ASYMPTOTIC GAUSSIANITY VIA COALESCENCE PROBABILITIES IN THE HAMMOND-SHEFFIELD URN

JAN LUKAS IGELBRINK AND ANTON WAKOLBINGER

ABSTRACT. For the renormalised sums of the random ± 1-colouring of the connected components of $\mathbb{Z}$ generated by the coalescing renewal processes in the "power law Pólya's urn" of Hammond and Sheffield [HS13] we prove functional convergence towards fractional Brownian motion, closing a gap in the tightness argument of their paper.

In addition, in the regime of the strong renewal theorem we gain insights into the coalescing renewal processes in the Hammond-Sheffield urn (such as the asymptotic depth of most recent common ancestors) and are able to control the coalescence probabilities of two, three and four individuals that are randomly sampled from $[n]$. This allows us to obtain a new, conceptual proof of the asymptotic Gaussianity (including the functional convergence) of the renormalised sums of more general colourings, which can be seen as an invariance principle beyond the main result of [HS13].

In this proof, a key ingredient of independent interest is a sufficient criterion for the asymptotic Gaussianity of the renormalised sums in randomly coloured random partitions of $[n]$, based on Stein's method.

Along the way we also prove a statement on the asymptotics of the coalescence probabilities in the long-range seedbank model of Blath, González Casanova, Kurt, and Spanò, see [BGKS13].

1

Contents

1. Introduction

We start with a brief description of the model of [HS13] and then state our main results together with a short outline of the paper.

For $0 < \alpha < \frac{1}{2}$ and a slowly varying function $L : \mathbb{R} \to \mathbb{R}^+$ let $\mu := \mu_{\alpha,L}$ be a probability measure on $\mathbb{N} = \{1, 2, \ldots\}$ having the power law tails

$$\mu\left(\{n, n+1, \ldots\}\right) \sim n^{-\alpha} L(n) \text{ as } n \to \infty, \tag{1.1}$$

with the usual convention that for two sequences $f(n)$, $g(n)$ of real numbers

$$f(n) \sim g(n) \text{ as } n \to \infty$$

means that $\lim_{n\to\infty} f(n)/g(n) = 1$. Throughout it will be assumed that

$$\text{the greatest common divisor of } \left\{n \in \mathbb{N} : \mu(n) > 0\right\} \text{ is one.} \tag{1.2}$$

Let R be an $\mathbb{N}$-valued random variable with distribution μ. A random directed graph G_μ with vertex set $\mathbb{Z}$ is generated in the following way: Let $(R_i)_{i\in\mathbb{Z}}$ be a family of independent copies of R. The random set of edges $E\left(\mathcal{G}_\mu\right)$ is then given by

$$E\left(\mathcal{G}_\mu\right) := \left\{(i, i - R_i) : i \in \mathbb{Z}\right\}.$$

This induces the random equivalence relation

$$i \sim j :\Longleftrightarrow i \text{ and } j \text{ belong to the same connected component of } \mathcal{G}_\mu. \tag{1.3}$$

Note that the symbol $\sim$ is used in (1.1) and (1.3) in two different meanings; this will cause no risk of confusion.

For $i \in \mathbb{Z}$ the connected component containing i is denoted by $\mathscr{C}_i$. The random variables $(R_i)_{i\in\mathbb{Z}}$ give rise to *coalescing renewal processes* starting from the integers; see Section 10 for an interpretation (and extension) in terms of the long-range seedbank model of [BGKS13]. In this terminology $\mathcal{G}_\mu$ is the graph of ancestral lineages of the individuals $i \in \mathbb{Z}$, and the component $\mathscr{C}_i$ consists of all $j \in \mathbb{Z}$ that are related to i, see Figure 1 for an illustration. The probability that 0 belongs to the ancestral lineage of n is thus given by the weight assigned to n by the renewal measure,

$$q_n := \mathbf{P}\left(\tilde{R}_1 + \ldots + \tilde{R}_j = n \text{ for some } j \geq 0\right) \tag{1.4}$$

with $\tilde{R}_1, \tilde{R}_2, \ldots$ being independent copies of R. (Note that $\mathbf{P}(0 \sim n)$ is in general larger than q_n because 0 and n may be related to each other even if 0 is not an ancestor of n.)
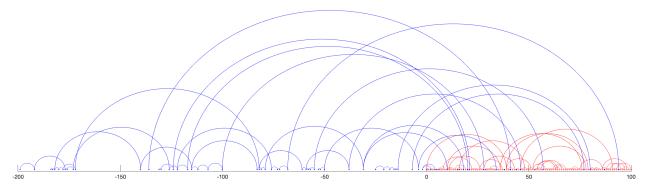
FIGURE 1. A realisation of the ancestral lineages of the individuals $\{0, \ldots, 100\}$ traced back till -200. Each of the arcs corresponds to an edge of G_μ. All the outgoing edges from $i = 0, \ldots, 100$ which map to an individual in $\{0, \ldots, 100\}$ are drawn (in red), whereas for i between -200 and -1 only those outgoing edges are drawn (in blue) that belong to an ancestral lineage of some $j \in \{0, \ldots, 100\}$. Here the exponent α in (1.1) was chosen as 0.39.

Hammond and Sheffield suggest the picture of an urn in which the types of the individuals i are determined recursively: each individual $i \in \mathbb{Z}$ inherits the type (or "colour") of its parent $i - R_i$. With $\{+1, -1\}$ as the set of colours, they show that the set of random colourings of $\mathbb{Z}$ that are consistent with G_μ has a Gibbs structure, with the extremal elements being given by i.i.d. assignments of colours to the connected components of G_μ. The main result of [HS13] concerns the asymptotics of the rescaled sum over the types of the individuals $1, \ldots, \lfloor tn \rfloor$, $t \geq 0$, which as $n \to \infty$ turns out to converge to fractional Brownian motion. The individuals' types arise as follows:

Assume that each component of G_μ gets coloured by an independent copy of a real-valued random variable Y. In the situation of [HS13], Y is a centered Rademacher(p) variable, i.e.

$$Y = \xi - (2p - 1) \text{ with } \mathbf{P}(\xi = +1) = p, \mathbf{P}(\xi = -1) = 1 - p. \tag{1.5}$$

For $i \in \mathbb{Z}$ the colour of the component $\mathscr{C}_i$ will be denoted by Y_i. Define the "random walk" (with dependent increments)

$$S_n := \sum_{i=1}^{n} Y_i, \quad n = 0, 1, \ldots. \tag{1.6}$$

By construction,

$$\sigma_n^2 := \mathbf{Var}[S_n] = \sum_{i,j \in [n]} \mathbf{Cov}[Y_i, Y_j] = \mathbf{E}[Y^2] \sum_{i,j \in [n]} \mathbf{P}(i \sim j). \tag{1.7}$$

[HS13, Lemma 3.1] show by Fourier and Tauberian arguments that

$$\sum_{i,j \in [n]} \mathbf{P}(i \sim j) \sim \frac{C_\alpha}{\alpha(2\alpha + 1)} \frac{n^{2\alpha+1}}{L(n)^2} \quad \text{as } n \to \infty, \tag{1.8}$$

with

$$C_\alpha := \frac{1}{\sum_{m \geq 0} q_m^2} \frac{\Gamma(1 - 2\alpha)}{\Gamma(\alpha)\Gamma(1 - \alpha)^3}. \tag{1.9}$$

We will obtain (1.8) as a corollary of Proposition 2.1 below, which requires the additional condition

$$q_n \sim \frac{1}{\Gamma(\alpha)\Gamma(1 - \alpha)} \frac{n^{\alpha-1}}{L(n)} \quad \text{as } n \to \infty. \tag{1.10}$$

This condition, which also appears in our Theorem 1.1, is equivalent to the validity of the Strong Renewal Theorem for the renewal process with an increment distribution μ satisfying (1.1) and

(1.2), see [CD19], whose Theorem 1.4 gives necessary and sufficient conditions in terms of μ for the validity of (1.10). A well-known sufficient condition for (1.10) is the criterion of [Don97]

$$\sup_{n \geq 1} \frac{n \, \mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty. \tag{1.11}$$

For $\frac{i-1}{n} \leq t \leq \frac{i}{n}$, $i, n \in \mathbb{N}$, let $S^{(n)}(t)$ be the linear interpolation of S_i/σ_n and S_{i+1}/σ_n. Because of (1.7) and (1.8), for all $t \geq 0$,

$$\mathbf{Var}\left[S^{(n)}(t)\right] \to t^{2\alpha+1} \quad \text{as } n \to \infty.$$

Since $(S_n)_{n \in \mathbb{N}_0}$ has stationary increments by construction, this implies the convergence

$$\mathbf{Cov}\left(S_s^{(n)}, S_t^{(n)}\right) \xrightarrow[n \to \infty]{} \frac{1}{2}\left(s^{2\alpha+1} + t^{2\alpha+1} - |t-s|^{2\alpha+1}\right), \quad s, t \geq 0.$$

The right-hand side is the covariance function of *fractional Brownian motion with Hurst parameter* $H = \frac{1}{2} + \alpha$, which is the unique centered Gaussian process with variance function t^{2H}, $t \geq 0$, stationary increments and a.s. continuous paths. The processes $S^{(n)}$ are centered as well. Thus, in order to prove that $S^{(n)}$ converges as $n \to \infty$ (in the sense of finite dimensional distributions) to fractional Brownian motion with Hurst parameter H, it only remains to show that the finite dimensional distributions of $S^{(n)}$ are asymptotically Gaussian. This is provided by

Theorem 1.1. *Let μ be a probability measure on $\mathbb{N}$ satisfying (1.1) and (1.2). Assume one of the following conditions (A) or (B):*

(A) *The colouring Y is given by (1.5).*
(B) *The weights q_n of the renewal measure specified in (1.4) satisfy the asymptotics (1.10), and the colouring Y obeys*

$$\mathbf{E}\left[Y\right] = 0 \text{ and } 0 < \mathbf{E}\left[Y^4\right] < \infty. \tag{1.12}$$

Then, for any fixed $d \in \mathbb{N}$ and fixed $0 < t_1 < \cdots < t_d < \infty$, the sequence $\left(S_{\lfloor t_1 n \rfloor}, \ldots, S_{\lfloor t_d n \rfloor}\right)_{n \in \mathbb{N}}$ is asymptotically Gaussian as $n \to \infty$.

Under assumption (A) of Theorem 1.1, for each fixed $t > 0$ asymptotic Gaussianity of $S_{\lfloor tn \rfloor}$ as $n \to \infty$ is proved in [HS13] via a martingale central limit theorem. The computations which ensure the applicability of the martingale CLT are quite subtle and involved, making substantial use of the specific form (1.5) of the colouring of the random graph $\mathcal{G}_\mu$. In [HS13] it is not explicitly discussed whether these arguments also carry over to the joint asymptotic Gaussianity of $S_{\lfloor t_1 n \rfloor}, \ldots, S_{\lfloor t_m n \rfloor}$ for fixed $t_1 < \cdots < t_m$. However, by applying the martingale CLT to linear combinations of these random variables one can check that this is indeed the case.

Under assumption (B) we give a new, conceptual proof of the asymptotic Gaussianity of the finite dimensional distributions of $S^{(n)}$. This proof, which is completed in Section 8, is based on insights into the structure of $\mathcal{G}_\mu$ which are stated in Section 2 and proved in Sections 4-7. A key ingredient in the new proof is Theorem 3.1, which provides a criterion for the asymptotic Gaussianity in randomly coloured random partitions also in a more general setting. Proposition 3.3, which is instrumental in the proof of Theorem 3.1, is based on Stein's method and yields the closeness of the distribution of S_n/σ_n to the standard normal distribution in terms of a bound that involves $\mathbf{Var}\left[Y^2\right]$; this explains the finiteness condition of $\mathbf{E}\left[Y^4\right]$ in (1.12).

Let us also mention that the loss of ground which comes with assuming the "strong renewal" condition (1.10) in addition to (1.1) and (1.2) seems rather minor. Indeed it becomes clear from the examples in [CD19, Section 10] that the class of measures μ which satisfy (1.1) and (1.2) but fail to satisfy (1.10) is rather special.

On the other hand, the benefit of assuming (1.10) is twofold. Firstly, it allows a direct analysis of asymptotic properties of the genealogy of the coalescing renewal processes in the Hammond-Sheffield urn, see Propositions 2.1, 2.3 and 2.5. Secondly, this opens the way to a two-step analysis (first of the random partition of $\mathbb{Z}$, then of its random colouring) which allows to derive the "invariance principle" stated in Theorem 1.1(B).

The following implication of Theorem 1.1 is immediate from its introductory discussion.

Corollary 1.2. *Under assumptions (A) or (B), $S^{(n)}$ converges as $n \to \infty$ in the sense of finite dimensional distributions to fractional Brownian motion with Hurst parameter $H = \frac{1}{2} + \alpha$.*

The next result, which will be proved in Section 9, amends the proof of [HS13, Lemma 4.1], see Remark 9.2. Here, for each $n \in \mathbb{N}$ and $T > 0$, $\left(S^{(n)}(t) \right)_{0 \leq t \leq T}$ is viewed as a random variable taking its values in $C\left([0, T]; \mathbb{R}\right)$, the space of continuous functions from $[0, T]$ to $\mathbb{R}$, equipped with the sup-norm.

Proposition 1.3. *Under the assumptions of Theorem 1.1, for all $T > 0$ the sequence of random variables $\left(S^{(n)}(t) \right)_{0 \leq t \leq T}$ is tight.*

A direct consequence of Corollary 1.2 and Proposition 1.3 is

Corollary 1.4. *Under the assumptions of Theorem 1.1, $S^{(n)}$ converges in distribution (with respect to the topology of locally uniform convergence) to fractional Brownian motion with Hurst parameter $H = \frac{1}{2} + \alpha$.*

2. Coalescence probabilities in the Hammond-Sheffield urn

In this section we will assume that the weights q_n of the renewal measure defined in (1.4) obey the asymptotics (1.10), see the discussion of this condition in Section 1.

Proposition 2.1. *The coalescence probabilities for the ancestral lineages obey the asymptotics*

$$\mathbf{P}(0 \sim i) \sim C_\alpha \frac{i^{2\alpha - 1}}{L(i)^2} \quad \text{as } i \to \infty, \tag{2.1}$$

with C_α as in (1.9).

Remark 2.2. (a) *The asymptotics (1.8) is a direct consequence of (2.1). Indeed, the latter implies*

$$\sum_{i \in [n]} (n - i) \mathbf{P}(i \sim 0) \sim n^{2\alpha + 1} \frac{C_\alpha}{L(n)^2} \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{i}{n} \right) \left(\frac{i}{n} \right)^{2\alpha - 1} \quad \text{as } n \to \infty, \tag{2.2}$$

with the limit of the Riemann sums being

$$\int_0^1 (1 - x) x^{2\alpha - 1} \mathrm{d}x = \frac{1}{2\alpha(2\alpha + 1)}.$$

Since the left-hand sides of (2.2) and (1.8) are equal, this shows the asserted implication.

(b) *In the light of the proof of Proposition 2.1 (carried out in Section 4) we conjecture that increment distributions μ that satisfy (1.1) and violate (1.10), generically also do not admit the asymptotics (2.1).*

The next result, Proposition 2.3, will be instrumental in the proof of Theorem 1.1 under Assumption (B). This proposition will consider the probability that three (respectively four) individuals that are randomly chosen from $[n]$ belong to the same component of $\mathcal{G}_\mu$.

Proposition 2.3. *Let $\mathscr{I}^{(n)}$, $\mathscr{J}^{(n)}$, $\mathscr{K}^{(n)}$ and $\mathscr{L}^{(n)}$ be independent and uniformly distributed on $[n]$, and independent of the random graph $\mathcal{G}_\mu$. Then for all $\delta > 0$, as $n \to \infty$,*

$$\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)} \sim \mathscr{K}^{(n)}\right) = O\left(n^{4\alpha-2+\delta}\right), \tag{2.3}$$

$$\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)} \sim \mathscr{K}^{(n)} \sim \mathscr{L}^{(n)}\right) = O\left(n^{6\alpha-3+\delta}\right). \tag{2.4}$$

Proposition 2.3 will be proved in Section 6. The following corollary will bound the triplet and quartet coalescence probabilities addressed in (2.3) and (2.4) asymptotically as $n \to \infty$ by powers of the pair coalescence probability. Estimates of this kind will be required in Theorem 3.2; note that for sufficiently small ε the powers guaranteed by Corollary 2.4 are strictly larger than those required in Theorem 3.2.

Corollary 2.4. *Let $\mathscr{I}^{(n)}$, $\mathscr{J}^{(n)}$, $\mathscr{K}^{(n)}$ and $\mathscr{L}^{(n)}$ be as in Proposition (2.3). Then for all $\varepsilon > 0$, as $n \to \infty$,*

$$\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)} \sim \mathscr{K}^{(n)}\right) = O\left(\left(\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)}\right)\right)^{2-\varepsilon}\right), \tag{2.5}$$

$$\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)} \sim \mathscr{K}^{(n)} \sim \mathscr{L}^{(n)}\right) = O\left(\left(\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)}\right)\right)^{3-\varepsilon}\right). \tag{2.6}$$

The proof of Corollary 2.4 is immediate from Proposition 2.3 together with (1.8). Indeed, (1.8) asserts that the order of $\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)}\right)$ is $\frac{n^{2\alpha-1}}{L(n)^2}$ as $n \to \infty$. With ε prescribed as in Corollary 2.4, it thus suffices to choose in Proposition 2.3 a positive δ that is smaller than $\varepsilon(1-2\alpha)$.

Although the next result, Proposition 2.5, will not be used explicitly in the proof of Theorem 1.1, it seems interesting in its own right and also gives an intuition why the estimates in Corollary 2.4 should hold. Qualitatively, Proposition 2.5 says that for large n the ancestral lineages of 0 and n with high probability either coalesce quickly (i.e. *on the scale n*) or never. This makes it believable that, as asserted in Corollary 2.4, the triplet coalescence probability should asymptotically be comparable to the square of the pair coalescence probability, and that the quartet coalescence probability should roughly be equal to the third power of the pair coalescence probability.

Proposition 2.5. *Let $\mathscr{M}(0,n) := \max\{j \le 0 : j \sim 0 \text{ and } j \sim n\}$ (with $\max \emptyset := -\infty$) be the most recent common ancestor of 0 and n, and put $D_n := -\mathscr{M}(0,n)$. Then, as $n \to \infty$, the sequence of random variables $\frac{D_n}{n}$, conditioned under $\{0 \sim n\}$, converges in distribution to the random variable D with density $B(\alpha, 1-2\alpha)^{-1}x^{\alpha-1}(1+x)^{\alpha-1}\,dx$, $x > 0$.*

Proposition 2.5 will be proved in Section 5. The distribution of the random variable D appearing in Proposition 2.5 is known as Beta prime distribution with parameters α and $1-2\alpha$; it arises as the distribution of $B/(1-B)$ where B is Beta$(\alpha, 1-2\alpha)$ distributed. See Figure 2 for simulations of the ancestral lineages, which also illustrate the depths of the most recent common ancestors.

The following lemma will also be important in the proof of Theorem 1.1 under Assumption (B).

Lemma 2.6. *Let $\sim$ be the random equivalence relation defined in (1.3). For $i, j, k, \ell \in \mathbb{Z}$,*

$$\mathbf{Cov}\left[I_{\{i \sim j\}}, I_{\{k \sim \ell\}}\right] \le \mathbf{P}\left(i \sim j \sim k \sim \ell\right). \tag{2.7}$$

Here and below, I_E denotes the indicator variable of an event E.

Remark 2.7. *The proof of Lemma 2.6 (given in Section 7) shows that (2.7) holds for general increment distributions μ, without the assumptions (1.1) and (1.2).*

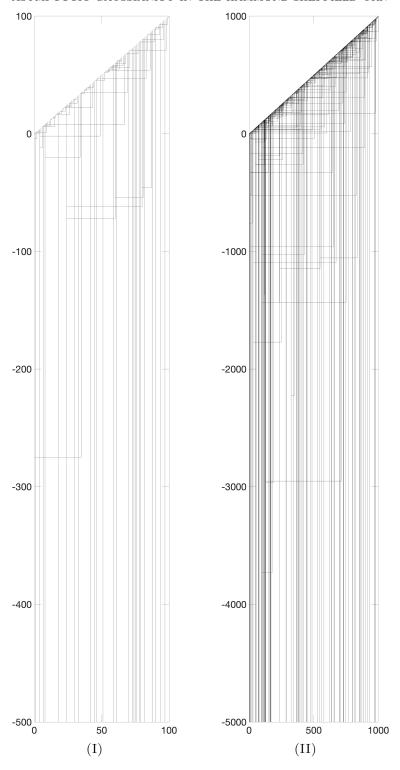FIGURE 2. This is a simulation of the ancestral lineages of the individuals $0, \ldots, n$, with $n = 100$ in panel (I) and $n = 1000$ in panel (II). Direction of time is vertical, and horizontal lines mark coalescence events. The two panels give an impression of how the genealogical forest of the individuals $i \in [n]$ scales with n, see e.g. Proposition 2.5. Like in Figure 1, the parameter α was chosen as 0.39.

3. ASYMPTOTIC GAUSSIANITY IN RANDOMLY COLOURED RANDOM PARTITIONS

In this section we consider a situation that is more general than the one described in Section 1. For $m \in \mathbb{N}$ let $\mathscr{P}^{(m)}$ be a random partition of $[m]$. The (random) equivalence relation on $[m]$ induced by $\mathscr{P}^{(m)}$ will be denoted by $\overset{m}{\sim}$, i.e.

$$i \overset{m}{\sim} j :\Longleftrightarrow i \text{ and } j \text{ belong to the same partition element of } \mathscr{P}^{(m)}. \tag{3.1}$$

The situation described in Section 1 fits into this framework, by choosing $\overset{m}{\sim}$ as the restriction to the set $[m]$ of the equivalence relation $\sim$ defined in (1.3). Note, however, that this kind of consistency of the relations $\overset{m}{\sim}$ is not required in the present section.

Let Y be a real valued random variable with $\mathbf{E}[Y] = 0$ and $0 < \mathbf{E}[Y^4] < \infty$. Thinking of each partition element being "coloured" by an independent copy of Y, we write $Y_i^{(m)}$ for the colour of the partition element in $\mathscr{P}^{(m)}$ to which $i \in [m]$ belongs. We then define for $k \in \mathbb{N}$

$$Z_k^{(m)} := \sum_{i=1}^{k} Y_i^{(m)}.$$

In the sequel we fix a natural number d and real numbers $0 = \rho_0 < \rho_1 < \cdots < \rho_d = 1$. The following theorem presents a sufficient criterion for the asymptotic normality of the sequence of $\mathbb{R}^d$-valued random variables

$$\mathcal{Z}^{(m)} := \left(Z_{\lfloor \rho_1 m \rfloor}^{(m)}, \ldots, Z_{\lfloor \rho_d m \rfloor}^{(m)} \right) \tag{3.2}$$

as $m \to \infty$. To prepare for this, let for all $m \in \mathbb{N}$ the random variables $\mathscr{I}^{(m)}, \mathscr{J}^{(m)}, \mathscr{K}^{(m)}$ and $\mathscr{L}^{(m)}$ be independent and uniformly distributed on $[m]$, and independent of $\mathscr{P}^{(m)}$ and of $\left(Y_i^{(m)} \right)_{i \in [m]}$.

Theorem 3.1. *The sequence of $\mathbb{R}^d$-valued random variables $\mathcal{Z}^{(m)}$ defined in (3.2) is asymptotically Gaussian as $m \to \infty$ provided the following conditions are satisfied:*

$$\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \overset{m}{\sim} \mathscr{K}^{(m)} \right) = o\left(\left(\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \right) \right)^{3/2} \right) \text{ as } m \to \infty, \tag{3.3}$$

$$\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \overset{m}{\sim} \mathscr{K}^{(m)} \overset{m}{\sim} \mathscr{L}^{(m)} \right) = o\left(\left(\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \right) \right)^{2} \right) \text{ as } m \to \infty, \tag{3.4}$$

$$\mathbf{Cov}\left[I_{\{i \overset{m}{\sim} j\}}, I_{\{k \overset{m}{\sim} \ell\}} \right] \leq \mathbf{P}\left(i \overset{m}{\sim} j \overset{m}{\sim} k \overset{m}{\sim} \ell \right) \quad \text{for all } m \in \mathbb{N} \text{ and } i, j, k, \ell \in [m], \tag{3.5}$$

and for all $(\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d \setminus \{(0, \ldots, 0)\}$ and

$$a_i^{(m)} := \alpha_g \text{ if } \lfloor \rho_{g-1} m \rfloor < i \leq \lfloor \rho_g m \rfloor, \quad i = 1, \ldots m; \quad g = 1, \ldots, d$$

there exists a constant $\widetilde{C} > 0$ (not depending on m) such that

$$\sum_{i,j=1}^{m} a_i^{(m)} a_j^{(m)} \mathbf{P}\left(i \overset{m}{\sim} j \right) \geq \widetilde{C} \sum_{i,j=1}^{m} \mathbf{P}\left(i \overset{m}{\sim} j \right), \quad m \in \mathbb{N}. \tag{3.6}$$

Remark 3.2.   (a) *Asymptotic Gaussianity of the univariate sequence $(Z_m^{(m)})_{m \in \mathbb{N}}$ is implied by the first three of the four conditions in Theorem 3.1. Indeed, for $d = 1$, condition (3.6) is automatically satisfied with $\widetilde{C} := \alpha_1$, since in that case $a_i^{(m)} = \alpha_1$ for $i = 1, \ldots, m$.*

(b) *If the partitions $\mathscr{P}^{(m)}$ are induced by a Hammond-Sheffield urn as described in Section 1, then* all *conditions of Theorem 3.1 (including condition (3.6) for all $d \in \mathbb{N}$) are implied by the assumption (1.10), see Section 8.*

Theorem 3.1 will be proved using the following proposition which, in turn, will be deduced from a theorem of Charles Stein, see [Ste86, Lecture X, Theorem 1]. Since m will be fixed in this proposition, we will write $\sim$ instead of $\overset{m}{\sim}$ for notational convenience and without any risk of confusion.

Proposition 3.3. *For fixed $m \in \mathbb{N}$ and $a_1, \ldots, a_m \in \mathbb{R}$ let*

$$\bar{S}_m := \sum_{i=1}^{m} a_i Y_i, \quad \bar{\sigma}_m^2 := \mathbf{Var}[\bar{S}_m], \tag{3.7}$$

with $Y_i := Y_i^{(m)}$ as defined at the beginning of this section. Let $\mathcal{N}$ be a standard normal random variable. Then for all continuously differentiable functions $h : \mathbb{R} \to \mathbb{R}$ with compact support

$$
\left| \mathbf{E}\left[h\left(\frac{\bar{S}_m}{\bar{\sigma}_m} \right) \right] - \mathbf{E}[h(\mathcal{N})] \right|
$$

$$
\leq \frac{c_1(h)}{\bar{\sigma}_m^2} \sqrt{ \mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 a_i a_j I_{\{i \sim j\}} \right] } + \frac{c_2(h)}{\bar{\sigma}_m^3} \mathbf{E}\left[\sum_{i,j,k=1}^{m} |Y_i|^3 |a_i| a_j a_k I_{\{i \sim j \sim k\}} \right]. \tag{3.8}
$$

where the finite numbers $c_1(h)$ and $c_2(h)$ are defined as

$$c_1(h) := 2 \sup |h - \mathbf{E}[h(\mathcal{N})]|, \quad c_2(h) := 2 \sup |h'|.$$

Proof. For $i \in [m]$ we put

$$X_i := \frac{a_i Y_i}{\bar{\sigma}_m}, \qquad M_i := \{ j \in [m] : j \sim i \}, \tag{3.9}$$

i.e., M_i is that element of the partition $\mathscr{P}^{(m)}$ which contains i. With $\mathscr{I}$ being a uniform pick from $[m]$ that is independent of $\mathscr{P}^{(m)}$, we write

$$M := (M_1, \ldots, M_m), \quad W := \sum_{i=1}^{m} X_i, \quad W^* := W - \sum_{j \in M_\mathscr{I}} X_j, \quad G := m X_\mathscr{I},$$

$$\mathcal{B} := \sigma\left(M, X_1, \ldots, X_m \right), \qquad \mathcal{C} := \sigma\left(M, \mathscr{I}, (X_j)_{j \not\sim \mathscr{I}} \right)$$

and note that $W = \mathbf{E}[G|\mathcal{B}]$ a.s. Now [Ste86, Lecture X, Theorem 1] asserts that

$$
\left| \mathbf{E}\left[h\left(W \right) \right] - \mathbf{E}[h(\mathcal{N})] \right|
$$

$$
\leq c_1(h) \left(\sqrt{ \mathbf{E}\left[\left(1 - \mathbf{E}\left[G(W - W^*)|\mathcal{B} \right] \right)^2 \right] } + c \mathbf{E}\left[\left| \mathbf{E}\left[G |\mathcal{C} \right] \right| \right] \right) \tag{3.10}
$$

$$
+ c_2(h) \mathbf{E}\left[|G|(W - W^*)^2 \right],
$$

with N, $c_1(h)$ and $c_2(h)$ as in (3.8) and a constant $c > 0$. Let us first turn to the term under the square root on the right-hand side of (3.10) and observe that

$$
\mathbf{E}\left[G(W - W^*)|\mathcal{B} \right] = \mathbf{E}\left[m X_\mathscr{I} \sum_{j \in M_\mathscr{I}} X_j \,\middle|\, \mathcal{B} \right] = \sum_{i=1}^{m} X_i \sum_{j \in M_i} X_j = \frac{1}{\bar{\sigma}_m^2} \sum_{i=1}^{m} a_i Y_i \sum_{j \in M_i} a_j Y_j.
$$

The expectation of this random variable is 1, since

$$\mathbf{E}\left[\mathbf{E}\left[\frac{1}{\bar{\sigma}_m^2}\sum_{i=1}^m a_i Y_i \sum_{j=1}^m a_j Y_j \,\middle|\, \mathscr{P}^{(m)}\right]\right] = \frac{1}{\bar{\sigma}_m^2}\mathbf{Var}[\bar{S}_m] = 1.$$

Hence the term under the square root in (3.10) equals

$$\mathbf{Var}\left[\sum_{i=1}^m X_i \sum_{j\in M_i} X_j\right].$$

The term $\mathbf{E}\left[\left|\mathbf{E}\left[G\,|\mathcal{C}\right]\right|\right]$ in the right-hand side of (3.10) vanishes, since the assumed independence of the colouring and the partitions together with the assumption $\mathbf{E}[X_i]=0$ implies

$$\mathbf{E}\left[\left|\mathbf{E}\left[G\,|\mathcal{C}\right]\right|\right] = \mathbf{E}\left[\left|\mathbf{E}\left[mX_{\mathscr{I}}\,|\,M,\mathscr{I}\right]\right|\right] = 0.$$

Finally, the rightmost term in (3.10) equals

$$\mathbf{E}\left[|G|\left(W-W^*\right)^2\right] = \mathbf{E}\left[|G|\left(\sum_{j\in M_{\mathscr{I}}}X_j\right)^2\right]$$

$$= \sum_{i=1}^m \mathbf{E}\left[I_{\{\mathscr{I}=i\}}|mX_i|\left(\sum_{j\in M_i}X_j\right)^2\right] = \sum_{i=1}^m \mathbf{E}\left[|X_i|\left(\sum_{j\in M_i}X_j\right)^2\right].$$

In summary we have shown that the right-hand side of (3.10) equals

$$c_1(h)\sqrt{\mathbf{Var}\left[\sum_{i=1}^m X_i \sum_{j\in M_i} X_j\right]} + c_2(h)\mathbf{E}\left[\sum_{i=1}^m |X_i|\left(\sum_{j\in M_i}X_j\right)^2\right], \tag{3.11}$$

which in turn is equal to the right-hand side of (3.8). This concludes the proof of Proposition 3.3.
□

Proof of Theorem 3.1. It suffices to show that for all $d\in\mathbb{N}$ and $(\alpha_1,\ldots,\alpha_d)\neq(0,\ldots,0)$, the linear combination

$$\bar{Z}_m := \alpha_0 Z^{(m)}_{\lfloor\rho_1 m\rfloor} + \alpha_1\left(Z^{(m)}_{\lfloor\rho_2 m\rfloor} - Z^{(m)}_{\lfloor\rho_1 m\rfloor}\right) + \cdots + \alpha_d\left(Z^{(m)}_m - Z^{(m)}_{\lfloor\rho_{d-1} m\rfloor}\right) \tag{3.12}$$

is asymptotically Gaussian as $m\to\infty$. For $m\in\mathbb{N}$ we put

$$a_i^{(m)} := \alpha_g \text{ if } \lfloor\rho_{g-1}m\rfloor < i \leq \lfloor\rho_g m\rfloor, \quad i=1,\ldots m; \quad g=1,\ldots,d. \tag{3.13}$$

It is then readily checked that $\bar{Z}_m$ defined in (3.12) satisfies

$$\bar{Z}_m = \sum_{i=1}^m a_i^{(m)} Y_i^{(m)}, \quad m\in\mathbb{N}, \tag{3.14}$$

and thus fits into the frame of Proposition 3.3, with $\bar{S}_m := \bar{Z}_m$. To use this proposition we will show that under the assumptions (3.3), (3.4), (3.5) and (3.6) and with $a_i = a_i^{(m)}$ from (3.13), both summands in the right-hand side of (3.8) converge to 0 as $m\to\infty$. For notational convenience we will for the rest of this proof suppress the superscript m in the equivalence relation $\overset{m}{\sim}$, in the coefficients $a_i^{(m)}$ and in the random variables $Y_i^{(m)}$, $\mathscr{I}^{(m)}$, $\mathscr{J}^{(m)}$, $\mathscr{K}^{(m)}$, $\mathscr{L}^{(m)}$.

For a constant C not depending on m we have

$$\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 a_i a_j I_{\{i\sim j\}}\right] \leq C\,\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 I_{\{i\sim j\}}\right], \tag{3.15}$$

$$\mathbf{E}\left[\sum_{i,j,k=1}^{m} |Y_i|^3 |a_i| a_j a_k I_{\{i\sim j\sim k\}}\right] \leq C\,\mathbf{E}\left[|Y|^3\right]\sum_{i,j,k=1}^{m} \mathbf{P}(i\sim j\sim k). \tag{3.16}$$

In order to bound $\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 I_{\{i\sim j\}}\right]$ from above, we decompose the variance with respect to $\mathscr{P}^{(m)}$ and first note that

$$\mathbf{E}\left[\sum_{i,j=1}^{m} Y_i^2 I_{\{i\sim j\}}\,\middle|\,\mathscr{P}^{(m)}\right] = \mathbf{E}\left[Y^2\right]\sum_{i,j=1}^{m} I_{\{i\sim j\}}.$$

The variance of the latter is

$$\mathbf{E}\left[Y^2\right]^2 \sum_{i,j,k,\ell\in[m]} \mathbf{Cov}\left[I_{\{i\sim j\}}, I_{\{k\sim \ell\}}\right]$$

which by assumption (3.5) is not larger than

$$\mathbf{E}\left[Y^2\right]^2 \sum_{i,j,k,\ell\in[m]} \mathbf{P}(i\sim j\sim k\sim \ell). \tag{3.17}$$

Next we note that

$$\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 I_{\{i\sim j\}}\,\middle|\,\mathscr{P}^{(n)}\right] = \sum_{i,j,k,\ell=1}^{m} \mathbf{Cov}\left[Y_i^2 I_{\{i\sim j\}}, Y_k^2 I_{\{k\sim \ell\}}\,\middle|\,\mathscr{P}^{(m)}\right]$$

$$= \mathbf{Var}\left[Y^2\right]\sum_{i,j,k,\ell=1}^{m} I_{\{i\sim j\sim k\sim \ell\}}.$$

Taking expectation of the latter and adding this to (3.17) we obtain

$$\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 I_{\{i\sim j\}}\right] \leq \mathbf{E}\left[Y^4\right]\sum_{i,j,k,\ell\in[m]} \mathbf{P}(i\sim j\sim k\sim \ell) = O\left(m^4\,\mathbf{P}\left(\mathscr{I}\sim\mathscr{J}\sim\mathscr{K}\sim\mathscr{L}\right)\right),$$

which because of (3.4) is $o\left(m^2\,\mathbf{P}\left(\mathscr{I}\sim\mathscr{J}\right)\right)^2$.

Likewise the right-hand side of (3.16) is $o\left(m^2\left(\mathbf{P}\left(\mathscr{I}\sim\mathscr{J}\right)\right)^{3/2}\right)$. From (3.7) and (3.14) we get

$$\bar{\sigma}_m^2 = \sum_{i,j\in[m]} a_i^{(m)} a_j^{(m)}\,\mathbf{Cov}\left[Y_i, Y_j\right] = \mathbf{Var}[Y]\sum_{i,j\in[m]} a_i^{(m)} a_j^{(m)}\,\mathbf{P}(i\sim j), \tag{3.18}$$

which due to assumption (3.6) is bounded from below by $\widetilde{C} m^2\,\mathbf{P}(\mathscr{I}\sim\mathscr{J})$. Thus under the assumptions of Theorem 3.1 the right-hand side of (3.8) converges to 0 as $m\to\infty$, which shows that Theorem 3.1 is a consequence of Proposition 3.3. $\qquad\square$

4. PAIR COALESCENCE PROBABILITIES: PROOF OF PROPOSITION 2.1

We now return to the setting of Section 1. Let $R_k^{(i)}$, $i \in \mathbb{Z}$, $k \in \mathbb{N}$ be independent copies of R, and define

$$A_i := \left\{ n \in \mathbb{Z} : i - R_1^{(i)} - \cdots - R_j^{(i)} = n \text{ for some } j \geq 0 \right\}. \tag{4.1}$$

Note that the A_i are independent and thus can be seen as decoupled versions of the ancestral lineages of the individuals $i \in \mathbb{Z}$. In particular they do not coalesce if they meet. Decomposing with respect to the most recent collision time one obtains immediately (cf. [HS13, p. 711]) that for $i > 0$

$$\mathbf{P}\left(A_0 \cap A_i \neq \emptyset\right) \sum_{m \geq 0} q_m^2 = \mathbf{E}\left[|A_0 \cap A_i|\right] = \sum_{m \geq 0} q_m q_{m+i}, \tag{4.2}$$

hence

$$\mathbf{P}(0 \sim i) = \frac{\sum_{m \geq 0} q_m q_{m+i}}{\sum_{m \geq 0} q_m^2}, \quad i \geq 0. \tag{4.3}$$

We will now assume (in accordance with the assumptions in Proposition 2.1) that the weights q_n of the renewal measure defined in (1.4) have the property (1.10). Under this condition we will prove

Proposition 4.1. *As $m \to \infty$,*

$$\sum_{j \geq 1} q_j q_{m+j} \sim \frac{1}{\Gamma(\alpha)^2 \Gamma(1-\alpha)^2} \frac{m^{2\alpha-1}}{L(m)^2} \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} \mathrm{d}x. \tag{4.4}$$

The asymptotics (2.1) claimed in Proposition 2.1 is immediate from (4.3) combined with (4.4).

The remainder of this section is devoted to the proof of Proposition 4.1. We will prove (4.4) first under a special assumption on the Karamata representation of the slowly varying function L.

Lemma 4.2. *Let $\alpha \in (0, \frac{1}{2})$, and consider*

$$r_n := n^{\alpha-1} K(n) \tag{4.5}$$

where $K(n)$ is of the form

$$K(n) = \exp\left(\int_B^n \frac{l(t)}{t} \, \mathrm{d}t\right) \tag{4.6}$$

with B a positive constant and $l(t)$, $t \geq B$, a bounded measurable function converging to 0 as $t \to \infty$. Then (r_n) is ultimately decreasing, and

$$\sum_{j \geq 1} r_j r_{i+j} \sim K(i)^2 i^{2\alpha-1} \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} \mathrm{d}x \quad \text{as } i \to \infty, \tag{4.7}$$

with

$$\int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx = B(\alpha, 1-2\alpha) = \frac{\Gamma(\alpha)\Gamma(1-2\alpha)}{\Gamma(1-\alpha)}. \tag{4.8}$$

Proof. a) The equality (4.8) is readily checked by substituting $y = \frac{x}{1+x}$.

b) The fact that (r_n) is ultimately decreasing follows from (4.5) together with the Karamata representation (4.6) of $K(n)$. To see this, we argue as follows, putting $\beta := 1 - \alpha$. Since $l(t)$ tends to zero for $t \to \infty$ we know that there exists $n_0 \in \mathbb{N}$ such that for all $t \geq n_0$ one has $l(t) < \beta$. This implies

$$\begin{aligned}
\frac{r_n}{r_{n+1}} &= \frac{n^{-\beta}}{(n+1)^{-\beta}} \cdot \frac{K(n)}{K(n+1)} \\
&= \frac{n^{-\beta}}{(n+1)^{-\beta}} \exp\left(-\int_n^{n+1} \frac{l(t)}{t} dt\right)
\end{aligned}$$

$$= \exp\left(\beta\left(\ln(n+1) - \ln(n)\right) - \int_n^{n+1} \frac{l(t)}{t} dt\right)$$

$$= \exp\left(\int_n^{n+1} \left(\frac{\beta}{t} - \frac{l(t)}{t}\right) dt\right).$$

Since by assumption the integrand on the right-hand side is strictly positive for $n \geq n_0$, we obtain that $(r_n)_n$ is decreasing for $n \geq n_0$.

c) In view of (4.5), the claimed asymptotics (4.7) is equivalent to

$$\lim_{i\to\infty} \frac{1}{i} \sum_{j\geq 1} \frac{r_j}{r_i} \frac{r_{i+j}}{r_i} = \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx. \tag{4.9}$$

We now set out to prove (4.9). To this purpose we show first that for all $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all sufficiently large i

$$\frac{1}{i} \sum_{j\geq Ni} \frac{r_j}{r_i} \frac{r_{i+j}}{r_i} < \varepsilon. \tag{4.10}$$

Since (r_n) is ultimately decreasing, (4.10) will follow if we can show that exists an $N \in \mathbb{N}$ such that for all sufficiently large i

$$\frac{1}{i} \sum_{j\geq Ni} \left(\frac{r_j}{r_i}\right)^2 < \varepsilon. \tag{4.11}$$

Again because of the ultimate monotonicity of (r_n), the left-hand side. of (4.11) is for sufficiently large i bounded from above by

$$\sum_{m=N}^\infty \left(\frac{r_{mi}}{r_i}\right)^2 = \sum_{m=N}^\infty m^{2\alpha-2} \left(\frac{K(mi)}{K(i)}\right)^2. \tag{4.12}$$

Using (4.6) one obtains that for any $\delta > 0$ and i so large that $l(t) < \delta$ for all $t \geq i$,

$$\frac{K(mi)}{K(i)} = \exp\left(\int_i^{mi} \frac{l(t)}{t} dt\right) \leq \exp\left(\delta\left(\ln(mi) - \ln(i)\right)\right) \leq m^\delta, \tag{4.13}$$

which implies by dominated convergence that for sufficiently large N the right-hand side of (4.12) is smaller than ε for all sufficiently large i. We have thus proved (4.10).

Next we show that for all $\varepsilon > 0$ there exists an $\eta \in \mathbb{N}$ such that for all sufficiently large i

$$\frac{1}{i} \sum_{j\leq \eta i} \frac{r_j}{r_i} \frac{r_{i+j}}{r_i} < \varepsilon. \tag{4.14}$$

Again by ultimate monotonicity of (r_n) which gives us $\frac{r_{i+j}}{r_i} \leq 1$ for i large enough, for this it suffices to show that for all $\varepsilon > 0$ there exists an $\eta > 0 \in \mathbb{N}$ such that for all sufficiently large i

$$\frac{1}{i} \sum_{j\leq \eta i} \frac{r_j}{r_i} < \varepsilon. \tag{4.15}$$

From [Fel71, Theorem 5 on p. 447] we obtain that

$$\sum_{j\leq \eta i} r_j \sim \frac{1}{\alpha} (\eta i)^\alpha K(\lfloor \eta i \rfloor) \quad \text{as } i \to \infty, \tag{4.16}$$

and hence

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j}{r_i} \sim \frac{1}{\alpha} \eta^\alpha \frac{K(\lfloor \eta i \rfloor)}{K(i)} \sim \frac{\eta^\alpha}{\alpha} \quad \text{as } i \to \infty, \tag{4.17}$$

which proves (4.15), and hence also (4.14). (The last asymptotic is by the fact that K is slowly varying.)

In view of (4.5), (4.10) and (4.14), for proving (4.7) it remains to show that

$$\lim_{i \to \infty} \frac{1}{i} \sum_{\eta i \leq j \leq N i} \frac{K(j)}{K(i)} \frac{K(i+j)}{K(i)} \left(1 + \frac{j}{i}\right)^{\alpha-1} \left(\frac{j}{i}\right)^{\alpha-1} = \int_\eta^N (1+x)^{\alpha-1} x^{\alpha-1} \mathrm{d}x. \tag{4.18}$$

From (4.6) one derives that

$$\lim_{i \to \infty} \sup_{\eta i \leq j \leq (N+1)i} \left| \frac{K(j)}{K(i)} - 1 \right| = 0.$$

Hence (4.18) boils down to a convergence of Riemann sums to its integral limit, and the proof of Lemma 4.2 is done. $\qquad\square$

Let us now complete the proof of Proposition 4.1.

Proof. The asymptotics (1.10) can be rewritten as

$$q_n = C_n n^{\alpha-1} \tilde{L}(n)$$

where $\tilde{L}(n)$ is a slowly varying function and $C_n \to \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} > 0$.

As in the proof of Lemma 4.2 it suffices to show that

$$\frac{1}{i} \sum_{j \geq 1} \frac{q_j}{q_i} \frac{q_{i+j}}{q_i} \to \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} \mathrm{d}x. \tag{4.19}$$

Because of the Karamata representation theorem (see e.g. Theorem 1.3.1 in [BGT87]) there exists a $K(n)$ satisfying (4.6) and a sequence D_n converging to a positive constant D such that

$$\tilde{L}(n) = D_n K(n), \quad n = 1, 2, \ldots \tag{4.20}$$

Defining r_n as in (4.5) we have

$$q_n = D_n C_n r_n, \quad n = 1, 2, \ldots$$

Since the asymptotics of neither the left-hand side of (4.9) nor that of the left-hand side of (4.19) reacts to the omission of a fixed finite number of summands, we see that (4.9) carries over to (4.19). $\qquad\square$

5. Depth of most recent common ancestor: Proof of Proposition 2.5

In this section we will assume that the weights q_n of the renewal measure defined in (1.4) obey the asymptotics (1.10), see the discussion after equation (1.10). For $i, m \in \mathbb{N}$ we set

$$f_i(m) = \mathbf{P}\left(\mathscr{M}(0, i) = -m\right), \quad F_i(m) = \mathbf{P}\left(\mathscr{M}(0, i) \geq -m\right).$$

For the independent couplings A_i, $i \in \mathbb{Z}$, of the ancestral lines of 0 and i as defined in (4.1) we have for all $r > 0$ and $i \in \mathbb{Z}$

$$\begin{aligned}
\sum_{k=0}^{ri} q_k q_{k+i} &= \mathbf{E}\left[\left|A_0 \cap A_i \cap \{0, \ldots, -ri\}\right|\right] \\
&= \sum_{k=0}^{ri} \left(f_i(k) \sum_{l=0}^{ri-k} q_l^2\right) \quad \leq \quad \left(\sum_{l=0}^{ri} q_l^2\right) F_i(ri),
\end{aligned}$$

and consequently

$$\mathbf{P}(D_i \leq ri) = F_i(ri) \geq \frac{\sum_{k=0}^{ri} q_k q_{k+i}}{\sum_{l=0}^{ri} q_l^2}. \tag{5.1}$$

As in the proof of Proposition 4.1 we obtain

$$\sum_{k=0}^{ri} q_k q_{k+i} \sim \frac{i^{2\alpha-1}}{L(i)^2} \frac{1}{\Gamma(1-\alpha)^2 \Gamma(\alpha)^2} \int_0^r x^{\alpha-1}(1+x)^{\alpha-1}\mathrm{d}x \quad \text{as } i \to \infty. \tag{5.2}$$

Together with (5.1) and the asymptotics (2.1) this gives

$$\liminf_{i\to\infty} \mathbf{P}\left(D_i \leq ri \,|\, 0 \sim i\right) \geq B(\alpha, 1-2\alpha)^{-1} \int_0^r x^{\alpha-1}(1+x)^{\alpha-1}\mathrm{d}x. \tag{5.3}$$

For the upper estimate we choose some arbitrary $\varepsilon > 0$ and observe

$$
\begin{aligned}
\sum_{k=0}^{(r+\varepsilon)i} q_k q_{k+i} &= \mathbf{E}\left[|A_0 \cap A_i \cap \{0,\ldots,-(r+\varepsilon)i\}|\right] \\
&= \sum_{k=0}^{(r+\varepsilon)i}\left(f_i(k) \sum_{l=0}^{(r+\varepsilon)i-k} q_l^2\right) \geq \sum_{k=0}^{ri}\left(f_i(k) \sum_{l=0}^{(r+\varepsilon)i-k} q_l^2\right) \\
&\geq \left(\sum_{l=0}^{\varepsilon i} q_l^2\right) \sum_{k=0}^{ri} f_i(k) = \left(\sum_{l=0}^{\varepsilon i} q_l^2\right) F_i(ri).
\end{aligned}
$$

Using (5.2), now with $r + \varepsilon$ instead of r, we get

$$\limsup_{i\to\infty} \mathbf{P}\left(D_i \leq ri \,|\, 0 \sim i\right) \leq B(\alpha, 1-2\alpha)^{-1} \int_0^{r+\varepsilon} x^{\alpha-1}(1+x)^{\alpha-1}\mathrm{d}x.$$

Since ε was arbitrary, this together with (5.3) gives the assertion of Proposition 2.5. $\square$

6. Triplet and quartet coalescence probabilities: Proof of Proposition 2.3

In this section we will assume that the weights q_n of the renewal measure defined in (1.4) obey the asymptotics (1.10). We now turn to the asymptotic analysis of triplet and quartet coalescence probabilities.

6.1. Triplet coalescence probabilities.

Lemma 6.1. *For all $r > 0$ and $i \in \mathbb{N}$ we have for a slowly varying function $\tilde{L}$ not depending on r*

$$\mathbf{P}\left(0 \sim i \sim \lfloor(1+r)i\rfloor\right) \leq \left(r^{\alpha-1} + r^{2\alpha-1}\right) \tilde{L}(i)\, i^{4\alpha-2}. \tag{6.1}$$

Proof. Let (A_k) be the independent (non-merging) ancestral lineages defined in (4.1). We set

$$
A_{k,\ell} := \begin{cases} A_{\max(A_k \cap A_\ell)} & \text{if } A_k \cap A_\ell \neq \emptyset, \\ \emptyset & \text{if } A_k \cap A_\ell = \emptyset, \end{cases}
$$

$$
B_{k,\ell} := \bigcup_{g \in A_k \cap A_\ell} A_g.
$$

In words, $B_{k,\ell}$ is the union of the (non-merging) ancestral lineages starting at all the points of intersection of A_k and A_ℓ. Distinguishing the 3 shapes of the ancestral tree of the individuals 0, i and $(1+r)i$ on the event $E := \{0 \sim i \sim (1+r)i\}$ (and omitting the Gauss brackets in $\lfloor(1+r)i\rfloor$ etc. for the sake of readability) we have by subadditivity

$$\mathbf{P}\left(0 \sim i \sim (1+r)i\right)$$

$$
\begin{aligned}
&\leq\quad \mathbf{P}\left(E; \mathscr{M}(0,i) \geq \mathscr{M}(0,(1+r)i)\right) + \mathbf{P}\left(E; \mathscr{M}(0,(1+r)i) \geq \mathscr{M}(0,i)\right) \\
&\qquad + \mathbf{P}\left(E; \mathscr{M}(i,(1+r)i) \geq \mathscr{M}(0,i)\right) \\
&\leq\quad \mathbf{P}\left(A_{0,i} \cap A_{(1+r)i} \neq \emptyset\right) + \mathbf{P}\left(A_{0,(1+r)i} \cap A_i \neq \emptyset\right) + \mathbf{P}\left(A_{i,(1+r)i} \cap A_0 \neq \emptyset\right) \\
&\leq\quad \mathbf{E}\left[|A_{0,i} \cap A_{(1+r)i}|\right] + \mathbf{E}\left[|A_{0,(1+r)i} \cap A_i|\right] + \mathbf{E}\left[|A_{i,(1+r)i} \cap A_0|\right] \\
&\leq\quad \mathbf{E}\left[|B_{0,i} \cap A_{(1+r)i}|\right] + \mathbf{E}\left[|B_{0,(1+r)i} \cap A_i|\right] + \mathbf{E}\left[|B_{i,(1+r)i} \cap A_0|\right]. \\
&=\quad \sum_{m\geq 0} q_m q_{m+i} \sum_{n\geq 0} q_n q_{n+(1+r)i+m} + \sum_{m\geq 0} q_m q_{m+(1+r)i} \sum_{n\geq 0} q_n q_{n+i+m} \\
&\qquad + \sum_{m\geq 0} q_m q_{ri+m} \sum_{n\geq 0} q_n q_{n+i-m}.
\end{aligned}
\tag{6.2}
$$

An inspection of the third summand on the right-hand side leads to

$$
\begin{aligned}
&\sum_{m\geq 0} q_m q_{ri+m} \sum_{n\geq 0} q_n q_{n+i-m} \\
&=\quad q_m q_{ri+i} \sum_{n\geq 0} q_n q_n + \sum_{m\neq i} q_m q_{ri+m} \sum_{n\geq 0} q_n q_{n+i-m} \\
&\leq\quad C\left(q_i q_{(1+r)i} + \sum_{m\neq i} \frac{m^{\alpha-1}}{L(m)} \frac{(m+ri)^{\alpha-1}}{L(m+ri)} \frac{|i-m|^{2\alpha-1}}{L(|i-m|)^2}\right) \\
&\sim\quad C\left(\frac{i^{2\alpha-2}}{L(i)^2} + \frac{i^{4\alpha-2}}{L(i)^4} \int_0^\infty x^{\alpha-1}(x+r)^{\alpha-1}|1-x|^{2\alpha-1}\mathrm{d}x\right) \\
&\leq\quad \tilde{L}(i) i^{4\alpha-2} r^{\alpha-1},
\end{aligned}
$$

where $\tilde{L}$ is a slowly varying function that dominates $C\left(\frac{1}{L^2} + \frac{1}{L^4}\right)$, and where the asymptotics is justified in the same way as (4.18) was derived first for slowly varying functions satisfying (4.6) and then, using the Karamata representation, for general slowly varying functions obeying (4.20). The first and the second summand on the r.h.s of (6.2) can be analysed in an analogous manner, leading to the bounds $\tilde{L}(i) i^{4\alpha-2} r^{2\alpha-1}$ and $\tilde{L}(i) i^{4\alpha-2} r^{\alpha-1}$, respectively. $\square$

Proof of Proposition 2.3 Part 1. We set out to show (2.3), and first observe that

$$
\sum_{i=0}^n \sum_{j=0}^n \mathbf{P}\left(0 \sim i \sim j\right) = \sum_{i=0}^n \mathbf{P}\left(0 \sim i\right) + 2 \sum_{i=1}^n \sum_{j=1}^{n-i} \mathbf{P}\left(0 \sim i \sim i+j\right).
\tag{6.3}
$$

By Lemma 6.1 we get, noting that $r^{\alpha-1} + r^{2\alpha-1} \leq 2r^{\alpha-1} + 1$,

$$
\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^{n-i} \mathbf{P}\left(0 \sim i \sim i+j\right) &\leq \sum_{i=1}^n \sum_{j=1}^{n-i} \tilde{L}(i) i^{4\alpha-2}\left(\left(\frac{j}{i}\right)^{\alpha-1} + 1\right) \\
&\leq \tilde{L}(n)\left(\sum_{i=1}^n i^{3\alpha-1}(n-i)^\alpha + n\sum_{i=1}^n i^{4\alpha-2}\right) = O\left(n^{4\alpha+\delta}\right)
\end{aligned}
\tag{6.4}
$$

for each $\delta > 0$. From (6.3) combined with (2.1) and (6.4) we obtain for each $\delta > 0$ the estimate

$$
\sum_{i=0}^n \sum_{j=0}^n \mathbf{P}\left(0 \sim i \sim j\right) \leq Cn^{4\alpha+\delta},
\tag{6.5}
$$

where the constant C depends on δ but not on n. An analogous calculation for k in place of 0 shows that also

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbf{P}\left(k \sim i \sim j\right) \leq Cn^{4\alpha+\delta}, \quad k \in [n], \tag{6.6}$$

where C can be chosen uniformly in n and in $k \in [n]$. (An intuitive reason for this uniformity comes from the fact that for each $k \in [n]$ and small $\varepsilon > 0$, the big majority of the pairs $(i,j) \in [n]^2$ leads to pairwise distances $|i-j|$, $|i-k|$ and $|j-k|$ that are all between εn and n.) With this uniformity in k, (2.3) follows directly from (6.6). $\qquad\square$

6.2. Quartet coalescence probabilities.

Lemma 6.2. *For all $r_2 > r_1 > 0$ and $i \in \mathbb{N}$ we have for a slowly varying function $\bar{L}$ not depending on r*

$$\mathbf{P}\left(0 \sim i \sim \lfloor(1+r_1)i\rfloor \sim \lfloor(1+r_2)i\rfloor\right)$$
$$\leq \quad \left(r_1^{\alpha-1} + r_1^{2\alpha-1} + r_2^{\alpha-1} + r_2^{2\alpha-1}\right)\bar{L}(i)i^{6\alpha-3}. \tag{6.7}$$

Proof. Again let (A_i) be the independent (non-merging) ancestral lineages defined in (4.1). Let $B_{k,\ell}$ be as in (6.1) and set

$$B_{[k,\ell]j} := \bigcup_{m \in B_{k,\ell}} \bigcup_{g \in A_m \cap A_j} A_g.$$

We fix $i \in \mathbb{N}$ and set with $j_1 := (1+r_1)i$, $j_2 := (1+r_2)i$,

$$\begin{aligned} Q := \quad &\{(0,i,j_1,j_2), (0,i,j_2,j_1), (0,j_1,i,j_2), (0,j_1,j_2,i),\\ &(0,j_2,i,j_1), (0,j_2,j_1,i), (i,j_1,0,j_2), (i,j_1,j_2,0),\\ &(i,j_2,0,j_1), (i,j_2,j_1,0), (j_1,j_2,0,i), (j_1,j_2,i,0)\} \end{aligned}$$

We now argue in a similar way as in the proof of Proposition 6.1. By subadditivity we get

$$\mathbf{P}\left(0 \sim i \sim (1+r_1)i \sim (1+r_2)\right) \leq \sum_{(k_1,k_2,k_3,k_4)\in Q} \mathbf{E}\left[\left|B_{[k_1,k_2]k_3} \cup A_{k_4}\right|\right].$$

By the very same arguments as in the proof of Lemma 6.1 one checks that each of the twelve summands is bounded by the right-hand side of (6.7). $\qquad\square$

Proof of Proposition 2.3 Part 2. We are now going to prove (2.4), and first set out to show that for all $\delta > 0$

$$\sum_{i=1}^{n}\sum_{j=1}^{n-i}\sum_{\ell=1}^{n-i-j}\mathbf{P}\left(0 \sim i \sim i+j \sim i+j+\ell\right) = O\left(n^{6\alpha+\delta}\right). \tag{6.8}$$

Setting $r_1 = \frac{j}{i}, r_2 = \frac{j+l}{i}$ we have $r_1^{\alpha-1} + r_1^{2\alpha-1} + r_2^{\alpha-1} + r_2^{2\alpha-1} \leq 4(r_1^{\alpha-1}+1)$ and by Lemma 6.2 the left-hand side of (6.8) is bounded from above by

$$C\sum_{i=1}^{n}\bar{L}(i)i^{6\alpha-3}\sum_{j=1}^{n-i}\sum_{\ell=1}^{n-i-j}\left[\left(\frac{j}{i}\right)^{\alpha-1}+1\right].$$

By arguments analogous to those leading to (6.3) in the proof of Part 1, this implies (6.8). As in the proof of Part 1 we can argue that, as $n \to \infty$ the terms

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{\ell=0}^{n}\mathbf{P}\left(k \sim i \sim j \sim \ell\right)$$

are of the same order uniformly in for all $k \in \{0, 1, \ldots, n\}$, so it is enough to look at the case $k = 0$. Also, from (2.1) and (2.3) it is clear that we may restrict to pairwise distinct i, j, ℓ. Thus, similar as in Part 1, (2.4) follows from (6.8). $\qquad\square$

7. A COVARIANCE ESTIMATE: PROOF OF LEMMA 2.6

For $i = j$ or $k = \ell$ the assertion of Lemma 2.6 is clearly true because then the left-hand side. of (2.7) vanishes. For $i = k$ we have

$$\mathbf{Cov}[I_{\{i \sim j\}}, I_{\{i \sim \ell\}}] = \mathbf{P}(i \sim j \sim \ell) - \mathbf{P}(i \sim j)\,\mathbf{P}(i \sim \ell) \le \mathbf{P}(i \sim j \sim \ell).$$

Hence we may assume without loss of generality that i, j, k, ℓ are pairwise distinct. We then have

$$\mathbf{Cov}\left[I_{i \sim k}, I_{j \sim \ell}\right] = \mathbf{P}\left(i \sim k \cap j \sim \ell\right) - \mathbf{P}\left(i \sim k\right)\mathbf{P}\left(j \sim \ell\right) \tag{7.1}$$

and

$$\mathbf{P}\left(i \sim k \cap j \sim \ell\right) = \mathbf{P}\left(i \sim k \not\sim j \sim \ell\right) + \mathbf{P}\left(i \sim k \sim j \sim \ell\right). \tag{7.2}$$

By (7.1) and (7.2), the inequality (2.7) is immediate from the following

Lemma 7.1. *For pairwise distinct i, j, k, ℓ we have*

$$\mathbf{P}\left(i \sim k \not\sim j \sim \ell\right) \le \mathbf{P}\left(i \sim k\right)\mathbf{P}\left(j \sim \ell\right)$$

Proof. Let A_g, $g \in \mathbb{Z}$, be defined as in (4.1). For $m \in \mathbb{N}$ and $i_1 < \ldots < i_m \in \mathbb{Z}$ we define an $(A_{i_1}, \ldots, A_{i_m})$-measurable random graph $G^{\{i_1,\ldots,i_m\}}$ which is equal in distribution to the subgraph of $\mathcal{G}_\mu$ that is formed by the (possibly coalescing) ancestral lineages of $i_1, i_2, \ldots, i_m$. The construction of $G^{\{i_1,\ldots,i_m\}}$ is done inductively in the following "lookdown" manner: the ancestral lineage of i_1 is taken as A_{i_1}, correspondingly, we put $G^{\{i_1\}} := A_{i_1}$. The ancestral lineage of i_{h+1} is given by $A_{i_{h+1}}$ as long as the latter did not meet $G^{\{i_1,\ldots,i_h\}}$. At the time of the first (seen in backward time direction) collision of $A_{i_{h+1}}$ with $G^{\{i_1,\ldots,i_h\}}$, the ancestral lineage of i_{h+1} is continued by the lineage in $G^{\{i_1,\ldots,i_h\}}$ that starts in the meeting point (and the continuation of $A_{i_{h+1}}$ from there on is erased). An inspection of $G^{\{i,j,k,\ell\}}$ reveals that

$$\begin{aligned}
&\mathbf{P}\left(\{i \sim k \not\sim j \sim \ell\}\right) \\
= \ &\mathbf{P}\left(\{A_i \cap A_k \neq \emptyset\} \cap \{A_j \cap A_\ell \neq \emptyset\} \cap \left\{G^{\{i,k\}} \cap G^{\{j,\ell\}} = \emptyset\right\}\right) \\
\le \ &\mathbf{P}\left(\{A_i \cap A_k \neq \emptyset\} \cap \{A_j \cap A_\ell \neq \emptyset\}\right),
\end{aligned}$$

which because of mutual independence of the A_g gives the assertion of the lemma. $\qquad\square$

8. ASYMPTOTIC GAUSSIANITY IN THE HAMMOND-SHEFFIELD URN: PROOF OF THEOREM 1.1(B)

We are going to apply Theorem 3.1, with $\mathscr{P}^{(m)}$ being the partition on $[m]$ that is generated by the Hammond-Sheffield urn, i.e. by the equivalence class $\sim$ defined in (1.3). For $d \in \mathbb{N}$ and $t_1 < \cdots < t_d$ as prescribed in Theorem 1.1 we apply Theorem 3.1 with $m = m(n) = \lfloor t_d n \rfloor$ and $\rho_g := \lfloor t_g n \rfloor / \lfloor t_d n \rfloor$, $g = 1, \ldots, d$. Under condition (B) of Theorem 1.1, Corollary 2.4 ensures the validity of assumptions (3.3) and (3.4), and Lemma 2.6 guarantees that (3.5) is fulfilled. It remains to check the assumption (3.6). Indeed, with

$$a(x) := \alpha_g \text{ if } \rho_{g-1} < x \le \rho_g, \quad x \in (0, 1], \quad g = 1, \ldots, d,$$

because of $\mathbf{P}(i \sim j) = \mathbf{P}(0 \sim |i - j|)$ and in view of Proposition 2.1 the left-hand side. of (3.6) has the asymptotics

$$m^{-2} \sum_{i,j \in [m]} a_i^{(m)} a_j^{(m)} \mathbf{P}(i \sim j) \sim \frac{Cm^{2\alpha-1}}{L(m)^2} \int_0^1 \int_0^1 a(x)a(y)|x - y|^{2\alpha-1}\mathrm{d}x\,\mathrm{d}y \quad \text{as } m \to \infty \tag{8.1}$$

The Riesz kernel $|x - y|^{2\alpha-1}$ is positive definite (see e.g. [Dos98]), hence the integral term in (8.1) is strictly positive, and consequently the left-hand side of (8.1) is of the order $m^{2\alpha+1}$ as $m \to \infty$. Because of (1.8), this is also the order of the right-hand side of (3.6).

9. Tightness: Proof of Proposition 1.3

Inspired by the proof of Theorem 1 in [Sot01], which shows tightness of a different approximation scheme for fractional Brownian motion, we will make use of the following

Lemma 9.1 ([Bil68, Theorem 13.5]). *Let $T > 0$ and $(\zeta^{(n)}(t))_{0 \leq t \leq T}$ be continuous processes that converge to a continuous process $(\zeta(t))_{0 \leq t \leq T}$ in the sense of finite dimensional distributions. Assume that for a nondecreasing continuous function F on $[0, T]$, for some $\gamma > 1$, for all $0 \leq s \leq t \leq u \leq T$, for some $N \in \mathbb{N}$ and all $n \geq N$*

$$\mathbf{E}\left[\left|\zeta^{(n)}(t) - \zeta^{(n)}(s)\right| \cdot \left|\zeta^{(n)}(u) - \zeta^{(n)}(t)\right|\right] \leq [F(u) - F(s)]^{\gamma}. \tag{9.1}$$

Then $(\zeta^n(t))_{0 \leq t \leq T}$ converges in distribution in $\left(C([0, T]), \|\cdot\|_{L^{\infty}([0,T])}\right)$ to $(\zeta(t))_{0 \leq t \leq T}$.

Proof of Proposition 1.3. We first note the following immediate consequence of (1.7) and (1.8): There exist constants $c, c' > 0$ such that

$$cn^{2\alpha+1}L(n)^{-2} \leq \sigma_n^2 \leq c'n^{2\alpha+1}L(n)^{-2}, \quad \forall n \in \mathbb{N}. \tag{9.2}$$

Next we fix $0 \leq s \leq t \leq u \leq T$ and $j, k, l \in \mathbb{N}$ satisfying

$$\frac{j}{n} \leq s < \frac{j+1}{n}, \quad \frac{k}{n} \leq t < \frac{k+1}{n}, \quad \frac{l}{n} \leq u < \frac{l+1}{n}.$$

The definition of $S^{(n)}$ as a linear interpolation gives

$$\sigma_n(S^{(n)}(t) - S^{(n)}(s)) = \sum_{m=j+1}^{k} Y_m + \left[1 - (sn - j)\right] Y_{j+1} + [tn - k] Y_{k+1}.$$

We note that $0 \leq [1 - (sn - j)] \leq 1$, $0 \leq [tn - k] \leq 1$ and get that for some constants c_1, c_2, c_3, c_4

$$
\begin{aligned}
\mathbf{Var}\left[S^{(n)}(t) - S^{(n)}(s)\right] &\leq \left[2\,\mathbf{Var}\left[S^{(k-j)}\right] + 2c_1\,\mathbf{Var}\left[Y_{j+1} + Y_{k+1}\right]\right]\sigma_n^{-2} \\
&\leq \left[2c'(k-j)^{2\alpha+1}L(k-j)^{-2} + c_2\right]\sigma_n^{-2} \\
&\leq \left[c_3(k-j)^{2\alpha+1}L(k-j)^{-2}\right]\sigma_n^{-2} \\
&\leq c_4\left(\frac{k-j}{n}\right)^{2\alpha+1}\left(\frac{L(n)}{L(k-j)}\right)^2.
\end{aligned} \tag{9.3}
$$

Here the first inequality holds because for any two square-integrable random variables G_1, G_2 one has $\mathbf{Var}[G_1 + G_2] \leq 2\,\mathbf{Var}[G_1] + 2\,\mathbf{Var}[G_2]$, the second and the last inequality hold because of (9.2), and the third one holds for some $c_3 > 0$, some $N_1 \in \mathbb{N}$ and all $n \geq N_1$ because L is a slowly varying function. (Remember that k, j, l depend on n for fixed time points s, t, u.) Analogously,

$$\mathbf{Var}\left[S^{(n)}(u) - S^{(n)}(t)\right] \leq c_3\left(\frac{l-k}{n}\right)^{2\alpha+1}\left(\frac{L(n)}{L(l-k)}\right)^2. \tag{9.4}$$

To use Lemma 9.1 we have to bound the expectation

$$\mathbf{E}\left[\left|S^{(n)}(t) - S^{(n)}(s)\right| \cdot \left|S^{(n)}(u) - S^{(n)}(t)\right|\right].$$

For $l \leq j+1$ we get that $|u-t| \leq \frac{2}{n}$ and $|t-s| \leq \frac{2}{n}$, so that by basic calculus $(u-t)(t-s) \leq (u-s)^{1+\alpha}$. By the definition of μ and the fact that L is slowly varying one has $L(n) \leq n^{-\alpha}$ for n large enough. So linear interpolation gives:

$$\mathbf{E}\left[|S^{(n)}(t) - S^{(n)}(s)| \cdot |S^{(n)}(u) - S^{(n)}(s)|\right]$$

$$\leq \left(\tilde{c} n^{-\frac{1}{2}-\alpha} L(n)\right)^2 \left((t-s) \cdot (u-s)\right)$$

$$\leq \left(\tilde{c} n^{-\frac{1}{2}}\right)^2 (u-s)^{1+2\alpha} = \tilde{c}^2 \frac{1}{n}(u-s)^{1+2\alpha}$$

$$\leq \tilde{c}^2 (u-s)^{1+2\alpha} = [c_0 u - c_0 s]^{1+2\alpha}$$

for some $c_0 > 0$. Now assume $l > j+1$. Cauchy-Schwarz and the estimates (9.3), (9.4) yield

$$\mathbf{E}\left[\left|S^{(n)}(t) - S^{(n)}(s)\right| \cdot \left|S^{(n)}(u) - S^{(n)}(t)\right|\right]$$

$$\leq \sqrt{\mathbf{Var}\left[S^{(n)}(t) - S^{(n)}(s)\right]} \sqrt{\mathbf{Var}\left[S^{(n)}(u) - S^{(n)}(t)\right]}$$

$$\leq \sqrt{c_3 \left(\frac{k-j}{n}\right)^{2\alpha+1} \left(\frac{L(n)}{L(k-j)}\right)^2} \sqrt{c_3 \left(\frac{l-k}{n}\right)^{2\alpha+1} \left(\frac{L(n)}{L(l-k)}\right)^2}$$

$$\leq c_3 \left(\frac{l-j}{n}\right)^{2\alpha+1} \left(\frac{L(n)}{L(k-j)L(l-k)}\right)^2$$

$$\leq 2c_3 \left(\frac{l-j}{n}\right)^{2\alpha+1} \leq \left(2c_3 \cdot 2(u-s)\right)^{2\alpha+1}.$$

(The third inequality holds because of $4|k-j| \cdot |l-k| \leq |l-j|^2$, and the fourth one holds for $n \geq N_2$ for some $N_2 \in \mathbb{N}$ because L is a slowly varying function.)

Thus, $\zeta^{(n)} := S^{(n)}$ fulfills condition (9.1) with with $N := \max\{N_1, N_2\}$, $\gamma := \alpha + 1/2$ and $F(x) := 4\max\{c_0, c_3\}x$. In view of Corollary 1.2 we can thus apply Lemma 9.1 and conclude the assertion of Proposition 1.3. □

Remark 9.2. *In* [HS13] *the functional convergence of $S^{(n)}$ was deduced from a tail estimate (uniform in n) on $\max_{t \in [0,1]} S_t^{(n)}$, stated in* [HS13, *Lemma 4.1]. The proof of this lemma given there relies on the statement that for a certain sequence (r_n) the inequalities (4.11) in* [HS13] *imply boundedness of (r_n). There are, however, examples of unbounded sequences which fulfill these inequalities. Still, things clear up nicely because the assertion of* [HS13, *Lemma 4.1] is a quick consequence of Corollary 1.4 and the Borell-TIS inequality.*

10. Coalescence probabilities in long-range seedbanks

In this section we will assume that the weights q_n of the renewal measure defined in (1.4) obey the asymptotics (1.10), see the discussion after equation (1.10). Following [BGKS13] we extend the model described in Section 1 as follows. For fixed $N \in \mathbb{N}$, the set of vertices of $\mathcal{G}_\mu$ is now $\mathbb{Z} \times [N]$. The set of those vertices whose first component is i constitutes the population of individuals living at time i. The parent of the individual (i, k) is $(i - R_{i,k}, H_{i,k})$, where the random variables $R_{i,k}$ are independent copies of R and the random variables $H_{i,k}$ are i.i.d. picks from $[N]$. In words, each individual chooses its parent uniformly from a previous time with delay (or dormancy) distribution μ. The corresponding urn model, which goes back to Kaj, Krone and Lascoux ([KKL01]), thus specialises to the Hammond-Sheffield urn for $N = 1$.

Again we write $(i, k) \sim (j, \ell)$ if the two individuals (i, k) and (j, ℓ) belong to the same connected component of $\mathcal{G}_\mu$. Thanks to Proposition 4.1, which also provides a proof of [BGKS13, Lemma 3.1 (c)], we arrive at the following analogue of Proposition 2.1 (see also [BGKS13, Theorem 3(c)]):

Proposition 10.1. *For all $k, \ell \in [N]$*

$$\mathbf{P}\left((0,k) \sim (i,\ell)\right) \sim C_{\alpha,N} \frac{i^{2\alpha-1}}{L(i)^2} \quad \text{as } i \to \infty, \tag{10.1}$$

where now

$$C_{\alpha,N} := \frac{1}{N + \sum_{m \geq 1} q_m^2} \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}.$$

Proof. Denote by $\tilde{A}_{0,k}$ and $\tilde{A}_{i,\ell}$ the decoupled ancestral lineages of the individuals $(0,k)$ and (i,ℓ) constructed in analogy to (4.1). Like in (4.2) we observe

$$\mathbf{E}\left[|\tilde{A}_{0,k} \cap \tilde{A}_{i,\ell}|\right] = \sum_{m \geq 0} \frac{1}{N} q_m q_{m+i}.$$

Decomposing at the most recent collision time of $\tilde{A}_{0,k}$ and $\tilde{A}_{i,\ell}$ we get

$$\mathbf{E}\left[\tilde{A}_{0,k} \cap \tilde{A}_{0,\ell}\right] = \mathbf{P}\left(\tilde{A}_{0,k} \cap \tilde{A}_{i,\ell} \neq \emptyset\right) \cdot \left(1 + \sum_{m \geq 1} \frac{1}{N} q_m^2\right).$$

The last two equalities combine to

$$\mathbf{P}\left(A_{0,k} \cap A_{i,\ell} \neq \emptyset\right) = \frac{\sum_{m \geq 0} q_m q_{m+i}}{N + \sum_{m \geq 1} q_m^2}.$$

The claimed asymptotics (10.1) is now immediate from Proposition 4.1. $\qquad\square$

Remark 10.2. *For given natural numbers n and N let $\mathscr{I}, \mathscr{J}, \mathscr{K}$ and $\mathscr{L}$ be independent and uniformly distributed on $[n] \times [N]$. In complete analogy to Proposition 2.3 one can derive that for all $\delta > 0$ and for a constant C not depending on n and N*

$$\mathbf{P}\left(\mathscr{I} \sim \mathscr{J} \sim \mathscr{K}\right) \quad \leq \quad \frac{C}{N^2} n^{4\alpha-2+\delta},$$

$$\mathbf{P}\left(\mathscr{I} \sim \mathscr{J} \sim \mathscr{K} \sim \mathscr{L}\right) \quad \leq \quad \frac{C}{N^3} n^{6\alpha-3+\delta}.$$

Consequently, along the lines of the proof of Theorem 1.1 one obtains the convergence of an analogue of $\left(S^{(n)}\right)$ towards fractional Brownian motion also in the long-range seedbank model of [BGKS13].

REFERENCES

[BGKS13] J. Blath, A. González Casanova, N. Kurt, and D. Spanò. The ancestral process of long-range seed bank models. *Journal of Applied Probability*, 50(3):741–759, 2013. URL http://www.jstor.org/stable/43283498.

[BGT87] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*. Cambridge University Press Cambridge [Cambridgeshire] ; New York, 1987. URL http://www.loc.gov/catdir/toc/cam027/86028422.html.

[Bil68] P. Billingsley. *Convergence of probability measures 2nd edition*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[CD19] F. Caravenna and R. Doney. Local large deviations and the strong renewal theorem. *Electronic Journal of Probability*, 24:1 − 48, 2019. doi:10.1214/19-EJP319.

[Don97] R. Doney. One-sided local large deviation and renewal theorems in the case of infinite mean. *Probability Theory and Related Fields*, 107:451–465, 04 1997. doi:10.1007/s004400050093.

[Dos98]   M. R. Dostanić. Spectral properties of the operator of riesz potential type. *Proceedings of the American Mathematical Society*, 126(8):2291–2297, 1998. URL http://www.jstor.org/stable/118743.

[Fel71]   W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[HS13]   A. Hammond and S. Sheffield. Power law Pólya's urn and fractional brownian motion. *Probability Theory and Related Fields*, 157(3):691–719, 2013. doi:10.1007/s00440-012-0468-6.

[KKL01]   I. Kaj, S. Krone, and M. Lascoux. Coalescent theory for seed bank models. *Journal of Applied Probability*, 38:285–300, 06 2001. doi:10.1017/S0021900200019860.

[Sot01]   T. Sottinen. Fractional Brownian motion, random walks and binary market models. *Finance Stoch.*, 5(3):343–355, 2001. doi:10.1007/PL00013536.

[Ste86]   C. Stein. Approximate computation of expectations. *Lecture Notes-Monograph Series*, 7:i–164, 1986. URL http://www.jstor.org/stable/4355512.

JAN LUKAS IGELBRINK, INSTITUT FÜR MATHEMATIK, JOHANNES GUTENBERG-UNIVERSITÄT MAINZ AND GOETHE-UNIVERSITÄT FRANKFURT, GERMANY.
    Email address: jigelbri@uni-mainz.de

ANTON WAKOLBINGER, INSTITUT FÜR MATHEMATIK, GOETHE-UNIVERSITÄT FRANKFURT, GERMANY.
    Email address: wakolbinger@math.uni-frankfurt.de