $\mathcal{C}^x : \mathcal{Z} \to \mathcal{Z}$, such that its ℓth coordinate is given by

$$\mathcal{C}^x(\mathbf{z})(\ell) = \mathbb{E}\left(x \exp(-|\mathbf{z}|\Gamma/x) \sum_{c \in \varphi^{-1}(\ell)} \prod_{i=1}^{\ell} \frac{(z(i)\Gamma/x)^{c(i)}}{c(i)!} \right). \tag{2.3}$$

From the heuristics of the previous section, $\mathcal{C}^x(\mathbf{z}_n)(\ell)$ is a natural candidate to approximate $\Lambda^{k,n}(\mathbf{z}_n)(\ell)$. As for $\Lambda^{k,n}$, we define the operator $\mathcal{C}^x$ on functions on on $\mathcal{Z}$ such that, for every function g bounded and measurable on $\mathcal{Z}$,

$$\mathcal{C}^x g(\mathbf{z}) = g\big(\mathcal{C}^x(\mathbf{z})\big).$$

We will show in due time that $\mathcal{C}^x(\mathbf{z}) \in \mathcal{Z}$, see Proposition 5.1.

Theorem 2.1 (Uniqueness of the martingale problem). *For every $\mathbf{z} \in \mathcal{Z}$ and $f \in \mathcal{T}$, the function*

$$x \to \big(\mathcal{C}^x f(\mathbf{z}) - f(\mathbf{z})\big)\rho x^{-\alpha}$$

is integrable on $(0, \infty)$. Define

$$\mathcal{A}f(\mathbf{z}) := \int_0^\infty \big(\mathcal{C}^x f(\mathbf{z}) - f(\mathbf{z})\big)\rho x^{-\alpha} dx. \tag{2.4}$$

There exists a unique càdlàg process $(\mu_t; t \geq 0)$ valued in $\mathcal{Z}$ with $\mu_0 = \mathbf{z}$ such that

$$\left(f(\mu_t) - \int_0^t \mathcal{A}f(\mu_s)ds;\ t \geq 0 \right) \quad \text{is a martingale for every } f \in \mathcal{T}. \tag{2.5}$$

Theorem 2.1 characterizes the limiting process in the following result.

Theorem 2.2. *Suppose that Assumptions 1.1 hold. If $\mu_0^n = \mathbf{z}_n \to \mathbf{z} \in \mathcal{Z}$, then for every $T > 0$,*

$$\big(\mu_t^n; t \geq 0\big) \implies (\mu_t; t \geq 0) \text{ in } D\big([0, T], \mathcal{Z}\big),$$

where the process $(\mu_t; t \geq 0)$ is the unique solution to the martingale problem (2.5) with initial condition $\mathbf{z}$.

2.4 Self-similarity

The second part of this paper is devoted to the study of the limiting process $(\mu_t; t \geq 0)$ characterized in Theorem 2.1. We prove that it is an infinite dimensional self-similar process, with negative index $\beta := \alpha - 1 \in (-1, 0)$ (Proposition 7.1). We can characterize its infinite dimensional Lamperti-Kiu transform.

The fact that $(\mu_t; t \geq 0)$ is self-similar is inherited from the regular tail behavior of $R(k)$ (which is reflected by the $x^{-\alpha}$ in the generator (equation (2.4)) together with

the fact that for any positive constant γ and for every $\ell \in \mathbb{N}$, $\gamma \mathcal{C}^x(\mathbf{z})(\ell) = \mathcal{C}^{\gamma x}(\gamma \mathbf{z})$ (see equation (2.3)).

To characterize this transformation, first consider the limiting block counting process $(|\mu_t|; t \geq 0)$. According to Proposition 7.1, $(|\mu_t|; t \geq 0)$ is a non-increasing self-similar positive Markov process with parameter β. The standard Lamperti transform tells us that such a process is identical in law to the exponential of a time-changed subordinator. Recall $(A_t; t > 0)$ and $(\xi_t; t > 0)$ defined in (1.3). Then $(\xi_t; t \geq 0)$ is a subordinator and $(|\mu_t|; t \geq 0)$ can be recovered by the relation $|\mu_t| = \exp(-\xi_{\tau_t})$, where $\tau_t := \inf\{u > 0 : \int_0^u \exp(\beta \xi_s) ds > t\}$.

Let us turn to the infinite dimensional self-similar process $(\mu_t; t \geq 0)$. Let $S := \{\mathbf{z} \in \ell^1(\mathbb{R}_+) : |\mathbf{z}| = 1\}$ be the unit sphere in $\ell^1(\mathbb{R}_+)$. The idea of the infinite dimensional Lamperti-Kiu transform is to decompose the process into its "radial part" $(|\mu_t|; t \geq 0)$ (the block counting process) and its "spherical part"

$$\left(\left(\frac{\mu_t(1)}{|\mu_t|}, \frac{\mu_t(2)}{|\mu_t|}, \dots\right); t \geq 0\right) \in S,$$

which encodes the evolution of the asymptotic frequencies of singletons, pairs, etc. In the spirit of the one-dimensional case, the process can be related to a time-changed Markov additive process (MAP, see [15]). Theorem 2.3 is a natural extension of Theorem 2.3. in [1], established in finite dimension.

Theorem 2.3 (Lamperti-Kiu transform)**.** *For every $t > 0$, define the stopping time $(A_t; t \geq 0)$ as in (1.3) and $((\xi_t, \theta_t); t \geq 0)$ as the process valued in $\mathbb{R}_+ \times S$ such that*

$$\xi_t := -\log(|\mu_{A_t}|), \quad \theta_t := \frac{1}{|\mu_{A_t}|} \mu_{A_t}.$$

Then, $((\xi_t, \theta_t); t \geq 0)$ is a MAP with respect to the filtration $(\mathcal{G}_t; t \geq 0) := (\mathcal{F}_{A_t}; t \geq 0)$, i.e. $(\xi_t; t \geq 0)$ is a subordinator and

$$\mathbb{E}_{\xi_0, \theta_0}\big(f(\xi_{t+s} - \xi_s, \theta_{t+s}) \mid \mathcal{G}_s\big) = \mathbb{E}_{0, \theta_s}\big(f(\xi_t, \theta_t)\big).$$

Further, $(\mu_t; t \geq 0)$ can be recovered from its Lamperti-Kiu transform through the formula

$$\mu_t = \theta_{\tau_t} \exp(-\xi_{\tau_t}) \tag{2.6}$$

where $\tau_t := \inf\{u > 0 : \int_0^u \exp(\beta \xi_s) ds > t\}$.

2.5 Site Frequency Spectrum

The third part of this work is devoted to the asymptotics of the SFS of the family of general Dirichlet Ξ-coalescents, in the limit of large n. Consider the infinite sites model, where it is assumed that mutations occur according to a Poisson Point Process of intensity $r > 0$ over the coalescent tree and that each new mutation falls in a new site so that all the mutations can be observed in the generic data. Define the rescaled SFS $F_n = (F_n(1), F_n(2), \dots)$ where

$$\forall i \leq n - 1, \quad F_n(i) = \frac{1}{n^\alpha} \#\{\text{mutations carried by } i \text{ individuals}\}.$$

Under the infinite sites model, there is a very close relation between the SFS and the block counting process of the coalescent tree. More precisely, conditional on the coalescent, the number of segregating mutations affecting i individuals of the sample is given by a Poisson random variable with parameter $r \int_0^{\hat{\mathbf{T}}_n} \hat{\mu}_s^n(i) ds$, where $\hat{\mathbf{T}}_n$ denotes the time to the most recent common ancestor (height) of the coalescent tree.

Theorem 2.4. *Let (ξ_t, θ_t) be defined as in Theorem 2.3. We have*

$$\lim_{n \to \infty} F_n = r \int_0^\infty \mu_s ds$$
$$= r \int_0^\infty \theta_u \exp\big((\alpha - 2)\xi_u\big) du,$$

where the convergence is meant in the weak sense with respect to the $\ell^1(\mathbb{R}_+)$ topology. In particular, the (rescaled) total number of mutations $|F_n|$ is asymptotically described by the exponential functional of a subordinator [11], i.e.,

$$\lim_{n \to \infty} |F_n| = r \int_0^\infty \exp\big((\alpha - 2)\xi_u\big) du.$$

Observe that a similar rescaling, in $n^\alpha, \alpha < 1$ appears for the lower part of the spectrum (small values of i) of Beta-coalescents coming down from infinity [5], although α was used in a different parametrization there. Also note that, in most coalescent models, the rescaling order is not the same all along the vector. As an example, four different renormalizations are listed in the study the SFS of the Bolthausen-Sznitman coalescent, [29].

2.6 Outline of the paper

The rest of the paper is organized as follows. In Section 3 we use Stein's method to derive bounds for the total variation distance between vectors obtained by throwing balls into urns and their Poisson approximations. These results are used in Section 4 to prove the convergence of the generator of the multidimensional block counting process $(\mu_t^n; t \geq 0)$ (defined in equation (2.2)) to the generator of the limiting process (defined in (2.4)). Section 5 is devoted to the study of the martingale problem (2.5). Before proving the uniqueness of its solution (Theorem 2.1), we analyze the coagulation operator $\mathcal{C}^x$ (some additional technical results can be found in Appendix B). In Section 6, we prove the convergence of $(\mu_t^n; t \geq 0)$ to the unique solution of the martingale problem (Theorem 2.2). In Section 7, we prove that the limiting process is self-similar and we characterize its Lamperti-Kiu transform (Theorem 2.3). We also provide an additional representation of the process using stochastic flows. Finally, in Section 8, we study some asymptotics on the branch lengths which allow us to prove Theorem 2.4. Appendix A contains some moment estimates on the mass partition components $p_j^{(k)}$ (defined in (1.1)) that are used in several proofs.

3 Urn estimates

Let E be a discrete space equipped with the usual σ-field $\mathcal{F}$ generated from the singletons. Recall that the total variation distance between two measures ν_1, ν_2 on E is given by

$$d_{TV}(\nu_1, \nu_2) = \sup_{A \in \mathcal{F}} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sum_{x \in E} |\nu_1(x) - \nu_2(x)|. \tag{3.1}$$

For a random variable X, we denote by $\mathfrak{L}(X)$ its law.

In this section we recall and establish some bounds for the total variation distance between binomial variables and vectors obtained by throwing balls into urns and their Poisson approximations. Those results are mainly obtained using Stein's method [40].

3.1 Undistinguishable balls

We start by considering n indistinguishable balls that are allocated at random to k urns. For $i \in [k]$, let p_i be the probability of being allocated to the ith urn. Let X_i be the

number of balls allocated to urn i so that X_i has a binomial distribution with parameters n and p_i and $\sum_{i=1}^{k} X_i = n$.

Lemma 3.1. *Let $y > 0$. Let Y be a Poisson distributed random variable with parameter yn/k. Then,*

$$d_{TV}\big(\mathfrak{L}(X_1),\ \mathfrak{L}(Y)\big) \leq \min\{p_1, p_1^2 n\} + \frac{n}{k}|y - kp_1|.$$

Proof. Let W be a Poisson random variable with parameter $p_1 n$. Using triangular inequality,

$$d_{TV}\big(\mathfrak{L}(X_1),\ \mathfrak{L}(Y)\big) \leq d_{TV}\big(\mathfrak{L}(X_1), \mathfrak{L}(W)\big) + d_{TV}\big(\mathfrak{L}(W), \mathfrak{L}(Y)\big).$$

For the first term in the RHS we use the celebrated Chen-Stein inequality for the approximation of the total variation between Poisson and binomial random variables (see for example Theorem 4.6 in [40]). For the second term we use an inequality for the total variation distance between Poisson random variables with different means, that can be found in equation (5) of [39]. $\square$

Now we consider balls that are allocated to two different urns.

Lemma 3.2. *Let (W_1, W_2) be a pair of independent Poisson distributed random variables with respective parameters $p_1 n$ and $p_2 n$. Then,*

$$d_{TV}\big(\mathfrak{L}(X_1, X_2),\ \mathfrak{L}(W_1, W_2)\big) \leq (p_1 + p_2)^2 n.$$

Proof. Our argument relies on the following observation

1. $Z := X_1 + X_2$ follows a binomial distribution of parameters $(n, p_1 + p_2)$.
2. Conditioned on $Z = z$, (X_1, X_2) has a multinomial law of parameters $(z, \frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.

Analogously, we can consider a pair of random variables constructed as follows. Let W be a Poisson random variable of parameter $(p_1 + p_2)n$. Conditionally on $W = z$, (W_1, W_2) has the same multinomial law of parameters $(z, \frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$. For $z \in \mathbb{N}_0$, denote by B_z a binomial random variable with parameters $(z, \frac{p_1}{p_1+p_2})$. Then

$$\sum_{i,j \geq 0} \big|\mathbb{P}\big((X_1, X_2) = (i,j)\big) - \mathbb{P}\big((W_1, W_2) = (i,j)\big)\big|$$

$$= \sum_{i,j \geq 0} \big|\mathbb{P}(Z = i+j)\mathbb{P}(B_{i+j} = i) - P(W = i+j)\mathbb{P}(B_{i+j} = i)\big|$$

$$= \sum_{z=0}^{n} \sum_{i=0}^{z} \big|\mathbb{P}(Z = z) - P(W = z)\big|\mathbb{P}(B_z = i) = \sum_{z=0}^{n} \big|\mathbb{P}(Z = z) - P(W = z)\big|.$$

So, using again Chen-Stein's inequality (Theorem 4.6 in [40]), we conclude that

$$d_{TV}\big(\mathfrak{L}(X_1, X_2),\ \mathfrak{L}(W_1, W_2)\big) = d_{TV}\big(\mathfrak{L}(Z),\ \mathfrak{L}(W)\big) \leq \min\big\{(p_1 + p_2), (p_1 + p_2)^2 n\big\}. \quad \square$$

3.2 Balls with distinct sizes

In this section we consider an urn problem where balls are distinguishable by their sizes. We start with a general result.

Lemma 3.3. *Let $\ell \in \mathbb{N}$ and let $(A^{(1)}, \ldots, A^{(\ell+1)})$ and $(B^{(1)}, \ldots, B^{(\ell+1)})$ be two vectors of independent random variables defined on the same discrete subset E. Then,*

$$d_{TV}\big(\mathfrak{L}\big(A^{(1)}, \ldots, A^{(\ell+1)}\big),\ \mathfrak{L}\big(B^{(1)}, \ldots, B^{(\ell+1)}\big)\big) \leq \sum_{i=1}^{\ell+1} d_{TV}\big(\mathfrak{L}\big(A^{(i)}\big), \mathfrak{L}\big(B^{(i)}\big)\big).$$

Proof. Let us denote by $\nu_{A^{(i)}}$ the distribution of $A^{(i)}$ and by $\nu_{B^{(i)}}$ the distribution of $B^{(i)}$, for $i \in [\ell + 1]$. Since the product function $f(x_1, \ldots, x_{\ell+1}) = x_1 \cdot \ldots \cdot x_{\ell+1}$ is 1-Lipshitz on $[0,1]^{\ell+1}$ for the L^1 norm on $\mathbb{R}^{\ell+1}$, the second definition of the total variation in (3.1) implies that

$$d_{TV}\big(\mathfrak{L}\big(A^{(1)}, \ldots, A^{(\ell+1)}\big),\ \mathfrak{L}\big(B^{(1)}, \ldots, B^{(\ell+1)}\big)\big)$$

$$= \frac{1}{2} \sum_{(x_1, \ldots, x_{\ell+1}) \in E^{\ell+1}} \big| \nu_{A^{(1)}}(x_1) \ldots \nu_{A^{(\ell+1)}}(x_{\ell+1}) - \nu_{B^{(1)}}(x_1) \ldots \nu_{B^{(\ell+1)}}(x_{\ell+1}) \big|$$

$$\leq \frac{1}{2} \sum_{(x_1, \ldots, x_{\ell+1}) \in E^{\ell+1}} \sum_{i=1}^{\ell+1} \big| \nu_{A^{(i)}}(x_i) - \nu_{B^{(i)}}(x_i) \big| = \sum_{i=1}^{\ell+1} d_{TV}\big(\mathfrak{L}\big(A^{(i)}\big), \mathfrak{L}\big(B^{(i)}\big)\big). \qquad \square$$

Fix $\ell \in \mathbb{N}$. We consider balls with $\ell + 1$ distinct sizes. For $i \in [\ell+1]$, suppose that there are N_i balls of size i. We allocate at random these balls into k urns, with probabilities $(p_1, \ldots, p_k)$. Let $(X_j^{(1)}, \ldots, X_j^{(\ell+1)})$ be the number of balls of each size that are allocated to urn j. The variables $(X_j^{(1)}, \ldots, X_j^{(\ell+1)})$ are independent binomial random variables with respective parameters (N_i, p_j). The following result is a direct consequence of Lemmas 3.1 and 3.3.

Corollary 3.4. *Let $(X^{(1)}, \ldots, X^{(\ell+1)})$ be independent binomial random variables with respective parameters $(N_1, p_1), \ldots, (N_{\ell+1}, p_1)$. Let $y_1 > 0$ and let $(Y^{(1)}, \ldots, Y^{(\ell+1)})$ be independent Poisson random variables with respective parameters $y_1 N_1/k, \ldots, y_1 N_{\ell+1}/k$. Let $\bar{N} = \sup_i N_i$. We have*

$$d_{TV}\big(\mathfrak{L}\big(X^{(1)}, \ldots, X^{(\ell+1)}\big),\ \mathfrak{L}\big(Y^{(1)}, \ldots, Y^{(\ell+1)}\big)\big)$$

$$\leq (\ell + 1)\left(\min\{p_1, p_1^2 \bar{N}\} + \frac{\bar{N}}{k} |y_1 - k p_1| \right).$$

For each $i \in [\ell + 1]$, let $\vec{X}^{(i)} := (X_1^{(i)}, X_2^{(i)})$ denote the number of balls of size i that are allocated to urn 1 and 2 respectively. The variable $X_1^{(i)}$ (resp. $X_2^{(i)}$) follows a binomial distribution of parameters (N_i, p_1) (resp., (N_i, p_2)). The following result is a direct consequence of Lemma 3.3 combined first with Lemma 3.2 and second with the Chen-Stein inequality.

Corollary 3.5. *Let $(\vec{W}^{(1)}, \ldots, \vec{W}^{(\ell+1)})$ be independent random vectors such that $\vec{W}^{(i)} = (W_1^{(i)}, W_2^{(i)})$ is a couple of independent Poisson random variables with respective parameters $p_1 N_i$, $p_2 N_i$. Then,*

$$d_{TV}\big(\mathfrak{L}\big(\vec{X}^{(1)}, \ldots, \vec{X}^{(\ell+1)}\big), \mathfrak{L}\big(\vec{W}^{(1)}, \ldots, \vec{W}^{(\ell+1)}\big)\big) \leq (\ell + 1)(p_1 + p_2)^2 \bar{N},$$

and, for $j \in \{1, 2\}$,

$$d_{TV}\big(\mathfrak{L}\big(X_j^{(1)}, \ldots, X_j^{(\ell+1)}\big), \mathfrak{L}\big(W_j^{(1)}, \ldots, W_j^{(\ell+1)}\big)\big) \leq (\ell + 1) p_j^2 \bar{N},$$

where $\bar{N} = \sup_i \{N_i\}$.

3.3 Coagulation operators defined from urn problems

Again, we fix $\ell \in \mathbb{N}$ and consider balls with $\ell + 1$ distinct sizes. Let $\vec{N} = (N_1, \ldots, N_{\ell+1})$ where N_i denotes the number of balls of size i. We allocate at random these balls into k urns, such that the probabilities of falling in different urns are given by $\vec{p} = (p_1, \ldots, p_k)$. We define a random coagulation of $\vec{N}$ by considering that balls that are assigned to the same urn are merged into one ball whose size is the sum of all of them, i.e., such that

$$\forall m \in \mathbb{N},\ \mathbf{C}^{\vec{p}}(\vec{N})(m) = \#\{j \leq k : \text{sum of the sizes of the balls falling in box } j \text{ is } m\}.$$

Lemma 3.6. *Fix a probability vector $\vec{p} = (p_1, \ldots, p_k) \in [0, 1]^k$ and consider the random coagulation associated to $\vec{p}$. For every $\vec{N} \in \mathbb{N}^{\ell+1}$,*

$$\forall m \in \mathbb{N}, \quad \left| N_m - \mathbf{C}^{\vec{p}}(\vec{N})(m) \right| \leq 2\left(|\vec{N}| - |\mathbf{C}^{\vec{p}}(\vec{N})| \right), \tag{3.2}$$

almost surely, and

$$\mathbb{E}\left(|\vec{N}| - |\mathbf{C}^{\vec{p}}(\vec{N})| \right) \leq |\vec{N}|^2 \sum_{j=1}^{k} p_j^2. \tag{3.3}$$

Proof. Let us first prove (3.2). The difference $N_m - \mathbf{C}^{\vec{p}}(\vec{N})(m)$ is the net gain or loss of balls of size m in the coagulation operation. This can be computed by taking the difference of the following two quantities

 (a) The number of balls of size m falling in an urn where another ball is assigned (regardless of its size).

 (b) The number of urns with more than two balls and whose sizes add up to m.

Suppose first that $N_m > \mathbf{C}^{\vec{p}}(\vec{N})(m)$. Thus, $N_m - \mathbf{C}^{\vec{p}}(\vec{N})(m)$ is smaller than (a) alone. Finally, (a) is smaller than the twice the total number of balls that are lost. As an illustrative example, consider the case when the N_m balls are assigned in pairs to $N_m/2$ different urns, coagulating into $N_m/2$ balls of size $2m$. Then the number of balls of size m that are lost is N_m and the total number of balls that are lost is $N_m/2$.

Now suppose that $N_m < \mathbf{C}^{\vec{p}}(\vec{N})(m)$. In this case $\mathbf{C}^{\vec{p}}(\vec{N})(m) - N_m$ is less than (b). In turn, this is less than the number of urns containing at least two balls, which is less than the total number of balls that are lost. This completes the proof of (3.2).

Now, let us turn to the proof of (3.3). Since for $x \in (0, 1)$, we have $\log(1 - x) \leq -x$ and $e^{-x} \leq 1 - x + x^2$, then

$$1 - (1 - p)^n = 1 - e^{n \log(1-p)} \geq 1 - e^{-np} \geq np - (np)^2.$$

Recall that $|\mathbf{C}^{\vec{p}}(\vec{N})|$ is the number of non-empty boxes when assigning $|\vec{N}|$ balls to k urns with probabilities $(p_i, \ldots, p_k)$. Using the previous inequality,

$$\mathbb{E}\left(|\vec{N}| - |\mathbf{C}^{\vec{p}}(\vec{N})| \right) = |\vec{N}| - \sum_{j=1}^{k} \left(1 - (1 - p_j)^{|\vec{N}|} \right)$$

$$\leq |\vec{N}| - \sum_{j=1}^{k} \left(|\vec{N}| p_j - |\vec{N}|^2 p_j^2 \right).$$

Observing that $\sum_{j=1}^{k} p_j = 1$ yields the desired result. $\square$

4 Sequence of urns. Convergence of the generator

Recall from Section 2 the continuous generator $\mathcal{A}$ defined in (2.4) and the discrete generator $\mathcal{A}_n$ defined in (2.2). We use the urn estimates obtained in Section 3 to prove the following result.

Proposition 4.1. *Let $f \in \mathcal{T}$. Consider a sequence $(\mathbf{z}_n; n \in \mathbb{N})$ with $\mathbf{z}_n \in \mathcal{Z}_n$ such that $\mathbf{z}_n \to \mathbf{z}$ in $\ell^1(\mathbb{R}_+)$. Then*

$$\mathcal{A}_n f(\mathbf{z}_n) \to \mathcal{A} f(\mathbf{z}).$$

4.1 First moment estimate

Recall the notations $\mathcal{C}^x$ from (2.3) and $\Lambda^{k,n}$ from (2.1). The main objective of this section is a careful justification of the approximation of $\Lambda^{k,n}(\mathbf{z}_n)(\ell)$ by $\mathcal{C}^{k/n}(\mathbf{z}_n)(\ell)$, as suggested in the heuristics provided in Section 2.2, together with a "rate of convergence" that will be needed to prove Proposition 4.1.

Lemma 4.2. *For every $\ell \in \mathbb{N}$ and $\mathbf{z}_n \in \mathcal{Z}_n$, we have*

$$\left|\mathbb{E}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) - \mathcal{C}^{k/n}(\mathbf{z}_n)(\ell)\right| \le 2(\ell+1)\frac{k}{n}\mathbb{E}\left(p_1^{(k)} + \frac{n}{k}|\Gamma - kp_1^{(k)}|\right)$$

and

$$\left|\mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)|\big) - |\mathcal{C}^{k/n}(\mathbf{z}_n)|\right| \le 2\frac{k}{n}\mathbb{E}\left(p_1^{(k)} + \frac{n}{k}|\Gamma - kp_1^{(k)}|\right),$$

where we recall that $p_1^{(k)} = w_1/\sum_{i=1}^k w_i$ and $\Gamma = w_1/\mathbb{E}(w_1)$ so that the random variables are coupled through the same w_1.

Proof. We start by proving the first inequality. Fix $\ell \in \mathbb{N}$. For $i \in [\ell]$, define $N_i = nz_n(i)$, and set $N_{\ell+1} = n(|\mathbf{z}_n| - \sum_{i=1}^\ell z_n(i))$. In terms of the urn problem of Section 3, N_i is the number of balls of size i, for $i \in [\ell]$, and $N_{\ell+1}$ is the number of balls of size strictly larger than ℓ. Let us now consider a partition $c \in \varphi^{-1}(\ell)$, i.e., a vector $c \in \ell^1(\mathbb{R}_+)$ such that $\sum ic(i) = \ell$. An urn containing $c(i)$ balls of size i for each $i \in [\ell]$ corresponds to the formation of a new block of size ℓ. Let $B(c)$ denote the number of urns containing balls given by the partition c (to ease the notation we do not indicate the dependence on $\mathbf{z}_n$ and k). We have

$$\Lambda^{k,n}(\mathbf{z}_n)(\ell) = \frac{1}{n}\sum_{c \in \varphi^{-1}(\ell)} B(c).$$

Mirroring the notation of Section 3.2, consider a vector of r.v.'s $(X^{(1)}, \ldots, X^{(\ell+1)})$ such that, conditional on $p_1^{(k)}$, the entries are independent and $X^{(i)}$ is distributed as a binomial r.v. with parameters $(N_i, p_1^{(k)})$. By exchangeability of the boxes, we have

$$\mathbb{E}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) = \frac{k}{n}\mathbb{E}\left(\sum_{c \in \varphi^{-1}(\ell)} b_1(c)\right),$$

where $b_1(c)$ is the indicator that $X^{(i)} = c(i)$ for every $i \in [\ell]$ and that $X^{(\ell+1)} = 0$.

Similarly, let $(Y^{(1)}, \ldots, Y^{(\ell+1)})$ such that, conditional on Γ, the entries are independent and $Y^{(i)}$ is distributed as a Poisson r.v. with parameter $\Gamma N_i/k$. Let $d_1(c)$ be the indicator that $Y^{(i)} = c(i)$ for every $i \in [\ell]$ and that $Y^{(\ell+1)} = 0$. A direct computation shows that

$$\mathcal{C}^{k/n}(\mathbf{z}_n)(\ell) = \frac{k}{n}\mathbb{E}\left(\sum_{c \in \varphi^{-1}(\ell)} d_1(c)\right).$$

It follows that

$$\left|\mathbb{E}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) - \mathcal{C}^{k/n}(\mathbf{z}_n)(\ell)\right| = \frac{k}{n}\left|\sum_{c \in \varphi^{-1}(\ell)} \big(\mathbb{E}(b_1(c)) - \mathbb{E}(d_1(c))\big)\right|$$

$$\le 2\frac{k}{n}d_{TV}\big(\mathfrak{L}(X^{(1)}, \ldots, X^{(\ell+1)}), \mathfrak{L}(Y^{(1)}, \ldots, Y^{(\ell+1)})\big).$$

The result follows by a direct application of Corollary 3.4 after conditioning on $p_1^{(k)}$ and Γ.

We now prove the second inequality. Mirroring the notation of Section 3.1, we consider the random variable X_1 such that conditional on $p_1^{(k)}$, X_1 is distributed as a binomial r.v. with parameters $(n|\mathbf{z}_n|, p_1^{(k)})$. Let b_1 be the indicator that box 1 contains at least one ball. Similarly, let Y_1 be the random variable such that, conditional on Γ, Y_1 is distributed as a Poisson r.v. with parameter $\Gamma n|\mathbf{z}_n|/k$. Let d_1 be the indicator that $Y_1 \geq 1$. We have

$$\mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)|\big) = \frac{k}{n}\mathbb{E}(b_1) \text{ and } \big|\mathcal{C}^{k/n}(\mathbf{z}_n)\big| = \frac{k}{n}\mathbb{E}(d_1)$$

and the result then follows from Lemma 3.1. $\qquad\square$

4.2 Second moment estimate

The aim of this section is to bound the variance of the operator $\Lambda^{k,n}(\mathbf{z})(\ell)$.

Lemma 4.3. *For every $\ell \in \mathbb{N}$, there exists a constant $C > 0$ such that for every $\mathbf{z}_n \in \mathcal{Z}_n$, we have*

$$\mathrm{Var}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) \quad \leq \quad C\left(\frac{1}{n} + \max\left\{1, \frac{k}{n}\right\} h(k)\right)$$

where h is a function of k, with no dependence in n, z_n which goes to 0 as $k \to \infty$. Further, the same property holds for $\mathrm{Var}(|\Lambda^{k,n}(\mathbf{z}_n)|)$.

Proof. We start by proving the inequality for $\mathrm{Var}(|\Lambda^{k,n}(\mathbf{z}_n)|)$. Fix $\ell \in \mathbb{N}$. In the following, we write $N_i = nz_n(i)$. We consider the vector $(\vec{X}^{(1)}, \dots, \vec{X}^{(\ell+1)})$ where the entries $\vec{X}^{(i)} = (X_1^{(i)}, X_2^{(i)})$ are, conditional on $(p_1^{(k)}, p_2^{(k)})$, independent random vectors such that $X_1^{(i)} + X_2^{(i)}$ is binomial with parameters $(N_i, p_1^{(k)} + p_2^{(k)})$ and, conditional on $\{X_1^{(i)} + X_2^{(i)} = z\}$, $\vec{X}^{(i)}$ is multinomial with parameters $(z, \frac{p_1^{(k)}}{p_1^{(k)}+p_2^{(k)}}, \frac{p_2^{(k)}}{p_1^{(k)}+p_2^{(k)}})$. Analogously to Lemma 4.2, we define $b_j(c)$ (with $j = 1, 2$) as the indicator that $X_j^{(i)} = c(i)$ for $i \in [\ell]$ and that $X_j^{(\ell+1)} = 0$. Adapting the notations in the proof of Lemma 4.2 and using again exchangeability between urns,

$$\mathrm{Var}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) = \frac{k(k-1)}{n^2} \sum_{c_1, c_2 \in \varphi^{-1}(\ell)} \mathrm{Cov}\big(b_1(c_1), b_2(c_2)\big)$$

$$+ \frac{k}{n^2} \sum_{c_1 \in \varphi^{-1}(\ell)} \mathrm{Cov}\big(b_1(c_1), b_1(c_1)\big). \tag{4.1}$$

Step 1. We start by considering the first term in (4.1). Define $(\vec{W}^{(1)}, \dots, \vec{W}^{(\ell+1)})$ such that the entries $\vec{W}^{(i)} = (W_1^{(i)}, W_2^{(i)})$ are, conditional on $(p_1^{(k)}, p_2^{(k)})$, independent random vectors such that $W_1^{(i)}$ and $W_2^{(i)}$ are independent Poisson variables with parameters $N_i p_1^{(k)}$ and $N_i p_2^{(k)}$. We define $d_j(c)$ (with $j = 1, 2$) as the indicator that $W_j^{(i)} = c(i)$ for $i \in [\ell]$ and that $W_j^{(\ell+1)} = 0$. We have

$$\big|\mathrm{Cov}\big(b_1(c_1), b_2(c_2)\big)\big| \leq \big|\mathbb{E}\big(b_1(c_1)b_2(c_2)\big) - \mathbb{E}\big(d_1(c_1)d_2(c_2)\big)\big|$$

$$+ \big|\mathbb{E}\big(d_1(c_1)d_2(c_2)\big) - \mathbb{E}\big(d_1(c_1)\big)\mathbb{E}\big(d_2(c_2)\big)\big|$$

$$+ \big|\mathbb{E}\big(d_1(c_1)\big)\mathbb{E}\big(d_2(c_2)\big) - \mathbb{E}\big(b_1(c_1)\big)\mathbb{E}\big(b_2(c_2)\big)\big|$$

$$\leq \big|\mathbb{E}\big(b_1(c_1), b_2(c_2)\big) - \mathbb{E}\big(d_1(c_1)d_2(c_2)\big)\big|$$

$$+ \big|\mathrm{Cov}\big(d_1(c_1), d_2(c_2)\big)\big| + \sum_{j=1}^{2}\big|\mathbb{E}\big(b_j(c_j)\big) - \mathbb{E}\big(d_j(c_j)\big)\big|, \tag{4.2}$$

where, for the last term of the second inequality, we used the fact that the product function $f(x_1, x_2) = x_1 x_2$ is 1-Lipshitz on $[0,1]^2$ in the L^1 norm of $\mathbb{R}^2$. We now consider each of the terms in the RHS of (4.2) separately.

We start with the third term. The result follows from the second item of Corollary 3.5 and point (ii) of Proposition A.1 (in Appendix A). More precisely, as in the proof Lemma 4.2,

$$
\begin{aligned}
\frac{k(k-1)}{n^2} &\sum_{c_1, c_2 \in \varphi^{-1}(\ell)} \sum_{j=1}^{2} \big| \mathbb{E}\big(b_j(c_j)\big) - \mathbb{E}\big(d_j(c_j)\big) \big| \\
&\leq 4 \frac{k(k-1)}{n^2} d_{TV}\big(\mathfrak{L}\big(X_1^{(1)}, \ldots, X_1^{(\ell+1)}\big), \mathfrak{L}\big(W_1^{(1)}, \ldots, W_1^{(\ell+1)}\big)\big) \\
&\leq \frac{4}{n}(\ell+1)\mathbb{E}\big(\big(kp_j^{(k)}\big)^2\big) \\
&\leq C\frac{1}{n},
\end{aligned}
\tag{4.3}
$$

where C is a positive constant.

For the first term, we observe that

$$
\begin{aligned}
\sum_{c_1, c_2 \in \varphi^{-1}(\ell)} &\big| \mathbb{E}\big(b_1(c_1)b_2(c_2)\big) - \mathbb{E}\big(d_1(c_1)d_2(c_2)\big) \big| \\
&\leq 2 d_{TV}\big(\mathfrak{L}\big(\vec{X}^{(1)}, \ldots, \vec{X}^{(\ell+1)}\big), \mathfrak{L}\big(\vec{W}^{(1)}, \ldots, \vec{W}^{(\ell+1)}\big)\big).
\end{aligned}
$$

We can then apply the first item of Corollary 3.5 to obtain the following bound

$$
\begin{aligned}
\frac{k(k-1)}{n^2} &\sum_{c_1, c_2 \in \varphi^{-1}(\ell)} \big| \mathbb{E}\big(b_1(c_1)b_2(c_2)\big) - \mathbb{E}\big(d_1(c_1)d_2(c_2)\big) \big| \\
&\leq \frac{2}{n}(\ell+1)\mathbb{E}\big(k^2\big(p_1^{(k)} + p_2^{(k)}\big)^2\big) \\
&\leq \frac{2}{n}(\ell+1)4\mathbb{E}\big(\big(kp_1^{(k)}\big)^2\big) \leq C\frac{1}{n},
\end{aligned}
\tag{4.4}
$$

where C is a positive constant and we used the fact that $\mathbb{E}((kp_1^{(k)})^2)$ is finite by point (ii) of Proposition A.1 (in Appendix A).

Finally, we deal with the second term of (4.2),

$$
\begin{aligned}
\big|\mathrm{Cov}\big(d_1(c_1), d_2(c_2)\big)\big| &= \big|\mathbb{E}\big(\mathrm{Cov}\big(d_1(c_1), d_2(c_2)\big) \mid p_1^{(k)}, p_2^{(k)}\big)\big) \\
&\quad + \mathrm{Cov}\big(\mathbb{E}\big(d_1(c_1) \mid p_1^{(k)}, p_2^{(k)}\big), \ \mathbb{E}\big(d_2(c_2) \mid p_1^{(k)}, p_2^{(k)}\big)\big)\big| \\
&= \big|\mathrm{Cov}\big(\mathbb{E}\big(d_1(c_1) \mid p_1^{(k)}, p_2^{(k)}\big), \mathbb{E}\big(d_2(c_2) \mid p_1^{(k)}, p_2^{(k)}\big)\big)\big| \\
&= \big|\mathrm{Cov}\big(g_{v,c_1}\big(kp_1^{(k)}\big), g_{v,c_2}\big(kp_2^{(k)}\big)\big)\big| \\
&\leq \max\left\{\frac{n^2}{k^2}, \frac{n}{k}\right\} h(k),
\end{aligned}
\tag{4.5}
$$

where $g_{v,c}$ and h are defined as in Lemma A.3 (in Appendix A) with $v = \frac{n}{k}(z_n(1), \ldots, z_n(\ell), |z_n|)$, where $v(\ell+1) = \frac{n}{k}|z_n| \leq \frac{n}{k}$. The inequality follows from that lemma.

Combining the three inequalities (4.3), (4.4) and (4.5), there exists a constant C such that the first term in the RHS of (4.1) can be bounded by

$$
\begin{aligned}
\frac{k(k-1)}{n^2} &\sum_{c_1 \in \varphi^{-1}(\ell_1), c_2 \in \varphi^{-1}(\ell_2)} \mathrm{Cov}\big(b_1(c_1), b_2(c_2)\big) \\
&\leq C\frac{1}{n} + h(k)\max\left\{1, \frac{k}{n}\right\},
\end{aligned}
$$

where $h(k) \to 0$ as $k \to \infty$.

Step 2. Now we consider the second term in (4.1),

$$\mathrm{Cov}\big(b_1(c_1), b_1(c_1)\big) = \mathrm{Var}\big(b_1(c_1)\big) = \mathbb{P}\big(b_1(c_1) = 1\big)\big(1 - \mathbb{P}\big(b_1(c_1) = 1\big)\big)$$
$$\leq \mathbb{P}\big(b_1(c_1) = 1\big) \leq \mathbb{E}\big(p_1^{(k)}\big),$$

where we used the fact that $\mathbb{P}(b_1(c_1) = 1)$ is bounded from above by the probability of the event of having at least one ball in the first box. So the second term in (4.1) can be bounded by

$$\frac{k}{n^2} \sum_{c_1 \in \varphi^{-1}(\ell)} \mathrm{Cov}\big(b_1(c_1), b_1(c_1)\big) \leq \frac{1}{n^2} \mathbb{E}\big(k p_1^{(k)}\big)\big|\varphi^{-1}(\ell)\big|.$$

This completes the proof of the first inequality of Lemma 4.3.

We now prove the inequality for $\mathrm{Var}(|\Lambda^{k,n}(\mathbf{z}_n)|)$. We write

$$\big|\Lambda^{k,n}(\mathbf{z}_n)\big| = \frac{1}{n}(k - B_0),$$

where B_0 is the number of empty boxes. The proof follows the same steps as the proof of the first inequality. If $b_{i,0}$ is the indicator that box i is empty

$$\mathrm{Var}\big(|\Lambda^{k,n}(\mathbf{z}_n)|\big) = \frac{k(k-1)}{n^2}\mathrm{Cov}(b_{1,0}, b_{2,0}) + \frac{k}{n^2}\,\mathrm{Var}(b_{1,0}, b_{1,0}).$$

In the first step, analogously to (4.2), we can write

$$\big|\mathrm{Cov}(b_{1,0}, b_{2,0})\big|$$
$$\leq \big|\mathbb{E}(b_{1,0}, b_{2,0}) - \mathbb{E}(d_{1,0}, d_{2,0})\big| + \big|\mathrm{Cov}(d_{1,0}, d_{2,0})\big|\sum_{j=1}^{2}\big|\mathbb{E}(b_{j,0}) - \mathbb{E}(d_{j,0})\big|,$$

where $d_{1,0}, i = \{1, 2\}$ is the indicator that a Poisson r.v. with parameter $n|\mathbf{z}_n|p_i^{(k)}$ is equal to 0. Conditioning on $(p_1^{(k)}, p_2^{(k)})$, $d_{1,0}$ and $d_{2,0}$ are independent. For the first term we use Lemma 3.2. For the second term, we use a similar bound to equation (4.5), where g_{v,c_i}, $i = 1, 2$ is replaced by g_{v,c_0} defined by $g_{v,c_0}(x) := \exp(-v(\ell + 1)x)$ that has Lipshitz constant $v(\ell + 1) = n/k$ and the inequality follows from Lemma A.2 instead of Lemma A.3. For the third term we use the Chen-Stein inequality. For the second step, the proof is analogous to the proof of the first inequality. $\square$

4.3 Convergence of the generators

The previous sections provide the main ingredients to prove Proposition 4.1. Before writing this proof, we still need one preliminary result.

Lemma 4.4. *Fix $f \in \mathcal{T}$. For any $A > 0$ and $n \in \mathbb{N}$, define*

$$\mathcal{R}_{n,A} := \sup_{\mathbf{z}_n \in \mathbf{z}_n}\left\{\sum_{k=An+1}^{\infty} \frac{R(k)}{n^{1-\alpha}}\big|\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big|\right\},$$

we have $\lim_{A \to \infty} \lim_{n \to \infty} \mathcal{R}_{n,A} = 0$.

Proof. Since $f \in \mathcal{T}$, there exists a Lipshitz function F and $\vec{\lambda} \in [0, 1]^K$ such that $f = F \circ \psi_{\vec{\lambda}}$. As a consequence, there exists $C > 0$ (that only depends on the choice of F), such that

$$\forall \mathbf{z}_n \in \mathcal{Z}_n, \quad \big|\Lambda^{k,n}f(\mathbf{z}_n) - f(\mathbf{z}_n)\big| \leq C \sum_{i=1}^{K}\big|\mathbb{E}\big(\Lambda^{k,n}\psi_{\lambda_i}(\mathbf{z}_n)\big) - \psi_{\lambda_i}(\mathbf{z}_n)\big|.$$

In addition, if $\lambda < 1$

$$\forall \mathbf{z}_n \in \mathcal{Z}_n, \ \ \big|\mathbb{E}\big(\Lambda^{k,n}\psi_\lambda(\mathbf{z}_n)\big) - \psi_\lambda(\mathbf{z}_n)\big| \le \sum_{\ell=1}^\infty \lambda^\ell \big|\mathbb{E}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell) - \mathbf{z}_n(\ell)\big)\big|$$

$$\le \mathbb{E}\big(|\mathbf{z}_n| - |\Lambda^{k,n}(\mathbf{z}_n)|\big) 2 \sum_{\ell=1}^\infty \lambda^\ell,$$

where in the last line we used the first part of Lemma 3.6. If $\lambda = 1$,

$$\forall \mathbf{z}_n \in \mathcal{Z}_n, \ \ \big|\mathbb{E}\big(\Lambda^{k,n}\psi_\lambda(\mathbf{z}_n)\big) - \psi_\lambda(\mathbf{z}_n)\big| = \mathbb{E}\big(|\mathbf{z}_n| - |\Lambda^{k,n}(\mathbf{z}_n)|\big).$$

Observe that $n(|\mathbf{z}_n| - |\Lambda^{k,n}(\mathbf{z}_n)|) \ge 0$ is the number of blocks that are lost in the coalescence event. Thus, we can apply (3.3) (with $\vec{N} = n\mathbf{z}_n$), and set the constant $C' = 2\lambda C/(1-\lambda)$ if $\lambda < 1$ and C if $\lambda = 1$ so that

$$\sum_{k=An+1}^\infty \frac{R(k)}{n^{1-\alpha}} \big|\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big| \le C' \sum_{k=An+1}^\infty \frac{R(k)}{n^{2-\alpha}} \left(n^2 |\mathbf{z}_n|^2 \mathbb{E}\left(\sum_{j=1}^k \big(p_j^{(k)}\big)^2 \right) \right)$$

$$\le \frac{C'}{n} \sum_{k=An+1}^\infty \frac{R(k)}{kn^{-\alpha-1}} \mathbb{E}\big(\big(kp_1^{(k)}\big)^2\big).$$

Notice that in the last line we used the exchangeability of the vector $(p_1^{(k)}, \ldots, p_k^{(k)})$ and that the bound does not depend on $\mathbf{z}_n$. By Assumptions 1.1, $\frac{1}{n} \sum_{k>An} \frac{R(k)}{kn^{-\alpha-1}} \to \int_A^\infty \rho \frac{dx}{x^{1+\alpha}}$, and by point (ii) of Proposition A.1 (in Appendix A), $\mathbb{E}\big(\big(kp_1^{(k)}\big)^2\big) \to \mathbb{E}(w_1^2)/\mathbb{E}(w_1)^2$, which yields the desired result. $\qquad\square$

As a consequence, we get a result that will be useful later on to obtain dominated convergence.

Corollary 4.5. *For every $f \in \mathcal{T}$, $\sup_{n\in\mathbb{N}, \mathbf{z}_n \in \mathcal{Z}_n} \mathcal{A}_n f(\mathbf{z}) < \infty$.*

Proof. First,

$$\big|\mathcal{A}_n f(\mathbf{z}_n)\big| \le \sum_{k=1}^n \frac{R(k)}{n^{1-\alpha}} \big|\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big| + \mathcal{R}_{n,1}.$$

The second term on the RHS is bounded by Lemma 4.4. Finally,

$$\sum_{k=1}^n \frac{R(k)}{n^{1-\alpha}} \big|\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big| \le 2\|f\|_\infty \sum_{k=1}^n \frac{R(k)}{n^{1-\alpha}}.$$

Since $R(k) \sim \rho k^{-\alpha}$, the RHS is converging to $\rho \int_0^1 \frac{dx}{x^\alpha} < \infty$. This completes the proof. $\quad\square$

Proof of Proposition 4.1. Let $f \in \mathcal{T}$. We divide $\mathcal{A}_n f(\mathbf{z}_n)$ in two parts.

$$\mathcal{A}_n f(\mathbf{z}_n) = \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big(\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big) + \sum_{k=An+1}^\infty \frac{R(k)}{n^{1-\alpha}} \big(\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big).$$

We proceed by taking successive limits, first when $n \to \infty$ and then when $A \to \infty$. By Lemma 4.4, the second term in the RHS vanishes and it remains to show that

$$\lim_{A\to\infty} \lim_{n\to\infty} \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big(\mathbb{E}\big(\Lambda^{k,n}f(\mathbf{z}_n)\big) - f(\mathbf{z}_n)\big) = \int_0^{+\infty} \big(\mathcal{C}^x f(\mathbf{z}) - f(\mathbf{z})\big)\rho x^{-\alpha} dx.$$

Next,

$$
\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big(\mathbb{E}\big(\Lambda^{k,n} f(\mathbf{z}_n)\big) - f(\mathbf{z}_n) \big) = \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big(\mathcal{C}^{k/n} f(\mathbf{z}_n) - f(\mathbf{z}_n) \big)
$$
$$
+ \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big(f\big(\mathbb{E}\big(\Lambda^{k,n} \mathbf{z}_n\big)\big) - \mathcal{C}^{k/n} f(\mathbf{z}_n) \big)
$$
$$
+ \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big(\mathbb{E}\big(\Lambda^{k,n} f(\mathbf{z}_n)\big) - f\big(\mathbb{E}\big(\Lambda^{k,n} \mathbf{z}_n\big)\big) \big), \quad (4.6)
$$

where the expectation is taken coordinatewise. The rest of the proof will be decomposed into three steps. In the first one, we will show that the first term converges to $\mathcal{A} f(\mathbf{z})$. In the second and third ones, we will show that the second and third terms on the RHS vanish. To do so, we will use the first and second moment estimates derived in this section.

Step 1. We have

$$
\frac{1}{n} \sum_{k=1}^{An} \frac{R(k)}{n^{-\alpha}} \big(\mathcal{C}^{k/n} f(\mathbf{z}_n) - f(\mathbf{z}_n) \big) = \frac{1}{n} \sum_{k=1}^{An} \frac{R(k)}{n^{-\alpha}} \big(\mathcal{C}^{k/n} f(\mathbf{z}) - f(\mathbf{z}) \big)
$$
$$
+ \frac{1}{n} \sum_{k=1}^{An} \frac{R(k)}{n^{-\alpha}} \big(\mathcal{C}^{k/n} f(\mathbf{z}_n) - f(\mathbf{z}_n) - \mathcal{C}^{k/n} f(\mathbf{z}) + f(\mathbf{z}) \big).
$$

For the first term, we first note that $\mathcal{C}^x(f)(\mathbf{z})$ is continuous in x. (This can be shown by a standard domination argument). This implies that

$$
\frac{1}{n} \sum_{k=1}^{An} \frac{R(k)}{n^{-\alpha}} \big(\mathcal{C}^{k/n} f(\mathbf{z}) - f(\mathbf{z}) \big) \to \int_0^A \big(\mathcal{C}^x f(\mathbf{z}) - f(\mathbf{z}) \big) \rho x^{-\alpha} dx.
$$

We now prove that the second term converges to 0. Since $f \in \mathcal{T}$, there exists a Lipshitz function F and $\vec{\lambda} \in [0,1]^K$ such that $f = F \circ \psi_{\vec{\lambda}}$, so there exists $C > 0$ such that

$$
\big| \mathcal{C}^{k/n} f(\mathbf{z}_n) - \mathcal{C}^{k/n} f(\mathbf{z}) \big| \le C \big| \big(\psi_{\vec{\lambda}}\big(\mathcal{C}^{k/n}(\mathbf{z}_n)\big) - \psi_{\vec{\lambda}}\big(\mathcal{C}^{k/n}(\mathbf{z})\big) \big|.
$$

It will be shown in Proposition 5.1 that for $\lambda \in [0, 1/4)$,

$$
\psi_\lambda\big(\mathcal{C}^{k/n}(\mathbf{z})\big) = \frac{k}{n} \mathbb{E}\bigg(\exp\bigg(-\big(|\mathbf{z}| - \psi_\lambda(\mathbf{z})\big) \frac{\Gamma n}{k} \bigg) - \exp\bigg(-|\mathbf{z}| \frac{\Gamma n}{k} \bigg) \bigg),
$$

where $|\mathbf{z}| - \psi_\lambda(\mathbf{z}) \ge 0$. Since the exponential function is Lipschitz on $(-\infty, 0)$, there exists a constant B such that

$$
\big| \psi_\lambda\big(\mathcal{C}^{k/n}(\mathbf{z}_n)\big) - \psi_\lambda\big(\mathcal{C}^{k/n}(\mathbf{z})\big) \big| \le B \mathbb{E}(\Gamma) \big(\big| |\mathbf{z}| - |\mathbf{z}_n| \big| + \big| \psi_\lambda(\mathbf{z}) - \psi_\lambda(\mathbf{z}_n) \big| \big).
$$

This bound is independent of k and goes to 0 as $n \to \infty$, which completes the proof.

Step 2. We prove that the absolute value of the second term in (4.6) converges to 0. Since $f \in \mathcal{T}$, the problem boils down to proving that for every $\lambda \in [0, 1]$,

$$
\lim_{n \to \infty} \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \sum_{\ell=1}^{\infty} \lambda^\ell \big| \mathbb{E}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) - \mathcal{C}^{k/n}(\mathbf{z}_n)(\ell) \big| = 0
$$

and

$$
\lim_{n \to \infty} \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big| \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)|\big) - \big| \mathcal{C}^{k/n}(\mathbf{z}_n) \big| \big| = 0.
$$

Using Lemma 4.2,

$$\text{for } \lambda < 1, \quad \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \sum_{\ell=1}^{\infty} \lambda^\ell \big| \mathbb{E}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) - \mathcal{C}^{\frac{k}{n}}(\mathbf{z}_n)(\ell)\big| \leq 2(I_{n,A} + J_{n,A}) \sum_{\ell=1}^{\infty} \lambda^\ell(\ell+1)$$

and

$$\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \big| \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)|\big) - |\mathcal{C}^{\frac{k}{n}}(\mathbf{z}_n)| \big| \leq 2(I_{n,A} + J_{n,A})$$

where

$$I_{n,A} := \sum_{k=1}^{An} \frac{R(k)}{n^{2-\alpha}} \mathbb{E}\big(k p_1^{(k)}\big) \text{ and } J_{n,A} := \sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \mathbb{E}\big(|\Gamma - k p_1^{(k)}|\big).$$

By Assumptions 1.1, and since $\mathbb{E}(k p_1^{(k)}) = 1$, we have $I_{n,A} \sim \frac{1}{n} \int_0^A \rho x^{-\alpha} dx \to 0$ as $n \to \infty$.

By point (ii) of Proposition A.1 in the Appendix, the sequence $(k p_1^{(k)}; k \in \mathbb{N})$ is uniformly integrable so that $\mathbb{E}(|\Gamma - k p_1^{(k)}|) \to 0$ and by a similar integral-sum comparison, $J_{n,A} \to 0$ as $n \to \infty$.

Step 3. Finally, we prove that the term on the third line of (4.6) converges to 0. As in the previous step, it is enough to prove that for every $\lambda < 1$,

$$\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \sum_{\ell=1}^{\infty} \lambda^\ell \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)(\ell) - \mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell))|\big) \to 0$$

and

$$\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \mathbb{E}\big(\big||\Lambda^{k,n}(\mathbf{z}_n)| - \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)|\big)\big|\big) \to 0.$$

We start by proving the first limit. Recall that $\mathbb{E}(|\Lambda^{k,n}(\mathbf{z}_n)(\ell) - \mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell))|) \leq 2\mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell)) \leq 2$ since $\Lambda^{k,n}(\mathbf{z}_n) \in \mathcal{Z}_n$, $\sum_{\ell=1}^{\infty} \mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell)) \leq 1$. Since $\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \to \rho \int_0^A \frac{dx}{x^\alpha}$, it is enough to prove that for every ℓ_0

$$\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \sum_{\ell=1}^{\ell_0} \lambda^\ell \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)(\ell) - \mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell))|\big) \to 0.$$

By applying succesively Cauchy-Schwarz and Jensen's inequality,

$$\sum_{\ell=1}^{\ell_0} \lambda^\ell \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)(\ell) - \mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell))|\big)$$

$$\leq \sqrt{\sum_{\ell=1}^{\ell_0} \lambda^\ell} \sqrt{\sum_{\ell=1}^{\ell_0} \lambda^\ell \mathbb{E}\big(|\Lambda^{k,n}(\mathbf{z}_n)(\ell) - \mathbb{E}(\Lambda^{k,n}(\mathbf{z}_n)(\ell))|\big)^2}$$

$$\leq \sqrt{\sum_{\ell=1}^{\ell_0} \lambda^\ell} \sqrt{\sum_{\ell=1}^{\ell_0} \lambda^\ell \mathrm{Var}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big)}.$$

Therefore, it is enough to prove that for every ℓ_0

$$\sum_{k=1}^{An} \frac{R(k)}{n^{1-\alpha}} \sqrt{\sum_{\ell=1}^{\ell_0} \lambda^\ell \mathrm{Var}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big)} \to 0. \tag{4.7}$$

Using the first item of Lemma 4.3, for every ℓ_0 there exists a constant C and a function $h(k) \to 0$ such that

$$\forall k \le An, \ \sum_{\ell=1}^{\ell_0} \lambda^\ell \operatorname{Var}\big(\Lambda^{k,n}(\mathbf{z}_n)(\ell)\big) \le C\left(\frac{1}{n} + h(k)\right)$$

with $h(k) \to 0$ as $k \to \infty$. Since $\sum_{k=1}^{An} \frac{R(k)k}{n^{2-\alpha}} \to \rho \int_0^A \frac{dx}{x^{1-\alpha}}$, it is easy to show (4.7) from there.

The second limit can be shown along the same lines. □

5 Martingale problem. Proof of Theorem 2.1

5.1 The coagulation operator

In this section, we study some properties of the coagulation operator $\mathcal{C}^x$ defined in (2.3).

Our results will be both based on the following interpretation of the operator $\mathcal{C}^x$. Conditional on a realization of the random variable $\Gamma = w_1/\mathbb{E}(w_1)$ and consider the sequence of random variables $\vec{N}_{x,\mathbf{z}} := (N_{x,\mathbf{z}}(i); i \in \mathbb{N})$ such that conditional on Γ, the $N_{x,\mathbf{z}}(i)$'s are independent and Poisson distributed with respective parameters $\Gamma \mathbf{z}(i)/x$. Define $\mathbf{C}^1(\vec{N}_{x,\mathbf{z}})$ as the random vector such that $\mathbf{C}^1(\vec{N}_{x,\mathbf{z}})(\ell) = \mathbb{1}_{\{\sum_i N_{x,\mathbf{z}}(i)=\ell\}}$. Using the notations of Lemma 3.6, $\mathbf{C}^1(\vec{N}_{x,\mathbf{z}})$ can be seen as the trivial coagulation operator associated to a single urn, applied to $\vec{N}_{x,\mathbf{z}}$. The following relation will be useful for the next results

$$\frac{1}{x}\mathcal{C}^x(\mathbf{z})(\ell) = \mathbb{E}\big(\mathbf{C}^1(\vec{N}_{x,\mathbf{z}})(\ell)\big). \tag{5.1}$$

Proposition 5.1. *Let $x > 0$ and $\mathbf{z} \in \mathcal{Z}$. The vector $\mathcal{C}^x(\mathbf{z})$ is in $\mathcal{Z}$ and for every $\lambda \in [0,1]$,*

$$\mathcal{C}^x \psi_\lambda(\mathbf{z}) = \mathbb{E}\big(x \exp(-|\mathbf{z}|\Gamma/x)\big(\exp(\psi_\lambda(\mathbf{z})\Gamma/x) - 1\big)\big). \tag{5.2}$$

Proof. We first prove that $\mathcal{C}^x(\mathbf{z})$ is in $\mathcal{Z}$. From (5.1), we have

$$\sum_{\ell=1}^\infty \ell \mathcal{C}^x(\mathbf{z})(\ell) = x \sum_{\ell=1}^\infty \ell \mathbb{P}\left(\sum_{i=1}^\infty i N_{x,\mathbf{z}}(i) = \ell\right)$$

$$= x \mathbb{E}\left(\sum_{i=1}^\infty i N_{x,\mathbf{z}}(i)\right) = x \frac{\mathbb{E}(\Gamma)}{x} \sum_{i=1}^\infty i \mathbf{z}(i) = 1.$$

Define

$$\rho(\lambda) := \mathbb{E}\big(x \exp(-|\mathbf{z}|\Gamma/x)\big(\exp(\psi_\lambda(\mathbf{z})\Gamma/x) - 1\big)\big).$$

According to Lemma B.1 (in Appendix B), for $\ell \in \mathbb{N}$,

$$\frac{1}{\ell!}\frac{d^\ell}{d\lambda^\ell}\rho(\lambda)\bigg|_{\lambda=0} = \mathbb{E}\left(x \exp(-|\mathbf{z}|\Gamma/x) \sum_{c \in \varphi^{-1}(\ell)} \prod_{i=1}^\ell \frac{(z(i)\Gamma/x)^{c(i)}}{c(i)!}\right),$$

whose expression coincides with the ℓth coordinate of $\mathcal{C}^x(\mathbf{z})$ in (2.3). In order to prove (5.2), it remains to show that the Mac-Laurin expansion of ρ converges to ρ pointwise on in a neighborhood of 0. The result for $\lambda \in [0,1]$ is obtained by standard analytic continuation. To do so, we use Taylor's theorem and prove that the remainder $R_\ell(\lambda)$ converges to 0. Let $\delta > 0$. Using Lemma B.2, for every $\lambda < \delta$ we have

$$R_\ell(\lambda) = \int_0^\lambda \frac{\rho^{(\ell+1)}(t)}{\ell!}(\lambda - t)^\ell dt \le \frac{x(2-\delta)^\ell}{(1-\delta)^{2(\ell+1)}}\int_0^\lambda (\ell+1)(\lambda-t)^\ell dt \le \frac{x}{2-\delta}\left(\frac{(2-\delta)\delta}{(1-\delta)^2}\right)^{\ell+1},$$

which converges to 0 as $\ell \to \infty$ for δ small enough. □

The next result is useful to study the integrability of the generator $\mathcal{A}$.

Lemma 5.2. *For every $\mathbf{z} \in \mathcal{Z}$ we have*

$$\big|\psi_\lambda(\mathbf{z}) - \mathcal{C}^x \psi_\lambda(\mathbf{z})\big| \leq \frac{1}{x}\frac{2\lambda}{1-\lambda} \ \ \text{if } \lambda \in [0,1),$$

and

$$\big|\psi_1(\mathbf{z}) - \mathcal{C}^x \psi_1(\mathbf{z})\big| \leq \frac{1}{x}.$$

Proof. Let $x > 0$ and condition on a realization of Γ. From the definition of our trivial coagulation operator, $|\mathbf{C}^1(\vec{N}_{x,z})| = \mathbb{1}_{\{|\vec{N}_{x,z}|>0\}}$,

$$\mathbb{E}\big(|\vec{N}_{x,z}| - |\mathbf{C}^1(\vec{N}_{x,z})|\big) = \mathbb{E}\big(|\vec{N}_{x,z}|\mathbb{1}_{\{|\vec{N}_{x,z}|\geq 2\}}$$
$$\leq \mathbb{E}(|\vec{N}_{x,z}|) = \frac{\mathbb{E}(\Gamma)|\mathbf{z}|}{x} \tag{5.3}$$

which is the desired result for $\lambda = 1$, since $\mathbb{E}(\Gamma) = 1$.

By Lemma 3.6, for $\ell \in \mathbb{N}$,

$$\mathbb{E}\big(\big|\vec{N}_{x,\mathbf{z}}(\ell) - \mathbf{C}^1(\vec{N}_{x,\mathbf{z}})(\ell)\big|\big) \leq 2\mathbb{E}\big(|\vec{N}_{x,\mathbf{z}}| - |\mathbf{C}^1(\vec{N}_{x,\mathbf{z}})|\big),$$

which, combined with (5.3), yields the result for $\lambda < 1$ by summing over ℓ. $\qquad\square$

5.2 Martingale problem

Proof of Theorem 2.1. Lemma 5.2, ensures that the integral with respect to x in the operator $\mathcal{A}$ is integrable at ∞. This, together with the fact that $x \to x^{-\alpha}$ is integrable at 0 shows that the operator $\mathcal{A}$ is well defined. We now proceed in three steps.

Step 1. Let $(\mu_t; t \geq 0)$ be a solution to the martingale problem. Fix $K \in \mathbb{N}$ and $\vec{\lambda} \in \{1\} \times [0,1)^{K-1}$. Define the projected process $(y_t^{\vec{\lambda}}; t \geq 0) := (\psi_{\vec{\lambda}}(\mu_t); t \geq 0)$. In Step 1, we are going to prove the uniqueness in law of the projected process. Notice that, since $\lambda_1 = 1$, the first coordinate of $y_t^{\vec{\lambda}}$ corresponds to $|\mu_t|$. Let $\mathcal{B}$ be the operator acting on $C^2([0,1]^K)$ such that for every $\vec{y} = (y_1, \ldots, y_K) \in [0,1]^K$

$$\mathcal{B}F(\vec{y}) := \int_0^\infty \big(F\big(\mathbb{E}\big(x\exp(-y_1\Gamma/x)\big(\exp(\vec{y}\Gamma/x) - 1\big)\big)\big) - F(\vec{y})\big)\rho x^{-\alpha}dx,$$

where $\exp(\vec{u})$ is the vector with coordinates $\{\exp(u_i)\}_{i=1}^K$, $\exp(\vec{u}) - 1$ is the vector with coordinates $(\exp(u_i) - 1)_{i=1}^K$ and the expected value is taken w.r.t. Γ. It is straightforward to see that Proposition 5.1 implies that $(y_t^{\vec{\lambda}}; t \geq 0)$ satisfies this martingale problem

$$\left(F\big(y_t^{\vec{\lambda}}\big) - \int_0^t \mathcal{B}\, F\big(y_s^{\vec{\lambda}}\big)ds; t \geq 0\right) \ \ \text{is a martingale for every } F \in C^2\big([0,1]^K\big).$$

To conclude, we now show that the solution to this problem is unique.

According to Theorem 5.1 in [3], we need to check that for every $\vec{\lambda} \in \{1\} \times [0,1]^{K-1}$ and every measurable set $B \subset \mathbb{R} \setminus \{0\}$, the function

$$\mathbf{z} \to G_B(\mathbf{z}) := \int_B g(\mathbf{z}, x)dx := \int_B \frac{|\psi_{\vec{\lambda}}(\mathbf{z}) - \mathcal{C}^x\psi_{\vec{\lambda}}(\mathbf{z})|^2}{|\psi_{\vec{\lambda}}(\mathbf{z}) - \mathcal{C}^x\psi_{\vec{\lambda}}(\mathbf{z})|^2 + 1}\rho x^{-\alpha}dx$$

is a continuous and bounded function. By standard continuity theorem under the integral, this boils down to proving that for every $x \in \mathbb{R} \setminus \{0\}$, $\mathbf{z} \to g(\mathbf{z}, x)$ is continuous and that there exists a function h satisfying $\int_B h(x)dx < \infty$ and such that for every $\mathbf{z} \in \mathcal{Z}$, $|g(\mathbf{z}, x)| \leq h(x)$. First, observe that $\mathbf{z} \to \mathcal{C}^x\psi_\lambda(\mathbf{z})$ is continuous. Since $\mathcal{C}^x(\mathbf{z})$ is defined as

an expectation with respect to Γ (see (2.3)), we use again a standard continuity under the integral theorem, by noticing that the quantity inside the expectation is bounded uniformly by x. This implies the continuity of $\mathbf{z} \to |\psi_{\vec{\lambda}}(\mathbf{z}) - \mathcal{C}^x \psi_{\vec{\lambda}}(\mathbf{z})|^2$ on $(0, \infty)$. The continuity of $\mathbf{z} \to g(\mathbf{z}, x)$ follows from there. The existence of the upper bound h follows from two observations. First,

$$\forall x \in (0, 1] \cap B, \ \ g(\mathbf{z}, x) \leq \rho x^{-\alpha}.$$

Second,

$$\forall x \in [1, \infty) \cap B, \ \ g(\mathbf{z}, x) \leq \sum_{i=1}^{K} \big|\psi_{\lambda_i}(\mathbf{z}) - \mathcal{C}^x \psi_{\lambda_i}(\mathbf{z})\big|^2 \rho x^{-\alpha}$$

which, combined with Lemma 5.2, implies the existence of a constant C such that for $\mathbf{z} \in \mathcal{Z}$

$$\forall x \in [1, \infty] \cap B, \ \ g(\mathbf{z}, x) \leq K \frac{C^2}{x^{2+\alpha}}.$$

Step 2. Let us study the uniqueness of the solution to our martingale problem. Fix $t_1 < \cdots < t_n$ and consider the multidimensional process $Z : \lambda \to (\psi_\lambda(\mu_{t_1}), \ldots, \psi_\lambda(\mu_{t_n}))$ on $[0, 1]$ (the "time" parameter is now λ). The previous step shows that the finite dimensional distributions of Z are uniquely determined. Since $|\mu_t| < 1$, the radius of convergence of $\sum_i \mu_t(i)\lambda^i$ is at least 1 and the process Z is continuous a.s. This implies that the distribution of $(\psi_\lambda(\mu_{t_1}), \ldots, \psi_\lambda(\mu_{t_n}); \lambda \in [0, 1))$ is uniquely determined.

Finally, we can differentiate $\psi_\lambda(\mu_t)$ under the sum at 0 infinitely many times to recover μ_t from its moment generating function, i.e.,

$$\forall k \in \mathbb{N}, \ \ \mu_t(k) = \frac{1}{k!} \frac{d^k}{d\lambda^k} \psi_\lambda(\mu_t)\bigg|_{\lambda=0}.$$

This shows that the finite dimensional distributions of μ_t are uniquely determined.

Step 3. The existence of a solution follows from our convergence result (Theorem 2.2). $\qquad\square$

6 Convergence to the limiting process. Proof of Theorem 2.2

In this section, we prove convergence in $D([0, T], \mathcal{Z})$ equipped with the Skorokhod M_1 topology. The proof is based on a useful characterization of tightness in M_1 (see Theorem 12.12.3 and Remark 12.3.2 in [45]). We work with M_1 instead of the more commonly used J_1 because, as far as we know, it is cumbersome to apply similar arguments for the J_1 topology.

Proposition 6.1. *For any $T > 0$, the sequence $(\mu^n; n \in \mathbb{N})$ is tight in $D([0, T], \mathcal{Z})$ equipped with the Skorokhod M_1 topology.*

Proof. Let us define the function $s : \ell^1(\mathbb{R}_+) \to \ell^\infty(\mathbb{R}_+)$ such that

$$s(x) = \left(\sum_{i=1}^{K} x(i) \right)_K.$$

We know that this is a continuous function. We consider the process $\bar{\mu}^n := s(\mu^n)$. It is sufficient to prove tightness of $\bar{\mu}^n$. Observe that every entry of $\bar{\mu}^n$ is decreasing.

We use Theorem 12.12.3 in [45]. Let us define the supremum norm on $\mathcal{Z}$

$$\forall x \in \mathcal{Z}, \ \|x\| := \sup_{0 \leq t \leq T} \|x(t)\| = \sup_{0 \leq t \leq T} \max_i |x_t(i)|.$$

We need to check that:

(i) For each $\epsilon > 0$, there exists c such that

$$\forall\, n \geq 1,\ \mathbb{P}\big(\|\bar{\mu}^n\| > c\big) < \epsilon.$$

(ii) For each $\epsilon > 0$ and $\eta > 0$, there exists δ such that

$$\forall\, n \geq 1,\ \mathbb{P}\big(w\big(\bar{\mu}^n, \delta\big) \geq \eta\big) < \epsilon,$$

where

$$w(x, \delta) := \sup_{t \in [0,T]} \left\{ \sup_{0 \vee (t-\delta) \leq t_1 < t_2 < t_3 \leq (t+\delta) \wedge T} \big\{ \|x_{t_2} - [x_{t_1}, x_{t_3}]\| \big\} \right\},$$

where the segment $[a, b]$ is defined as $[a, b] := \{\alpha a + (1 - \alpha)b : 0 \leq \alpha \leq 1\}$ and the difference between x_{t_2} and $[x_{t_1}, x_{t_3}]$ is the smallest difference between x_{t_2} and any point in the segment $[x_{t_1}, x_{t_3}]$, where we use the standard definition of segments in Banach spaces.

The first condition is trivial (since $\sum_{i=1}^{K} z(i) < 1$ for any $z \in \mathcal{Z}$). To check the second condition, we first notice that

$$\|\bar{\mu}_t^n\| = |\mu_t^n|.$$

Since every coordinate of $(\bar{\mu}_t^n; t \geq 0)$ is decreasing, for any $t \in [0, T]$ and any $(t - \delta) \leq t_1 < t_2 < t_3 \leq (t + \delta)$,

$$\big\| \bar{\mu}_{t_2}^n - \big[\bar{\mu}_{t_1}^n, \bar{\mu}_{t_3}^n\big] \big\| \leq \|\bar{\mu}_{t_1}^n\| - \|\bar{\mu}_{t_3}^n\|$$
$$\leq |\mu_{t_1}^n| - |\mu_{t_3}^n|.$$

Using Markov's inequality,

$$\mathbb{P}\big((|\mu_{t_1}^n| - |\mu_{t_3}^n|) \geq \eta\big) \leq \frac{1}{\eta} \mathbb{E}\big(|\mu_{t_1}^n| - |\mu_{t_3}^n|\big)$$
$$= \frac{1}{\eta} \sum_{k=1}^{\infty} \frac{R(k)}{n^{1-\alpha}} \mathbb{E}\left(\int_{t_1}^{t_3} \big(|\mu_u^n| - \Lambda^{k,n}|\mu_u^n|\big) du \right).$$

Using Corollary 4.5 and the fact that $t_3 - t_1 \leq \delta$, this quantity tends to 0 as $\delta \to 0$ which completes the proof. $\qquad\square$

In the following, we denote by μ any subsequential limit of $(\mu^n; n \in \mathbb{N})$ in $D([0, T], \mathcal{Z})$. It remains to prove that μ is the (unique) solution to the martingale problem. We start by showing that the limiting process μ has no fixed point of discontinuity.

Lemma 6.2. *For any $t \in [0, T]$, and $t_p \downarrow t$ or $t_p \uparrow t$, $\mu_{t_p} \Longrightarrow \mu_t$.*

Proof. Since the function $(\nu_t := \mathbb{E}(|\mu_t|); t \geq 0)$ is non-increasing and valued in $[0, 1]$, it has at most a countable set D of discontinuity points. We aim at showing that the set D is empty. Let $t \in [0, T]$ and $t_p \downarrow t$. Since D is countable, we can always take a sequence $\bar{t}_p > t_p$ s.t, $\bar{t}_p \notin D$ and $\bar{t}_p \to t$. By monotonicity,

$$0 \leq \nu_t - \nu_{t_p} \leq \nu_t - \nu_{\bar{t}_p}.$$

We have

$$\mathbb{E}\big(|\mu_t^n| - |\mu_{\bar{t}_p}^n|\big) = \int_t^{\bar{t}_p} \sum_{k=1}^{\infty} \frac{R(k)}{n^{1-\alpha}} \mathbb{E}\big(|\mu_u^n| - \Lambda^{k,n}|\mu_u^n|\big)) du.$$

We now let $n \to \infty$ (at fixed p). Since $\bar{t}_p \notin D$, we have $\lim_{n \to \infty} \mathbb{E}(|\mu_t^n| - |\mu_{\bar{t}_p}^n|) = \nu_t - \nu_{\bar{t}_p}$. On the other hand, Corollary 4.5 implies the existence of a constant C such that

$$\lim_{n \to \infty} \mathbb{E}\big(|\mu_t^n| - |\mu_{\bar{t}_p}^n|\big) \leq C(\bar{t}_p - t).$$

As a consequence, $\nu_t - \nu_{\bar{t}_p} \to 0$ as $p \to \infty$.

Let us now consider the functions $(\nu_t^K := \mathbb{E}(\sum_{i=1}^K \mu_t(i)); t \geq 0)$, which are also non-increasing and valued in $[0, 1]$. By (3.2) in Lemma 3.6, we have

$$\mathbb{E}\left(\sum_{i=1}^K \mu_t^n(i) - \sum_{i=1}^K \mu_{t_p}^n(i)\right) \leq 2K\mathbb{E}\big(|\mu_t^n| - |\mu_{t_p}^n|\big),$$

and we can apply the same reasoning as above to prove that for every $K \in \mathbb{N}$, $\nu_t^K - \nu_{\bar{t}_p}^K \to 0$ as $p \to \infty$.

This implies that μ_{t_p} converges to μ_t in distribution coordinatewise. By Sheffe's lemma, μ_{t_p} converges to μ_t in distribution in $\ell^1(\mathbb{R}_+)$.

The proof for $t_p \uparrow t$ follows along the same lines. $\qquad\square$

Proof of Theorem 2.2. We need to show that the process $(\mu_t; t \geq 0)$ is the (unique) solution to the martingale problem. Let $f \in \mathcal{T}$. Let $p \in \mathbb{N}$. Let $h_1, \ldots, h_p$ be continuous and bounded functions from $\ell^1(\mathbb{R}_+)$ to $\mathbb{R}_+$. Let $t_1 < \cdots < t_p \leq t$ and $s \geq 0$. Recall that $\mathcal{A}_n$ refers to the generator of the rescaled process μ^n, so that it remains to prove that for such choice of times and test functions, we have

$$\lim_{n \to \infty} \mathbb{E}\left(\left(f(\mu_{t+s}^n) - f(\mu_t^n) - \int_t^{t+s} \mathcal{A}_n f(\mu_u^n) du\right) \prod_{i=1}^p h_i(\mu_{t_i}^n)\right)$$

$$= \mathbb{E}\left(\left(f(\mu_{t+s}) - f(\mu_t) - \int_t^{t+s} \mathcal{A}f(\mu_u) du\right) \prod_{i=1}^p h_i(\mu_{t_i})\right). \tag{6.1}$$

Let us now consider a coupling such that μ^n converges to μ a.s. in $D([0, T], \mathcal{Z})$. In virtue of Lemma 6.2, the times $t, t+s$ and t_i's are a.s. continuity points for the limiting process so that $\mu_u^n \to \mu_u$ for $u \in \{s, s+t, t_1, \ldots, t_k\}$ a.s. Further, by monotonicity of each coordinate, the set of discontinuities for the functions $(|\mu_t|; t \geq 0)$ and $(\sum_{i=1}^\ell \mu_t(i)); t \geq 0)$ is a (random) countable set a.s. This implies that the set of discontinuity points for the limiting process $(\mu_t; t \geq 0)$ has a.s. null Lebesgue measure i.e., for every fixed t, $\mathbb{P}(\mu$ is continuous at $t) = 1$.

Now, in virtue of Corollary 4.5, we can use the bounded convergence theorem (to pass the limit inside $\mathbb{E}$ and the time integral). (6.1) follows from Proposition 4.1 and the fact that a.s. the set of discontinuities for the limiting process has null Lebesgue measure. $\qquad\square$

7 Self-similarity

In this section, we show that the limiting process $(\mu_t; t \geq 0)$ is a self-similar Markov process. The self-similarity property provides a natural Lamperti representation of the process, given in (2.6). This representation allows to construct the process μ as the flow induced by a SDE driven by a Lévy noise, see Theorem 7.2.

Let us first generalize the concept of self-similarity in $\mathbb{R}_+^d$ [30, 1] to $\ell^1(\mathbb{R}_+)$. Let $(\nu_t; t \geq 0)$ be a process valued in $\ell^1(\mathbb{R}_+)$. Let $\mathbb{P}_{\mathbf{z}}$ be the law of ν starting from the initial condition $\mathbf{z}$. We say that ν is self-similar with parameter β if for every $\gamma > 0$

$$\big((\nu_t; t \geq 0), \mathbb{P}_{\mathbf{z}}\big) = \big((\gamma\nu_{t\gamma^{-\beta}}; t \geq 0), \mathbb{P}_{\gamma^{-1}\mathbf{z}}\big).$$

Proposition 7.1. *The process $(\mu_t; t \geq 0)$ is a self-similar process with parameter $\beta = \alpha - 1$.*

Proof. Fix $\gamma > 0$ and consider the rescaled process $\mu^{(\gamma)} := (\gamma\mu_{t\gamma^{-\beta}}; t \geq 0)$. By uniqueness of the solution of the martingale problem introduced in Theorem 2.1, it is sufficient to check that $\mu^{(\gamma)}$ is also a solution. Fix an integer K, a function $F \in C^2([0,1]^K)$ and a vector $\vec{\lambda} \in [0,1]^K$. Let $F^{(\gamma)}(x) = F(\gamma x)$. Since $(\mu_s; s \geq 0)$ is a solution of the martingale problem, then

$$M_t := F^{(\gamma)}\big(\psi_{\vec{\lambda}}(\mu_t)\big) - \int_0^t \int_{\mathbb{R}_+} \big(F^{(\gamma)}\big(\psi_{\vec{\lambda}}\big(\mathcal{C}^x(\mu_s)\big)\big) - F^{(\gamma)}\big(\psi_{\vec{\lambda}}(\mu_s)\big)\big)\rho x^{-\alpha}dxds$$

$$= F\big(\psi_{\vec{\lambda}}(\gamma\mu_t)\big) - \int_0^t \int_{\mathbb{R}_+} \big(F\big(\gamma\psi_{\vec{\lambda}}\big(\mathcal{C}^x(\mu_s)\big)\big) - F\big(\psi_{\vec{\lambda}}(\gamma\mu_s)\big)\big)\rho x^{-\alpha}dxds$$

is a martingale. Next, we observe that for every $\mathbf{z} \in \ell^1(\mathbb{R}_+)$

$$\gamma\mathcal{C}^x\big(\psi_{\vec{\lambda}}(\mathbf{z})\big) = \mathcal{C}^{\gamma x}\big(\psi_{\vec{\lambda}}(\gamma\mathbf{z})\big).$$

This implies

$$M_{t\gamma^{-\beta}} = F\big(\psi_{\vec{\lambda}}\big(\mu_t^{(\gamma)}\big)\big) - \int_0^{t\gamma^{-\beta}} \int_{\mathbb{R}_+} \big(F\big(\psi_{\vec{\lambda}}\big(\mathcal{C}^{\gamma x}(\gamma\mu_s)\big)\big) - F\big(\psi_{\vec{\lambda}}(\gamma\mu_s)\big)\big)\rho x^{-\alpha}dxds.$$

Changing the variables $\bar{s}\gamma^{-\beta} = s$ and $\bar{x} = \gamma x$ in the latter integral yields

$$M_{t\gamma^{-\beta}} = F\big(\psi_{\vec{\lambda}}\big(\mu_t^{(\gamma)}\big)\big) - \underbrace{\gamma^{-\beta+\alpha-1}}_{=1}\int_0^t \int_{\mathbb{R}_+} \big(F\big(\psi_{\vec{\lambda}}\big(\mathcal{C}^{\bar{x}}\big(\mu_{\bar{s}}^{(\gamma)}\big)\big)\big) - F\big(\psi_{\vec{\lambda}}\big(\mu_{\bar{s}}^{(\gamma)}\big)\big)\big)\rho\bar{x}^{-\alpha}d\bar{x}d\bar{s}.$$

This shows that

$$F\big(\psi_{\vec{\lambda}}\big(\mu_t^{(\gamma)}\big)\big) - \int_0^t \int_{\mathbb{R}_+} \big(F\big(\psi_{\vec{\lambda}}\big(\mathcal{C}^{\bar{x}}\big(\mu_{\bar{s}}^{(\gamma)}\big)\big)\big) - F\big(\psi_{\vec{\lambda}}\big(\mu_{\bar{s}}^{(\gamma)}\big)\big)\big)\rho\bar{x}^{-\alpha}d\bar{x}d\bar{s}$$

defines a martingale for every $F \in C^2([0,1]^K)$ so that $\mu^{(\gamma)}$ is also a solution of the martingale problem introduced in Theorem 2.1. $\qquad\square$

Proof of Theorem 2.3. By Proposition 7.1 the process is self-similar in $\ell^1(\mathbb{R}_+)$. Theorem 2.3 in [1] is analogous to Theorem 2.3 for self-similar processes valued in $\mathbb{R}^d$. The proof goes verbatim for our state space. $\qquad\square$

Now, let us provide an alternative representation of the MAP. Consider the generating function of θ_t (the component encoding the asymptotic frequencies introduced in Theorem 2.3), i.e.,

$$\forall t \geq 0, \quad x_t^\lambda := \psi_\lambda(\theta_t) = \sum_{i \geq 1} \lambda^i \theta_t(i). \tag{7.1}$$

For every $\lambda \in [0,1)$, $\vec{\lambda} \in [0,1)^K$, define the processes $x^\lambda := (x_t^\lambda; t \geq 0)$ and $(x_t^{\vec{\lambda}}; t \geq 0) := (\psi_{\vec{\lambda}}(\theta_t); t \geq 0)$. By construction, we have $x_0^\lambda = \lambda$. We consider the flow of stochastic processes indexed by λ, $(x^\lambda; \lambda \in [0,1))$. In Theorem 7.2 we prove that $((\xi_t, x_t^\lambda); t \geq 0)$ has the same law as the flow of stochastic processes $((\bar{\xi}_t, \bar{x}_t^\lambda); t \geq 0)$ defined in (7.2) below.

Let N be a Ppp in $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity measure $dt \otimes \rho a^{-\alpha}da$. We define

$$H(x,a) := \frac{\mathbb{E}(\exp(-\Gamma/a)(\exp(\Gamma x/a) - 1))}{\mathbb{E}(1 - \exp(-\Gamma/a))}, \quad g(a) = -\log\mathbb{E}\big[a\big(1 - \exp(-\Gamma/a)\big)\big],$$

where the expected value is taken w.r.t. Γ. For every $\lambda \in [0,1)$, we consider the unique weak solution to the following stochastic flow

$$\begin{cases} \bar{x}_t^\lambda = \lambda + \int_{\mathbb{R}_+^*} \int_0^t [H(\bar{x}_{s-}^\lambda, a) - \bar{x}_{s-}^\lambda]N(ds \times da) \\ \bar{\xi}_t = \qquad \int_{\mathbb{R}_+^*} \int_0^t g(a)N(ds \times da). \end{cases} \tag{7.2}$$

For every $\vec{\lambda} \in [0,1)^K$, define the projected process $(\bar{x}_t^{\vec{\lambda}}; t \geq 0) := ((\bar{x}_t^{\lambda_1}, \ldots, \bar{x}_t^{\lambda_K}); t \geq 0)$. Note that since $\mathbb{E}(\Gamma) = 1$, $(\bar{\xi}_t; t \geq 0)$ is a spectrally positive subordinator. The existence and uniqueness of such a weak solution is proved along the exact same lines as in [9] Theorem 2, where they considered qualitatively similar stochastic flows driven by a Levy noise.

Theorem 7.2. *Let $K \geq 2$. For every $\vec{\lambda}' \in [0,1)^{K-1}$, let $((\xi_t, x_t^{\vec{\lambda}'}); t \geq 0)$ be defined as in Theorem 2.3 and (7.1) and let $((\bar{\xi}_t, \bar{x}_t^{\vec{\lambda}'}); t \geq 0)$ be the unique weak solution to the stochastic flow defined in (7.2). Then*

$$\left((\xi_t, x_t^{\vec{\lambda}'}); t \geq 0\right) = \left((\bar{\xi}_t, \bar{x}_t^{\vec{\lambda}'}); t \geq 0\right) \;\; \text{in law}.$$

This result allows us to identify the jump measure of the subordinator ξ_t as follows. Consider the measure on $\mathbb{R}_+$ defined by $\rho x^{-\alpha} dx$. The jump measure of the subordinator is the pushforward of this measure by the function g (as defined above). This is a direct consequence of Theorem 7.2.

Proof. Let $\vec{\lambda} = (1, \vec{\lambda}') \in \{1\} \times [0,1)^{K-1}$. Let $(\mu_t; t \geq 0)$ be a solution of the martingale problem defined in (2.5). Recall the definition of

$$y_t^{\vec{\lambda}} := \psi_{\vec{\lambda}}(\mu_t) = \left(|\mu_t|, \psi_{\vec{\lambda}'}(\mu_t)\right)$$

(as in the proof of Theorem 2.1). For every $\vec{y} \in [0,1]^K$, recall the definition of

$$\mathcal{B}F(\vec{y}) := \int_0^\infty \left(F\left(\mathbb{E}\left(x\exp(-\vec{y}_1\Gamma/x)\left(\exp(\vec{y}\Gamma/x) - 1\right)\right)\right) - F(\vec{y})\right)\rho x^{-\alpha}dx,$$

where the expected value is taken w.r.t. Γ. As in the proof of Theorem 2.1, $y^{\vec{\lambda}}$ is the unique solution of the martingale problem

$$\left(F\left(y_t^{\vec{\lambda}}\right) - \int_0^t \mathcal{B}\, F\left(y_s^{\vec{\lambda}}\right)ds; t \geq 0\right) \;\; \text{is a martingale for every } F \in C^2\left([0,1]^K\right).$$

Define the process $(\bar{w}_t^{\vec{\lambda}}; t \geq 0) := (\exp(-\bar{\xi}_t), \exp(-\bar{\xi}_t)\bar{x}_t^{\vec{\lambda}'}; t \geq 0)$. Let $(\bar{\tau}_t; t \geq 0)$ be the Lamperti change of time in (2.6) defined w.r.t. $\bar{\xi}$. Since $\bar{\tau}$ is the inverse time change defined in Theorem 2.3, we need to prove that $(\bar{w}_{\bar{\tau}_t}^{\vec{\lambda}}; t \geq 0) = (y_t^{\vec{\lambda}}; t \geq 0)$ in law. The strategy consists in showing that the time changed process $w_{\bar{\tau}_t}^{\vec{\lambda}}$ solves the same martingale problem.

For every $\vec{w} = (w_1, \ldots, w_K) \in [0,1]^K$, define

$$\mathcal{D}F(\vec{w}) := \int_0^\infty \left(F\left(w_1\exp\left(-g(a)\right)\bar{H}(\vec{w}, a)\right) - F(\vec{w})\right)\rho a^{-\alpha}da,$$

where $\bar{H}(\vec{w}, a) := (1, H(w_2/w_1, a), \ldots, H(w_K/w_1, a))$. From the definition of the stochastic flow (7.2), at every jump time (t, a), we have the following transitions

$$\bar{\xi}_{t^+} = \bar{\xi}_{t^-} + g(a), \;\; \bar{x}_{t^+}^\lambda = H\left(\bar{x}_{t^-}^\lambda, a\right).$$

By applying Ito's formula in the discontinuous case (see [36])

$$\left(F\left(\bar{w}_t^{\vec{\lambda}}\right) - \int_0^t \mathcal{D}\, F\left(\bar{w}_s^{\vec{\lambda}}\right)ds; t \geq 0\right) \;\; \text{is a martingale for every } F \in C^2\left([0,1]^2\right).$$

Since $\bar{\tau}_t$ is a stopping time, the time changed process

$$\left(F\left(\bar{w}_{\bar{\tau}_t}^{\vec{\lambda}}\right) - \int_0^{\bar{\tau}_t} \mathcal{D}\, F\left(\bar{w}_s^{\vec{\lambda}}\right)ds, t \geq 0\right) \tag{7.3}$$

is also a martingale with respect to the time-changed filtration $\bar{\mathcal{G}}_{\tau_t}$, where $\bar{\mathcal{G}}_t$ is the filtration of the original weak solution. We now make the change of variable $s = \bar{\tau}_u = \inf\{t > 0 : \int_0^t \exp(\beta \xi_v) dv > u\}$ in the latter integral. Since $u = \int_0^s \exp((\alpha - 1)\bar{\xi}_v) dv$, we have

$$\exp\big(-(\alpha - 1)\bar{\xi}_{\bar{\tau}_u}\big) du = ds. \tag{7.4}$$

Thus, we get that

$$\int_0^{\bar{\tau}_t} \mathcal{D}F\big(\bar{w}_s^{\vec{\lambda}}\big) ds = \int_0^t \int_0^\infty \exp\big(-(\alpha - 1)\bar{\xi}_{\bar{\tau}_u}\big)\big(F\big(e^{-\bar{\xi}_{\bar{\tau}_u} - g(a)}\bar{H}\big(\bar{w}_{\bar{\tau}_u}^{\vec{\lambda}}, a\big)\big) - F\big(\bar{w}_{\bar{\tau}_u}^{\vec{\lambda}}\big)\big)\rho a^{-\alpha} da du.$$

Making the change of variables $\bar{a} = a/e^{-\bar{\xi}_{\tau_s}}, s = u$, we get that

$$\int_0^{\bar{\tau}_t} \mathcal{D}F\big(\bar{w}_s^{\vec{\lambda}}\big) ds$$

$$= \int_0^t \int_0^\infty \big(F\big(e^{-\bar{\xi}_{\bar{\tau}_u} - g(\bar{a}e^{\bar{\xi}_{\bar{\tau}_u}})}\bar{H}\big(\bar{w}_{\bar{\tau}_u}^{\lambda}, \bar{a}e^{\bar{\xi}_{\bar{\tau}_u}}\big)\big) - F\big(\bar{w}_{\bar{\tau}_u}^{\vec{\lambda}}\big)\big)\rho \bar{a}^{-\alpha} d\bar{a} du$$

$$= \int_0^t \int_0^\infty \bigg(F\bigg(\mathbb{E}\bigg(a \exp\bigg(-\frac{e^{-\bar{\xi}_{\bar{\tau}_u}}\Gamma}{a}\bigg)\bigg(\exp\bigg(\frac{\bar{w}_{\bar{\tau}_u}^{\vec{\lambda}}\Gamma}{a}\bigg) - 1\bigg)\bigg)\bigg) - F\big(\bar{w}_{\bar{\tau}_u}^{\vec{\lambda}}\big)\bigg)\rho a^{-\alpha} da du$$

$$= \int_0^t \mathcal{B}\, F\big(\bar{w}_{\bar{\tau}_u}^{\vec{\lambda}}\big) du,$$

where the second equality is obtained by using the definitions of $\bar{H}$ and g. This completes the proof of the proposition since (7.3) is a martingale. $\qquad\square$

8 Site frequency spectrum: proof of Theorem 2.4

In this section, we use the Poissonian construction of a general Dirichlet coalescent. Namely, we consider a Ppp on $\mathbb{R}_+ \times \mathbb{N}$ with intensity measure $dt \otimes \sum_{k \in \mathbb{N}} R(k)\delta_k$. For $k \geq 1$, an atom (t, k) of the point process corresponds to an event with k urns in the paintbox construction. It will be referred to as a k-event. We recall the definition of the non-rescaled process $(\hat{\mu}_t^n; t \geq 0)$, where $\hat{\mu}_t^n(i)$ counts the number of blocks of size i at time t and $\mu_t^n = \frac{1}{n}\hat{\mu}_{tn^{\alpha-1}}^n$ in (1.2). As a consequence of Theorem 2.2, for any $T > 0$,

$$\big(\mu_t^n; t \geq 0\big) \Longrightarrow (\mu_t; t \geq 0) \text{ in } D\big([0, T], \mathcal{Z}\big).$$

(In particular, $(|\mu_t^n|; t \geq 0)$ converges to the total mass of the limiting process $(|\mu_t|; t \geq 0)$.) By the Skorokhod representation theorem (Theorem 6.7 in [12]), we assume without loss of generality that the convergence holds almost surely. Under this coupling, we aim at proving the following result.

Proposition 8.1. *Let* $\hat{\mathbf{T}}_n = \inf\{t > 0 : |\hat{\mu}_t^n| = 1\}$ *be the time to the MRCA of the population and let* $\mathbf{T}_n = n^{1-\alpha}\hat{\mathbf{T}}_n$ *be its renormalized version. As* $n \to \infty$,

$$\int_0^{\mathbf{T}_n} \mu_s^n ds \Longrightarrow \int_0^\infty \mu_s ds \quad \text{in } \ell^1\big(\mathbb{R}^+\big).$$

The latter mostly relies on the following technical lemma.

Lemma 8.2. *We have that*

$$\limsup_n \mathbb{E}\bigg(\int_0^{\mathbf{T}_n} |\mu_s^n| ds\bigg) < \infty.$$

Proof. Set $\tau_0^n = 0$ and define inductively

$$\forall j \geq 1, \ \tau_j^n = \inf\bigg\{t > \tau_{j-1}^n : \text{a } k\text{-event occurs with } k \leq \frac{n}{2^j}\bigg\}.$$

To simplify the notation, we drop the dependence in n and write $\tau_j^n \equiv \tau_j$. By the strong Markov property, $\tau_j - \tau_{j-1}$ is an exponential random variable with parameter $\sum_{k \leq \frac{n}{2^j}} R(k)$. Recall that $\lim_{k \to \infty} k^\alpha R(k) = \rho$. The limiting behavior of R implies that

$$\lim_{\ell \to \infty} \frac{1}{\ell^{1-\alpha}} \sum_{k \leq \ell} R(k) = \int_0^1 \frac{\rho dx}{x^\alpha} = \frac{\rho}{1-\alpha}.$$

Let $\ell_0 = \inf\{k \geq 0 : R(k) > 0\}$. Since $\sum_{k \leq \ell} R(k) > 0$ for every $\ell \geq \ell_0$, there exists $C > 0$ such that for every $\ell \geq \ell_0$

$$\sum_{k \leq \ell} R(k) \geq C^{-1} \ell^{1-\alpha}.$$

Define $j_0^n \equiv j_0 = [\log_2(\frac{n}{\ell_0})]$. For every $j \leq j_0$, we have $n/2^j \geq \ell_0$ so that

$$\sum_{k \leq \frac{n}{2^j}} R(k) \geq C^{-1} \left(\frac{n}{2^j} \right)^{1-\alpha} \quad \text{and thus} \quad \mathbb{E}(\tau_j - \tau_{j-1}) \leq C \left(\frac{2^j}{n} \right)^{1-\alpha}.$$

By construction, $|\hat{\mu}_t^n| \leq \frac{n}{2^j}$ for every $t > \tau_j$. As a consequence, for every $j \leq j_0$, we have

$$\begin{aligned}
\mathbb{E}\left(\int_0^{\tau_j} |\hat{\mu}_t^n| \right) &\leq \sum_{i=1}^j \mathbb{E}(\tau_i - \tau_{i-1}) \frac{n}{2^{i-1}} \\
&\leq C \sum_{i=1}^j \left(\frac{n}{2^{i-1}} \right)^\alpha \\
&\leq 2Cn^\alpha. \hspace{5cm} (8.1)
\end{aligned}$$

Next, we use the following change of variables

$$\frac{1}{n^\alpha} \int_0^{\hat{\mathbf{T}}^n} |\hat{\mu}_s^n| ds = \frac{1}{n^\alpha} \int_0^{n^{1-\alpha} \hat{\mathbf{T}}^n} |\hat{\mu}_{\bar{s}/n^{1-\alpha}}^n| \frac{1}{n^{1-\alpha}} d\bar{s} = \int_0^{\mathbf{T}^n} |\mu_{\bar{s}}^n| d\bar{s}. \hspace{2cm} (8.2)$$

Using the fact that there are at most ℓ_0 lineages remaining at time τ_{j_0}, we have

$$\begin{aligned}
\mathbb{E}\left(\int_0^{\mathbf{T}^n} |\mu_t^n| dt \right) &= \frac{1}{n^\alpha} \mathbb{E}\left(\int_0^{\hat{\mathbf{T}}^n} |\hat{\mu}_t^n| dt \right) \\
&\leq \frac{1}{n^\alpha} \mathbb{E}\left(\int_0^{\tau_{j_0}} |\hat{\mu}_t^n| dt \right) + \frac{1}{n^\alpha} \mathbb{E}\left(\int_{\tau_{j_0}}^{\hat{\mathbf{T}}^n} |\hat{\mu}_t^n| dt \right) \\
&\leq 2C + \frac{1}{n^\alpha} \mathbb{E}\left(\int_0^{\hat{\mathbf{T}}_{\ell_0}} |\hat{\mu}_t^{\ell_0}| dt \right) \\
&\leq 2C + \frac{\ell_0^\alpha}{n^\alpha} \mathbb{E}(\hat{\mathbf{T}}_{\ell_0}),
\end{aligned}$$

where in the last line we used (8.1). It remains to show that the expectation on the RHS is finite. In order to see that, we note that successive ℓ_0-events are separated by independent exponential r.v.'s with the same parameter $R(\ell_0) > 0$. Further, at any of those events, there is a strictly positive probability p to go from n lineages to a single lineage. By a simple coupling argument, one can bound from above the r.v. $\hat{\mathbf{T}}_{\ell_0}$ by $\sum_{i=1}^X e_i$ where the e_i's are i.d.d. exponential r.v.'s with parameter $R(\ell_0)$ and X is an independent geometric r.v. with parameter p. Since the upper bound has mean $\frac{1}{R(\ell_0)p} < \infty$, $\mathbb{E}(\hat{\mathbf{T}}_{\ell_0}) < \infty$. $\hspace{2cm} \square$

Proof of Proposition 8.1. We need to show that for every $A > 0$,

$$\limsup_{A \to \infty} \limsup_{n \to \infty} \mathbb{E}\left(\int_A^{\mathbf{T}_n} |\mu_s^n| ds\right) = 0, \tag{8.3}$$

where, implicitly, $\int_A^{\mathbf{T}_n} |\mu_s^n| ds = 0$ on the event $\{\mathbf{T}_n < A\}$ (which is highly unlikely in the limit); for every $d \in \mathbb{N}$,

$$\int_0^A \big(|\mu_s^n|, \mu_s^n(1), \ldots, \mu_s^n(d)\big) ds \Longrightarrow \int_0^A \big(|\mu_s|, \mu_s(1), \ldots, \mu_s(d)\big) ds, \tag{8.4}$$

and

$$\mathbb{E}\left(\int_0^\infty |\mu_s| ds\right) < \infty. \tag{8.5}$$

We start with (8.4). Let $K \in \mathbb{N}$. By monotonicity, we have

$$\int_0^A |\mu_s^n| ds \leq \frac{1}{K} \sum_{i=0}^{\lfloor AK \rfloor + 1} |\mu_{\frac{i}{K}}^n|.$$

Since the limiting process $(\mu_t; t \geq 0)$ has no fixed point of discontinuity (see Lemma 6.2), the RHS is converging to $K^{-1} \sum_{i=0}^{\lfloor AK \rfloor + 1} |\mu_{i/K}|$ for every $K \in \mathbb{N}$. As a consequence, $\lim \int_0^A |\mu_s^n| ds$ is bounded from above by $\int_0^A |\mu_s| ds$. A similar argument shows the reverse bound. This proves that $\lim_n \int_0^A |\mu_s^n| ds = \int_0^A |\mu_s| ds$. Next, for every $J, n \in \mathbb{N}$, the process $(\sum_{i=1}^J \mu_s^n(i); s \geq 0)$ is also monotone in s, the exact same argument shows convergence of $\sum_{i=1}^J \int_0^A \mu_s^n(i) ds$ to $\sum_{i=1}^J \int_0^A \mu_s(i) ds$. Finally, the proof of (8.4) is complete by noting that all the previous convergence statements hold jointly for every $J \in \mathbb{N}$.

Let us now proceed with the first property (8.3). The Markov property, together with the change of variables (8.2) implies that

$$\mathbb{E}\left(\int_A^{\mathbf{T}_n} |\mu_s^n| ds\right) = \frac{1}{n^\alpha} \sum_{j=1}^n \mathbb{P}\left(|\mu_A^n| = \frac{j}{n}\right) \mathbb{E}\left(\int_0^{\hat{\mathbf{T}}_j} |\hat{\mu}_s^j| ds\right)$$

$$\leq \frac{C}{n^\alpha} \sum_{j=1}^n j^\alpha \mathbb{P}\left(|\mu_A^n| = \frac{j}{n}\right)$$

where C is a constant and the inequality follows from (8.1). Let $\varepsilon \in (0, 1)$. Then

$$\mathbb{E}\left(\int_A^{\mathbf{T}_n} |\mu_s^n| ds\right) \leq \frac{C}{n^\alpha} \left(\sum_{j=\lfloor \varepsilon n \rfloor + 1}^n j^\alpha \mathbb{P}\left(|\mu_A^n| = \frac{j}{n}\right) + \sum_{j=1}^{\lfloor \varepsilon n \rfloor} j^\alpha \mathbb{P}\left(|\mu_A^n| = \frac{j}{n}\right)\right)$$

$$\leq C\big(\mathbb{P}\big(|\mu_A^n| \in [\varepsilon, 1]\big) + \varepsilon^\alpha\big).$$

For every arbitrary $\varepsilon \in (0, 1)$,

$$\limsup_{A \to \infty} \limsup_{n \to \infty} \mathbb{E}\left(\int_A^{\mathbf{T}_n} |\mu_s^n| ds\right) \leq C\left(\lim_{A \to \infty} \mathbb{P}\big(|\mu_A| \in [\varepsilon, 1]\big) + \varepsilon^\alpha\right)$$

$$= C\varepsilon^\alpha,$$

where we used the fact that $|\mu_t| \to 0$ as $t \to \infty$ (this can easily be proved from the Lamperti transform). This ends the proof of (8.3).

We now turn to (8.5). First, recall that $\mathbf{T}_n > \mathbf{T}_2 > 0$ then $\hat{\mathbf{T}}_n \to \infty$. Furthermore, the fact that $|\mu_t^n| \to |\mu_t|$ and Fatou's lemma yield

$$\mathbb{E}\left(\int_0^\infty |\mu_s| ds\right) \leq \lim_{n \to \infty} \mathbb{E}\left(\int_0^{\hat{\mathbf{T}}_n} |\mu_s^n| ds\right).$$

So, (8.5) follows from Lemma 8.2. $\qquad\square$

Proof of Theorem 2.4. By Proposition 8.1,

$$\int_0^{\mathbf{T}_n} \mu_s^n ds \Longrightarrow \int_0^\infty \mu_s ds \quad \text{in } \ell^1(\mathbb{R}_+).$$

The relation

$$\int_0^\infty \mu_s ds = \int_0^\infty \theta_{\tau_s} \exp(-\xi_{\tau_s}) ds = \int_0^\infty \theta_u \exp\big((\alpha - 2)\xi_u\big) du$$

follows from the change of variable $\tau_s = u$ and (7.4).

Given a realization of the coalescent, the number of segregating mutations in a branch sub-tending i leaves is given by a Poisson r.v. with parameter $r \int_0^{\hat{\mathbf{T}}_n} \hat{\mu}_s^n(i) ds$. The total number of mutations is given by a Poisson r.v. with parameter $r \int_0^{\hat{\mathbf{T}}_n} |\hat{\mu}_s^n| ds$. Using the change of variables (8.2),

$$\frac{r}{n^\alpha} \int_0^{\hat{\mathbf{T}}_n} \hat{\mu}_s^n ds = r \int_0^{n^{1-\alpha}\hat{\mathbf{T}}_n} \mu_{\bar{s}}^n d\bar{s} \Longrightarrow r \int_0^\infty \mu_{\bar{s}} d\bar{s},$$

which implies the desired result. $\qquad\square$

A Moment estimates of the partition mass components

In this section we provide some useful asymptotic results for the lengths of the unit mass partition when they are obtained from i.i.d. random variables with a finite second moment.

Proposition A.1. *Consider a family of i.i.d. random variables $(w_1, \ldots, w_k)$ on $(0, \infty)$ such that $\mathbb{E}(w_1^2) < \infty$. For $j \in [k]$, set $p_j^{(k)} = w_j / \sum_{i=1}^k w_i$ and $\bar{p}_j^{(k)} = w_j / k\mathbb{E}(w_1)$. Then we have the following asymptotics:*

(i) $\lim_{k\to\infty} kp_1^{(k)} = \Gamma := \frac{w_1}{\mathbb{E}(w_1)}$ *almost surely.*

(ii) $\lim_{k\to\infty} \mathbb{E}((kp_1^{(k)})^2) = \frac{\mathbb{E}(w_1^2)}{\mathbb{E}(w_1)^2}.$

(iii) $\lim_{k\to\infty} \mathbb{E}((k\bar{p}_1^{(k)} - kp_1^{(k)})^2) = 0.$

Proof. By the law of large numbers we obtain (i). The result in (ii) can be found in Lemma 4.3 of [17].

To prove (iii), we write

$$\mathbb{E}\big((k\bar{p}_1^{(k)} - kp_1^{(k)})^2\big) = \mathbb{E}\bigg(\frac{k^2 w_1^2}{(\sum_{i=1}^k w_i)^2}\bigg) - \mathbb{E}\bigg(\frac{2k w_1^2}{\mathbb{E}(w_1)\sum_{i=1}^k w_i}\bigg) + \mathbb{E}\bigg(\frac{w_1^2}{\mathbb{E}(w_1)^2}\bigg).$$

Using (ii), it is enough to prove that

$$\mathbb{E}\bigg(\frac{k w_1^2}{\mathbb{E}(w_1)\sum_{i=1}^k w_i}\bigg) \to \frac{\mathbb{E}(w_1^2)}{\mathbb{E}(w_1)^2}.$$

We follow the proof of Lemma 4.3 of [17]. Let $\delta_1 > 0$, then

$$\mathbb{E}\bigg(\frac{k w_1^2}{\mathbb{E}(w_1)\sum_{i=1}^k w_i}\bigg) \geq \mathbb{E}\bigg(\frac{w_1^2}{\delta_1 + \mathbb{E}(w_1)(k^{-1}\sum_{i=1}^k w_i)}\bigg).$$

Since $w_1 > 0$ almost surely, the bounded convergence theorem yields

$$\liminf_{k\to\infty} \mathbb{E}\bigg(\frac{k w_1^2}{\mathbb{E}(w_1)\sum_{i=1}^k w_i}\bigg) \geq \frac{\mathbb{E}(w_1^2)}{\delta_1 + \mathbb{E}(w_1)^2}.$$

The inequality holds for every $\delta_1 > 0$, which implies that the above limit is larger than $\mathbb{E}(w_1^2)/\mathbb{E}(w_1)^2$.

To prove an upper bound, we use the Cauchy-Schwarz inequality,

$$\mathbb{E}\left(\frac{kw_1^2}{\mathbb{E}(w_1)\sum_{i=1}^k w_i}\right) \leq \frac{1}{\mathbb{E}(w_1)}\sqrt{\mathbb{E}(w_1^2)}\sqrt{\mathbb{E}\left(\frac{w_1^2}{(k^{-1}\sum_{i=1}^k w_i)^2}\right)}$$

and the result follows from (ii). $\qquad\square$

Lemma A.2. *Let g_1, g_2 be two Lipschitz functions with respective Lipschitz constants $G_1, G_2 \geq 1$. Also suppose that $\|g_1\|_\infty, \|g_2\|_\infty \leq 1$. Define $p_1^{(k)}, p_2^{(k)}$ as in Proposition A.1. Then,*

$$\left|\mathrm{Cov}\left(g_1\left(kp_1^{(k)}\right), g_2\left(kp_2^{(k)}\right)\right)\right| \leq \max\{G_1, G_2, G_1 G_2\}h(k),$$

where h is a function of k, independent on g_1, g_2 going to 0 as $k \to \infty$.

Proof. Set $\bar{p}_1^{(k)} := \frac{w_1}{k\mathbb{E}(w_1)}$ and $\bar{p}_2^{(k)} := \frac{w_2}{k\mathbb{E}(w_2)}$. We have

$$\left|\mathrm{Cov}\left(g_1\left(kp_1^{(k)}\right), g_2\left(kp_2^{(k)}\right)\right)\right|$$
$$= \left|\mathrm{Cov}\left(g_1\left(k\bar{p}_1^{(k)}\right) + \left(g_1\left(kp_1^{(k)}\right) - g_1\left(k\bar{p}_1^{(k)}\right)\right), g_2\left(k\bar{p}_2^{(k)}\right) + \left(g_2\left(kp_2^{(k)}\right) - g_2\left(k\bar{p}_2^{(k)}\right)\right)\right)\right|$$
$$\leq \underbrace{\left|\mathrm{Cov}\left(g_1\left(kp_1^{(k)}\right) - g_1\left(k\bar{p}_1^{(k)}\right), g_2\left(kp_2^{(k)}\right) - g_2\left(k\bar{p}_2^{(k)}\right)\right)\right|}_{A}$$
$$+ \underbrace{\left|\mathrm{Cov}\left(g_1\left(k\bar{p}_1^{(k)}\right), g_2\left(kp_2^{(k)}\right) - g_2\left(k\bar{p}_2^{(k)}\right)\right)\right|}_{B} + \underbrace{\left|\mathrm{Cov}\left(g_2\left(k\bar{p}_2^{(k)}\right), g_1\left(kp_1^{(k)}\right) - g_1\left(k\bar{p}_1^{(k)}\right)\right)\right|}_{C}.$$

To bound A, observe that $p_1^{(k)}$ (resp. $\bar{p}_1^{(k)}$) has the same law as $p_2^{(k)}$ (resp. $\bar{p}_2^{(k)}$). So, by the Cauchy-Schwarz inequality,

$$A \leq \sqrt{\mathrm{Var}\left(g_1\left(kp_1^{(k)}\right) - g_1\left(k\bar{p}_1^{(k)}\right)\right)}\sqrt{\mathrm{Var}\left(g_2\left(kp_1^{(k)}\right) - g_2\left(k\bar{p}_1^{(k)}\right)\right)}$$
$$\leq \sqrt{\mathbb{E}\left(\left(g_1\left(kp_1^{(k)}\right) - g_1\left(k\bar{p}_1^{(k)}\right)\right)^2\right)}\sqrt{\mathbb{E}\left(\left(g_2\left(kp_1^{(k)}\right) - g_2\left(k\bar{p}_1^{(k)}\right)\right)^2\right)}$$
$$\leq G_1 G_2 \mathbb{E}\left(\left(k\bar{p}_1^{(k)} - kp_1^{(k)}\right)^2\right) = G_1 G_2 h(k),$$

where in the last line we used point (iii) of Proposition A.1.

Let us turn to B (the bound of C will be obtained in the same way).

$$B \leq \sqrt{\mathrm{Var}\left(g_1\left(k\bar{p}_1^{(k)}\right)\right)}\sqrt{\mathrm{Var}\left(g_2\left(kp_2^{(k)}\right) - g_2\left(k\bar{p}_2^{(k)}\right)\right)}$$
$$\leq \sqrt{\mathbb{E}\left(g_1\left(k\bar{p}_1^{(k)}\right)^2\right)}\sqrt{\mathbb{E}\left(\left(g_2\left(kp_2^{(k)}\right) - g_2\left(k\bar{p}_2^{(k)}\right)\right)^2\right)}$$
$$\leq \|g_1\|_\infty G_2 \sqrt{E\left(\left(k\bar{p}_2^{(k)} - kp_2^{(k)}\right)^2\right)} = G_2 h(k),$$

where in the last line we used point (iii) of Proposition A.1. $\qquad\square$

Lemma A.3. *For every partition c of ℓ, define the function $g_{v,c}$ on $\mathbb{R}_+$ by*

$$g_{v,c}(x) := x^{|c|}e^{-v(\ell+1)x}\prod_{i=1}^{\ell}\frac{v(i)^{c(i)}}{c(i)!}.$$

For every $c_1, c_2 \in \varphi^{-1}(\ell)$, there exists h such that for every $v = (v(1), \ldots, v(\ell), v(\ell+1)) \in \mathbb{R}_+^{\ell+1}$ with $v(\ell+1) \geq \sum_{i \leq \ell} v(i)$

$$\left|\mathrm{Cov}\left(g_{v,c_1}\left(kp_1^{(k)}\right), g_{v,c_2}\left(kp_2^{(k)}\right)\right)\right| \leq \max\{v(\ell+1), v(\ell+1)^2\}h(k),$$

where h is going to 0 as $k \to \infty$.

Proof. We use Lemma A.2. It is easy to see that $\|g_{v,c}\|_\infty \leq 1$. We just need to prove that the function $g_{v,c}$ is Lipshitz with constant $v(\ell+1)K$, for some K that only depends on c. To compute the Lipshitz constant, set $\alpha := v(\ell+1)$. Then

$$g'_{v,c}(x) = (|c| - \alpha x)x^{|c|-1}e^{-\alpha x}\prod_{i=1}^{\ell}\frac{v(i)^{c(i)}}{c(i)!}$$

and

$$g''_{v,c}(x) = \left(-\alpha x + (|c| - \alpha x)(|c| - 1 - \alpha x)\right)x^{|c|-2}e^{-\alpha x}\prod_{i=1}^{\ell}\frac{v(i)^{c(i)}}{c(i)!},$$

which switches sign when $|c|(|c| - 1) - 2|c|\alpha x + \alpha^2 x^2 = 0$, that is when $x = \frac{|c|+\epsilon\sqrt{|c|}}{\alpha}, \epsilon \in \{-1, 1\}$. Since $\lim_{x\to\infty} g'_{v,c}(x) = 0$ and $g'_{v,c}(0) = 0$, the maximum of $|g'_{v,c}|$ on $\mathbb{R}_+$ is attained for $\epsilon = +1$ or $\epsilon = -1$. Thus, using $v(i) \leq \alpha$, we have

$$\max_{x\in\mathbb{R}_+}|g'_{v,c}(x)| = \max_{-1,+1}\sqrt{|c|}(|c| + \epsilon\sqrt{|c|})^{|c|-1}e^{-(|c|+\epsilon\sqrt{|c|})}\frac{1}{\alpha^{|c|-1}}\prod_{i=1}^{\ell}\frac{v(i)^{c(i)}}{c(i)!}$$

$$\leq \alpha \max_{-1,+1}e^{-(|c|+\epsilon\sqrt{|c|})}\sqrt{|c|}(|c| + \epsilon\sqrt{|c|})^{|c|-1} \times \frac{1}{c!},$$

which completes the proof. $\square$

B Derivatives for Proposition 5.1

In this section, the values of x and $\mathbf{z}$ will remain fixed and for the sake of clarity, we will forget the dependence in those two constants in the next results.

Lemma B.1. *Fix* $x > 0$, $\delta \in (0, 1)$ *and* $\mathbf{z} \in \ell^1(\mathbb{R}_+)$. *For every* $\lambda \in [0, \delta]$, *define*

$$\rho(\lambda) := \mathbb{E}\big(x\exp(-|\mathbf{z}|\Gamma/x)\big(\exp\big(\psi_\lambda(\mathbf{z})\Gamma/x\big) - 1\big)\big).$$

Then ρ *is infinitely differentiable on* $[0, \delta]$ *and for every* $\ell \in \mathbb{N}$, *its* ℓ^{th} *derivative is such that*

$$\frac{1}{\ell!}\rho^{(\ell)}(\lambda) = \sum_{c\in\varphi^{-1}(\ell)}\mathbb{E}\left(x\exp(-|\mathbf{z}|\Gamma/x)\prod_{i=1}^{\ell}\frac{(\Gamma/x)^{c(i)}}{c(i)!}\right)\prod_{i=1}^{\ell}\left(\sum_{k=i}^{\infty}\binom{k}{i}z(k)\lambda^{k-i}\right)^{c(i)} \quad \text{(B.1)}$$

with the convention $0^0 = 1$.

Proof. Let f, g be infinitely differentiable on their respective spaces. Faà di Bruno's formula states that

$$\frac{1}{\ell!}\frac{d^\ell}{d\lambda^\ell}(f \circ g)(\lambda) = \sum_{c\in\varphi^{-1}(\ell)}f^{(|c|)}\big(g(\lambda)\big)\prod_{i=1}^{\ell}\frac{1}{c(i)!}\left(\frac{g^{(i)}(\lambda)}{i!}\right)^{c(i)}. \quad \text{(B.2)}$$

Let $f(y) = e^y - 1$ and $g(\lambda) = \psi_\lambda(\mathbf{z})\Gamma/x$ for some fixed value of Γ. Since $\mathbf{z} \in \ell^1(\mathbb{R}_+)$, the function g is a power series with a radius of convergence larger or equal to 1. Thus g is infinitely differentiable on $[0, \delta]$ and

$$g^{(i)}(\lambda) = \frac{\Gamma}{x}\sum_{k=i}^{\infty}(k)_i z(k)\lambda^{k-i},$$

where $(k)_i = k \ldots (k-i+1)$. For our choice of f and g, (B.2) translates into the following formula

$$\frac{1}{\ell!}\frac{d^\ell}{d\lambda^\ell}(f\circ g)(\lambda)$$

$$= \sum_{c\in\varphi^{-1}(\ell)}\left(x\exp\big((\psi_\lambda(\mathbf{z})-|\mathbf{z}|)\Gamma/x\big)\prod_{i=1}^\ell\frac{(\Gamma/x)^{c(i)}}{c(i)!}\right)\prod_{i=1}^\ell\left(\sum_{k=i}^\infty\binom{k}{i}z(k)\lambda^{k-i}\right)^{c(i)}.$$

Recall that $\rho(\lambda) = \mathbb{E}(f\circ g(\lambda))$. In order to complete the proof, we need to justify the derivation inside the expected value. Since the coefficients of $\mathbf{z}$ are non-negative then $\psi_\lambda(\mathbf{z})-|\mathbf{z}|<0$ for every $\lambda\in[0,1)$. Furthermore, if $\lambda\in[0,\delta]$, then

$$\psi_\lambda(\mathbf{z})-|\mathbf{z}| \leq \psi_\delta(\mathbf{z})-|\mathbf{z}| =: -b.$$

As a consequence,

$$\frac{1}{\ell!}\frac{d^\ell}{d\lambda^\ell}(f\circ g)(\lambda) \leq \sum_{c\in\varphi^{-1}(\ell)}\left(x\exp(-b\Gamma/x)\prod_{i=1}^\ell\frac{(\Gamma/x)^{c(i)}}{c(i)!}\right)\prod_{i=1}^\ell\left(\sum_{k=i}^\infty\binom{k}{i}z(k)\delta^{k-i}\right)^{c(i)}.$$

The right hand side of the inequality does not depend on λ, and one can check that it has a finite expectation. Since $\rho(\lambda)=\mathbb{E}(f\circ g(\lambda))$ (where the expectation is taken w.r.t. Γ), by standard derivation theorem under the integral, $\rho(\lambda)$ is infinitely differentiable and the derivatives can be calculated by differentiating inside the expectation. $\square$

Lemma B.2. *Let $\rho(\lambda)$ be the function introduced in Lemma B.1. For every $\ell\in\mathbb{N}$ and $\lambda\in[0,\delta]$, with $\delta<1$, we have*

$$\frac{1}{\ell!}|\rho^{(\ell)}(\lambda)| < x\frac{(2-\delta)^{\ell-1}}{(1-\delta)^{2\ell}}.$$

Proof. Fix $\ell\in\mathbb{N}$. We start by recalling that for $\lambda\in[0,\delta]$,

$$\sum_{k=i}^\infty\binom{k}{i}\lambda^{k-i} = \frac{1}{(1-\lambda)^{i+1}} \leq \frac{1}{(1-\delta)^{i+1}}.$$

Observe that if $c\in\varphi^{-1}(\ell)$, $\sum ic(i)=\ell$, so

$$\prod_{i=1}^\ell\left(\sum_{k=i}^\infty\binom{k}{i}z(k)\lambda^{k-i}\right)^{c(i)} \leq \prod_{i=1}^\ell|\mathbf{z}|^{c(i)}\left(\frac{1}{(1-\delta)^{i+1}}\right)^{c(i)}$$

$$\leq \frac{|\mathbf{z}|^{|c|}}{(1-\delta)^{\ell+|c|}}.$$

Finally, replacing in (B.1),

$$\frac{1}{\ell!}\rho^{(\ell)}(\lambda) = \sum_{j=1}^\ell\sum_{\substack{c\in\varphi^{-1}(\ell)\\|c|=j}}\mathbb{E}\left(x\exp(-|\mathbf{z}|\Gamma/x)\prod_{i=1}^\ell\frac{(\Gamma/x)^{c(i)}}{c(i)!}\right)\prod_{i=1}^\ell\left(\sum_{k=i}^\infty\binom{k}{i}z(k)\lambda^{k-i}\right)^{c(i)}$$

$$\leq \frac{x}{(1-\delta)^\ell}\sum_{j=1}^\ell\mathbb{E}\big(\exp(-|\mathbf{z}|\Gamma/x)\big(|\mathbf{z}|\Gamma/x(1-\delta)\big)^j\big)\sum_{\substack{c\in\varphi^{-1}(\ell)\\|c|=j}}\prod_{i=1}^\ell\frac{1}{c(i)!}$$

$$= \frac{x}{(1-\delta)^\ell}\sum_{j=1}^\ell\mathbb{E}\left(\exp(-|\mathbf{z}|\Gamma/x)\frac{(|\mathbf{z}|\Gamma/x(1-\delta))^j}{j!}\right)\binom{\ell-1}{j-1}$$

$$\leq \frac{x}{(1-\delta)^\ell}\sum_{j=1}^\ell\binom{\ell-1}{j-1}\frac{1}{(1-\delta)^j} = x\frac{(2-\delta)^{\ell-1}}{(1-\delta)^{2\ell}}. \qquad\square$$

References

[1] L. Alili, L. Chaumont, P. Garczyk and T. Żak. Inversion, duality and Doob h-transforms for self-similar Markov processes. *Electron. J. Probab.*, **22**:20, 2017. MR3622890

[2] A.L. Basdevant and C. Goldschmidt. Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent. *Electron. J. Probab.*, **13**, 486–512, 2008. MR2386740

[3] R.F. Bass. Stochastic differential equations with jumps. *Probab. Surveys*, **1**, 1–19, 2004. MR2095564

[4] J. Berestycki, N. Berestycki and V. Limic. The Λ-coalescent speed of coming down from infinity. *Ann. Probab.*, **38**:1, 207–233, 2010. MR2599198

[5] J. Berestycki, N. Berestycki and J. Schweinsberg. Small-time behavior of β coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.*, **44**:2, 214—238, 2008. MR2446321

[6] J. Bertoin. Random fragmentation and coagulation processes. Cambridge University Press, 2006. MR2253162

[7] J. Bertoin and M.E. Caballero, Entrance from $0+$ for increasing semi-stable Markov processes. *Bernoulli*, **8**, 2, 195–205, 2002. MR1895890

[8] J. Bertoin and J.F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields*, **126**:2, 261–288, 2003. MR1990057

[9] J. Bertoin and J.F. Le Gall. Stochastic flows associated to coalescent processes II. *Annales de l'I.H.P. Probabilités et statistiques*, **21**:3, 307–333, 2005. MR2139022

[10] J. Bertoin and M. Yor. The entrance laws of self-similar Markov processes and exponential functionals of Lévy processes. *Potential Anal.*, **17**:4, 389–400, 2002. MR1918243

[11] J. Bertoin and M. Yor. Exponential functionals of Lévy processes., *Probab. Surveys* **2**, 191–212, 2005. MR2178044

[12] P. Billingsley. Convergence of probability measures. Wiley, 1999. MR1700749

[13] M. Birkner, I. Dahmer, C.S. Diehl and G. Kersting. The joint fluctuations of the lengths of the Beta$(2 - \alpha, \alpha)$-coalescents. arXiv:2009.13642.

[14] J. Blath, M.C. Cronjäger, B. Eldon and M. Hammer. The site-frequency spectrum associated with Ξ-coalescents. *Theor. Popul. Biol.*, **110**, 36–50, 2016.

[15] E. Çinlar. Markov additive processes. I, II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **24**, 1972. MR0329047

[16] E. Cinlar, Entrance-exit distributions for Markov additive processes. *Math. Programming Stud.*, **15**, 22–38, 1976. MR0445621

[17] A. Cortines. The genealogy of a solvable population model under selection with dynamics related to directed polymers. *Bernoulli.*, **22**:4, 2209–2236, 2016. MR3498028

[18] I. Dahmer and G. Kersting. The internal branch lengths of the Kingman coalescent. *Ann. Appl. Prob.*, **25**, 2015. MR3325275

[19] S. Dereich, L. Döring and A.E. Kyprianou. Real self-similar processes started from the origin, *Ann. Probab.*, **45**:3, 1952–2003, 2017. MR3650419

[20] C.S. Diehl and G. Kersting. Tree lengths for general Λ-coalescents and the asymptotic site frequency spectrum around the Bolthausen-Sznitman coalescent. *Ann. Appl. Prob.*, **29**:5, 2700–2743, 2019. MR4019873

[21] S. Ethier and T. G. Kurtz. Markov Processes: Characterization and Convergence, 1986. MR0838085

[22] Y.X. Fu and W.H. Li. Statistical tests of neutrality of mutations. *Genetics*, **133**:3, 693–709, 1993.

[23] F. Gaiser and M. Möhle. On the block-counting process and the fixation line of exchangeable coalescents. *ALEA, Lat. Am. J. Probab. Math. Stat.*, **13**, 809–833, 2016. MR3546382

[24] A. González Casanova, V. Miró Pina and A. Siri-Jégousse. The symmetric coalescent and Wright-Fisher models with bottlenecks. *Ann. Appl. Probab.*, **32**:1, 235–268, 2022. MR4386526

[25] B. Haas and G. Miermont. Self-similar scaling limits of non-increasing Markov chains. *Bernoulli*, **17**:4, 1217–1247, 2011. MR2854770

[26] A. Hobolth, A. Siri-Jégousse and M. Bladt. Phase-type distributions in population genetics. *Theor. Pop. Biol.*, **127**, 16–32, 2019.

[27] T. Huillet and M. Möhle. Asymptotic genealogies for a class of generalized Wright-Fisher models. *Modern Stoch. Theory Appl.*, **9**:1, 17–43, 2021. MR4388708

[28] G. Kersting, J.C. Pardo and A. Siri-Jégousse. Total internal and external lengths of the Bolthausen-Sznitman coalescent. *J. Appl. Probab.*, **51A**, 73–86, 2014. MR3317351

[29] G. Kersting, A. Siri-Jégousse and A. H. Wences. Site frequency spectrum of the Bolthausen-Sznitman coalescent. *ALEA, Lat. Am. J. Probab. Math. Stat.*, **18**, 1483–1505, 2021. MR4282195

[30] S.W. Kiu. Semistable Markov processes in $\mathbb{R}^n$. *Stochastic Process. Appl.*, **10**:2, 183–191, 1980. MR0587423

[31] A.E. Kyprianou, V. Rivero, B. Sengul and T. Yang. Entrance laws at the origin of self-similar Markov processes in high dimensions. *Trans. Amer. Math. Soc.*, **373**, 6227–6299, 2020. MR4155177

[32] J. Lamperti. Semi-stable stochastic processes. *Trans. Am. Math. Soc.*, **104**:1, 62–78, 1962. MR0138128

[33] J. Lamperti. Semi-stable Markov processes. I. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **22**, 205–225, 1972. MR0307358

[34] V. Limic. On the speed of coming down from infinity for Ξ-coalescent processes. *Electron. J. Probab.*, **15**:8, 2010. MR2594877

[35] M. Möhle. Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent. *Stochastic Process. Appl.*, **120**, 2159–2173, 2010. MR2684740

[36] P.A. Meyer. Un cours sur les intégrales stochastiques, in: Séminaire de Probabilit'es X, Lecture Notes Math., **511**, Springer, Berlin, 1976. MR0448511

[37] M. Möhle. The rate of convergence of the block counting process of exchangeable coalescents with dust. *ALEA, Lat. Am. J. Probab. Math. Stat.*, **18**, 1195–1220, 2021. MR4282186

[38] M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, **29**, 1547–1562, 2001. MR1880231

[39] B. Roos. Improvements in the Poisson approximation of mixed Poisson distributions. *J. Stat. Plan. Inference*, **8**, 210–293, 2003. MR1965122

[40] N. Ross. Fundamentals of Stein's method. *Probab. Surveys*, **113**, 467–483, 2011. MR2861132

[41] J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, **5**:12, 2000. MR1781024

[42] J.P. Spence, J.A. Kamm and Y.S. Song. The site frequency spectrum for general coalescents. *Genetics*, **202**:4, 1549–1561, 2016.

[43] A.V. Skorokhod. Limit theorems for stochastic processes. *Th. Probab. Appl.*, **1**, 261–290, 1956. MR0084897

[44] A.H. Wences and A. Siri-Jégousse. Exchangeable coalescents beyond the Cannings class. arXiv:2212.02154, 2022.

[45] W. Whitt. Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues. Springer, 2002. MR1876437

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

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

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

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