Random Genealogies: Examples, Scalings and Long-time behaviour

Dissertation
zur Erlangung des Grades “Doktor der Naturwissenschaften”
am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

Jan Lukas Igelbrink
geb. in Frankfurt am Main

Mainz, den 01. Dezember 2024

Datum der Disputation: 08. November 2024

Abstract

This thesis consists of three parts. Part I is dedicated to what we call the *Hammond-Sheffield urn*. This model is built from a collection of coalescing renewal processes with an increment distribution having power-law tails. Depending on the exponent of the power-law the random graph, which is induced on $\mathbb{Z}$ by the coalescing renewal processes, consists of infinitely many or one component almost surely. In the first regime rescaled sums over $\mathbb{R}$-valued random colourings of the connected components of this graph give rise to a discrete approximation of fractional Brownian motion. This is the main result of [HS13]. The core step is to prove that the scaling limit is a Gaussian process. This thesis gives a new proof based on Stein's method. The analysis that is required for this yields new insights into the genealogical structure of the random graph. This thesis then continues with a further analysis of the second regime in which the random graph is a tree.

Part II is dedicated to the analysis of the ancestral lineages of randomly sampled individuals in Galton-Watson trees according to different sampling schemes. The first sampling scheme relates to the uniform choice and was discussed by Cheek and Johnston in [CJ23]. This sampling scheme induces an *ancestral selection bias* of which our alternative proof provides a convenient interpretation. We analyse two other sampling schemes with the same new method.

The topic of Part III is the analysis of a variant of Muller's ratchet, a well-known model from population genetics. In the classical variant of Muller's ratchet individuals acquire deleterious mutations and have a selective advantage over individuals with less mutations, which is proportional to the difference in the number of mutations. After some time the type carrying the smallest mutational load gets lost from the population, this is called a click of the ratchet. In the model variant introduced by González Casanova, Smadi and Wakolbinger in [GSW23] the fitness advantage is just binary. Here we study the near-critical case in which the mutation-selection ratio is close to one, and we find a close correspondence to the so called Poisson profile approximation of the classical ratchet that was studied by Etheridge, Pfaffelhuber and Wakolbinger in [EPW09]. Our focus is then on the asymptotics of the clicktime of this variant of the ratchet.

Contents

Introduction

This thesis consists of three parts. Part I is dedicated to what we call the *Hammond-Sheffield urn*, a model introduced by Alan Hammond and Scott Sheffield in [HS13], there called a *power law Pólya's urn*. This model is built from a collection of coalescing renewal processes with an increment distribution having power-law tails. Specifically, every individual $i \in \mathbb{Z}$ selects its parent $i - R_i$ independently with a distance $R_i > 0$ following a power-law. This induces a random graph on $\mathbb{Z}$. Depending on the exponent of the tails of the distribution of R_i this random graph consists of infinitely many or one component almost surely.
In the first case rescaled sums over random colourings of this graph give rise to a discrete approximation of fractional Brownian motion (with Hurst parameter $H \in \left(\frac{1}{2}, 1\right)$), that is the unique centered Gaussian process with stationary increments having covariance function

$$(s,t) \mapsto \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right), \quad s,t \geq 0.$$

This is the main result of [HS13]. The core step is to prove that the scaling limit is a Gaussian process. In Chapter 1 of this thesis a new proof is given: Based on Stein's method, a normal approximation theorem for random colourings of random graphs is established and then applied to this model. The analysis that is required for the application of this theorem yields new insights into the genealogical structure of the random graph. In Chapter 2 this construction is then used to construct discrete approximations of *branching fractional Brownian motion*, of which the asymptotic speed of the maximum particle is analysed.
Sections S.1.1-S.1.3 and S.1.6-S.1.8 give a synopsis of Chapters 1 and 2. Section S.1.4, part of Section S.1.5.1, and Section S.1.5.2 present still unpublished work. In Section S.1.4 we show that the random graph has no bi-infinite path. In Section S.1.5.1 we analyse the depths of the most recent common ancestor of a pair of individuals in the HS-urn. In Section S.1.5.2 we study the scale of the depth of the MRCA of many individuals in the regime where the graph is an infinite tree and the ancestry of a block of individuals can be read as a coalescent. Proofs of the unpublished results of Section S.1.4 and S.1.5 can be found in Chapter 3. This chapter also outlines an alternative approach to the scale of the most recent common ancestor of a large block of individuals. In the preamble of Section S.1 we give references to some of the literature related to the HS-model.

Part II is dedicated to the analysis of the ancestral lineages of randomly sampled individuals in Galton-Watson trees (or more general Bellmann-Harris trees) according to different sampling schemes. The first sampling scheme relates to the uniform choice and was introduced by Cheek and Johnston in [CJ23]. This sampling scheme induces an *ancestral selection bias* of which our alternative proof provides a convenient interpretation. Another sampling scheme was introduced by Geiger in [Gei99]. Here one embeds the tree into the plane and traces back the left-most ancestral lineage still alive. The third sampling scheme was introduced by Chauvin, Rouault and Wakolbinger in [CRW91] and considers a tree-indexed continuous (in time and space) Markov process. Here one conditions on one particle being at a specific location at a specific time and analyses the conditional tree-indexed process. We show that our general approach can be used to analyse all three sampling schemes in a unified way. A more comprehensive introduction into Part II is given in Section S.2.

The topic of Part III is the analysis of a variant of Muller's ratchet, a well-known model from population genetics introduced by Herbert Muller in [Mul32] and [Mul64] to describe the accumulation of mutations in an asexually reproducing population. This model has also been used to explain the existence of sexual reproduction despite of its high cost. In the classical variant of Muller's ratchet individuals acquire deleterious mutations and have a selective advantage over individuals with less mutations, which is proportional to the difference in the number of mutations. After some time the type carrying the smallest mutational load gets lost from the population, this is called a click of the ratchet. In the model variant introduced by González Casanova, Smadi and Wakolbinger in [GSW23] the fitness advantage is just binary. In this work we study the near-critical case in which the mutation-selection ratio is close to one, and we find a close correspondence to the so called Poisson profile approximation of the classical ratchet that was studied in [EPW09]. Our focus is then on the asymptotics of the clicktime of this variant of the ratchet. In Section S.3.4 we give a short outlook on future work concerning the type frequency profile. A brief sketch of this future work is then found in Chapter 6. A more detailed introduction into Part III is given in Section S.3.

So far four chapters of this thesis have been published, and three of those have appeared in peer-reviewed journals:

- Chapter 1: J. L. Igelbrink and A. Wakolbinger. Asymptotic Gaussianity via coalescence probabilities in the Hammond-Sheffield urn. *ALEA Lat. Am. J. Probab. Math. Stat.*, 20(1):53–74, 2023. http://doi.org/10.30757/alea.v20-0. https://arxiv.org/abs/2201.06576

- Chapter 2: A. González Casanova and J. L. Igelbrink. Branching fractional Brownian motion: discrete approximations and maximal displacement. *arXiv preprint* arXiv:2310.04386, submitted. https://arxiv.org/abs/2310.04386.

- Chapter 4: J. L. Igelbrink and J. Ischebeck. Ancestral reproductive bias in continuous time branching trees under various sampling schemes. *J. Math. Biol.*, 89(11), 2024. https://doi.org/10.1007/s00285-024-02105-9. https://arxiv.org/abs/2309.05998.

- Chapter 5: J. L. Igelbrink, A. González Casanova, C. Smadi, and A. Wakolbinger. Muller's ratchet in a near-critical regime: tournament versus fitness proportional selection. *Theoretical Population Biology*, 158:121–138, 2024. https://doi.org/10.1016/j.tpb.2024.06.001. https://arxiv.org/abs/2306.00471.

My **share in the co-publications** included in the thesis is as follows.

Part I: All proofs in the journal publication [IW23] (Chapter 1 of the thesis, which is summarised in Sections S.1.1-S.1.3 of the synopsis) have been carried out by myself, prepared by discussions with my co-author Anton Wakolbinger. The conceptual novelties in [IW23] are due to both authors in equal parts.
The arXiv preprint [GI23] (Chapter 2 of the thesis, summarised in Sections S.1.6-S.1.8) was initiated by the idea of Adrián González Casanova to apply the approach from [IW23] to tree-indexed processes. He (as well my PhD advisors Professors Birkner and Wakolbinger) accompanied the genesis of this paper, whose proofs have been achieved in largest parts by myself.
The results contained in the thesis' unpublished part Sections S.1.4, S.1.5 and Chapter 3 have been achieved by myself, under advice of Professors Birkner and Wakolbinger. The idea of the second proof of Proposition S.1.5, which is presented in the last section of Chapter 3, is due to Jan Swart.

Part II: The journal publication [II24] (Chapter 4 of the thesis, summarised in Section S.2) is due to both authors in equal parts. The second author of [II24], Jasper Ischebeck, is a PhD student at Goethe-University Frankfurt.

Part III: The conceptual planning and design of the journal publication [IGSW24] (Chapter 5 of the thesis, summarised in Sections S.3.1-S.3.3) is due to its 4 co-authors in equal share. The proof of the main result of [IGSW24] (Section 5.4, 23 pages) has mostly been carried out by Charline Smadi and myself.
Section S.3.4 and Chapter 6 of the thesis give an outlook on current joint work with Charline Smadi and Anton Wakolbinger on aspects of random genealogy that complement [IGSW24]. This is based on Sections 4 and 5.2 of the arXiv preprint [IGSW23] and joint discussions during a research stay in Grenoble in May 2024. Here the share is equal between the three of us.

Synopsis

S.1 The Hammond-Sheffield urn: genealogy and scaling

In the paper [HS13] entitled *Power law Pólya's urn and fractional Brownian motion* Alan Hammond and Scott Sheffield propose a way how to successively colour the elements $i \in \mathbb{Z}$ by $+1$ or -1, given a configuration in $\{-1, +1\}^{\{k \in \mathbb{Z}: k < i\}}$, which describes the colouring of the integers below i.
The rule is as follows: Looking back from i, let i adopt the colour of $i - R_i$, where the R_i are a family of independent random variables with one and the same distribution μ. As proved in [HS13], this gives rise to a convex family of probability distributions on $\{-1, +1\}^{\mathbb{Z}}$ whose extremal elements are parametrised by the elements p of the unit interval, and in fact arise by a colouring (in terms of a ± 1 p-coin tossing) of the connected components of the graph that is generated by the parental relation stemming from this urn model. While [HS13] focuses from the very beginning on the dynamics of the random colourings, we tried to pursue the principle "first the genealogy, then the colouring". For probability distributions μ with power-tails, this led to a new proof of the main result of [HS13] (see Theorem S.1.1 below) as well as to interesting questions related to properties of the genealogical graph associated with the Hammond-Sheffield urn, whose analysis constitutes a substantial part of Chapters 1-3 of this thesis.

With a different focus Blath, González Casanova, Kurt, and Spanò, [BGKS13] consider an extension of this model: For fixed $N \in \mathbb{N}$, the set of vertices is now $\mathbb{Z} \times [N]$. Individual (i, k) is understood as the individual with number $i \in [N]$ living at time $k \in \mathbb{Z}$. The parent of the individual (i, k) is $(i - R_{i,k}, H_{i,k})$, where the random variables $R_{i,k}$ are independent with distribution μ and the random variables $H_{i,k}$ are independent uniform picks from $[N]$. In words, each individual chooses its parent uniformly from a previous time with delay (or dormancy) distribution μ. For μ with compact support this goes back to Kaj, Krone and Lascoux [KKL01]. Note that this specialises to the Hammond-Sheffield urn for $N = 1$, see Section 1.10 for a more elaborate discussion and analysis of this extension.
A similar model has been studied in [CKT20] by Chierichetti, Kumar and Tomkins. The difference is the following: While a HS-urn always has an infinite history, [CKT20] consider a model starting with a finite history of size h, where individual i selects its parent according to some (power-)law μ conditioned on $\{1, 2, \ldots, i + h - 1\}$. They consider the individuals $-h + 1, \ldots, 0$ having certain types and ask questions about the loss of types. Anderson,

Kumar, Tomkins, Andrew and Vassilvitskii [AKTV14] describe this model as useful for modeling reconsumption likelihoods. [CKT20] even highlights the usefulness for the modeling of the world-wide-web graph.

S.1.1 The Hammond-Sheffield urn as a random graph on $\mathbb{Z}$

For $0 < \alpha < 1$ 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

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

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

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

Let R be an $\mathbb{N}$-valued random variable with distribution μ and let $(R_i)_{i\in\mathbb{Z}}$ be a family of independent copies of R. With these ingredients, let $\mathcal{G}_\mu$ be the random directed graph with vertex set $\mathbb{Z}$ and edge set $E\left(\mathcal{G}_\mu\right)$ given by

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

This induces the random equivalence relation

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

Note that the symbol $\sim$ is used in (S.1.1) and (S.1.3) in two different meanings; this will cause no risk of confusion. The Hammond-Sheffield urn induces a natural definition of ancestral

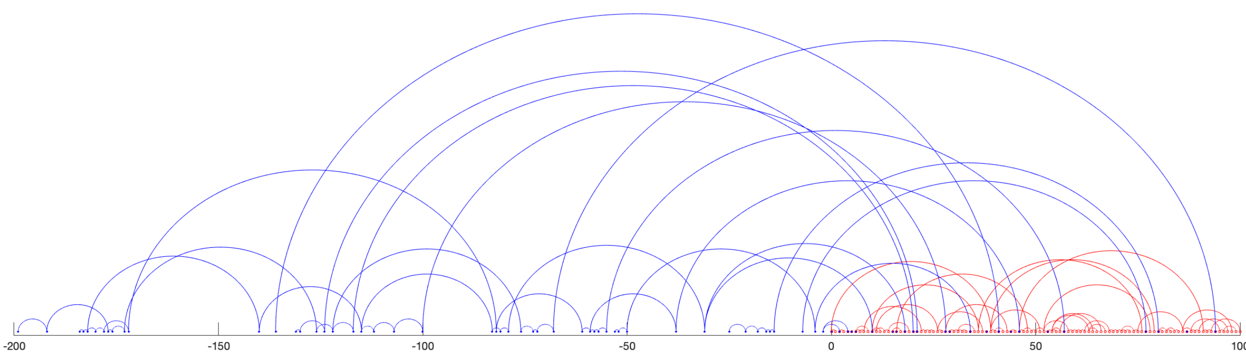


Figure S.1: A realisation of the ancestral lineages of the individuals $\{0, \ldots, 100\}$ traced back till -200. Each of the arcs corresponds to an edge of $\mathcal{G}_\mu$. All the outgoing edges from $i = 0, \ldots, 100$ which map to an individual in $\{0, \ldots, 100\}$ are drawn (in red), whereas 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 (S.1.1) was chosen as 0.39.

lineages, see also [HS13, Lemma 2.1 and its proof]: For $i \in \mathbb{Z}$ we call $i - R_i$ the parent of i.

In correspondence to this we call any $j \in \mathbb{Z}$ with $j - R_j = i$ a child of i and define the k-th ancestor of i recursively by

$$A_i(0) := i, \qquad A_i(k) := A_i(k-1) - R_{A_i(k-1)}. \tag{S.1.4}$$

The ancestral lineage of i is defined by

$$A_i := \left\{ A_i(\ell) : \ell \in \mathbb{N}_0 \right\}. \tag{S.1.5}$$

This gives rise to *coalescing renewal processes* starting from the integers.

For $i \in \mathbb{Z}$ the connected component containing i is denoted by $\mathscr{C}_i$. 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 relatives of i, see Figure S.1 for an illustration.

For $n \in \mathbb{N}$ the probability that 0 belongs to the ancestral lineage of n is thus given by the weight assigned to n by the renewal measure,

$$q_n := \mathbf{P}\left(\tilde{R}_1 + \ldots + \tilde{R}_j = n \text{ for some } j \geq 0\right) \tag{S.1.6}$$

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.)

We have the following dichotomy, see [HS13, Proposition 2]: If $\alpha > \frac{1}{2}$, then $\mathcal{G}_\mu$ has one component almost surely, while if $\alpha < \frac{1}{2}$ it has infinitely many components almost surely. (The case $\alpha = \frac{1}{2}$ depends on the slowly varying function L.) In the first regime we refer to $\mathcal{G}_\mu$ as the HS-tree. A synopsis about unpublished results on this can be found in Section S.1.5. In the second regime the random graph is a forest. In this regime [HS13] suggests the picture of an urn in which types get copied. A synopsis about our results published so far can be found in Sections S.1.2, S.1.3 and Sections S.1.6-S.1.8. Results about pair coalescences in both regimes are stated in Section S.1.5.

S.1.2 Discrete approximation of fractional Brownian motion via the HS-urn

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 (FBM), which is the unique centered Gaussian process with variance function t^{2H}, stationary increments and a.s. continuous paths.

As indicated at the beginning of this section, the individuals' types arise as follows: Assume that each component of $\mathcal{G}_\mu$ 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.

$$Y = \xi - (2p-1) \text{ with } \mathbf{P}(\xi = +1) = p, \mathbf{P}(\xi = -1) = 1 - p. \tag{S.1.7}$$

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)

$$S_n := \sum_{i=1}^{n} Y_i, \quad n = 0, 1, \ldots. \tag{S.1.8}$$

By construction,

$$\sigma_n^2 := \mathbf{Var}[S_n] = \sum_{i,j\in[n]} \mathbf{Cov}[Y_i, Y_j] = \mathbf{E}[Y^2] \sum_{i,j\in[n]} \mathbf{P}(i \sim j). \tag{S.1.9}$$

For $\frac{i-1}{n} \le t \le \frac{i}{n}$, $i, n \in \mathbb{N}$, let $S^{(n)}(t)$ be the linear interpolation of S_i/σ_n and S_{i+1}/σ_n. Then $S^{(n)}(t)$ approximates fractional Brownian motion in the sense of the following Theorem.

Theorem S.1.1 (Theorem 1.1.1, Corollary 1.1.2, Proposition 1.1.3, Corollary 1.1.4)**.** *Let μ be a probability measure on $\mathbb{N}$ satisfying* (S.1.1) *and* (S.1.2) *with $0 < \alpha < \frac{1}{2}$. Assume one of the following conditions (A) or (B):*

(A) The colouring Y is given by (S.1.7)*.*

(B) The weights q_n of the renewal measure specified in (S.1.6) *satisfy the asymptotics*

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

and the colouring Y obeys

$$\mathbf{E}\left[Y\right] = 0 \textit{ and } 0 < \mathbf{E}\left[Y^4\right] < \infty. \tag{S.1.11}$$

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

While part (A) corresponds to [HS13, Theorem 1.1], we relax in part (B) the assumption of (S.1.7) on the colouring to (S.1.11), but need to assume that the asymptotics (S.1.10) hold. This condition is equivalent to the validity of the Strong Renewal Theorem for the renewal process with an increment distribution μ satisfying (S.1.1) and (S.1.2), see Caravenna and Doney [CD19], whose Theorem 1.4 gives necessary and sufficient conditions in terms of μ for the validity of (S.1.10). A well-known sufficient condition for (S.1.10) is the criterion of Doney [Don97]

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

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

We now give a sketch of proof of Theorem S.1.1. [HS13, Lemma 3.1] show by Fourier and Tauberian arguments that

$$\sum_{i,j\in[n]} \mathbf{P}(i \sim j) \sim \frac{C_\alpha}{\alpha(2\alpha+1)} \frac{n^{2\alpha+1}}{L(n)^2} \quad \text{as } n \to \infty, \tag{S.1.12}$$

with

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

In the proof of Proposition 1.2.1 we will obtain (S.1.12) in an alternative way by analysing $\mathbf{P}(i \sim j)$ under the condition (S.1.10). Because of (S.1.9) and (S.1.12), 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* $H = \frac{1}{2} + \alpha$. 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. Indeed, this is the most intricate part of the proof. What is then left for obtaining a functional limit theorem is a tightness argument. For the latter we use the same proof strategy as Sottinen in [Sot01, Proof of Theorem 1] and apply [Bil68, Theorem 13.5] by Billingsley. See Section 1.9 for the details.

Under assumption (A) of Theorem S.1.1, for each fixed $t > 0$ the 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; from the very beginning they make substantial use of the specific form (S.1.7) of the colouring of the random graph $\mathcal{G}_\mu$. In [HS13] it is not explicitly discussed whether these arguments also carry over from $S_{\lfloor tn \rfloor}$ to $S_{\lfloor t_1 n \rfloor}, \ldots, S_{\lfloor t_m n \rfloor}$. However, again thanks to the specific assumption (S.1.7) on Y one can check that this is indeed the case, thus rendering the asserted asymptotic Gaussianity of the finite dimensional distributions of $S^{(n)}$.

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 1.8, is based on insights into the structure of $\mathcal{G}_\mu$. A key ingredient is a Stein-Theorem for sequences of coloured partitions (that can be more general than the ones obtained from the HS urn), see Theorem S.1.2.

S.1.3 Asymptotic Gaussianity in randomly coloured random partitions

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.

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

Let Y be a real valued random variable obeying (S.1.11). Thinking again 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. In correspondence to S_n, see (S.1.8), we then define

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

Now 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

$$\mathscr{Z}^{(m)} := \left(Z_{\lfloor \rho_1 m \rfloor}^{(m)}, \ldots, Z_{\lfloor \rho_d m \rfloor}^{(m)} \right) \tag{S.1.13}$$

as $m \to \infty$.

Theorem S.1.2 (Theorem 1.3.1)**.** *For all $m \in \mathbb{N}$ let 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 $\left(Y_i^{(m)}\right)_{i \in [m]}$. The sequence of $\mathbb{R}^d$-valued random variables $\mathscr{Z}^{(m)}$ defined in (S.1.13) is asymptotically Gaussian as $m \to \infty$ provided that as $m \to \infty$*

$$\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \overset{m}{\sim} \mathscr{K}^{(m)}\right) = o\left(\left(\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)}\right)\right)^{3/2}\right), \tag{S.1.14}$$

$$\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \overset{m}{\sim} \mathscr{K}^{(m)} \overset{m}{\sim} \mathscr{L}^{(m)}\right) = o\left(\left(\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)}\right)\right)^{2}\right), \tag{S.1.15}$$

and

$$\mathbf{Cov}\left[I_{\{i \overset{m}{\sim} j\}}, I_{\{k \overset{m}{\sim} \ell\}}\right] \leq \mathbf{P}\left(i \overset{m}{\sim} j \overset{m}{\sim} k \overset{m}{\sim} \ell\right) \quad \text{for all } m \in \mathbb{N} \text{ and } i, j, k, \ell \in [m], \tag{S.1.16}$$

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

$$\sum_{i,j=1}^{m} a_i^{(m)} a_j^{(m)} \mathbf{P}\left(i \overset{m}{\sim} j\right) \geq \widetilde{C} \sum_{i,j=1}^{m} \mathbf{P}\left(i \overset{m}{\sim} j\right), \quad m \in \mathbb{N}. \tag{S.1.17}$$

The proof of this theorem as well as that of the key Proposition 1.3.3 is based on Stein's method [Ste86]. Proposition 1.3.3 bounds the distance of the distribution of

$$\frac{Z_m^{(m)}}{\sqrt{\mathbf{Var}\left[Z_m^{(m)}\right]}}$$

and 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 (S.1.11).

To apply Theorem S.1.2 to the HS-urn conditions, (S.1.14) and (S.1.15) are then checked by computing pair, triplet and quadruple coalescence probabilities in Proposition 1.2.1, Lemma 1.6.1 and Lemma 1.6.2. In essence, this is achieved by computing the expected number of common ancestors for independent ancestral lineages. Condition (S.1.16) is then checked in Lemma 1.2.5 by a qualitative analysis of the graph structure of $\mathcal{G}_\mu$. Condition 1.3.6 is checked in Section 1.8 with a general statement about Riesz kernels, see also [Dos98].

Remark S.1.3. In [Dro23] Drogin applies a method similar as the one of [HS13] to long-range (power-law) voter models on $\mathbb{Z}$. In a suitable scaling limit he obtains a Gaussian field, which can be understood as a family of fractional Brownian motions, there called *dynamical fractional Brownian motion*. Applying Theorem S.1.2 to d-dimensional voter models is the topic of ongoing work with González Casanova and Hammond.

S.1.4 The HS-trees do not contain a spine

We show that (informally spoken) the random graph $\mathcal{G}_\mu$ does not contain a bi-infinite path. In view of Lemma S.1.7 and Proposition S.1.23 this can be seen as a qualitative pendant to the assumption (S.1.43) made in Proposition S.1.27.

Definition S.1.4. A set $S \subset \mathcal{G}_\mu$ is called a *spine* if $\sup S = \infty$ and for all $i \in S$ the parent of i is $\max\{j \in S : j < i\}$.

Proposition S.1.5. *Let $\alpha \in (0,1)$. Almost surely the random graph $\mathcal{G}_\mu$ does not contain a spine.*

Remark S.1.6. As the proof below shows in fact Proposition S.1.5 would also hold for more general μ as long as $\sum_k k\mu\left(\{k\}\right)$ is not finite.

We will prove this in two ways, see below and Section 3.2.1 for a proof via a martingale suggested by Jan Swart. Broutin, Devroye, Lugosi and Oliveira [BDLO23] analysed similar models applying a method that bears similarities to the method in our proof below and also the one applied in our alternative proof in Section 3.3. In fact, Proposition S.1.5 can also be obtained as a corollary from [BDLO23, Theorem 3.ii)]. The proof strategy there is to show that there is almost surely no more than one spine, and that the existence of one spine implies the existence of two spines. The latter then gives a contradiction since there exists no stationary renewal process with increment distribution μ, but every ancestral lineage backwards is a renewal process with increment distribution μ. The uniqueness argument there and here in our proof is in the style of the arguments used by Burton and Keane in [BK89] to prove the uniqueness of an infinite percolation cluster.

Set

$$t(\ell) := \mathbf{P}\left(0 \notin A_\ell \text{ and } \exists y > \ell : 0 \sim y\right) \tag{S.1.18}$$

and set

$$d(x) = \frac{1}{x}\sum_{\ell=1}^{x} t(\ell). \tag{S.1.19}$$

The quantity $t(\ell)$ and similar ones will play an essential role in analysing the most recent common ancestor of the HS-tree later on, see Section S.1.5. Proposition S.1.5 is then enough to show the following Lemma.

Lemma S.1.7. *For $\alpha \in (0,1)$ we have $d(x) \to 0$ for $x \to \infty$.*

Essentially this means that in a large interval in $-\mathbb{N}_0$ no macroscopic fraction of individuals has a descendant to the right of zero.

Proof. We have

$$t(x) \leq \mathbf{P}\left(\exists y > x : 0 \sim y\right),$$

which by Proposition S.1.5 converges to zero as $x \to \infty$. This implies

$$d(x) \leq \frac{1}{x}\mathbf{E}\left[\#\left\{-x \leq z \leq 1 : \exists y > 0 \text{ such that } z \sim y\right\}\right] = \frac{1}{x}\sum_{\ell=1}^{x} t(\ell) \to 0. \tag{S.1.20}$$

□

The key for proving Proposition S.1.5 via shift invariance of the distribution of the random graph $\mathcal{G}_\mu$ is the following lemma.

Lemma S.1.8. *Almost surely the random graph $\mathcal{G}_\mu$ does not contain infinitely many spines.*

Proof. Let G_n be the number of children of $\{0, \ldots, n\}$ above n, namely

$$G_n := \sum_{k=n+1}^{\infty} \mathbf{1}_{R_k \in [k-n,k]}. \tag{S.1.21}$$

Then

$$\begin{aligned}
\mathbf{E}\left[G_n\right] &= \sum_{k=n+1}^{\infty} \sum_{j=k-n}^{k} \mathbf{P}\left(R = j\right) \\
&= \sum_{k=1}^{\infty} \sum_{j=k}^{n+k} \mathbf{P}\left(R = j\right) \\
&= \sum_{k=1}^{\infty} \sum_{j=0}^{n} \mathbf{P}\left(R = j + k\right) \\
&= \sum_{j=0}^{n} \sum_{k=1}^{\infty} \mathbf{P}\left(R = j + k\right) \\
&\leq \text{const} \cdot \sum_{j=0}^{n} (j+1)^{-\alpha} \\
&\leq \text{const} \cdot (n+1)^{1-\alpha}.
\end{aligned}$$

So

$$\frac{G_n}{n} \xrightarrow{\mathcal{L}_1} 0 \text{ for } n \to \infty. \tag{S.1.22}$$

This gives that the number of splitting points of spines (i.e. points from which there exist two disjoint paths to $+\infty$) does not have a positive density. By stationarity there are none. This gives the assertion. □

Proof of Proposition S.1.5. Set

$$X_m := \mathbf{1}_{m \text{ lies in a spine}} \cdot \tag{S.1.23}$$

The sequence $(X_m)_{m\in\mathbb{Z}}$ is stationary. Let $\mathcal{I}$ be the shift-invariant σ-algebra, then by Birkhoff's ergodic theorem (see [Kle08, Satz 20.14])

$$\mathbf{E}\left[X_0|\mathcal{I}\right] = \lim_{K\to\infty} \frac{1}{2K+1} \sum_{\ell=-K}^{K} X_\ell. \tag{S.1.24}$$

So if there exists one ore more spines they have a positive density. If there exists more than one spine, they coalesce. Set

$$Y_m := \mathbf{1}_{m \text{ is a splitting point of two spines}}, \tag{S.1.25}$$

then also $(Y_m)_{m\in\mathbb{Z}}$ is a stationary sequence. And with the same arguments we get that also the splitting points of spines must have a positive density, if there exists more than one spine. Since we do not have infinitely many spines by Lemma S.1.8 we can assume that there exists either one or no spine. Assume that there exists a spine S, then

$$\left(\mathbf{P}\left(B \subset S\right)\right)_{B\subset\mathbb{Z}} \tag{S.1.26}$$

characterises its distribution. Using the notation

$$B + x := \{y + x : y \in B\} \tag{S.1.27}$$

we get for $B = \{k_1, \ldots, k_n\} \subset \mathbb{Z}$, $k_1 < \ldots < k_n$

$$\begin{aligned}
\mathbf{P}\left(B \subset S\right) &= \mathbf{P}\left(k_1, \ldots, k_{n-1} \in S | k_n \in S\right) \mathbf{P}\left(k_n \in S\right) \\
&= \mathbf{P}\left(k_1, \ldots, k_{n-1} \in A_{k_n}\right) \mathbf{P}\left(k_n \in S\right) \\
&= \mathbf{P}\left(k_{1+x}, \ldots, k_{n-1+x} \in A_{k_n+x}\right) \mathbf{P}\left(k_n + x \in S\right) \\
&= \mathbf{P}\left(k_{1+x}, \ldots, k_{n-1+x} \in S | k_{n+x} \in S\right) \mathbf{P}\left(k_{n+x} \in S\right) \\
&= \mathbf{P}\left(B + x \subset S\right) \\
&= \mathbf{P}\left(B \subset S - x\right).
\end{aligned}$$

So S is stationary. Conditional on $\{x \in S\}$ the set $S \cap [-\infty, x]$ is equal in distribution to the range of a renewal process with increment distribution μ. Since there exists no stationary renewal process with increment distribution μ there exists no spine. □

S.1.5 Depth of most recent common ancestors in the Hammond-Sheffield urn

As outlined at the end of Section S.1.3, the application of Stein's method to derive Theorem S.1.1 required an analysis of coalescence probabilities in the regime $\alpha \in \left(0, \frac{1}{2}\right)$. We now continue with an analysis of the genealogical structure of the random graph $\mathcal{G}_\mu$. Let μ again satisfy (S.1.1) and (S.1.2).

Recall from (S.1.6) that the probability that 0 belongs to the ancestral lineage of n is given by the weight assigned to n by the renewal measure,

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

with $\tilde{R}_1, \tilde{R}_2, \ldots$ being independent copies of R.
We are interested in the question how fast the ancestral lineages $A_1, \ldots, A_n$, see (S.1.5), coalesce. For this purpose we set

$$\mathrm{MRCA}(i, j) := \max\left(A_i \cap A_j\right)$$

and for any non-empty $B \subset \mathbb{Z}$

$$\mathrm{MRCA}(B) := \max\left(\bigcap_{i \in B} A_i\right).$$

We here stick to the convention that the maximum of an empty set is $-\infty$.

S.1.5.1 Most recent common ancestor of a pair

We start with a result on the most recent common ancestor of two individuals in the regime $\alpha \in \left(0, \frac{1}{2}\right)$. This result was an ingredient in the proof of Theorem S.1.1; we include it here also for comparison with the next theorem in the other regime $\alpha \in \left(\frac{1}{2}, 1\right)$, to which it has striking parallels.

Proposition S.1.9 (Proposition 1.2.4)**.** *Let $\alpha \in \left(0, \frac{1}{2}\right)$. Conditioned under $\{0 \sim n\}$ the sequence of random variables*

$$-\frac{\mathrm{MRCA}(0, n)}{n}$$

converges as $n \to \infty$ in distribution to a Beta'-distribution with parameters α and $1 - 2\alpha$, that is a random variable having density

$$\frac{x^{\alpha-1}(1+x)^{\alpha-1}}{\mathrm{Beta}(\alpha, 1-2\alpha)}\,\mathrm{d}x, \qquad x > 0.$$

See Section 1.5 for a proof.

Remark S.1.10. Let $\beta, \gamma > 0$. The distribution with density

$$\frac{x^{\beta-1}(1+x)^{-\beta-\gamma}}{\mathrm{Beta}(\beta, \gamma)}\,\mathrm{d}x, \qquad x > 0 \tag{S.1.28}$$

is called Beta'-distribution (pronounced *Beta-prime*) with parameters (β, γ). It arises as the distribution of

$$Y := \frac{X}{1-X}$$

provided X follows a Beta-distribution with parameters (β, γ). Naturally this gives, that $1/Y$ is Beta'-distributed with parameters (γ, β).

Recall that for $\alpha \in \left(\frac{1}{2}, 1\right)$ the random graph $\mathcal{G}_\mu$ has only one component almost surely. In this case $\mathcal{G}_\mu$ is a tree, and we call it the Hammond-Sheffield Tree (HS-Tree). For the latter we can prove a statement very similar to Proposition S.1.9.

Theorem S.1.11. *For $\alpha \in \left(\frac{1}{2}, 1\right)$ the distribution of*

$$-\frac{\mathrm{MRCA}(0,n)}{n}$$

converges to a Beta'-distribution with parameters $(1-\alpha, 2\alpha-1)$ as $n \to \infty$, that is a random variable having density

$$\frac{x^{-\alpha}(1+x)^{-\alpha}}{\mathrm{Beta}\left(1-\alpha, 2\alpha-1\right)}\,\mathrm{d}x, \qquad x > 0.$$

This is proved in Section 3.1.2.

Remark S.1.12. Note that in both regimes $\alpha \in \left(\frac{1}{2}, 1\right)$ and $\alpha \in \left(0, \frac{1}{2}\right)$ we obtain a Beta'-distribution, once with parameters $(\alpha, 1-2\alpha)$ and once with parameters $(1-\alpha, 2\alpha-1)$. The parameters are mirrored around $\frac{1}{2}$ in the sense that for $\alpha = \frac{1}{2} - \varepsilon$ the asymptotic distribution of $\mathrm{MRCA}(0,n)/n$ conditioned on $\{0 \sim n\}$ is asymptotically equal to the asymptotic distribution of $\mathrm{MRCA}(0,n)/n$ for $\alpha = \frac{1}{2} + \varepsilon$. See also Remark 3.2.14 for an interpretation of this in a setting of Bessel random walks.

Remark S.1.13. In the case $\alpha \in \left(\frac{1}{2}, 1\right)$ the most recent common ancestor of a finite set $B \subset \mathbb{Z}$ is almost surely finite. Hence for all $n \in \mathbb{N}$ the tree of the ancestors of $[n]$ coalesces to one single ancestral lineage. See Figure S.2 for an illustration.

Both Proposition S.1.9 and Theorem S.1.11 analyse the depth of the most recent common ancestor of two individuals with a large distance. The following gives a similar result for two individuals with a fixed distance.

Proposition S.1.14. *For $\alpha \in (\frac{1}{2}, 1)$ and $n \in \mathbb{N}$ we have the asymptotics*

$$\mathbf{P}\left(-\frac{\mathrm{MRCA}(0,n)}{n} > x\right) \sim \widetilde{C}(n,\alpha)x^{1-2\alpha} \qquad \text{as } x \to \infty$$

for a constant $\tilde{C}(n,\alpha)$.

This is proved in Section 3.1.3. Note that the tails have the same exponent as in Theorem S.1.11.

Remark S.1.15. We now give a heuristic explanation for Proposition S.1.14 and Theorem S.1.11. Let n be fixed as in Proposition S.1.14. Typically an individual has of order x^α ancestors in an interval of length x. If now one of the two ancestral lineages behaves typically, the other one makes at most of order $x^{1-\alpha}$ jumps and one of them is of order x. This happens with probability of order $x^{1-\alpha} \cdot x^{-\alpha} = x^{1-2\alpha}$. More jumps would lead to a coalescence event since the probability to hit one of the ancestors of the other line is roughly $x^{\alpha-1}$ for every jump. This explains the tail probability obtained in Proposition S.1.14 heuristically.
Now consider the event $\left\{\mathrm{MRCA}(0,n)/n \leq -x\right\}$ as $n \to \infty$: In order to arrive at Theorem S.1.11 (and also the following Proposition S.1.16) we think of the tree in a rescaled

version and appeal to self-similarity: One individual will make the typical amount of $(xn)^\alpha$ many jumps, while the other individual will make of order $x^{1-\alpha} \cdot n^\alpha$ many jumps of which one needs to be of size xn. This happens thus with probability of order $x^{1-\alpha} \cdot x^{-\alpha} = x^{1-2\alpha}$, which corresponds to the tails of the Beta'-distribution in Theorem S.1.11 (and the moderate deviations result of Proposition S.1.16 below).

In the regime $\alpha \in \left(0, \frac{1}{2}\right)$, the proof of Proposition S.1.9 makes use of the fact that for two independent ancestral lineages A_0 and A_n, starting in 0 respectively n, we have

$$\begin{aligned}\sum_{k=0}^{xn} q_k q_{k+n} &= \mathbf{E}\left[|A_0 \cap A_n \cap [-xn, 0]|\right] \\ &= \sum_{k=0}^{xn} \left(\mathbf{P}\left(\mathrm{MRCA}(0,n) = -k\right) \sum_{\ell=0}^{xn-k} q_\ell^2 \right).\end{aligned}$$

This will be then enough to suitably bound $\mathbf{P}\left(\mathrm{MRCA}(0,n) \geq -k\right)$ from below and above. In the regime $\alpha \in \left(\frac{1}{2}, 1\right)$ the proofs of Theorem S.1.11 and Proposition S.1.14 given in Section 3.1 use a method similar to the one applied by Bertoin in [Ber99, Corollary 13] for the ranges of α-stable subordinators: For

$$T \overset{(d)}{=} \mathrm{Exp}(\lambda)$$

and two independent ancestral lineages A_0 and A_n, starting in 0 respectively n, we observe that the Laplace transform of the random variable $|\mathrm{MRCA}(0,n)|$ obeys

$$\begin{aligned}\mathcal{L}\left(\left|\mathrm{MRCA}(0,n)\right|\right)(\lambda) &= \mathbf{E}\left[e^{-\lambda \cdot \left|\max\{A_0 \cap A_n\}\right|}\right] \\ &= \mathbf{E}\left[\mathbf{P}\left(T > \left|\max\left\{A_0 \cap A_n\right\}\right| \Big| \max\left\{A_0 \cap A_n\right\}\right)\right] \\ &= \mathbf{E}\left[\mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset \middle| \max\left\{A_0 \cap A_n\right\}\right)\right] \\ &= \mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset\right),\end{aligned}$$

and for an independent copy $\tilde{A}_0$ of A_0

$$\mathbf{E}\left[\left|A_0 \cap A_n \cap [-T, 0]\right|\right] \tag{S.1.29}$$

$$= \mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset\right) \mathbf{E}\left[\left|\left|A_0 \cap \tilde{A}_0 \cap [-T, 0]\right|\right|\right]. \tag{S.1.30}$$

Since the expectation in (S.1.29) and (S.1.30) can be expressed in terms of the renewal function $(q_k)_{k\geq 0}$ we are able to obtain precise asymptotics. This is enough to conclude the main result by essentially inverting the Laplace transform. See Section 3.1.2 for more details. Analysing the asymptotics of $\mathcal{L}\left(\max\left\{A_0 \cap A_n\right\}\right)(\lambda)$ in more detail and using Tauberian Theorems by Feller [Fel71, Chapter XIII.5] (or equivalently by Bingham, Goldie and Teugels [BGT87, Chapter 1.7]), reveals then the following result about moderate deviations (whose proof will be given in Section 3.1.4).

Proposition S.1.16. *Let $(\theta_n)_{n\in\mathbb{N}}$ be a positive sequence converging to zero such that*

$$\frac{1}{C_q^2}\sum_{\ell\geq 0} q_\ell\left(q_\ell - q_{\ell+n}\right) e^{-\frac{\theta_n \ell}{n}} \sim n^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(x+1)^{\alpha-1}\right)\mathrm{d}x \qquad \text{as } n\to\infty \tag{S.1.31}$$

and

$$\tilde{C}(\alpha) := \frac{2-2\alpha}{(2\alpha-1)}\cdot\frac{\Gamma(\alpha)}{\Gamma(1-\alpha)\cdot\Gamma(2\alpha-1)}. \tag{S.1.32}$$

Then for $\alpha\in(\frac{1}{2},1)$ and all $x>0$

$$\mathbf{P}\left(-\frac{\mathrm{MRCA}(0,n)}{n} > \frac{x}{\theta_n}\right) \sim \tilde{C}(\alpha) x^{1-2\alpha}\theta_n^{2\alpha-1} \qquad \text{as } n\to\infty.$$

In particular for $\theta_n = n^{-\delta}$ this yields

$$\mathbf{P}\left(-\mathrm{MRCA}(0,n) > n^{1+\delta}\right) \sim \tilde{C}(\alpha)\left(n^\delta\right)^{1-2\alpha}. \tag{S.1.33}$$

Note that for $\theta_n \equiv 0$ the condition (S.1.31) is just the standard Riemann sum approximation. Essentially (S.1.31) ensures that θ_n converges to zero sufficiently fast.

S.1.5.2 Scale of the most recent common ancestor of $[n]$

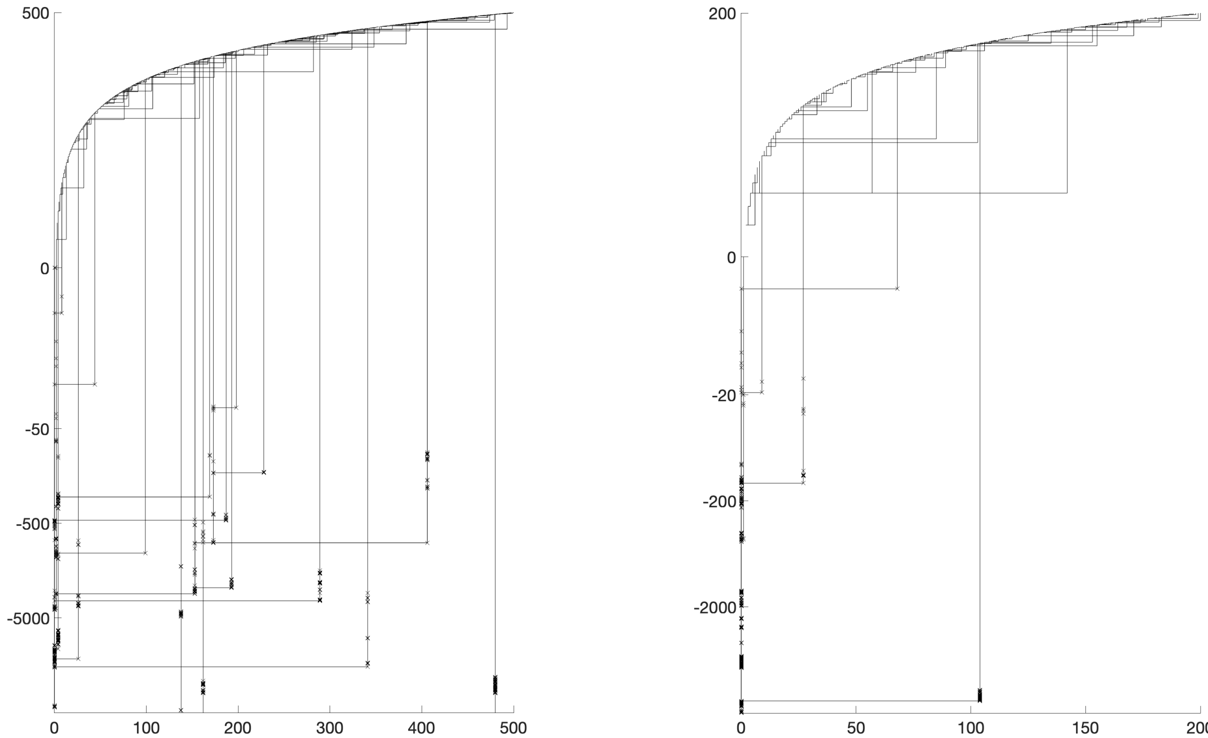
Let $\alpha > \frac{1}{2}$. On which scale does the bulk of individuals $[n] = \{1,2,\dots,n\}$ find its common ancestor as $n\to\infty$? This can be seen as a question on the asymptotics of the depth of the HS-tree (or HS-coalescent) with leaves $1,2,\dots n$. We know that the number of ancestral lineages of $[n]$ that arrive in the negative integers (as well as of those that arrive below $-n$) is of the order $n^{1-\alpha}$, and we will give arguments which make it plausible that the number of ancestral lineages that arrive below $-n^\beta$ is of order $n^{1-\beta\alpha}$ for $1\leq\beta\leq\frac{1}{\alpha}$. Since also the number of individuals in $[n]$ whose parent is below $-n^\beta$ is of the order $n\cdot(n^\beta)^{-\alpha} = n^{1-\beta\alpha}$, this (roughly) amounts to the conjecture that for any $\beta\in(1,1/\alpha)$ the lion's share of the number of ancestral lineages that intersect the interval $(-n^\beta,0)$ coalesce with each other still on scale n^β for $1<\beta<\frac{1}{\alpha}$.

Definition S.1.17. Let $(s_n)_n$ be a positive sequence with $s_n\uparrow\infty$. We say that a sequence of random variables $(X_n)_n$ *lives on scale* $(s_n)_n$ if $(|X_n|/s_n)_n$ is tight and stochastically bounded away from 0.

Conjecture S.1.18. *The sequence $\left(\mathrm{MRCA}\left([n]\right)\right)_n$ lives on the scale $(n^{1/\alpha})_n$.*

Remark S.1.19. The sequence $\left(\mathrm{MRCA}\left([n]\right)/n^{1/\alpha}\right)_n$ is stochastically bounded away from zero. Indeed, we know that

$$\begin{aligned}&\liminf_{n\to\infty}\mathbf{P}\left(\mathrm{MRCA}\left([n]\right)\leq -n^{1/\alpha}\right)\\ \geq\ &\liminf_{n\to\infty}\mathbf{P}\left(\exists \text{ an individual in } [n] \text{ whose parent is below } -n^{1/\alpha}\right).\end{aligned} \tag{S.1.34}$$



(A) **(B)**

Figure S.2: This is a simulation of the ancestral lineages of the individuals $0, \ldots, n$, with $n = 500, \alpha = 0.65$ in panel (A) and $n = 200, \alpha = 0.85$ in panel (B). Direction of time is vertical on logarithmic scale, and horizontal lines mark coalescence events. The two panels give an impression how the genealogy of the individuals $i \in [n]$ scales with n.

Every individual $i \in [n]$ now has a chance greater than $(2n^{1/\alpha})^{-\alpha}$ to have its parent below $n^{1/\alpha}$. So (S.1.34) is bounded from below by

$$\liminf_{n\to\infty} \left(1 - \left(2n^{1/\alpha}\right)^{-\alpha}\right)^n = \liminf_{n\to\infty} \left(1 - \frac{2^{-\alpha}}{n}\right)^n,$$

which is bounded from below by $\frac{1}{2}e^{2^{-\alpha}}$.

Therefore the sequence $\left(\text{MRCA}\left([n]\right)/n^{1/\alpha}\right)_n$ is stochastically bounded away from 0.

Remark S.1.20. A stronger common stochastic bound than the one constructed above in Remark S.1.19 can be constructed in the following way: The most recent common ancestor of $[n]$ must be below $X = -\max_{i=1,\ldots,n-1} R_i$, which has a distribution $\hat{\mu}$ on scale $n^{\frac{1}{\alpha}}$, where $\hat{\mu}$ has the same tail behaviour as μ. Now X and n still have to find their most recent common ancestor, which leads back to a *pairwise* situation. Indeed Theorem S.1.11 gives that the so constructed bound on $(\text{MRCA}([n])/n^{1/\alpha})_n$ is equal in distribution to

$$-\hat{R} \cdot B \quad \text{where } \hat{R} \text{ has distribution } \hat{\mu} \text{ and } B \text{ has distribution } \text{Beta}'\left(1-\alpha, 2\alpha-1\right). \tag{S.1.35}$$

In view of Remark S.1.19, Conjecture S.1.18 amounts to claiming the tightness of the

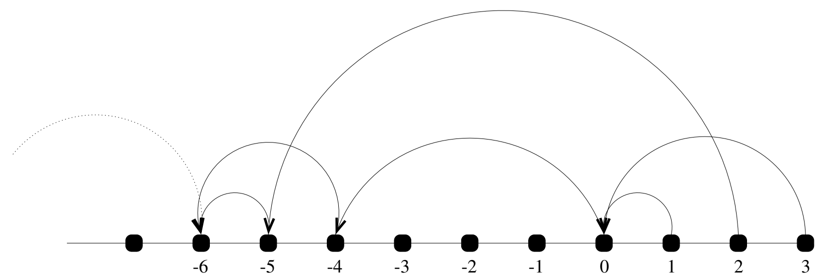


Figure S.3: Depicted is a sample realisation of a part of the ancestral lineages of $[3] = \{1, 2, 3\}$. In this example $A_3(0) = 3$, $A_3(1) = 0 = A_1(1)$, $A_3(2) = -4 = A_1(2)$, $A_3(3) = -6 = A_1(3) = A_2(2)$ and $A_2(0) = 2$, $A_2(1) = -5$, $A_1(0) = 1$.
So $\mathrm{MRCA}(1, 2) = -6$, $\mathrm{MRCA}(1, 3) = 0$, $\mathrm{MRCA}(2, 3) = -6$ and $\mathrm{MRCA}([3]) = -6$.
For U and N we have $U(3, 0) = 3$, $U(3, -1) = 1$ and $U(3, -5) = 1$, $U(3, -6) = 0$ as well as $N(3, 0) = 2$, $N(3, -5) = 2$, $N(3, -6) = 1$.

sequence $(\mathrm{MRCA}([n])/n^{1/\alpha})_n$. The results stated and explained in this section are steps towards a proof of the following statement.

Statement S.1.21. *The sequence of random variables*

$$\left(\frac{\mathrm{MRCA}\left([n]\right)}{n^{\frac{1}{\alpha}}}\right)_n \tag{S.1.36}$$

is tight.

We now make the heuristic rigorous in the following way: For $y \in -\mathbb{N}$ set

$$N(n, y) := \#\Big\{i \in [n] : A_i \cap A_j \cap \{y, \ldots, n\} = \emptyset \text{ for all } j = 1, \ldots, i-1\Big\}, \tag{S.1.37}$$

the cardinality of the partition of $[n]$ that is generated by the equivalence relation

$$i \sim_y j :\Leftrightarrow \mathrm{MRCA}(i, j) > y,$$

or intuitively spoken, the number of remaining ancestral lines of $[n]$ at $y \in -\mathbb{N}$. Moreover, set

$$U(n, y) := \#\Big\{i \in [n] : i - R_i \leq y\Big\},$$

the number of individuals $i \in [n]$ with a parent below y. As indicated above, the intuition will now be that

$$N\left(n, -n^{\beta}\right) \approx U\left(n, -n^{\beta}\right) \qquad \beta \geq 1 \tag{S.1.38}$$

holds, which means that individuals having a parent on scale $-n^{\beta}, \beta \geq 1$ will typically find their most recent common ancestor already on this scale. This is made believable by the above results about pairwise coalescence events. Denote the ancestral lineage of k by A_k and the *ancestry* of a set $B \subset \mathbb{N}$ by

$$A_B := \bigcup_{k \in B} A_k. \tag{S.1.39}$$

We then define the *pruned ancestry* of $B \subset \mathbb{N}$ by

$$\hat{A}_B := A_B \backslash A_0. \tag{S.1.40}$$

See Figure S.4 for an illustration of the pruned ancestry of [3].

Definition S.1.22. Let $\mathcal{A}$ be a random subset of $\mathbb{Z}$. Then

$$x \mapsto \mathbf{P}\left(x \in \mathcal{A}\right), \qquad x \in \mathbb{Z} \tag{S.1.41}$$

is called the (expected local) *density* of $\mathcal{A}$.

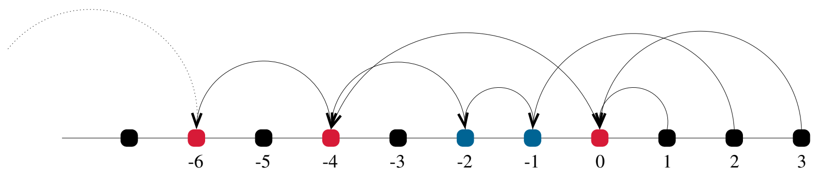


Figure S.4: An illustration of the pruned ancestry $\hat{A}_{[3]}$ of [3]: The ancestry of 0 in red, the pruned ancestry $\hat{A}_{[3]}$ of [3] in blue.

The following proposition, which is proved in Section 3.2.1, now bounds $N(n, y)$ under an assumption on the density of the pruned ancestry of $[n]$.

Proposition S.1.23. *Let $\varepsilon, \xi, \eta > 0$. Assume that there exists $\beta \in \left(1, \frac{1}{\alpha}\right)$, such that*

$$t^{(n)}(x) := \mathbf{P}\left(-x \in \hat{A}_{[n]}\right) = O\left(x^{-\frac{\alpha}{2}}\right) \qquad \text{for } x := x(n) \gg n^{\beta}. \tag{S.1.42}$$

Then for any non-decreasing sequence $(a_n)_n$ with $n^{\beta-1} \ll a_n \ll n^{\frac{1}{\alpha}-1}$

$$\begin{aligned}&\mathbf{P}\left(N(n, -a_n \cdot n) \geq (1+\varepsilon)\frac{1}{1-\alpha} \cdot (a_n n)^{1-\alpha}\left[\left(\frac{1}{a_n}+1\right)^{1-\alpha} - 1\right]\right) \\ \leq\;& \text{const} \cdot \left[a_n^{1-\frac{\alpha}{2}+\xi} n^{1-\frac{3}{2}\alpha+\xi} + a_n^{\alpha-\xi+\eta} n^{-\xi+\eta}\right].\end{aligned} \tag{S.1.43}$$

Note that as shown below in Remark S.1.26 the term $(1+\varepsilon)\frac{1}{1-\alpha}\cdot(a_n n)^{1-\alpha}\left[\left(\frac{1}{a_n}+1\right)^{1-\alpha} - 1\right]$ is just $(1+\varepsilon)$ times the expectation of $U(n, -a_n \cdot n)$.

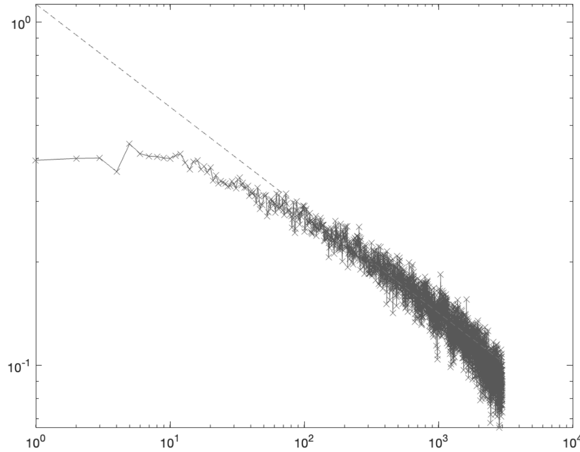
Conjecture S.1.24. *We conjecture that (S.1.42) is valid for $\beta = 1$. In fact, we conjecture that even the following stronger statement is valid:*

$$t(\ell) = \mathbf{P}\left(-\ell \in \hat{A}_{\mathbb{N}}\right) = O\left(\ell^{-\frac{\alpha}{2}}\right) \qquad \text{as } \ell \to \infty. \tag{S.1.44}$$

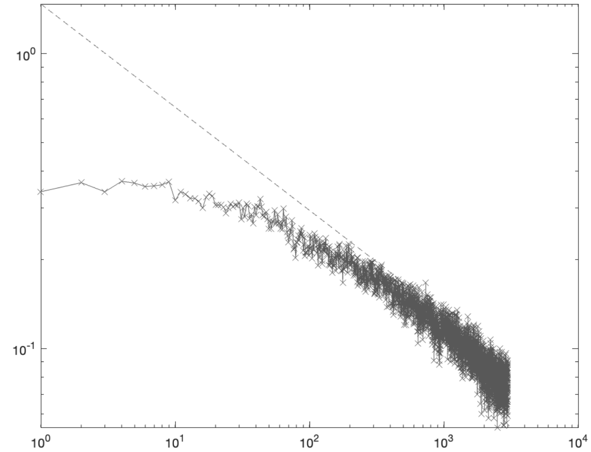
Remark S.1.25. Recall $t(x)$ was defined in (S.1.18). Since

$$\mathbf{P}\left(-x \in \hat{A}_{[n]}\right) \leq \mathbf{P}\left(-x \in \hat{A}_{\mathbb{N}}\right) = t(x) \tag{S.1.45}$$

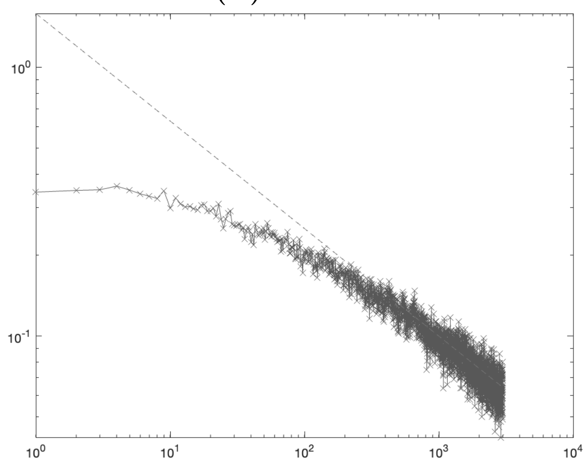
it is clear that (S.1.44) implies (S.1.42).



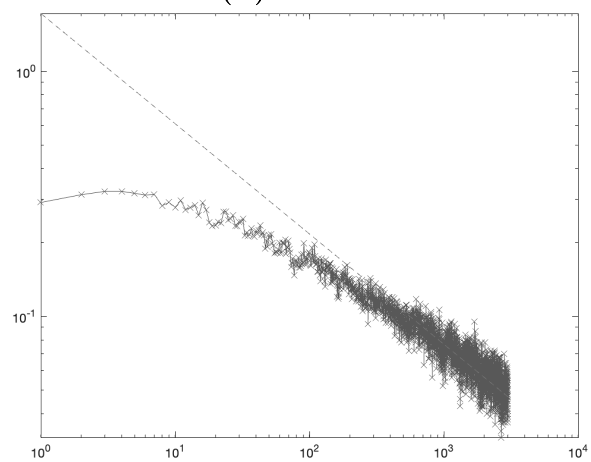
(A) $\alpha = 0.6$



(B) $\alpha = 0.7$



(C) $\alpha = 0.8$



(D) $\alpha = 0.9$

Figure S.5: We estimate $t^{(n)}(x) := \mathbf{P}\left(0 \notin A_\ell \text{ and } \exists n \geq y > \ell : 0 \sim y\right)$ via simulations for $n = 3000$ and various values of α and for $x \in \{1, \ldots, n\}$ with 1000 simulations and compare it to const $\cdot |x|^{-\frac{\alpha}{2}}$ (dotted line).

Simulations supporting (S.1.42) can be found in Figure S.5 as well as Section 3.2.6. (S.1.44) is heuristically justified in Section 3.2.5 by the construction and analysis of a critical branching random walk.

Remark S.1.26. Indeed (S.1.43) implies a statement that corresponds to (S.1.38): Since

$$U(n, -y) \overset{(d)}{=} \sum_{i=1}^{n} B_i(y)$$

for $B_i(y)$ independent Ber $\Big(\mu\left(\{y+i+1, y+i+2, \ldots\}\right)\Big)$-distributed, the random variable $U(n, -y)$ has variance

$$\mathbf{Var}[U(n,-y)] \leq \sum_{i=1}^{n} \mu\left(\{y+i+1, y+i+2, \ldots\}\right) = \mathbf{E}\left[U(n,-y)\right] = o\Big(\mathbf{E}\left[U(n,-y)\right]^2\Big),$$

giving that it concentrates around its expectation

$$\sum_{i=1}^{n} \mu\left(\{y+i+1, y+i+2, \ldots\}\right) \sim \sum_{i=1}^{n} (y+i+1)^{-\alpha} \sim \frac{n^{1-\alpha}}{1-\alpha}\left[\left(\frac{y}{n}+1\right)^{1-\alpha} - \left(\frac{y}{n}\right)^{1-\alpha}\right]. \tag{S.1.46}$$

So (S.1.43) implies that for $\varepsilon, \xi, \eta > 0$, $\delta \in \left(0, n^{\frac{1}{\alpha}-1}\right)$, $\gamma = 1+\delta$, and $y \gg n^{\beta}$

$$\mathbf{P}\left(N(n, -yn^{\gamma}) > (1+\varepsilon) \cdot U(n, -yn^{\gamma} + \varepsilon n^{\gamma})\right) \leq \text{const} \cdot \left[n^{-\alpha-\gamma\left(\frac{1}{2}\alpha-1-\xi\right)} + n^{\gamma\alpha-\alpha-\gamma\xi+\eta}\right]. \tag{S.1.47}$$

Since the right hand side converges to zero for $\alpha \in \left(\frac{1}{8}\left(1+\sqrt{33}\right), 1\right)$ as $n \to \infty$, and since $N(n, -yn^{\gamma}) \geq U(n, -yn^{\gamma})$ by definition, this justifies the non-formal statement (S.1.38).

The central heuristic behind Proposition S.1.23 is as follows: The bound (S.1.42) implies, that there is not enough space for the ancestral lineages to avoid each other once they are close. This is illustrated by Figure S.2: The small dots mark an ancestor in an ancestral lineage. Once we see one of these ancestors, then it usually does not take long for its ancestral lineage to coalesce with one of the others.
From the estimate (S.1.43) in Proposition S.1.23 we can then (at least for sufficiently large α), get to Statement S.1.21:

Proposition S.1.27. *Let* $\alpha \in \left(\frac{1}{8}\left(1+\sqrt{33}\right), 1\right)$. *Assume that there exists* $\beta \in \left(1, \frac{1}{\alpha}\right)$, *such that for* $\varepsilon, \xi, \eta > 0$ *and any non-decreasing sequence* $(a_n)_n$ *with* $n^{\beta-1} \ll a_n \ll n^{\frac{1}{\alpha}-1}$ *the estimate* (S.1.43) *is true. Then Statement S.1.21 is true, namely the family of random variables*

$$\left(\frac{\mathrm{MRCA}\left([1,n]\right)}{n^{\frac{1}{\alpha}}}\right)_n$$

is tight.

Remark S.1.28. The arbitrary looking value of $\frac{1}{8}\left(1+\sqrt{33}\right) \approx 0.8431$ in Proposition S.1.27 is obtained by noting that only for $\alpha > \frac{1}{8}\left(1+\sqrt{33}\right)$ we can for all $a_n := n^{\delta}, 0 \leq \delta < \frac{1}{\alpha} - 1$ choose $\xi > 0$ and $\eta > 0$, such that the upper bound (S.1.43) is decreasing in n.

Since the assumption of the validity of (S.1.43) is in essence a quantification of (S.1.38) it suffices for a proof of Proposition S.1.27 to show that (roughly spoken) the ancestral lineages of those individuals from $[n]$ that have their parent at a scale slightly smaller than $n^{1/\alpha}$ coalesce on their way to scale $n^{1/\alpha}$. This is provided by Lemma 3.2.1, which plays a key role in the proof of Proposition S.1.27 that is given in Section 3.2.2.

Proposition S.1.27, Proposition S.1.23 and Remark S.1.19 together imply the following theorem.

Theorem S.1.29. *Let* $\alpha \in \left(\frac{1}{8}\left(1+\sqrt{33}\right), 1\right)$. *Assume that for some* $\beta \in \left(1, \frac{1}{\alpha}\right)$ *the condition* (S.1.43) *holds true. Then* $\left(\mathrm{MRCA}([0,n])\right)_n$ *lives on scale* $n^{1/\alpha}$.

In Section 3.2 we outline a second route and its connections to the above route to a proof of Statement S.1.21 which is based on an assumption of the behaviour of many disjoint renewal chains. We will obtain that this assumption also gives evidence in the direction of (S.1.42). In Section 3.2.7 we then highlight a connection to discrete Bessel random walks, which might open another route to a proof of the validity of Statement S.1.21.

S.1.6 Fractional Brownian motion: Autoregression versus moving average

This discrete construction of fractional Brownian motion in Section S.1.2 is in the spirit of an autoregression. In contrast to this, the well-known Mandelbrot-van-Ness representation of fractional Brownian motion has the flavour of a moving average. Indeed, Mandelbrot and van Ness [MVN68] show that fractional Brownian motion has a kernel representation in terms of Wiener increments $\mathrm{d}W(s)$:

$$B^H(t) = \int_{(-\infty,t]} K(s,t)\,\mathrm{d}W(s), \quad t \in \mathbb{R} \tag{S.1.48}$$

with

$$K(s,t) := K^H(s,t) := \frac{1}{C_H}\left[\mathbf{1}_{s\le 0}\left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right) + \mathbf{1}_{t\ge s>0}(t-s)^{H-\frac{1}{2}}\right].$$

Using this representation and fractional calculus Gripenberg and Norros [GN96] derived the following prediction formula for FBM.

Proposition S.1.30 (Proposition 2.3.4)**.** *For a fractional Brownian motion B with Hurst parameter $\frac{1}{2} < H < 1$ for all $t > 0$ we have*

$$\mathbf{E}\left[B_t | \sigma\left(B_s, s \le 0\right)\right] = \int_{-\infty}^{0} g(t,s)\,\mathrm{d}B_s$$

with

$$g(t,-s) := \frac{\sin\left(\pi\left(H-\frac{1}{2}\right)\right)}{\pi} \cdot t^{-H+\frac{1}{2}} \cdot \int_0^t \frac{\xi^{H-\frac{1}{2}}}{\xi+s}\,\mathrm{d}\xi.$$

Using the HS approximations one obtains a "microscopic" interpretation of this formula in Chapter 2. The main idea is that by analysing the asymptotics of

$$b_{n,-k} := \mathbf{P}\left(\max\left(\text{ ancestral lineage of } n \cap \{-\infty,\dots,-2,-1\}\right) = -k\right), \tag{S.1.49}$$

one can compute the conditional expectation of Y_n given the colours $Y_{-1}, Y_{-2}, \dots$. See Section 2.4 for details.

S.1.7 Tree-indexed fractional Brownian motion and its discrete approximation

Chapter 2 constructs and analyses a *branching* version of fractional Brownian motion, more specifically *fractional Brownian motion indexed by a Yule tree* $\mathfrak{Y}$ whose root branch is augmented by the time interval $(-\infty, 0)$ (In a Yule tree every individual splits into two after an Exp(1)-distributed waiting time, where the time origin is in 0 and the waiting times are independent.). The representation (S.1.48) lends itself to a construction of the $\mathfrak{y}$-indexed FBM. This works in two steps: the first is for a given realisation $\mathfrak{y}$ of $\mathfrak{Y}$, the second is a randomisation over $\mathfrak{y}$. For this purpose we want to describe a binary splitting tree $\mathfrak{y}$ in the

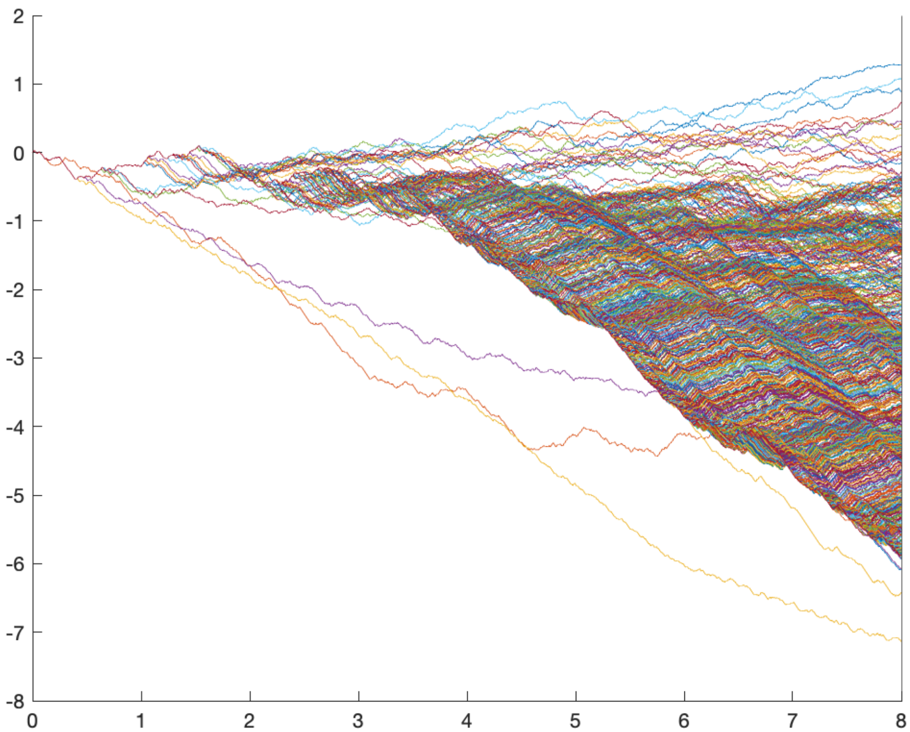


Figure S.6: This shows a simulation of the branching fractional Brownian motion with Hurst parameter $H = 0.85$. The y-axis is measured in units of $\left(\sum_{\ell\geq 0} q_\ell^2\right)^{-\frac{1}{2}}$ for q_ℓ defined by (S.1.6) and $\alpha = H - \frac{1}{2}$. The simulation is based on the discrete approximations (S.1.51).

following way: The branching points are indexed by $h \in \mathcal{V} := \bigcup_{m\in\mathbb{N}_0} \{0,1\}^m$, the vertices of the binary Ulam-Harris tree (see [Har63, Section 2.1]). For $h, h' \in \mathcal{V}$ we say that h' is a child of h if $h' = h0$ or $h' = h1$. The tree $\mathfrak{y}$ is represented by

$$\mathfrak{y} = \bigcup_{h\in\mathcal{V}} \{h\} \times (\ell_h, r_h]\,,$$

where the lifetime intervals $(\ell_h, r_h]$ are requested to satisfy

a) $\ell_\emptyset = -\infty$ and $r_\emptyset > 0$,

b) for $h \in \mathcal{V}$ with children h', h'' we have $r_h = \ell_{h'} = \ell_{h''}$,

c) for all sequences $(h_0, h_1, \ldots) \in \mathcal{V}^{\mathbb{N}}$ with $h_0 = \emptyset$ and h_{m+1} being a child of h_m for all $m \geq 0$, we have that $(\ell_{h_m})_{m\in\mathbb{N}_0}$ has no accumulation points.

Denote by $\mathcal{Y}$ the set of binary splitting trees whose lifetime intervals satisfy the conditions a)-c). For each $v = (h,t) \in \mathfrak{y}$, let $t(v) := t$. In this way, the *ancestral lineage* $\mathfrak{a}(v) \subset \mathfrak{y}$ of each $v \in \mathfrak{y}$ is in isometric correspondence with $(-\infty, t(v)]$. Let $dW_\mathfrak{y}(v)$, $v \in \mathfrak{y}$, be a standard Gaussian white noise on $\mathfrak{y}$. Then the $\mathfrak{y}$*-indexed FBM* has the representation

$$B^H_\mathfrak{y}(v) := \int_{\mathfrak{a}(v)} K\left(t(u), t(v)\right) \mathrm{d}W_\mathfrak{y}(u), \quad v \in \mathfrak{y}. \tag{S.1.50}$$

This construction is implicit in work by Adler and Samorodnitsky, see [AS95], who introduced branching fractional Brownian motion as a starting point for their construction of *historical super fractional Brownian motion* as a process of measures on the path space that can be seen as a high density limit of critically branching fractional Brownian motions. For related approaches to tree-indexed processes with memory see Kouritzin, Lê and Sezer [KLS19], who studied the long-term behaviour of the cloud of particles with a focus on the bulk. In order to describe the discrete approximations of $\mathfrak{y}$-indexed FBM we introduce a branching version of the HS-model. Let $\mathfrak{y} \in \mathcal{Y}$ (for example a realisation of a Yule tree), see Figure S.7.

For $\nu \in \{0,1\}^{\mathbb{N}}$ let $h_0 := \emptyset$ and $h_i := (\nu_1, \ldots, \nu_i)$, $i \in \mathbb{N}$. The *branch of* $\mathfrak{y}$ *corresponding to* ν is defined as

$$b = \bigcup_{i=1}^{\infty} \{h_i\} \times \left(\ell_{h_i}, r_{h_i}\right].$$

Let $i_1 < i_2 < \ldots$ be the indices for which $\nu_{i_k} = 1$, $k = 1, 2, \ldots$. The branch of $\mathfrak{y}$ corresponding to ν is then given the name $0\ell_{h_{i_1}} \ldots \ell_{h_{i_k}} \ldots$. This means that the branch named $0s_1s_2\ldots$ splitted off from the branch named 0 at time s_1, from the branch named $0s_1$ at time s_2, and so on, see Figure S.7. In particular, the branch corresponding to the sequence $(0,0,\ldots)$, which we call the *main branch*, is named 0. We then denote by $\mathcal{B}$ the collection of branches of the tree. For two branches $b, \tilde{b}$ we denote by $b \wedge \tilde{b}$ the time when they split, for example $0 \wedge 0rs = r$ and $0r \wedge 0rs = s$, as depicted in Figure S.7. The point on the branch b at time t on the tree is denoted by $(b,t)_\mathfrak{y}$. We say that the branch b is older than the branch $\tilde{b}$ if the last digit ($\in \mathbb{R}$) of $\tilde{b}$ is larger than the last digit of b. For example, 0 is the oldest branch and b is older than bs. In order to construct the n-th discrete approximation of $\mathfrak{y}$-indexed H-fractional Brownian motion we need to define a sequence of branching HS-models. In the n-th model by $(bs,k)_{(n)} = (bs, k/n)_\mathfrak{y}$ we denote individual no. $k \in \mathbb{Z}$ in the branch bs, which branched off from branch b at time s in the n-th discrete approximation. If $k \leq \lfloor sn \rfloor$ this is the corresponding individual in branch b, if $k > \lfloor sn \rfloor$ it is the $\left[k - \lfloor sn \rfloor\right]$-th individual after the branch point at which bs branched off b. Especially this means that $(b,0)$ corresponds to the same individual for all branches $b \in \mathcal{B}$, see Figure S.8A. We will mostly omit the subscripts and just write (bs,k), when it is clear that we a referring to the n-th discrete approximation.

We are now ready to construct the n-th discrete approximation of $\mathfrak{y}$-indexed H-fractional Brownian motion:

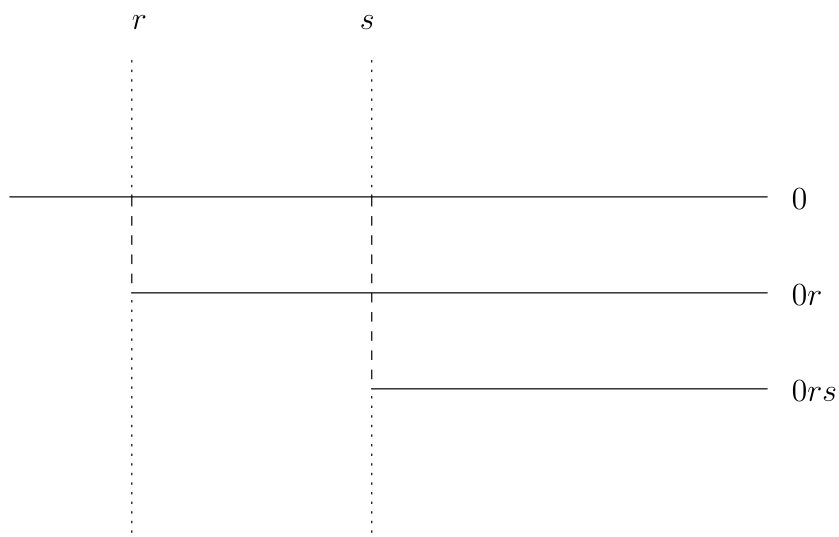
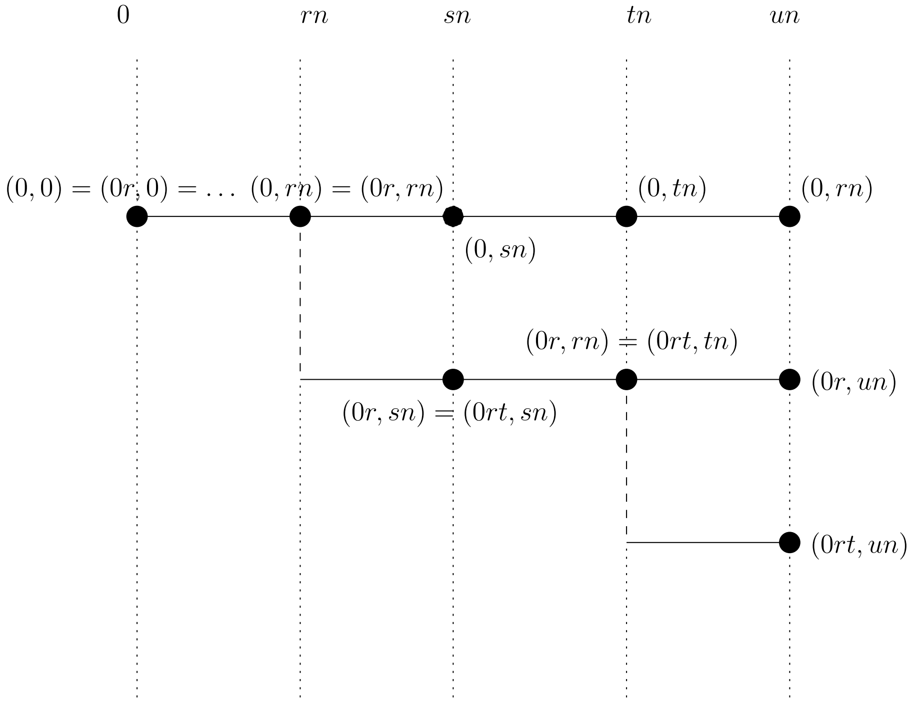


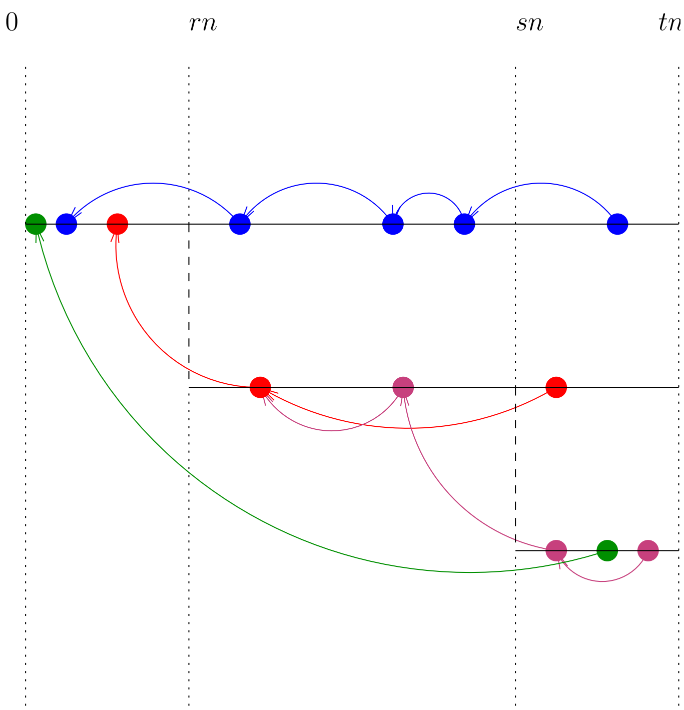
Figure S.7: Three branches of the tree $\mathfrak{y}$ are shown. The ancestral branch is denoted by 0. The branch splitting off from the ancestral branch at time r is denoted by $0r$, and the branch splitting off from the branch $0r$ at time s is denoted by $0rs$. Time runs from left to right, and distances in $\mathfrak{y}$ are measured horizontally. Each branch b is conceived as a copy of $\mathbb{R}$, with common ancestries being glued together.

1. Sample a realisation $\mathfrak{y}$ of a Yule tree.
2. Sample a HS genealogy along the discretisation of the branch 0. More specifically, sample an HS genealogy over $\mathbb{Z}$, and identify the integer $i \in \mathbb{Z}$ with the point $(0, i/n)_{\mathfrak{y}}$ in the tree $\mathfrak{y}$. This is done as described in Section S.1.2, using the branch 0 instead of the real numbers.
3. We proceed recursively. Let the $R_i^{(b)}$ be independent and have distribution μ. Note that each individual $(b, i)_{(n)}$ has an ancestor at distance $R_i^{(b)}$ to the left to which it is connected regardless of the branch it lies in.
   See Section 2.2.2 for a more formal description and Figure S.8 for an illustration.

Note that due to the assumption $0 < \alpha < \frac{1}{2}$ the above procedure produces a random graph with almost surely infinitely many connected components. See Figure S.8B for an illustration of the random graph. In the same spirit as in Section S.1.2 and Chapter 1 we assign a type ± 1 to each component independently with probability $\frac{1}{2}$. (We want to emphasize that we only choose this colouring Y to ease notation and enhance the readability of computations. In general colourings obeying (S.1.7) or (S.1.11) work.)
For an individual (b, k) we denote by $Y_{(b,k)}$ the type of its component, where we often will omit the b if $b = 0$. In correspondence to (S.1.8) the random walk for the n-th discrete



(A) An illustration of a $\mathfrak{y}$-indexed HS-urn and the multiple names an individual can have.



(B) An illustration of a $\mathfrak{y}$-indexed HS-model in which we follow three ancestral lines. Two of them coalesce. Red and purple belong to the same component.

Figure S.8

approximation along the main branch 0 is defined by

$$S_0^{(n)}(t) := \frac{1}{c(n)} \left[\sum_{\ell=1}^{\lfloor tn \rfloor} Y_\ell + \big[tn - \lfloor tn \rfloor\big]\, Y_{\lfloor tn \rfloor} + \big[\lceil tn \rceil - tn\big]\, Y_{\lceil tn \rceil} \right], \qquad t \geq 0$$

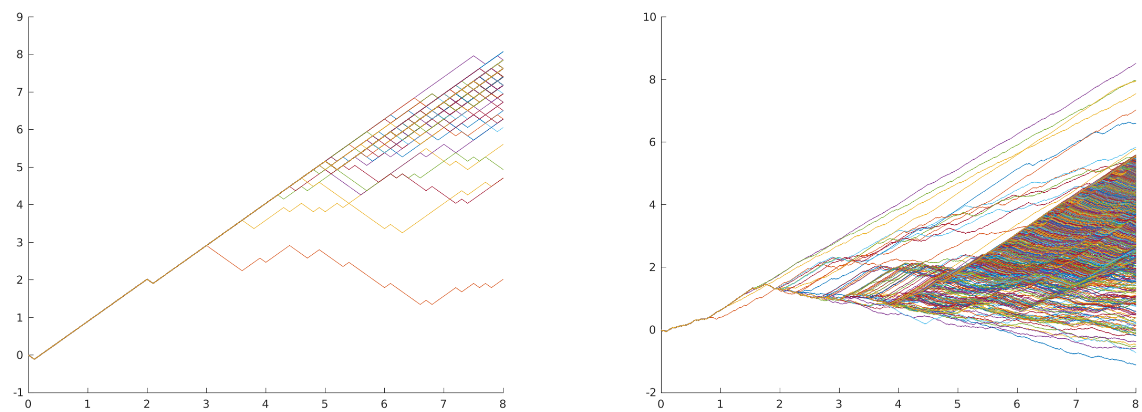
for the scaling function

$$c(n) := \left(n^{2\alpha+1} \frac{1}{\sum_{\ell \geq 0} q_\ell^2} \cdot \frac{1}{\alpha(2\alpha+1)} \cdot \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3} \right)^{\frac{1}{2}}.$$

Now define the random walk along a branch bs inductively by

$$S_{bs}^{(n)}(t) := \mathbf{1}_{t \leq s} S_b^{(n)}(t) + \mathbf{1}_{t > s} \left[S_b^{(n)}(s) + \frac{\sum_{\ell=\lfloor ns \rfloor+1}^{\lfloor tn \rfloor} Y_{(bs,\ell)} + \big[tn - \lfloor tn \rfloor\big] Y_{(bs,\lfloor tn \rfloor)} + \big[\lceil tn \rceil - tn\big] Y_{(bs,\lceil tn \rceil)}}{c(n)} \right]. \tag{S.1.51}$$

Observe that this means that for two branches b and $\tilde{b}$ the processes $S_b^{(n)}$ and $S_{\tilde{b}}^{(n)}$ are equal till $b \wedge \tilde{b}$ and share some common memory afterwards.



(A) $n = 10$ **(B)** $n = 300$

Figure S.9: This is a simulation of $S_{\mathfrak{y}}^{(n)} := \left(\left(S_b(t) \right)_{0 \leq t \leq 8} \right)_{b \in \mathcal{B}}$ for $\alpha = 0.45$. The y-axis is measured in units of $\left(\sum_{\ell \geq 0} q_\ell^2 \right)^{-\frac{1}{2}}$ for q_ℓ defined by (S.1.6).

See Figure S.9 for a simulation using different values of n.
In the sequel we denote by $\mathbf{P}_{\mathfrak{y}}, \mathbf{E}_{\mathfrak{y}}$ the law of the $\mathfrak{y}$-indexed random walks given the tree $\mathfrak{y}$. Using the results from Section S.1.2 (see also Chapter 1) and some general theory about $\mathfrak{y}$-indexed processes (with memory), see Section 2.5, we can then prove the following Theorem.

Theorem S.1.31 (Theorem 2.3.8)**.** *Let $T > 0$. Let $\mathfrak{y} \in \mathcal{Y}$. Then*

$$S_{\mathfrak{y}}^{(n)} := \left(\left(S_b^{(n)}(t) \right)_{[0,T]} \right)_{b \in \mathcal{B}}$$

converges in distribution to $\mathfrak{y}$-indexed fractional Brownian motion.

The simulation depicted in Figure S.6 is based on the discrete approximation provided by Theorem 2.3.2.

S.1.8 The right-most particle of branching fractional Brownian motion

We now turn to the analysis of the speed of the right-most particle of branching fractional Brownian motion, in the spirit of McKean's celebrated work [McK75] (see also the earlier work [INW68, Example 3.4.A] by Ikeda, Nagasawa and Watanabe) about branching Brownian motion, which was later substantially refined by Bramson [Bra78] and Lalley and Sellke [LS87]. Specifically we will show that the maximum $M^H(t)$ of a BFBM with Hurst parameter $H > \frac{1}{2}$ behaves asymptotically like

$$m(t) := m(H,t) := t^{H+\frac{1}{2}} \sqrt{\frac{\sqrt{\pi} 2^{2H+1} H}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)\left(H+\frac{1}{2}\right)^2}} \tag{S.1.52}$$

in the sense of the following Theorem.

Theorem S.1.32 (Theorem 2.3.8)**.** *For all* $\varepsilon > 0$

$$\mathcal{P}\left(\left|\frac{M^H(t)}{m(t)} - 1\right| > \varepsilon\right) \to 0 \quad \textit{as } t \to \infty.$$

See Figure S.10 for an illustration of $m(t)/t^{H+1/2}$ for varying H. Note that for $H = \frac{1}{2}$ the value of $m(t)$ equals $\sqrt{2}t$, which is consistent with the results about branching Brownian motion.

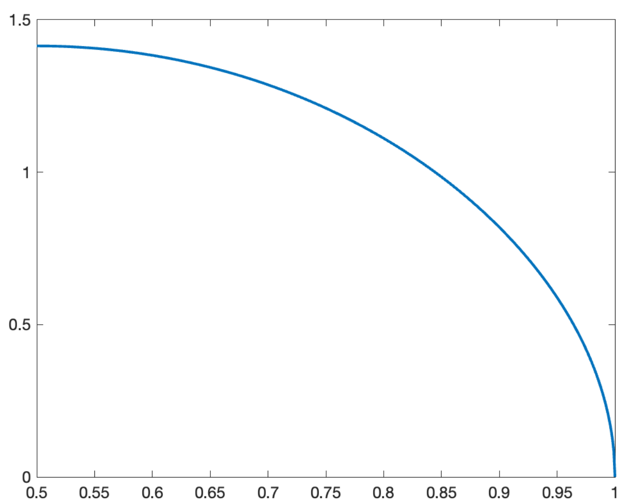


Figure S.10: Graph of $m(t)/t^{H+\frac{1}{2}}$ as a function of H.

Remark S.1.33. Here we point to a connection with Theorem S.2.3, which gives a result about the ancestral lineage of a particle conditioned to be at point s at time t. [CRW91, p. 124, Corollary] highlights that in the light of precisely this theorem Chauvin and Rouault [CR88] analyse the maximum particle of a branching Brownian motion conditioned on being unusually far away. They obtain that the ancestral lineage on any fixed compact time interval behaves asymptotically as $t \to \infty$ like a Brownian motion with drift.

Up to the constant factor in (S.1.52), the leading order $t^{H+\frac{1}{2}}$ of the maximum can be easily explained. Indeed, for a number $\lfloor e^t \rfloor$ of independent normally distributed random variables with variance t^{2H} the leading order of the maximum would be given by

$$\mathfrak{m}(t) := \sqrt{2}\, t^{H+\frac{1}{2}}$$

as suggested by the estimate

$$e^t\, \mathbf{P}\left(\mathcal{N}\left(0, t^{2H}\right) \geq \mathfrak{m}(t)\right) \approx e^t \exp\left(-\frac{\mathfrak{m}(t)^2}{2t^{2H}}\right) = 1.$$

Due to correlations we obtain (S.1.50) which only differs by the constant prefactor depending on H.

The proof technique of this theorem relies heavily on exploring the connection between branching random walks and the so-called generalised random energy models (GREM): A precursor (and a special case) of the GREM is the REM, Random Energy Model, introduced by Derrida in [Der80], continued in [Der81]. This is the Gaussian random field $(X_\sigma)_{\sigma=1,\dots,2^N}$ for independent X_σ. An overview and analysis of this can be found in lecture notes by Bolthausen and Sznitman, see [BS02].
In [Der85] Derrida introduced the GREM, a REM with multiple levels, which induces a hierarchical structure. For example a GREM with two levels is a correlated random field

$$(X_\sigma)_{\sigma\in\left\{1,\dots,2^{N/2}\right\}\times\left\{1,\dots,2^{N/2}\right\}}$$

with

$$X_\sigma \equiv Y^{(1)}_{\sigma_1} + Y^{(2)}_{\sigma_1,\sigma_2}$$

for independent collections of independent random variables

$$\left(Y^{(1)}_{\sigma_1}\right)_{\sigma_1\in\left\{1,\dots,2^{N/2}\right\}} \qquad \text{and} \qquad \left(Y^{(2)}_{\sigma_1,\sigma_2}\right)_{\sigma_1\in\left\{1,\dots,2^{N/2}\right\},\sigma_2\in\left\{1,\dots,2^{N/2}\right\}}.$$

The covariance of the random field then depends on the so called overlap. See the lecture notes by Kistler [Kis15, Section 2.2] for a more formal introduction.

We first give a short outline of our proof and then continue with a variation of this result for deterministic binary branching trees, which can be obtained as a corollary of results by Bovier and Kurkova [BK04b]. Afterwards we give a remark concerning the literature exploring the connection between branching random walks and GREMs.

Our proof strategy consists in a discrete approximation, but in contrast to the HS-approximation described above only in an approximation of $\left(B_b(t)\right)_{b\in\mathcal{B}}$ for one point in time t by a GREM.

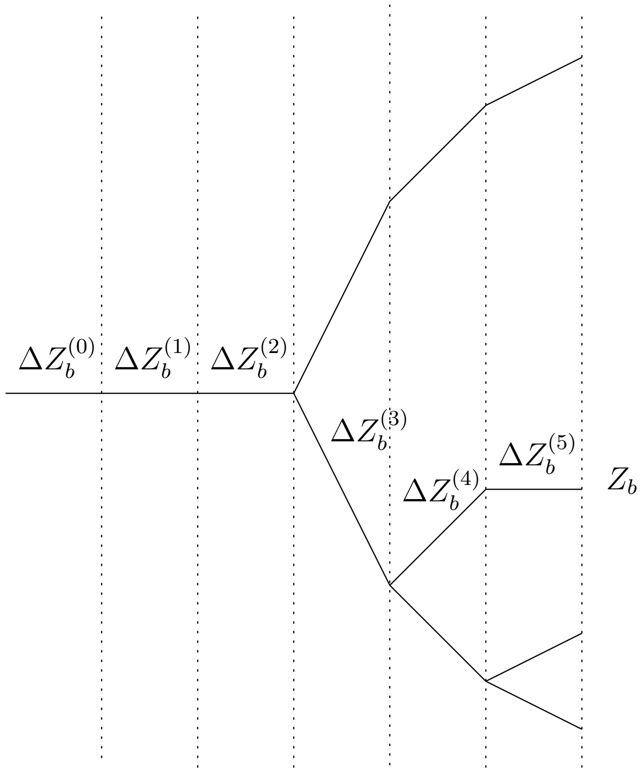


Figure S.11: A GREM on a discretised Yule tree. For illustration purposes we assumed that $1/K$ is so small compared to the time horizon depicted that we only see binary splits.

The core idea of the proof is that the maximum of a BFBM can only be attained by a trajectory staying very close to the maximum all along the way.

We start with a brief sketch on how to prove the lower bound. Let $K \in \mathbb{N}$. For all $i = 1, \dots, K-1$ we shift all branching events in $\left[\frac{i}{K}t, \frac{i+1}{K}t\right]$ of the tree $\mathfrak{y}$ to $\frac{i}{K}t$, such that we can only have branching events at $0, \frac{t}{K}, \frac{2t}{K}, \dots, \frac{(K-1)t}{K}$. This approximation to the Yule tree realisation $\mathfrak{y}$ will be denoted by $\mathfrak{y}_{K,t}$, and its collection of branches by $\mathcal{B}_{K,t}$. In the GREM we set

$$Z_b^{(i)} := \sum_{\ell=0}^{i} \Delta Z_b^{(\ell)},$$

where the $\Delta Z_b^{(\ell)}$ are the increments along the branch b, see Figure S.11. For two branches b and $\tilde{b}$ with $b \wedge \tilde{b} = s$ we get

$$\mathbf{Cov}\left[B_b(t), B_{\tilde{b}}(t)\right] = t^{2H}\left[1 - C_\rho\left(1 - \frac{s}{t}\right)^{2H}\right] =: \rho(t,t,s),$$

where

$$C_\rho = \frac{\sqrt{\pi}2^{2H-1}}{\Gamma\left(1-H\right)\Gamma\left(H+\frac{1}{2}\right)},$$

see (2.3.6). Thus, in order to represent the $\mathfrak{y}_{K,t}$-indexed fractional Brownian motion as a

GREM, we choose, for each $b \in \mathcal{B}_{K,t}$, the distribution of the increment $\Delta Z_b^{(i)}$ as

$$\mathcal{N}\left(0, \rho\left(t, t, \frac{i}{K}t\right) - \rho\left(t, t, \frac{i-1}{K}t\right)\right) \qquad \text{if } i \geq 1$$

and

$$\Delta Z_b^{(0)} \equiv Z_b^{(0)} \sim \mathcal{N}\left(0, \rho\left(0, t, t\right)\right) = \mathcal{N}\left(0, \left(1 - C_\rho\right) t^{2H}\right).$$

Remember that in a GREM the increments $\Delta Z_b^{(i)}$ and $\Delta Z_{\tilde{b}}^{(i)}$ coincide if the branches b and $\tilde{b}$ did not separate till $\frac{iT}{K}$, meaning $b \wedge \tilde{b} > \frac{iT}{K}$.

Since we shifted all branching events in the time intervals of length $\frac{1}{K}$ to the left an application of Slepian's Lemma, see Lemma 2.7.1, shows that the order of the maximum of the approximating GREM gives an upper bound for the order of the maximum of BFBM. Now denote the rescaled standard variations of $\Delta Z_0^{(i)}, i \geq 1$ as

$$\begin{aligned}
\Delta f_i &:= \sqrt{2\frac{t}{K}}\sqrt{\rho\left(t, t, \frac{i}{K}t\right) - \rho\left(t, t, \frac{i-1}{K}t\right)} \\
&= \sqrt{2\frac{t}{K}} t^H \left[-C_\rho\left(1 - \frac{i}{K}\right)^{2H} + C_\rho\left(1 - \frac{i-1}{K}\right)^{2H}\right]^{\frac{1}{2}}
\end{aligned}$$

and the rescaled standard variation of $\Delta Z_0^{(0)}$

$$\Delta f_0 := \sqrt{\rho\left(t, t, 0\right)}.$$

The naive bound

$$\sum_{i=0}^{K} \Delta f_i \sim \frac{2t^{H+\frac{1}{2}}\sqrt{C_\rho H}}{K} \sum_{i=1}^{\ell} \left(1 - \frac{i}{K}\right)^{H-\frac{1}{2}} \sim 2t^{H+\frac{1}{2}}\sqrt{C_\rho H} \int_0^1 \left(1-y\right)^{H-\frac{1}{2}} \mathrm{d}y = m(t)$$

corresponds to a *greedy strategy* in the GREM. We show with Markov's inequality that it is indeed an upper bound. By using a GREM in which we shifted the branching events to the right and applying the Payley-Zygmund inequality we show that this is also a lower bound. See Section 2.8 for more details.

In Remark 2.3.11 we give another proof of Theorem S.1.32, which combines results of Arguin, Bovier and Kistler [ABK11] with the Mandelbrot-van-Ness-representation of fractional Brownian motion in terms of Wiener integrals and the Payley-Wiener partial integration formula. This proof was suggested by an anonymous referee of a previous version of [GI23]. The idea is to plug in a path of the maximum order of branching Brownian motion into the Mandelbrot-van-Ness-representation and to notice that this already gives the maximum order of branching fractional Brownian motion. However, this can not be used to make statements about the subleading order since in a BBM the particle of the maximum order changes through time.

Of a GREM on a deterministic binary branching tree, which is considered the classical form, Bovier and Kurkova analyse the leading order of the maximum in [BK04b]. Their results lead to the following version of the above Theorem.

Theorem S.1.34 (Theorem 2.3.6)**.** *Let $\mathfrak{y}_{\text{bin}}$ be a deterministic binary branching tree (every branch branches into two after time 1). The leading order of the maximum of a BFBM*

$$\left(\left(B_b(t) \right)_{t \geq 0} \right)_{b \in \mathcal{B}}$$

with Hurst parameter $H \in \left(\frac{1}{2}, 1\right)$ is

$$m^{\text{bin}}(t) := m^{\text{bin}}(H, t) := t^{H+\frac{1}{2}} \sqrt{\frac{\log(2)\sqrt{\pi}2^{2H+1}H}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)\left(H+\frac{1}{2}\right)^2}}$$

in the sense that

$$\mathbf{E}_{\mathfrak{y}_{\text{bin}}}\left[\frac{\max_{b \in \mathcal{B}} B_b(t)}{m^{\text{bin}}(t)}\right] \to 1 \quad \text{for } t \to \infty.$$

Note that in comparison to (S.1.52) only a factor $\sqrt{\log(2)}$ appears. This factor is easily explained by the fact that the underlying tree has 2^t instead of e^t many leaves.

For subleading orders work by Maillard and Zeitouni [MZ16] allows some conjectures in our regime, see Remark 2.3.12.

The method above and the connection between GREMs and branching random walks used above has first been explored by Arguin, Bovier and Kistler in [ABK11] for branching Brownian motion. In contrast to Brownian motion fractional Brownian motion is a process with memory. As we have seen above the corresponding GREM is then a continuous GREM with decreasing variance. In the following we give a few remarks concerning the connection between branching random walks and the GREM, see Remark 2.3.13 for a more detailed remark about the historical developments. In [ABK11] Arguin, Bovier and Kistler were able to analyse the full extremal process, including the maximal particle, the second maximal, and so on. They showed that extremal particles descend from ancestors that split either shortly after zero or just before the observed time. In [ABK13a] the authors analysed the empirical distribution of the maximal displacement and showed that a Gumbel distribution with a random shift occurs as a limit thus proving a conjecture of Lalley and Sellke. In [ABK13b] they continued the study and showed that the extremal process converges in law to a Poisson cluster point process. Maillard and Zeitouni then studied the maximum of BBM with decaying variance, which is close to our setup and allows some conjectures on the subleading order behaviour of the maximum particle, recall Remark 2.3.12 for a more detailed explanation.

S.2 Ancestral reproductive bias in branching processes

Chapter 4 works in the setting of continuous-time branching processes. Consider a continuous-time branching process with N_t individuals alive at time t, started with one individual at time 0. At the end of its lifetime, an individual is replaced by a random number of independent offspring with distribution $(p_k)_{k \geq 0}$. When lifetimes of the individuals are i.i.d. with an arbitrary distribution μ on $\mathbb{R}_+$, the resulting process is called a *Bellman-Harris* process [BH48]. In the special case of exponentially distributed lifetimes, this process is a continuous-time (Bienaymé-)Galton-Watson process, which is also called *one-dimensional continuous-time Markov branching process*, see [AN72, Chapter 3]. For those processes, Cheek and Johnston [CJ23] study the process of reproduction times and family sizes along the ancestral lineage of an individual sampled from all those alive at a given time $T > 0$, conditioned on the event $\{N_T > 0\}$. They obtain a bias along this lineage, which they call *ancestral reproductive bias*.

In Chapter 4 we give a short and conceptual probabilistic proof of the main results of [CJ23] in the more general Bellman-Harris setting. We model the uniform sampling by giving independent markers distributed uniformly on $[0, 1]$ to all individuals being alive at time T and selecting the individual with the largest marker. Our proof strategy based on this point of view then explains that this bias can in some sense be interpreted as competition, see also the sketch of proof in Section S.2.1 and the proof in Section 4.4.

Chapter 4 will then make use of the developed technique to give a proof of results by Chauvin, Rouault and Wakolbinger [CRW91]. They consider a continuous (in time and space) Markov process indexed by a continuous-time Galton-Watson tree, think of branching Brownian motion, and condition it on having a particle at point s at time t (see Section S.2.2 and Section 4.5 for more details). By distributing appropriate markers along the individuals alive at time T, we will obtain a new derivation of the distribution of reproductive events along the individual's ancestral lineage.

We then continue with an application to results by Geiger [Gei99], who obtained that the left most ancestral lineage in a planar embedding of the tree has a different *ancestral reproductive bias*. We will model the embedding by giving independent markers distributed uniformly on $[0, 1]$ to all individuals. The individuals can then be thought of as embedded into the plane according to their lexicographic order. This point of view again allows an elegant proof based on properties of families of independent uniformly on $[0, 1]$ distributed random variables. See Section S.2.3 and Section 4.6.

S.2.1 Sampling an ancestral line at random

Associate a uniform marker in $[0, 1]$, independently to each individual alive at time T. On the event $\{N_T > 0\}$, let the individual V be the individual with the largest marker, and let S be its mark. This corresponds to a uniformly distributed random pick from all the individuals alive at time T. We define the total population size process $(N_t)_{t \geq 0}$ to be right continuous with left limits. As a consequence, if T_1 is the lifetime of the root individual, then N_{T_1} has distribution $(p_k)_{k \geq 0}$. Let J be the random number of reproduction events and $0 < T_1 < T_2 < \cdots < T_J \leq T$ be the random times of reproduction events along the ancestral lineage of V. Let $L_1, \ldots, L_J$ be the offspring sizes in these reproduction events and let $0 < \tau_1 < \tau_2 < \cdots$ be the random arrival times in a renewal process with interarrival time distribution μ. See Figure 5.1 for a sample realisation.
Denote by $\mathbf{P}$ and $\mathbf{E}$ the probability measure and expectation for $N_0 = 1$. We will prove the following extension of [CJ23, Theorems 2.3 and 2.4].

Theorem S.2.1 (Theorem 4.2.1)**.** *For $j \geq 0$, $0 < t_1 < \ldots < t_j \leq T \in \mathbb{R}$ and $\ell_1, \ldots, \ell_j \in \mathbb{N}$ we have*

$$\begin{aligned} &\mathbf{P}\left(N_T > 0, J = j,\, T_1 \in \mathrm{d}t_1, \ldots T_j \in \mathrm{d}t_j,\, L_1 = \ell_1, \ldots, L_j = \ell_j,\, S \in \mathrm{d}s\right) \\ &= \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \ldots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} > T\right) \prod_{i=1}^{j} \left(\ell_i p_{\ell_i}\, \mathbf{E}\left[s^{N_{T-t_i}}\right]^{\ell_i - 1}\right) \mathrm{d}s. \end{aligned} \tag{S.2.1}$$

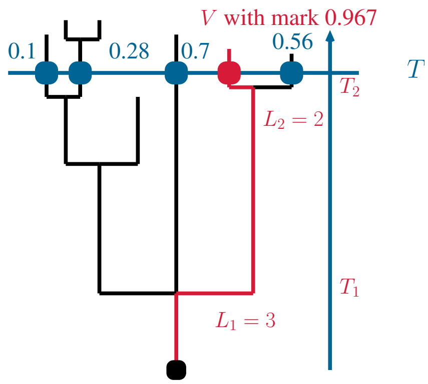


Figure S.12: An example for a realisation of the random variables S, L_1, L_2, T_1, T_2 in the first sampling scheme.

As we will see, the argument s of the generating functions that appear in the (S.2.1) corresponds to the realisation of the largest marker. Thus the term $\prod_{i=1}^{j} \mathbf{E}\left[s^{N_{T-t_i}}\right]^{\ell_i - 1}$ in (S.2.1) can be interpreted as stemming from the competition for the largest marker.

The main ingredient for the proof is then the following observation: For $\ell \in \mathbb{N}$, let $\widetilde{S}$ be the maximum of ℓ independent Unif$[0,1]$-distributed random variables $U_1, \ldots, U_\ell$. Then the density of $\widetilde{S}$ is

$$\mathbf{P}\left(\widetilde{S} \in \mathrm{d}s\right) = \ell s^{\ell - 1}\, \mathrm{d}s, \quad 0 \leq s \leq 1. \tag{S.2.2}$$

Indeed, because of exchangeability,

$$\mathbf{P}\left(\widetilde{S} \in \mathrm{d}s\right) = \ell\, \mathbf{P}\left(U_1 \in \mathrm{d}s\right) \mathbf{P}\left(U_2 < s, \ldots, U_\ell < s\right),$$

which equals the r.h.s. of (S.2.2).

The following specialises to (S.2.2) when putting $\tilde{N} \equiv 1$:

Lemma S.2.2 (Lemma 4.3.1.)**.** *Let $\widetilde{N}$ be an $\mathbb{N}_0$-valued random variable, and $\widetilde{N}_1, \widetilde{N}_2, \ldots$ be i.i.d. copies of $\tilde{N}$. Given $\widetilde{N}_1, \widetilde{N}_2, \ldots$ let*

$$U_{1,1}, \ldots U_{1,\widetilde{N}_1}, U_{2,1}, \ldots U_{2,\widetilde{N}_2}, \ldots$$

be independent Unif$[0,1]$-distributed random variables, and write

$$S_k \;\; := \;\; \max\left\{U_{k,1}, \ldots, U_{k,\widetilde{N}_k}\right\}, \quad k = 1, 2, \ldots$$

$$S^{(\ell)} \;:=\; \max\{S_1,\ldots,S_\ell\}, \quad \ell\in\mathbb{N},$$

where we put $\max(\emptyset) := -\infty$. *Then, for all* $\ell\in\mathbb{N}$, *the density of* $S^{(\ell)}$ *is*

$$\mathbf{P}\left(\widetilde{N}_1+\ldots+\widetilde{N}_\ell>0,\, S^{(\ell)}\in \mathrm{d}s\right) = \ell\,\mathbf{E}\left[s^{\widetilde{N}}\right]^{\ell-1}\mathbf{P}\left(\widetilde{N}_1>0,\, S_1\in\mathrm{d}s\right), \quad 0\le s\le 1.$$

Together with an induction argument this lemma is the backbone of our proof.

S.2.2 Conditioning on a marker value

Relating to work of Chauvin, Rouault and Wakolbinger [CRW91] also the following setup is analysed in Chapter 4: [CRW91] consider a Markov process with an atomless transition probability indexed by a continuous-time Galton-Watson-tree and they then condition on an individual at time T to be at a given location. To relate this to the framework described above, we assume that each individual alive at time T in the Bellmann-Harris tree carries a marker in some standard Borel space E and these random marks have the following properties:

(M1) Their marginal distributions (denoted by ν) are identical and do not depend on the reproduction events

(M2) A.s. no pair of marks is equal.

Think for example of branching Brownian motion or fractional branching Brownian motion as discussed in Section S.1.7 and Chapter 2. The positions of the different particles clearly depend on each other via the genealogy, however, at time t the marginal distribution of the position of each particle (in a branching fractional Brownian motion with Hurst parameter H) is a centered Gaussian random variable with variance t^{2H}, irrespective of its past genealogical events in the underlying continuous-time Galton-Watson tree. Thus (M1) is fulfilled. Since two jointly Gaussian random variables whose correlation coefficient is not equal to one are a.s. not equal, (M2) is also fulfilled. We now condition on $\{N_T>0\}$ and, for given $s\in E$, on

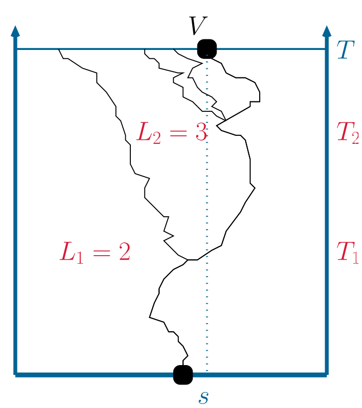


Figure S.13: An example for a realisation of the random variables L_1, L_2, T_1, T_2 in the second sampling scheme.

one of the N_T individuals having marker value s. Remember the previous notation: Denote by V the individual having marker s. Figure S.13 depicts a sample realisation.
The following Theorem generalises (part of) [CRW91, Theorem 2] to general lifetime time distributions.

Theorem S.2.3 (Theorem 4.5.1)**.** *For $j \geq 0$, $0 < t_1 < \ldots < t_j < T$ and $\ell_1, \ldots, \ell_j \in \mathbb{N}$ we have for ν-almost all s*

$$\mathbf{P}\left(J = j,\, T_1 \in \mathrm{d}t_1, \ldots, T_j \in \mathrm{d}t_j,\, L_1 = \ell_1, \ldots, L_j = \ell_j \middle| N_T > 0, \exists\, \text{marker} \in \mathrm{d}s\right)$$

$$= \frac{1}{\mathbf{E}[N_T]} \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \ldots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} \geq T\right) \prod_{i=1}^{j} \ell_i p_{\ell_i}.$$

The proof strategy mainly consists in using the fact that conditioning on at least one individual having marker s is equivalent to the condition that exactly one individual has marker s.

S.2.3 Sampling the left-most ancestral lineage

Third, a sampling scheme introduced by Geiger [Gei99] is analysed; the leftmost surviving ancestral lineage in a planar embedding of the tree: At any reproduction event we assign independent uniformly on $[0, 1]$ distributed markers to all children. An individual can now be uniquely determined by the markers along its ancestral lineage. And the markers can be interpreted as an encoding of the embedding by drawing the trees of children from left to the right in the lexicographic order of their markers. On the event $\{N_T > 0\}$, let V be the individual whose markers along the entire ancestral lineage comes first in the lexicographic ordering among all individuals alive at time T. Denote by K_i the number of siblings born at reproduction event number i along the ancestral lineage of V which have a lower lexicographic order than V and whose descendants hence die out before time T. Figure S.14 shows a realisation for this sampling rule.

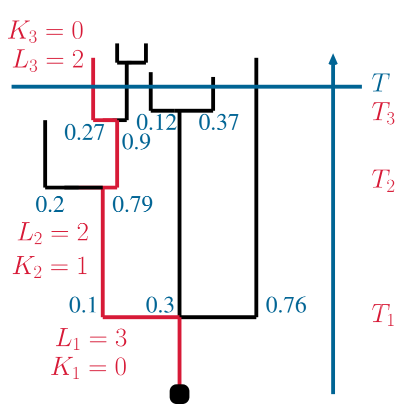


Figure S.14: An example for a realisation of markers and random variables $L_1, L_2, K_1, K_2, T_1, T_2$ in the third sampling scheme.

Theorem S.2.4 (Theorem 4.6.1)**.** *For $j \geq 0$, $0 < t_1 < \ldots < t_j < T$, $\ell_1, \ldots, \ell_j \in \mathbb{N}$ and $k_i \in \{1, \ldots, \ell_i - 1\}$ we have*

$$\begin{aligned}
&\mathbf{P}\left(N_T > 0, J = j, T_1 \in \mathrm{d}t_1, \ldots, T_j \in \mathrm{d}t_j, L_1 = \ell_1, \ldots, L_j = \ell_j, K_1 = k_1, \ldots, K_j = k_j\right) \\
&= \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \ldots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} \geq T\right) \prod_{i=1}^{j} \left(p_{\ell_i} \mathbf{P}\left(N_{T-t_i} = 0\right)^{k_i}\right).
\end{aligned}$$

The proof strategy is again to use an elementary fact about families of Unif $[0,1]$-distributed random variables, which itself can again be shown via exchangeability.

Lemma S.2.5 (Lemma 4.6.2)**.** *Let $\widetilde{N}$ be an $\mathbb{N}_0$-valued random variable, and $\widetilde{N}_1, \widetilde{N}_2, \ldots$ be i.i.d. copies of $\tilde{N}$. Given $\widetilde{N}_1, \widetilde{N}_2, \ldots$ let $U_1, U_2, \ldots$ be independent Unif$[0,1]$-distributed random variables, and write*

$$\begin{aligned}
S^{(\ell)} &:= \min\left\{U_k \mid \widetilde{N}_k \geq 1, k = 1, \ldots, \ell\right\}, \\
K^{(\ell)} &:= \left|\left\{U_k \mid U_k < S^{(\ell)}, k = 1, \ldots, \ell\right\}\right|,
\end{aligned}$$

where we put $\min(\emptyset) := +\infty$. Then, for all $k < \ell \in \mathbb{N}$ we have

$$\mathbf{P}\left(\widetilde{N}_1 + \ldots + \widetilde{N}_\ell > 0,\ K^{(\ell)} = k\right) = \mathbf{P}\left(\widetilde{N} = 0\right)^k \mathbf{P}\left(\widetilde{N} > 0\right).$$

S.3 A variant of Muller's ratchet in a near-critical parameter regime

S.3.1 The classical variant of Muller's ratchet

Muller's ratchet is a prototype model in population genetics. Originally it was conceived to explain the ubiquity of sexual reproduction among eukaryotes despite its many costs, see the work by Muller and also Felsenstein [Mul64, Fel74]. In its bare bones version, Muller's ratchet models a haploid, asexual population whose size N is constant over the generations. The neutral part of the random reproduction is given by Wright-Fisher or Moran dynamics. Slightly deleterious mutations are acquired along the lineages at a rate m, and individuals carrying fewer mutations have a selective advantage. The classical variant of Muller's ratchet considers *fitness proportional* selection, where the selective advantage of an individual carrying κ deleterious mutations over a contemporanean that carries a larger number κ' of deleterious mutations is $\frac{\mathfrak{s}}{N}(\kappa' - \kappa)$. Since the mutation mechanism is assumed to be unidirectional, every once in a while the type with the currently smallest number of mutations κ will disappear from the population. As Herbert Muller puts it in his pioneering paper [Mul64], "*an irreversible ratchet mechanism exists in the non-recombining species . . . that prevents selection, even if intensified, from reducing the mutational loads below the lightest . . . , whereas, contrariwise, 'drift', and what might be called 'selective noise' must allow occasional slips of the lightest loads in the direction of increased weight.*"

It is these "slips of the lightest loads" which are called *clicks of the ratchet*. The question "How often does the ratchet click?" was asked by Etheridge, Pfaffelhuber and Wakolbinger in [EPW09], and there it was found that

$$\gamma := \frac{m}{\mathfrak{s} \log(Nm)} \tag{S.3.1}$$

is "*an important factor in determining the rate of the ratchet*". Specifically, under the assumption $1 \ll Nm \ll N$, [EPW09] states the following *Rule of Thumb* for the classical ratchet:

(RTC) *The rate of the (classical) ratchet is of the order $N^{\gamma-1}m^{\gamma}$ for $\gamma \in (\frac{1}{2}, 1)$, whereas it is exponentially slow in $(Nm)^{1-\gamma}$ for $\gamma < \frac{1}{2}$.*

With the *mutation-selection ratio*

$$\theta := \frac{m}{\mathfrak{s}},$$

(RTC) predicts the expected interclick time in the case $\gamma \in (\frac{1}{2}, 1)$ as

$$N(Nm)^{-\gamma} = Ne^{-\theta}.$$

As observed by John Haigh ([Hai78]), in the deterministic limit ($N \to \infty$ and $m, \mathfrak{s}$ not depending on N) the type frequency profile in equilibrium becomes Poisson with parameter θ. Consequently, for $\gamma \in (\frac{1}{2}, 1)$ the rule (RTC) goes along with Haigh's prediction that the rate of the ratchet should be proportional to the inverse of the size of the best class.

For a polynomial mutation rate $m = N^{-\beta}$, $0 < \beta < 1$, the condition that γ remains constant (or at least bounded away from 0 and ∞) as $N \to \infty$ amounts to the requirement that the mutation-selection ratio θ is of the order $\log N$ as $N \to \infty$.

For the purpose of illustration we will consider a family of parameter scalings for (m, θ) which we call the (β, δ)*-scaling of the classical ratchet*:

$$m = N^{-\beta}, \qquad \theta = \delta \log N. \tag{S.3.2}$$

This amounts to *moderate mutation-selection*, with the mutation-selection ratio θ diverging logarithmically in N. The factor δ in front of $\log N$ turns out to be critical for the click rate. Indeed, in the (β, δ)-scaling, (S.3.1) takes the form

$$\gamma(\beta, \delta) = \frac{\delta}{1-\beta}.$$

The condition $0 < \gamma < 1$ from (RTC) restricts the pair (β, δ) to the triangle

$$\Delta := \{(\beta, \delta) : 0 < \beta,\ 0 < \delta < 1 - \beta\}.$$

The *polynomial* and the *exponential regime* predicted by (RTC) correspond to

$$\mathcal{P} := \{\tfrac{1}{2} < \gamma(\beta, \delta) < 1\} = \{(\beta, \delta) \in \Delta : \tfrac{1}{2}(1-\beta) < \delta < 1 - \beta\},$$

$$\mathcal{E} := \{0 < \gamma(\beta, \delta) < \tfrac{1}{2}\} = \{(\beta, \delta) \in \Delta : 0 < \delta < \tfrac{1}{2}(1-\beta)\},$$

and the predictions for the orders of magnitude of the expected interclick times take the form

$$N(Nm)^{-\gamma} = N^{1-\delta} \qquad \text{for} \quad \gamma \in \left(\frac{1}{2}, 1\right), \tag{S.3.3}$$

$$\exp\left(\text{const}(Nm)^{1-\gamma}\right) = \exp\left(\text{const}\, N^{1-\beta-\delta}\right) \qquad \text{for} \quad \gamma \in \left(0, \frac{1}{2}\right) \tag{S.3.4}$$

In view of the predicted transition from polynomial to exponential click rates we refer to $\mathcal{P} \cup \mathcal{E}$ as a *near-critical regime*.

The evidence for (RTC) that is given in [EPW09] is based on a diffusion approximation for the evolution of the relative size X_0 of the *best class* (which consists of the individuals that carry the least amount of mutations in the current population).

S.3.2 Tournament versus fitness proportional selection

Chapter 5 considers a variant of Muller's ratchet in which fitness proportional selection is replaced by *(binary) tournament selection*. This kind of selection has been studied in the context of evolutionary computation ([BT96, BFM18]) and has found attention also in the biological literature [PBB+15]. In the ratchet's context this means that selective advantage of an individual carrying κ deleterious mutations over a contemporanean that carries a larger number κ' of deleterious mutations is constant (say $\frac{s}{N}$ for some $s = s_N > 0$), irrespective of the value of the difference $\kappa' - \kappa$. For the Moran version of the tournament ratchet, which was introduced in [GSW23] and whose definition we recall now, see also Section 5.2, this means that "pairwise selective fights" are always won by the fitter individual.

We assume that the mutation rate m is equal for both ratchets, but the selection coefficients $\mathfrak{s}$ for the classical ratchet and s for the tournament ratchet are different; note the typographical difference between s and $\mathfrak{s}$. The mutation-selection ratio is given by

$$\begin{cases} \theta := \dfrac{m}{\mathfrak{s}} & \text{for the classical ratchet} \\ \rho := \dfrac{m}{s} & \text{for the tournament ratchet.} \end{cases}$$

The following definition gives the jump rates for the type frequencies of the two ratchets.

Definition S.3.1 (Definition 5.2.1)**.** Writing N_κ for the current number of individuals of type κ, the jump rates are specified as follows:

- Resampling: for $\kappa \neq \kappa'$,
  $(N_\kappa, N_{\kappa'})$ jumps to $(N_\kappa + 1, N_{\kappa'} - 1)$ at rate $\frac{1}{2N} N_\kappa N_{\kappa'}$
- Mutation: for κ,
  $(N_\kappa, N_{\kappa+1})$ jumps to $(N_\kappa - 1, N_{\kappa+1} + 1)$ at rate $m N_\kappa$
- Selection: for $\kappa < \kappa'$,

$$(N_\kappa, N_{\kappa'}) \text{ jumps to } (N_\kappa + 1, N_{\kappa'} - 1) \text{ at rate } \begin{cases} \dfrac{\mathfrak{s}}{N} N_\kappa N_{\kappa'} (\kappa' - \kappa) & \text{for the classical ratchet} \\ \dfrac{s}{N} N_\kappa N_{\kappa'} & \text{for the tournament ratchet.} \end{cases}$$

The currently best type is

$$K^*(t) := \min \left\{ \kappa \in \mathbb{N}_0 : N_\kappa(t) > 0 \right\}.$$

Other than in the classical ratchet, the size of the (m, s)-tournament ratchet's best class follows an autonomous dynamics *up to its time of extinction*; at this time the class which was so far the second-best becomes the best one. As explained in Section 5.3, this dynamics is *equal* to that of the so called Poisson profile approximation (see Section 5.3.2) of the size of the classical $(m, \mathfrak{s})$-ratchet's best class, provided that the mutation-selection ratios ρ and θ of the tournament and the classical ratchet are related by

$$\rho := \frac{m}{s} = 1 - \exp(-m/\mathfrak{s}) = 1 - e^{-\theta}. \tag{S.3.5}$$

Chapter 5 focuses on the conjecture that the following

Rule of thumb for the near-critical tournament ratchet (RTT) should hold: *As $N \to \infty$, the expected time between clicks is*

$$\begin{aligned} &\asymp \sqrt{\frac{N}{m}} && \text{if} \quad Nm(1-\rho)^2 \to 0, && \text{(S.3.6)} \\ &\asymp \exp\left(Nm(1-\rho)^2\right) && \text{if} \quad Nm(1-\rho)^2 \to \infty. \end{aligned}$$

Here and below, $\asymp$ stands for logarithmic equivalence, i.e. $a_N \asymp b_N$ means $\log a_N \sim \log b_N$, or equivalently $\frac{\log a_N}{\log b_N} \to 1$. For $k = 0, 1, \ldots$ let $Y_k^{\mathrm{C}}(t) = N_{K^*+k}^{\mathrm{C}}(t)$ and $Y_k^{\mathrm{T}}(t) = N_{K^*+k}^{\mathrm{T}}(t)$ be the sizes of the $(k+1)^{\mathrm{st}}$-best class of the classical and the tournament ratchet, where $(N_\kappa^{\mathrm{C}})_{\kappa\in\mathbb{N}_0}$ and $(N_\kappa^{\mathrm{T}})_{\kappa\in\mathbb{N}_0}$ follow the dynamics specified in the above Definition. The jump rates from n to $n-1$ are given for both Y_0^{C} and Y_0^{T} by

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+m\right), \tag{S.3.7}$$

but the jump rates from n to $n+1$ are different: those of Y_0^{T} are

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+\frac{m}{\rho}\left(1-\frac{n}{N}\right)\right), \tag{S.3.8}$$

while those of Y_0^{C} are

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+\frac{m}{\theta}\sum_{k=1}^{\infty} kY_k^{\mathrm{C}}\right).$$

An inspection of the jump rates in Definition S.3.1 reveals that for each $k \in \mathbb{N}$ the process $(Y_0^{\mathrm{T}}, \ldots, Y_k^{\mathrm{T}})$ obeys an autonomous dynamics up to the extinction time of Y_0^{T}; for $k = 0$ this is evident from (S.3.7) and (S.3.8).

The jump rates reveal that $(Y_0^{\mathrm{T}}, Y_1^{\mathrm{T}})$ has, asymptotically as $N \to \infty$, the center of attraction

$$(\mathfrak{a}, \mathfrak{b}) := \left(N(1-\rho), N\sqrt{1-\rho}\right) \tag{S.3.9}$$

provided $Nm \to \infty$ and $\rho \to 1$. To see this, note that the dynamics of $(Y_0^{\mathrm{T}}, Y_1^{\mathrm{T}})$ is autonomous up to the first hitting of $\{0\} \times \{0, \ldots, N\}$, and that the states of $(Y_0^{\mathrm{T}}, Y_1^{\mathrm{T}})$ for which the upward jump rates are asymptotically equal to the downward jump rates have the asymptotic (Np_0, Np_1), with (p_0, p_1) given by $(1-\rho, \sqrt{1-\rho})$.

S.3.3 Asymptotics of the clickrate in a near-critical regime

The above discussion suggests that in order to prove (RTT) we need to analyse the time to extinction of Y_0^{T} depending on its initial state. This is provided by the following theorem. In Remark 5.3.5 we discuss what are the ingredients missing to go from Theorem S.3.2 to a proof of (RTT), and we also indicate a different route to the proof of (RTT), using the technique developed in [GSW23]. Here and below we write $a_n \ll b_n$ for two sequences if $a_n/b_n \to 0$ as $n \to \infty$.

Theorem S.3.2 (Theorem 5.3.4)**.** *Let T_0 be the extinction time of the birth-and-death process Y_0^{T} with jump rates (S.3.7) and (S.3.8), let $1 \gg m \gg \frac{1}{N}$, and let ρ be a sequence in $[\rho_0, 1)$ for some fixed $\rho_0 \in (0,1)$. Here and below we suppress the subscript N in m_N, ρ_N, etc.*

1. [Polynomial regime] *Assume* $Nm(1-\rho)^2 \to 0$ *as* $N \to \infty$. *Let* (j_N) *be a sequence of natural numbers in* $[N]$. *If* $j_N \ll \sqrt{\frac{N/m}{\log(N/m)}}$, *then*

$$\mathbf{E}_{j_N}[T_0] \sim 2j_N \left(\log\sqrt{\frac{N}{m}} - \log j_N\right),$$

whereas if $j_N \gg \sqrt{\frac{N}{m}}$, *then*

$$\mathbf{E}_{j_N}[T_0] \sim \frac{\pi^{3/2}}{2}\sqrt{\frac{N}{m}}.$$

The expected number of returns of the process Y_0^{T} *to* $\lceil \mathfrak{a} \rceil$, *when starting above* $\mathfrak{a} = (1-\rho)N$, *is asymptotically equivalent to* $\frac{1}{m(1-\rho)}$ *as* $N \to \infty$.

2. [Exponential regime] *Assume* $Nm(1-\rho)^2 \to \infty$ *and* $1 \ll j_N \le N$ *as* $N \to \infty$. *Then*

$$\mathbf{E}_{j_N}[T_0] \sim \left(1 - \exp\left(-2m\left(\frac{1}{\rho}-1\right)j_N\right)\right)\sqrt{\frac{\pi}{mN}}\,v_N,$$

with

$$v_N := \frac{1}{m\left(\frac{1}{\rho}-1\right)} \exp\left(2Nm(1-\rho)^2\eta(m,\rho)\right), \tag{S.3.10}$$

$$\eta(m,\rho) := -\frac{1}{2m}\left[\frac{1}{1-\rho}\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1 - \frac{1}{(1+2m)^{\ell}}\right)\frac{(1-\rho)^{\ell-1}}{\ell(\ell+1)}\right].$$

In particular, with

$$e_N := \frac{1}{1-\rho}\sqrt{\frac{\pi}{mN}}\,v_N$$

one has

$$\mathbf{E}_{j_N}[T_0] \sim \begin{cases} e_N & \text{if } j_N \gg \frac{1}{m(1-\rho)} \\ e_N(1-\exp(-2C/\rho)) & \text{if } j_N \sim \frac{C}{m(1-\rho)} \\ e_N\, 2j_N\, m(1/\rho - 1) & \text{if } j_N \ll \frac{1}{m(1-\rho)} \end{cases}.$$

The expected number of returns of the process Y_0^{T} *to* $\lfloor \mathfrak{a} \rfloor$, *when starting above* $\mathfrak{a} = (1-\rho)N$, *is asymptotically equivalent to* (S.3.10) *as* $N \to \infty$.

In view of (S.3.5) we define, in analogy to (S.3.2), the (β, δ)-scaling for the tournament ratchet as

$$m = N^{-\beta}, \qquad \rho = \frac{m}{s} = 1 - N^{-\delta}.$$

With this scaling, (RTT) takes the following form: *As* $N \to \infty$, *the expected time between clicks is*

$$\asymp N^{\frac{1+\beta}{2}} \qquad \text{if} \quad (\beta,\delta) \in \mathcal{P}, \tag{S.3.11}$$

$$\asymp \exp\left(N^{1-\beta-2\delta}\right) \qquad \text{if} \quad (\beta,\delta) \in \mathcal{E}. \tag{S.3.12}$$

Let us give a heuristics for the long-term behaviour of Y_0^{T}, which also points towards (RTT) as well as part of (RTC). The rates (S.3.7) and (S.3.8) consist of three parts: the fluctuation terms $\pm\frac{n}{2}\left(1-\frac{n}{N}\right)$, the net linear birth rate $n\frac{m}{\rho}(1-\rho)$ and the quadratic death rate $\frac{m}{\rho}\frac{n^2}{N}$. As long as Y_0^{T} is below $\mathfrak{a}/2$, it is stochastically bounded from below by a binary Galton-Watson process Y^ℓ with supercriticality $m(1-\rho)/2$, and stochastically bounded from above by a binary Galton-Watson process Y^u with supercriticality $m(1-\rho)$. Haldane's formula, which in this case coincides with the formula for the escape probability of a simple random walk with constant drift, gives that the survival probability of the offspring of one individual in Y^ℓ (resp Y^u) is $\sim m(1-\rho) \sim N^{-\beta-\delta}$ (resp. $\sim 2m(1-\rho) \sim 2N^{-\beta-\delta}$). Hence the probability that Y_0^{T} when starting in $\mathfrak{a}/4$ hits 0 before reaching $\mathfrak{a}/2$, is asymptotically between

$$\big(1-2m(1-\rho)\big)^{\frac{\mathfrak{a}}{4}} \sim \left(1-2N^{-\beta-\delta}\right)^{N^{1-\delta}/4} \quad \text{and} \quad \big(1-2m(1-\rho)\big)^{\frac{\mathfrak{a}}{4}} \sim \left(1-N^{-\beta-\delta}\right)^{N^{1-\delta}/4},$$

which both converge to 0 if and only if $1-\beta-2\delta>0$, i.e. $\gamma>\frac{1}{2}$. In this case the number of excursions which Y_0^{T} makes from $\mathfrak{a}/4$ up to $\mathfrak{a}/2$ before going extinct is geometric with expectation asymptotically between

$$\exp\left(\frac{1}{4}N^{1-\beta-2\delta}\right) \qquad \text{and} \qquad \exp\left(\frac{1}{2}N^{1-\beta-2\delta}\right).$$

This gives an intuitive explanation why $\gamma=\frac{1}{2}$ is the boundary between the exponential and the polynomial regime and also for the exponential rate stated above.

The reason why this exponent is different from the one appearing in (S.3.4) is that [EPW09] work here not with the Poisson profile approximation, but with (a rescaling of the diffusion approximation of) the so-called *relaxed Poisson profile approximation*, see Remark 5.3.2.

In the case $\gamma>\frac{1}{2}$, the center of attraction plays a negligible role. What becomes relevant then is the threshold for n above which the quadratic death rate $\frac{m}{\rho}\frac{n^2}{N}$ becomes large. Obviously, the order of magnitude of this threshold is $\sqrt{\frac{N}{m}} = N^{\frac{1+\beta}{2}}$. Above this threshold, Y_0^{T} is strongly pushed downwards, making the time spent above the threshold negligible. Below the threshold, Y_0^{T} behaves similar to a driftless linear birth-and-death process with upward and downward jump rates given by (S.3.7). This gives a qualitative explanation of the orders of magnitude of the expected times to extinction for the polynomial regime.

While both (RTC) and (RTT) state the same boundary ($\gamma=\frac{1}{2}$) between the polynomial and the exponential regime, the exponents differ between (S.3.3) and (S.3.11) as well as between (S.3.4) and (S.3.12). Besides that there is also a more structural difference. While (S.3.11) and (S.3.12) somehow suggest a kind of continuous crossover from the polynomial regime to the exponential regime, the rule (RTC), see (S.3.3) and (S.3.4), predicts a jump, since (S.3.4) is still exponentially large for $\gamma=\frac{1}{2}$.

Also note that in the polynomial regime $\mathcal{P}$ the exponent $\frac{1+\beta}{2}$ for the tournament ratchet is larger than the exponent $1-\delta$ for the classical ratchet. This can be explained in the following way. The centers of attraction of the equilibrium profile weights of the best and the second best class differ asymptotically by the factor $\sqrt{1-\rho} = N^{\frac{\delta}{2}}$ for the tournament ratchet (see (S.3.9)), while they are given by the Poisson weights $e^{-\theta}$ and $\theta e^{-\theta}$ for the classical ratchet and hence for the latter differ only by the factor $\theta = \delta \log N$ (and thus have the same polynomial order $N^{1-\delta}$). This latter factor is only logarithmic in N; therefore, when starting the "new best class" at the time of a click in its "old" center of attraction, the tournament ratchet has a longer way to go than the classical ratchet. The exponent $\frac{1+\beta}{2}$ in (S.3.6) will be obtained by a Green function analysis in the proof of Theorem 5.3.4. We give a short outline of this analysis below. This analysis will also rigorously explain the exponent $1-\delta$ in (S.3.3), which corresponds to Haigh's prediction, saying that "the interclick times are of the order of the size of the best class".

We now give a short outline of the proof of the above Theorem. The proof is based on an asymptotic analysis of the Green function of $Y =: Y_0^{\mathrm{T}}$,

$$G(j,n) := G^N(j,n) = \mathbf{E}_j \left[\int_0^{T_0} I_{\{Y_t^N = n\}} \, \mathrm{d}t \right], \quad 1 \le j, n \le N$$

as $N \to \infty$. By assumption the upward and downward jump rates of Y from n are given by

$$\lambda_n := n \left(\frac{1}{2} \left(1 - \frac{n}{N}\right) + \frac{m}{\rho} \left(1 - \frac{n}{N}\right)\right)$$

and

$$\mu_n := n \left(\frac{1}{2} \left(1 - \frac{n}{N}\right) + m \right).$$

The following lemma, see e.g. [SaSha13, (2.4)] or [DSS05, (15)], expresses the Green function in terms of the *oddsratio products*

$$r_0 := 1, \quad r_k := \prod_{\ell=1}^{k} \frac{\mu_\ell}{\lambda_\ell}, \qquad k \in \{1, \dots, N-1\}.$$

Lemma S.3.3 (Lemma 5.4.1)**.** *For $1 \le j, n \le N$,*

$$G(j,n) = \frac{1}{\mu_n} \sum_{\ell=0}^{j-1 \wedge n-1} \prod_{k=\ell+1}^{n-1} \frac{\lambda_k}{\mu_k}. \tag{S.3.13}$$

With

$$R_k := \sum_{i=0}^{k-1} r_i, \qquad k \in \{1, \dots, N\},$$

we obtain from (S.3.13):

$$G(j,n) = \begin{cases} \frac{R_{j \wedge n}}{\lambda_n r_n} \text{ if } n < N, \\ \frac{R_j}{\mu_N r_{N-1}} \text{ if } n = N. \end{cases}$$

Consequently,

$$\mathbf{E}_j[T_0] = \sum_{n=1}^{N} G(j,n) = \sum_{n=1}^{N-1} \frac{R_{n\wedge j}}{\lambda_n r_n} + \frac{R_j}{\mu_N r_{N-1}}. \tag{S.3.14}$$

Note that

$$U(j) := \log r_j$$

(sometimes also referred to as *potential*, cf. [DSS05, (16)]) is an additive functional, and (S.3.13) translates into

$$G(j,n) = \frac{1}{\mu_n} \sum_{\ell=0}^{j-1\wedge n-1} e^{-(U(n-1)-U(\ell))}.$$

In order to obtain asymptotics for (S.3.14) we need to find asymptotics for the terms r_k and R_k as $N \to \infty$.

We can express $\log r_j$ as

$$\log r_j = \sum_{k=1}^{j} \log\left(\frac{\mu_k}{\lambda_k}\right) = j \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{k=1}^{j} \log\left(\frac{1-k/((1+2m)N)}{1-k/N}\right).$$

This expression allows us the following asymptotic description of r_j.

Lemma S.3.4 (Lemma 5.4.2)**.** *Let $\xi = \xi_N$ be a sequence converging to 0 so slowly that $\xi \gg m$. Then for N large enough and $j \le (1-\xi)N$*

$$0 \le \log r_j - j \log\left(\frac{1+2m}{1+2m/\rho}\right) - \sum_{\ell=1}^{\infty}\left(1 - \frac{1}{(1+2m)^\ell}\right)\frac{1}{\ell(\ell+1)}\frac{j^{\ell+1}}{N^\ell} \le \text{const}\cdot\frac{m}{\xi}.$$

Our analysis, see Lemmas 5.4.3 and 5.4.7, then shows that, as j increases, r_j is essentially constant on a large interval, before it starts to decrease as j approaches the center of attraction $N(1-\rho)$ of the best class. This allows an analysis of the cumulated oddsratio products R_j and of the terms $G(j,n)$, which is then enough to prove the Theorem in both regimes.

S.3.4 Outlook: The sample genealogy in the regime of exponentially small click rates

In Chapter 6 an outlook on current work with Charline Smadi and Anton Wakolbinger is presented. The ideas are based on Sections 4 and 5.2 of the preprint [IGSW23], which are not contained in the publication [IGSW24]. There we consider the tournament ratchet with the dynamics specified in Definition S.3.1 in the near-critical exponential regime, that is $m(1-\rho)^2 N \to \infty$ with $\rho \uparrow 1$ as $N \to \infty$. We analyse the type frequency profile of a sample of individuals at times $t \gg 1/s$ when all individuals at time zero are of the same type. This is achieved via a representation of the sample genealogy in terms of an ancestral selection graph decorated with a Poisson process of mutation events.

Part I

The Hammond-Sheffield urn

Chapter 1

Asymptotic Gaussianity via coalescence probabilities in the Hammond-Sheffield urn[1]

[1] Appeared as [IW23]

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].

1.1 Introduction

We start with a brief description of the model of Hammond and Sheffield [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

$$\mu\left(\{n, n+1, \ldots\}\right) \sim n^{-\alpha} L(n) \text{ as } n \to \infty, \tag{1.1.1}$$

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

$$\text{the greatest common divisor of } \left\{n \in \mathbb{N} : \mu(n) > 0\right\} \text{ is one.} \tag{1.1.2}$$

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

$$i \sim j :\Longleftrightarrow i \text{ and } j \text{ belong to the same connected component of } \mathcal{G}_\mu. \tag{1.1.3}$$

Note that the symbol $\sim$ is used in (1.1.1) and (1.1.3) 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 1.10 for an interpretation (and extension) in terms of the long-range seedbank model of Blath et al. [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 1.1 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,

$$q_n := \mathbf{P}\left(\tilde{R}_1 + \ldots + \tilde{R}_j = n \text{ for some } j \geq 0\right) \tag{1.1.4}$$

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.)

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 $\mathcal{G}_\mu$ has a Gibbs structure, with the extremal elements being given by i.i.d. assignments of colours to the connected components of $\mathcal{G}_\mu$. The main result of [HS13] concerns the asymptotics of the rescaled sum over the types of the individuals

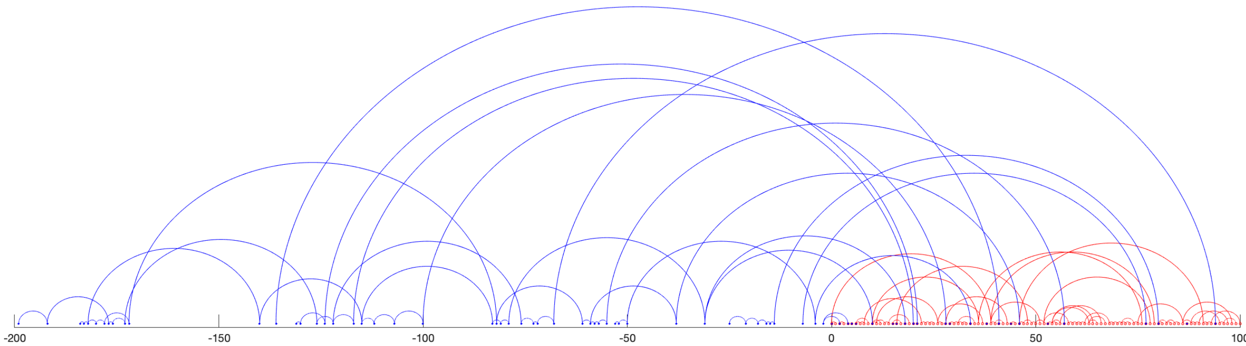


Figure 1.1: A realisation of the ancestral lineages of the individuals $\{0, \dots, 100\}$ traced back till -200. Each of the arcs corresponds to an edge of G_μ. All the outgoing edges from $i = 0, \dots, 100$ which map to an individual in $\{0, \dots, 100\}$ are drawn (in red), whereas for i between -200 and -1 only those outgoing edges are drawn (in blue) that belong to an ancestral lineage of some $j \in \{0, \dots, 100\}$. Here the exponent α in (1.1.1) was chosen as 0.39.

$1, \dots, \lfloor tn \rfloor$, $t \geq 0$, which as $n \to \infty$ turns out to converge to fractional Brownian motion. The individuals' types arise as follows:

Assume that each component of $\mathcal{G}_\mu$ 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.

$$Y = \xi - (2p - 1) \text{ with } \mathbf{P}(\xi = +1) = p, \mathbf{P}(\xi = -1) = 1 - p. \tag{1.1.5}$$

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)

$$S_n := \sum_{i=1}^{n} Y_i, \quad n = 0, 1, \dots. \tag{1.1.6}$$

By construction,

$$\sigma_n^2 := \mathbf{Var}[S_n] = \sum_{i,j \in [n]} \mathbf{Cov}[Y_i, Y_j] = \mathbf{E}[Y^2] \sum_{i,j \in [n]} \mathbf{P}(i \sim j). \tag{1.1.7}$$

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

$$\sum_{i,j \in [n]} \mathbf{P}(i \sim j) \sim \frac{C_\alpha}{\alpha(2\alpha + 1)} \frac{n^{2\alpha+1}}{L(n)^2} \quad \text{as } n \to \infty, \tag{1.1.8}$$

with

$$C_\alpha := \frac{1}{\sum_{m \geq 0} q_m^2} \frac{\Gamma(1 - 2\alpha)}{\Gamma(\alpha)\Gamma(1 - \alpha)^3}. \tag{1.1.9}$$

We will obtain (1.1.8) as a corollary of Proposition 1.2.1 below, which requires the additional condition

$$q_n \sim \frac{1}{\Gamma(\alpha)\Gamma(1 - \alpha)} \frac{n^{\alpha - 1}}{L(n)} \quad \text{as } n \to \infty. \tag{1.1.10}$$

This condition, which also appears in our Theorem 1.1.1, is equivalent to the validity of the Strong Renewal Theorem for the renewal process with an increment distribution μ satisfying (1.1.1) and (1.1.2), see Caravenna and Doney [CD19], whose Theorem 1.4 gives necessary and sufficient conditions in terms of μ for the validity of (1.1.10). A well-known sufficient condition for (1.1.10) is the criterion of Doney [Don97]

$$\sup_{n\geq 1} \frac{n\,\mathbf{P}(R=n)}{\mathbf{P}(R>n)} < \infty. \tag{1.1.11}$$

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 (1.1.7) and (1.1.8), 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

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

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 1.1.1. *Let μ be a probability measure on $\mathbb{N}$ satisfying (1.1.1) and (1.1.2). Assume one of the following conditions (A) or (B):*

(A) The colouring Y is given by (1.1.5).

(B) The weights q_n of the renewal measure specified in (1.1.4) satisfy the asymptotics (1.1.10), and the colouring Y obeys

$$\mathbf{E}[Y] = 0 \text{ and } 0 < \mathbf{E}\left[Y^4\right] < \infty. \tag{1.1.12}$$

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 1.1.1, 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; from the very beginning they make substantial use of the specific form (1.1.5) of the colouring of the random graph $\mathcal{G}_\mu$. In [HS13] it is not explicitly discussed whether these arguments also carry over to $S_{\lfloor t_1 n\rfloor}, \ldots, S_{\lfloor t_m n\rfloor}$. However, again thanks to the specific assumption (1.1.5) on Y one can check that this is indeed the case, thus rendering the asserted asymptotic Gaussianity of the finite dimensional distributions of $S^{(n)}$.

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 1.8, is based on insights into the structure of $\mathcal{G}_\mu$ which are stated in Section 1.2 and proved in Sections 1.4-1.7. A key ingredient in the new proof is Theorem 1.3.1, which provides a criterion for the asymptotic Gaussianity in randomly coloured random partitions also in a more general setting. Proposition 1.3.3, which is instrumental in the proof of Theorem 1.3.1, 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 (1.1.12).

Let us also mention that the loss of ground which comes with assuming the "strong renewal" condition (1.1.10) in addition to (1.1.1) and (1.1.2) seems rather minor. Indeed it becomes clear from the examples in [CD19, Section 10] that the class of measures μ which satisfy (1.1.1) and (1.1.2) but fail to satisfy (1.1.10) is rather special.

On the other hand, the benefit of assuming (1.1.10) 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 1.2.1, 1.2.3 and 1.2.4. 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 1.1.1(B).

The following implication of Theorem 1.1.1 is immediate from its introductory discussion.

Corollary 1.1.2. *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 1.9, amends the proof of [HS13, Lemma 4.1], see Remark 1.9.2. Here, for each $n \in \mathbb{N}$ and $T > 0$, $\left(S^{(n)}(t)\right)_{0\le t\le T}$ 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 1.1.3. *Under the assumptions of Theorem 1.1.1, for all $T > 0$ the sequence of random variables $\left(S^{(n)}(t)\right)_{0\le t\le T}$ is tight.*

A direct consequence of Corollary 1.1.2 and Proposition 1.1.3 is

Corollary 1.1.4. *Under the assumptions of Theorem 1.1.1, $S^{(n)}$ converges in distribution (with respect to the topology of locally uniform convergence) to fractional Brownian motion with Hurst parameter $H = \frac{1}{2} + \alpha$.*

1.2 Coalescence probabilities in the Hammond-Sheffield urn

In this section we will assume that the weights q_n of the renewal measure defined in (1.1.4) obey the asymptotics (1.1.10), see the discussion of this condition in Section 1.1.

Proposition 1.2.1. *The coalescence probabilities for the ancestral lineages obey the asymptotics*

$$\mathbf{P}(0 \sim i) \sim C_\alpha \frac{i^{2\alpha-1}}{L(i)^2} \qquad \textit{as } i \to \infty, \tag{1.2.1}$$

with C_α as in (1.1.9).

Remark 1.2.2. (a) The asymptotics (1.1.8) is a direct consequence of (1.2.1). Indeed, the latter implies

$$\sum_{i\in[n]} (n-i)\,\mathbf{P}(i \sim 0) \sim n^{2\alpha+1} \frac{C_\alpha}{L(n)^2} \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{i}{n}\right) \left(\frac{i}{n}\right)^{2\alpha-1} \qquad \text{as } n \to \infty, \tag{1.2.2}$$

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 (1.2.2) and (1.1.8) are equal, this shows the asserted implication.

(b) In the light of the proof of Proposition 1.2.1 (carried out in Section 1.4) we conjecture that increment distributions μ that satisfy (1.1.1) *and* violate (1.1.10), generically also do not admit the asymptotics (1.2.1).

The next result, Proposition 1.2.3, will be instrumental in the proof of Theorem 1.1.1 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 $\mathcal{G}_\mu$, and will bound these probabilities asymptotically as $n \to \infty$ by powers of the "pair coalescence probability". Note that for sufficiently small δ the powers guaranteed by Proposition 1.2.3 are strictly larger than those required in Theorem 1.3.2.

Proposition 1.2.3. *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 $n \to \infty$,*

$$\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)} \sim \mathscr{K}^{(n)}\right) = O\left(n^{4\alpha-2+\delta}\right) = O\left(\left(\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)}\right)\right)^{2-\frac{\delta}{1-2\alpha}}\right), \tag{1.2.3}$$

$$\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)} \sim \mathscr{K}^{(n)} \sim \mathscr{L}^{(n)}\right) = O\left(n^{6\alpha-3+\delta}\right) = O\left(\left(\mathbf{P}\left(\mathscr{I}^{(n)} \sim \mathscr{J}^{(n)}\right)\right)^{3-\frac{\delta}{1-2\alpha}}\right). \tag{1.2.4}$$

Proposition 1.2.3 will be proved in Section 1.6. Although the next result, Proposition 1.2.4, will not be used explicitly in the proof of Theorem 1.1.1, it seems interesting in its own right and also gives an intuition why the estimates in Proposition 1.2.3 should hold. Qualitatively, Proposition 1.2.4 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 Proposition 1.2.3, 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 1.2.4. *Let $\mathscr{M}(0,n) := \max\{j \le 0 : j \sim 0$ 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 1.2.4 will be proved in Section 1.5. The distribution of the random variable D appearing in Proposition 1.2.4 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 1.2 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 1.1.1 under Assumption (B).

Lemma 1.2.5. *Let $\sim$ be the random equivalence relation defined in (1.1.3). For $i, j, k, \ell \in \mathbb{Z}$,*

$$\mathbf{Cov}\left[I_{\{i\sim j\}}, I_{\{k\sim \ell\}}\right] \le \mathbf{P}\left(i \sim j \sim k \sim \ell\right). \tag{1.2.5}$$

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

Remark 1.2.6. The proof of Lemma 1.2.5 (given in Section 1.7) shows that (1.2.5) holds for general increment distributions μ, without the assumptions (1.1.1) and (1.1.2).

1.3 Asymptotic Gaussianity in randomly coloured random partitions

In this section we consider a situation that is more general than the one described in Section 1.1. 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.

$$i \overset{m}{\sim} j :\Longleftrightarrow i \text{ and } j \text{ belong to the same partition element of } \mathscr{P}^{(m)}. \tag{1.3.1}$$

The situation described in Section 1.1 fits into this framework, by choosing $\overset{m}{\sim}$ as the restriction to the set $[m]$ of the equivalence relation $\sim$ defined in (1.1.3). 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

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

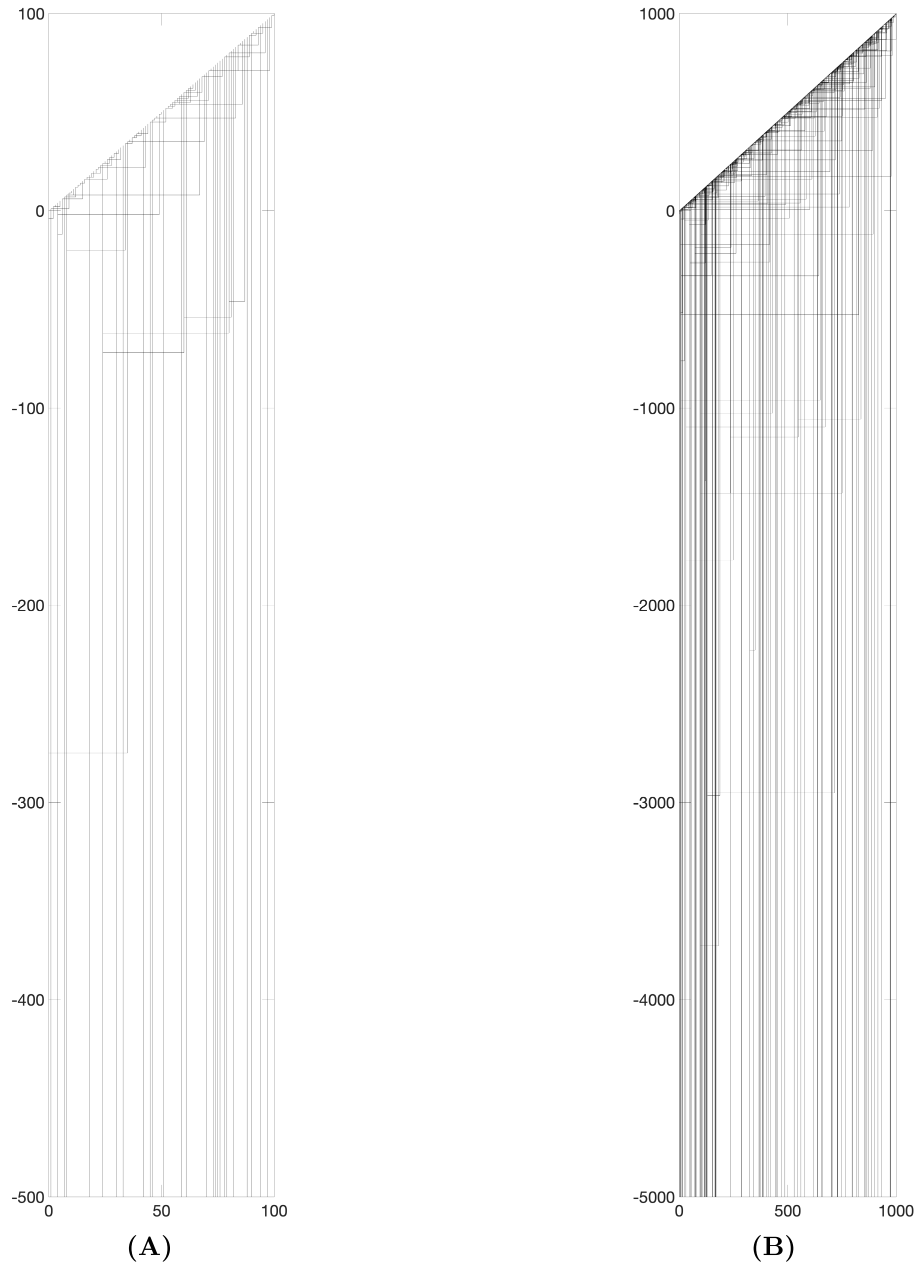


Figure 1.2: This is a simulation of the ancestral lineages of the individuals $0, \ldots, n$, with $n = 100$ in panel (A) and $n = 1000$ in panel (B). Direction of time is vertical, and horizontal lines mark coalescence events. The two panels give an impression of how the genealogical forest of the individuals $i \in [n]$ scales with n, see e.g. Proposition 1.2.4. Like in Figure 1.1, the parameter α was chosen as 0.39.

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

$$\mathscr{Z}^{(m)} := \left(Z^{(m)}_{\lfloor \rho_1 m \rfloor}, \ldots, Z^{(m)}_{\lfloor \rho_d m \rfloor} \right) \tag{1.3.2}$$

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 1.3.1. *The sequence of $\mathbb{R}^d$-valued random variables $\mathscr{Z}^{(m)}$ defined in (1.3.2) is asymptotically Gaussian as $m \to \infty$ provided the following conditions are satisfied:*

$$\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \overset{m}{\sim} \mathscr{K}^{(m)}\right) = o\left(\left(\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)}\right)\right)^{3/2}\right), \tag{1.3.3}$$

$$\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)} \overset{m}{\sim} \mathscr{K}^{(m)} \overset{m}{\sim} \mathscr{L}^{(m)}\right) = o\left(\left(\mathbf{P}\left(\mathscr{I}^{(m)} \overset{m}{\sim} \mathscr{J}^{(m)}\right)\right)^{2}\right), \tag{1.3.4}$$

as $m \to \infty$, and

$$\mathbf{Cov}\left[I_{\{i \overset{m}{\sim} j\}}, I_{\{k \overset{m}{\sim} \ell\}}\right] \le \mathbf{P}\left(i \overset{m}{\sim} j \overset{m}{\sim} k \overset{m}{\sim} \ell\right) \quad \text{for all } m \in \mathbb{N} \text{ and } i,j,k,\ell \in [m], \tag{1.3.5}$$

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

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

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

$$\sum_{i,j=1}^{m} a_i^{(m)} a_j^{(m)} \mathbf{P}\left(i \overset{m}{\sim} j\right) \ge \widetilde{C} \sum_{i,j=1}^{m} \mathbf{P}\left(i \overset{m}{\sim} j\right), \quad m \in \mathbb{N}. \tag{1.3.6}$$

Remark 1.3.2. (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 1.3.1. Indeed, for $d = 1$, condition (1.3.6) is automatically satisfied with $\widetilde{C} := \alpha_1^2$, 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 1.1, then *all* conditions of Theorem 1.3.1 (including condition (1.3.6) for *all* $d \in \mathbb{N}$) are implied by the assumption (1.1.10), see Section 1.8.

Theorem 1.3.1 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 1.3.3. *For fixed $m \in \mathbb{N}$ and $a_1, \dots, a_m \in \mathbb{R}$ let*

$$\bar{S}_m := \sum_{i=1}^{m} a_i Y_i, \quad \bar{\sigma}_m^2 := \mathbf{Var}[\bar{S}_m], \tag{1.3.7}$$

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

$$\begin{aligned} &\left| \mathbf{E}\left[h\left(\frac{\bar{S}_m}{\bar{\sigma}_m} \right) \right] - \mathbf{E}[h(\mathcal{N})] \right| \\ &\leq \frac{c_1(h)}{\bar{\sigma}_m^2} \sqrt{\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 a_i a_j I_{\{i \sim j\}} \right]} + \frac{c_2(h)}{\bar{\sigma}_m^3} \mathbf{E}\left[\sum_{i,j,k=1}^{m} |Y_i|^3 |a_i| a_j a_k I_{\{i \sim j \sim k\}} \right]. \end{aligned} \tag{1.3.8}$$

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

$$X_i := \frac{a_i Y_i}{\bar{\sigma}_m}, \qquad M_i := \{j \in [m] : j \sim i\}, \tag{1.3.9}$$

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

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

$$\mathcal{B} := \sigma(M, X_1, \dots, X_m), \qquad \mathcal{C} := \sigma\left(M, \mathscr{I}, (X_j)_{j \nsim \mathscr{I}}\right)$$

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

$$\begin{aligned} &\left| \mathbf{E}\left[h(W) \right] - \mathbf{E}[h(\mathcal{N})] \right| \\ \leq\ & c_1(h) \left(\sqrt{\mathbf{E}\left[\left(1 - \mathbf{E}\left[G(W - W^*) | \mathcal{B} \right] \right)^2 \right]} + c\, \mathbf{E}\left[\left| \mathbf{E}\left[G \,|\mathcal{C} \right] \right| \right] \right) \\ & + c_2(h)\, \mathbf{E}\left[|G|(W - W^*)^2 \right], \end{aligned} \tag{1.3.10}$$

with $\mathcal{N}$, $c_1(h)$ and $c_2(h)$ as in (1.3.8) and a constant $c > 0$. Let us first turn to the term under the square root on the right-hand side of (1.3.10) 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

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

Hence the term under the square root in (1.3.10) 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 (1.3.10) 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 (1.3.10) equals

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

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

In summary we have shown that the right-hand side of (1.3.10) equals

$$c_1(h)\sqrt{\mathbf{Var}\left[\sum_{i=1}^m X_i\sum_{j\in M_i}X_j\right]}+c_2(h)\,\mathbf{E}\left[\sum_{i=1}^m|X_i|\left(\sum_{j\in M_i}X_j\right)^2\right],\tag{1.3.11}$$

which in turn is equal to the right-hand side of (1.3.8). This concludes the proof of Proposition 1.3.3. □

Proof of Theorem 1.3.1. It suffices to show that for all $d\in\mathbb{N}$ and $(\alpha_1,\dots,\alpha_d)\neq(0,\dots,0)$, the linear combination

$$\bar{Z}_m:=\alpha_1Z^{(m)}_{\lfloor\rho_1 m\rfloor}+\alpha_2\left(Z^{(m)}_{\lfloor\rho_2 m\rfloor}-Z^{(m)}_{\lfloor\rho_1 m\rfloor}\right)+\dots+\alpha_d\left(Z^{(m)}_m-Z^{(m)}_{\lfloor\rho_{d-1}m\rfloor}\right)\tag{1.3.12}$$

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

$$a_i^{(m)}:=\alpha_g\text{ if }\lfloor\rho_{g-1}m\rfloor<i\le\lfloor\rho_g m\rfloor,\quad i=1,\dots m;\quad g=1,\dots,d.\tag{1.3.13}$$

It is then readily checked that $\bar{Z}_m$ defined in (1.3.12) satisfies

$$\bar{Z}_m = \sum_{i=1}^{m} a_i^{(m)} Y_i^{(m)}, \quad m \in \mathbb{N}, \tag{1.3.14}$$

and thus fits into the frame of Proposition 1.3.3, with $\bar{S}_m := \bar{Z}_m$. To use this proposition we will show that under the assumptions (1.3.3), (1.3.4), (1.3.5) and (1.3.6) and with $a_i = a_i^{(m)}$ from (1.3.13), both summands in the right-hand side of (1.3.8) 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

$$\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 a_i a_j I_{\{i\sim j\}}\right] \le C\,\mathbf{Var}\left[\sum_{i,j=1}^{m} Y_i^2 I_{\{i\sim j\}}\right], \tag{1.3.15}$$

$$\mathbf{E}\left[\sum_{i,j,k=1}^{m} |Y_i|^3 |a_i| a_j a_k I_{\{i\sim j\sim k\}}\right] \le C\,\mathbf{E}\left[|Y|^3\right] \sum_{i,j,k=1}^{m} \mathbf{P}(i\sim j\sim k). \tag{1.3.16}$$

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 (1.3.5) is not larger than

$$\mathbf{E}\left[Y^2\right]^2 \sum_{i,j,k,\ell\in[m]} \mathbf{P}(i\sim j\sim k\sim \ell). \tag{1.3.17}$$

Next we note that

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

Taking expectation of the latter and adding this to (1.3.17) we obtain

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

which in view of (1.3.4) is $o\left(m^2\,\mathbf{P}\left(\mathscr{I}\sim\mathscr{J}\right)\right)^2$.

Likewise the right-hand side of (1.3.16) is $o\left(m^2\left(\mathbf{P}\left(\mathscr{I}\sim\mathscr{J}\right)\right)^{3/2}\right)$. From (1.3.7) and (1.3.14) we get

$$\bar{\sigma}_m^2 = \sum_{i,j\in[m]} a_i^{(m)} a_j^{(m)}\,\mathbf{Cov}\left[Y_i, Y_j\right] = \mathbf{Var}[Y] \sum_{i,j\in[m]} a_i^{(m)} a_j^{(m)}\,\mathbf{P}(i\sim j), \tag{1.3.18}$$

which due to assumption (1.3.6) is bounded from below by $\widetilde{C}m^2\,\mathbf{P}(\mathscr{I}\sim\mathscr{J})$. Thus under the assumptions of Theorem 1.3.1 the right-hand side of (1.3.8) converges to 0 as $m\to\infty$, which shows that Theorem 1.3.1 is a consequence of Proposition 1.3.3. □

1.4 Pair coalescence probabilities: Proof of Proposition 1.2.1

Let $R_k^{(i)}$, $i\in\mathbb{Z}$, $k\in\mathbb{N}$ be independent copies of R, and define

$$A_i := \left\{n\in\mathbb{Z} : i - R_1^{(i)} - \cdots - R_j^{(i)} = n \text{ for some } j\ge 0\right\}. \tag{1.4.1}$$

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 $i>0$

$$\mathbf{P}\left(A_0\cap A_i\neq\emptyset\right)\sum_{m\ge 0} q_m^2 = \mathbf{E}\left[|A_0\cap A_i|\right] = \sum_{m\ge 0} q_m q_{m+i}, \tag{1.4.2}$$

hence

$$\mathbf{P}(0\sim i) = \frac{\sum_{m\ge 0} q_m q_{m+i}}{\sum_{m\ge 0} q_m^2}, \quad i\ge 0. \tag{1.4.3}$$

We will now assume (in accordance with the assumptions in Proposition 1.2.1) that the weights q_n of the renewal measure defined in (1.1.4) have the property (1.1.10). Under this condition we will prove

Proposition 1.4.1. *As $m\to\infty$,*

$$\sum_{j\ge 1} q_j q_{m+j} \sim \frac{1}{\Gamma(\alpha)^2\Gamma(1-\alpha)^2}\frac{m^{2\alpha-1}}{L(m)^2}\int_0^\infty (1+x)^{\alpha-1}x^{\alpha-1}\mathrm{d}x. \tag{1.4.4}$$

The asymptotics (1.2.1) claimed in Proposition 1.2.1 is immediate from (1.4.3) combined with (1.4.4).

The remainder of this section is devoted to the proof of Proposition 1.4.1. We will prove (1.4.4) first under a special assumption on the Karamata representation of the slowly varying function L.

Lemma 1.4.2. *Let $\alpha \in (0, \frac{1}{2})$, and consider*

$$r_n := n^{\alpha-1} K(n) \tag{1.4.5}$$

where $K(n)$ is of the form

$$K(n) = \exp\left(\int_B^n \frac{\ell(t)}{t}\,\mathrm{d}t\right) \tag{1.4.6}$$

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

$$\sum_{j\geq 1} r_j r_{i+j} \sim K(i)^2 i^{2\alpha-1} \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} \mathrm{d}x \quad \text{as } i \to \infty, \tag{1.4.7}$$

with

$$\int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx = B(\alpha, 1-2\alpha) = \frac{\Gamma(\alpha)\Gamma(1-2\alpha)}{\Gamma(1-\alpha)}. \tag{1.4.8}$$

Proof. a) The equality (1.4.8) is readily checked by substituting $y = \frac{x}{1+x}$.

b) The fact that (r_n) is ultimately decreasing follows from (1.4.5) together with the Karamata representation (1.4.6) of $K(n)$. To see this, we argue as follows, putting $\beta := 1-\alpha$. Since $\ell(t)$ 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 $\ell(t) < \beta$. This implies

$$\begin{aligned}
\frac{r_n}{r_{n+1}} &= \frac{n^{-\beta}}{(n+1)^{-\beta}} \cdot \frac{K(n)}{K(n+1)} \\
&= \frac{n^{-\beta}}{(n+1)^{-\beta}} \exp\left(-\int_n^{n+1} \frac{\ell(t)}{t}\mathrm{d}t\right) \\
&= \exp\left(\beta\left(\ln(n+1) - \ln(n)\right) - \int_n^{n+1} \frac{\ell(t)}{t}\mathrm{d}t\right) \\
&= \exp\left(\int_n^{n+1} \left(\frac{\beta}{t} - \frac{\ell(t)}{t}\right)\mathrm{d}t\right).
\end{aligned}$$

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 (1.4.5), the claimed asymptotics (1.4.7) is equivalent to

$$\lim_{i\to\infty} \frac{1}{i} \sum_{j\geq 1} \frac{r_j}{r_i} \frac{r_{i+j}}{r_i} = \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} \mathrm{d}x. \tag{1.4.9}$$

We now set out to prove (1.4.9). 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

$$\frac{1}{i} \sum_{j \geq Ni} \frac{r_j}{r_i} \frac{r_{i+j}}{r_i} < \varepsilon. \tag{1.4.10}$$

Since (r_n) is ultimately decreasing, (1.4.10) will follow if we can show that exists an $N \in \mathbb{N}$ such that for all sufficiently large i

$$\frac{1}{i} \sum_{j \geq Ni} \left(\frac{r_j}{r_i} \right)^2 < \varepsilon. \tag{1.4.11}$$

Again because of the ultimate monotonicity of (r_n), the left-hand side of (1.4.11) is for sufficiently large i bounded from above by

$$\sum_{m=N}^{\infty} \left(\frac{r_{mi}}{r_i} \right)^2 = \sum_{m=N}^{\infty} m^{2\alpha-2} \left(\frac{K(mi)}{K(i)} \right)^2. \tag{1.4.12}$$

Using (1.4.6) one obtains that for any $\delta > 0$ and i so large that $\ell(t) < \delta$ for all $t \geq i$,

$$\frac{K(mi)}{K(i)} = \exp\left(\int_i^{mi} \frac{\ell(t)}{t} \mathrm{d}t \right) \leq \exp\left(\delta \left(\ln(mi) - \ln(i) \right) \right) \leq m^{\delta}, \tag{1.4.13}$$

which implies by dominated convergence that for sufficiently large N the right-hand side of (1.4.12) is smaller than ε for all sufficiently large i. We have thus proved (1.4.10).

Next we show that for all $\varepsilon > 0$ there exists an $\eta \in (0, \infty)$ such that for all sufficiently large i

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j}{r_i} \frac{r_{i+j}}{r_i} < \varepsilon. \tag{1.4.14}$$

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

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j}{r_i} < \varepsilon. \tag{1.4.15}$$

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

$$\sum_{j \leq \eta i} r_j \sim \frac{1}{\alpha} (\eta i)^{\alpha} K(\lfloor \eta i \rfloor) \quad \text{as } i \to \infty, \tag{1.4.16}$$

and hence

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j}{r_i} \sim \frac{1}{\alpha} \eta^{\alpha} \frac{K(\lfloor \eta i \rfloor)}{K(i)} \sim \frac{\eta^{\alpha}}{\alpha} \quad \text{as } i \to \infty, \tag{1.4.17}$$

which proves (1.4.15), and hence also (1.4.14). (The last asymptotic is by the fact that K is slowly varying.)

In view of (1.4.5), (1.4.10) and (1.4.14), for proving (1.4.7) it remains to show that

$$\lim_{i\to\infty}\frac{1}{i}\sum_{\eta i\le j\le Ni}\frac{K(j)}{K(i)}\frac{K(i+j)}{K(i)}\left(1+\frac{j}{i}\right)^{\alpha-1}\left(\frac{j}{i}\right)^{\alpha-1}=\int_{\eta}^{N}(1+x)^{\alpha-1}x^{\alpha-1}\mathrm{d}x. \tag{1.4.18}$$

From (1.4.6) one derives that

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

Hence (1.4.18) boils down to a convergence of Riemann sums to its integral limit, and the proof of Lemma 1.4.2 is done. □

Let us now complete the proof of Proposition 1.4.1.

Proof. The asymptotics (1.1.10) can be rewritten as

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

where $\tilde{L}(n)$ is a slowly varying function and $C_n \to C > 0$.

As in the proof of Lemma 1.4.2 it suffices to show that

$$\frac{1}{i}\sum_{j\ge 1}\frac{q_j}{q_i}\frac{q_{i+j}}{q_i}\to\int_0^{\infty}(1+x)^{\alpha-1}x^{\alpha-1}\mathrm{d}x. \tag{1.4.19}$$

Because of the Karamata representation theorem (see e.g. Theorem 1.3.1 in [BGT87]) there exists a $K(n)$ satisfying (1.4.6) and a sequence D_n converging to a positive constant D such that

$$\tilde{L}(n) = D_n K(n), \quad n = 1, 2, \ldots \tag{1.4.20}$$

Defining r_n as in (1.4.5) 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 (1.4.9) nor that of the left-hand side of (1.4.19) reacts to the omission of a fixed finite number of summands, we see that (1.4.9) carries over to (1.4.19). □

1.5 Depth of most recent common ancestor: Proof of Proposition 1.2.4

In this section we will assume that the weights q_n of the renewal measure defined in (1.1.4) obey the asymptotics (1.1.10), see the discussion after equation (1.1.10). For $i, m \in \mathbb{N}$ we set

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

For the independent couplings A_i, $i \in \mathbb{Z}$, of the ancestral lines of 0 and i as defined in (1.4.1) we have for all $r > 0$ and $i \in \mathbb{Z}$

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

and consequently

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

As in the proof of Proposition 1.4.1 we obtain

$$\sum_{k=0}^{ri} q_k q_{k+i} \sim i^{2\alpha-1} \frac{1}{\Gamma(1-\alpha)^2 \Gamma(\alpha)^2} \int_0^r x^{\alpha-1}(1+x)^{\alpha-1} \mathrm{d}x \quad \text{as } i \to \infty. \tag{1.5.1}$$

Together with the asymptotics (1.2.1) this gives

$$\liminf_{i\to\infty} \mathbf{P}\left(D_i \leq ri \,|\, 0 \sim i\right) \geq B(\alpha, 1-2\alpha)^{-1} \int_0^r x^{\alpha-1}(1+x)^{\alpha-1} \mathrm{d}x. \tag{1.5.2}$$

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

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

Using (1.5.1), 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 (1.5.2) gives the assertion of Proposition 1.2.4.

1.6 Triplet and quartet coalescence probabilities: Proof of Proposition 1.2.3

In this section we will assume that the weights q_n of the renewal measure defined in (1.1.4) obey the asymptotics (1.1.10). We now turn to the asymptotic analysis of triplet and quartet coalescence probabilities.

1.6.1 Triplet coalescence probabilities

Lemma 1.6.1. *For all $r > 0$ and $i \in \mathbb{N}$ we have for a slowly varying function $\tilde{L}$ not depending on r*

$$\mathbf{P}\left(0 \sim i \sim \lfloor (1+r)i \rfloor\right) \leq \left(r^{\alpha-1} + r^{2\alpha-1}\right) \tilde{L}(i)\, i^{4\alpha-2}. \tag{1.6.1}$$

Proof. Let (A_k) be the independent (non-merging) ancestral lineages defined in (1.4.1). We set

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

$$B_{k,\ell} := \bigcup_{g \in A_k \cap A_\ell} A_g. \tag{1.6.2}$$

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

$$\begin{aligned}
& \mathbf{P}\left(0 \sim i \sim (1+r)i\right) \\
\leq\ & \mathbf{P}\left(E; \mathscr{M}(0,i) \geq \mathscr{M}(0,(1+r)i\right) + \mathbf{P}\left(E; \mathscr{M}(0,(1+r)i) \geq \mathscr{M}(0,i)\right) \\
& + \mathbf{P}\left(E; \mathscr{M}(i,(1+r)i) \geq \mathscr{M}(0,i)\right) \\
\leq\ & \mathbf{P}\left(A_{0,i} \cap A_{(1+r)i} \neq \emptyset\right) + \mathbf{P}\left(A_{0,(1+r)i} \cap A_i \neq \emptyset\right) + \mathbf{P}\left(A_{i,(1+r)i} \cap A_0 \neq \emptyset\right) \\
\leq\ & \mathbf{E}\left[|A_{0,i} \cap A_{(1+r)i}|\right] + \mathbf{E}\left[|A_{0,(1+r)i} \cap A_i|\right] + \mathbf{E}\left[|A_{i,(1+r)i} \cap A_0|\right] \qquad (1.6.3) \\
\leq\ & \mathbf{E}\left[|B_{0,i} \cap A_{(1+r)i}|\right] + \mathbf{E}\left[|B_{0,(1+r)i} \cap A_i|\right] + \mathbf{E}\left[|B_{i,(1+r)i} \cap A_0|\right]. \\
=\ & \sum_{m \geq 0} q_m q_{ri+m} \sum_{n \geq 0} q_n q_{n+i-m} + \sum_{m \geq 0} q_m q_{m+(1+r)i} \sum_{n \geq 0} q_n q_{n+i+m} \\
& + \sum_{m \geq 0} q_m q_{ri+m} \sum_{n \geq 0} q_n q_{n+i-m}
\end{aligned}$$

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

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

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 (1.4.18) was derived first for slowly varying functions satisfying (1.4.6) and then, using the Karamata representation, for general slowly varying functions obeying (1.4.20). The first and the second summand on the r.h.s of (1.6.3) 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. □

Proof of Proposition 1.2.3 Part 1. We set out to show (1.2.3), and first observe that

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbf{P}\left(0\sim i\sim j\right)\leq\sum_{i=0}^{n}\mathbf{P}\left(0\sim i\right)+2\sum_{i=1}^{n}\sum_{j=1}^{n-i}\mathbf{P}\left(0\sim i\sim i+j\right) \tag{1.6.4}$$

By Lemma 1.6.1 we get, noting that $r^{\alpha-1}+r^{2\alpha-1}\leq 2r^{\alpha-1}+1$,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n-i}\mathbf{P}\left(0\sim i\sim i+j\right)&\leq\sum_{i=1}^{n}\sum_{j=1}^{n-i}\tilde{L}(i)i^{4\alpha-1}\left(\left(\frac{j}{i}\right)^{\alpha-1}+1\right)\\&\leq\tilde{L}(n)\left(\sum_{i=1}^{n}i^{3\alpha-1}(n-i)^{\alpha}+\sum_{i=1}^{n}i^{4\alpha-1}\right)=O\left(n^{4\alpha+\delta}\right)\end{aligned} \tag{1.6.5}$$

for each $\delta>0$. From (1.6.4) combined with (1.2.1) and (1.6.5) we obtain for each $\delta>0$ the estimate

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbf{P}\left(0\sim i\sim j\right)\leq Cn^{4\alpha+\delta}, \tag{1.6.6}$$

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

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbf{P}\left(k\sim i\sim j\right)\leq Cn^{4\alpha+\delta},\quad k\in[n], \tag{1.6.7}$$

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.)

The first equality in (1.2.3) now follows directly from (1.6.7), and the second equality in (1.2.3) is immediate from (1.1.8). □

1.6.2 Quartet coalescence probabilities

Lemma 1.6.2. *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*

$$\begin{aligned}&\mathbf{P}\left(0\sim i\sim\lfloor(1+r_1)i\rfloor\sim\lfloor(1+r_2)i\rfloor\right)\\\leq\;&\left(r_1^{\alpha-1}+r_1^{2\alpha-1}+r_2^{\alpha-1}+r_2^{2\alpha-1}\right)\bar{L}(i)i^{6\alpha-3}.\end{aligned} \tag{1.6.8}$$

Proof. Again let (A_i) be the independent (non-merging) ancestral lineages defined in (1.4.1). Let $B_{k,\ell}$ be as in (1.6.2) 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 $j_1 := (1+r_1)i, \quad j_2 := (1+r_2)i$,

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

We now argue in a similar way as in the proof of Proposition 1.6.1. By subadditivity we get

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

By the very same arguments as in the proof of Lemma 1.6.1 one checks that each of the twelve summands is bounded by the right-hand side of (1.6.8). □

Proof of Proposition 1.2.3 Part 2. We are now going to prove (1.2.4), and first set out to show that for all $\delta > 0$

$$\sum_{i=1}^{n} \sum_{j=1}^{n-i} \sum_{\ell=1}^{n-i-j} \mathbf{P}\left(0 \sim i \sim i+j \sim i+j+\ell\right) = O\left(n^{6\alpha+\delta}\right). \tag{1.6.9}$$

By Lemma 1.6.2, and since $r_1^{\alpha-1} + r_1^{2\alpha-1} + r_2^{\alpha-1} + r_2^{2\alpha-1} \leq 4(r_1^{\alpha-1}+1)$, the left-hand side. of (1.6.9) 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 (1.6.4) in the proof of Part 1, this implies (1.6.9). 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 (1.2.1) and (1.2.3) it is clear that we may then restrict to pairwise distinct i, j, ℓ. The first equality in (1.2.4) thus follows from (1.6.9), and (like in Part 1) the second equality in (1.2.4) is clear from (1.1.8). □

1.7 A covariance estimate: Proof of Lemma 1.2.5

For $i = j$ or $k = \ell$ the assertion of Lemma 1.2.5 is clearly true because then the left-hand side of (1.2.5) vanishes. For $i = k$ we have

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

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

$$\mathbf{Cov}\left[I_{i\sim k}, I_{j\sim \ell}\right] = \mathbf{P}\left(i \sim k \cap j \sim \ell\right) - \mathbf{P}\left(i \sim k\right)\mathbf{P}\left(j \sim \ell\right) \tag{1.7.1}$$

and

$$\mathbf{P}\left(i \sim k \cap j \sim \ell\right) = \mathbf{P}\left(i \sim k \not\sim j \sim \ell\right) + \mathbf{P}\left(i \sim k \sim j \sim \ell\right). \tag{1.7.2}$$

By (1.7.1) and (1.7.2), the inequality (1.2.5) is immediate from the following

Lemma 1.7.1. *For pairwise distinct i, j, k, ℓ we have*

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

Proof. Let A_g, $g \in \mathbb{Z}$, be defined as in (1.4.1). 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

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

which because of mutual independence of the A_g gives the assertion of the lemma. □

1.8 Asymptotic Gaussianity in the Hammond-Sheffield urn: Proof of Theorem 1.1.1(B)

We are going to apply Theorem 1.3.1, 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 (1.1.3). For $d \in \mathbb{N}$ and $t_1 < \cdots < t_d$ as prescribed in Theorem 1.1.1 we apply Theorem 1.3.1 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 1.1.1, Proposition 1.2.3 ensures the validity of assumptions (1.3.3) and (1.3.4), and Lemma 1.2.5 guarantees that (1.3.5) is fulfilled. It remains to check the assumption (1.3.6). Indeed, with

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

because of $\mathbf{P}(i \sim j) = \mathbf{P}(0 \sim |i-j|)$ and in view of Proposition 1.2.1 the left-hand side. of (1.3.6) has the asymptotics

$$m^{-2} \sum_{i,j \in [m]} a_i^{(m)} a_j^{(m)} \mathbf{P}(i \sim j) \sim \frac{Cm^{2\alpha-1}}{L(m)^2} \int_0^1 \int_0^1 a(x)a(y)|x-y|^{2\alpha-1} \mathrm{d}x \, \mathrm{d}y \quad \text{as } m \to \infty \tag{1.8.1}$$

The Riesz kernel $|x-y|^{2\alpha-1}$ is positive definite (see e.g. [Dos98]), hence the integral term in (1.8.1) is strictly positive, and consequently the left-hand side of (1.8.1) is of the order $m^{2\alpha+1}$ as $m \to \infty$. Because of (1.1.8), this is also the order of the right-hand side of (1.3.6).

1.9 Tightness: Proof of Proposition 1.1.3

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 1.9.1 ([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 $n \geq N$*

$$\mathbf{E}\left[\left| \zeta^{(n)}(t) - \zeta^{(n)}(s) \right| \cdot \left| \zeta^{(n)}(u) - \zeta^{(n)}(t) \right| \right] \leq \left[F(u) - F(s) \right]^{\gamma}. \tag{1.9.1}$$

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

Proof of Proposition 1.1.3. We first note the following immediate consequence of (1.1.7) and (1.1.8): There exist constants $c, c' > 0$ such that

$$cn^{2\alpha+1} L(n)^{-2} \leq \sigma_n^2 \leq c' n^{2\alpha+1} L(n)^{-2}, \quad \forall n \in \mathbb{N}. \tag{1.9.2}$$

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

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

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

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

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

$$
\begin{aligned}
\mathbf{Var}\left[S^{(n)}(t)-S^{(n)}(s)\right] &\leq \left[2\,\mathbf{Var}\left[S^{(k-j)}\right]+2c_1\,\mathbf{Var}\left[Y_{j+1}+Y_{k+1}\right]\right]\sigma_n^{-2} \\
&\leq \left[2c'(k-j)^{2\alpha+1}L(k-j)^{-2}+c_2\right]\ \sigma_n^{-2} \\
&\leq \left[c_3(k-j)^{2\alpha+1}L(k-j)^{-2}\right]\sigma_n^{-2} \\
&\leq c_4\left(\frac{k-j}{n}\right)^{2\alpha+1}\left(\frac{L(n)}{L(k-j)}\right)^2. \qquad (1.9.3)
\end{aligned}
$$

Here the first inequality holds because for any two square-integrable random variables G_1, G_2 one has $\mathbf{Var}\left[G_1+G_2\right] \leq 2\,\mathbf{Var}[G_1]+2\,\mathbf{Var}[G_2]$, the second and the last inequality hold because of (1.9.2), 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,

$$
\mathbf{Var}\left[S^{(n)}(u)-S^{(n)}(t)\right] \leq c_3\left(\frac{\ell-k}{n}\right)^{2\alpha+1}\left(\frac{L(n)}{L(\ell-k)}\right)^2. \qquad (1.9.4)
$$

To use Lemma 1.9.1 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 $\ell \leq j+1$ 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:

$$
\begin{aligned}
&\mathbf{E}\left[|S^{(n)}(t)-S^{(n)}(s)|\cdot|S^{(n)}(u)-S^{(n)}(s)|\right] \\
\leq\ & \left(\tilde{c}n^{-\frac{1}{2}-\alpha}L(n)\right)^2\left((t-s)\cdot(u-s)\right) \\
\leq\ & \left(\tilde{c}n^{-\frac{1}{2}}\right)^2(u-s)^{1+2\alpha} = \tilde{c}^2\frac{1}{n}(u-s)^{1+2\alpha} \\
\leq\ & \tilde{c}^2(u-s)^{1+2\alpha} = \left[c_0u-c_0s\right]^{1+2\alpha}
\end{aligned}
$$

for some $c_0>0$. Now assume $\ell > j+1$. Cauchy-Schwarz and the estimates (1.9.3),(1.9.4) yield

$$
\begin{aligned}
&\mathbf{E}\left[\left|S^{(n)}(t)-S^{(n)}(s)\right|\cdot\left|S^{(n)}(u)-S^{(n)}(t)\right|\right] \\
\leq\ & \sqrt{\mathbf{Var}\left[S^{(n)}(t)-S^{(n)}(s)\right]}\sqrt{\mathbf{Var}\left[S^{(n)}(u)-S^{(n)}(t)\right]} \\
\leq\ & \sqrt{c_3\left(\frac{k-j}{n}\right)^{2\alpha+1}\left(\frac{L(n)}{L(k-j)}\right)^2}\sqrt{c_3\left(\frac{\ell-k}{n}\right)^{2\alpha+1}\left(\frac{L(n)}{L(\ell-k)}\right)^2} \\
\leq\ & c_3\left(\frac{\ell-j}{n}\right)^{2\alpha+1}\frac{L(n)^2}{L(k-j)L(\ell-k)}
\end{aligned}
$$

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

(The third inequality holds because of $4|k-j|\cdot|\ell-k| \leq |\ell-j|^2$, 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 (1.9.1) 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 1.1.2 we can thus apply Lemma 1.9.1 and conclude the assertion of Proposition 1.1.3. □

Remark 1.9.2. 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 1.1.4 and the Borell-TIS inequality.

1.10 Coalescence probabilities in long-range seedbanks

In this section we will assume that the weights q_n of the renewal measure defined in (1.1.4) obey the asymptotics (1.1.10), see the discussion after equation (1.1.10). Following [BGKS13] we extend the model described in Section 1.1 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 1.4.1, which also provides a proof of [BGKS13, Lemma 3.1 (c)], we arrive at the following analogue of Proposition 1.2.1 (see also [BGKS13, Theorem 3(c))]:

Proposition 1.10.1. *For all $k, \ell \in [N]$*

$$\mathbf{P}\big((0,k) \sim (i,\ell)\big) \sim C_{\alpha,N} \frac{i^{2\alpha-1}}{L(i)^2} \qquad \text{as } i \to \infty, \tag{1.10.1}$$

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 (1.4.1). Like in (1.4.2) 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 (1.10.1) is now immediate from Proposition 1.4.1. □

Remark 1.10.2. 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 1.2.3 one can derive that for all $\delta > 0$ and for a constant C not depending on n and N

$$\begin{aligned} \mathbf{P}\left(\mathscr{I} \sim \mathscr{J} \sim \mathscr{K}\right) &\leq \frac{C}{N^2} n^{4\alpha-2+\delta}, \\ \mathbf{P}\left(\mathscr{I} \sim \mathscr{J} \sim \mathscr{K} \sim \mathscr{L}\right) &\leq \frac{C}{N^3} n^{6\alpha-3+\delta}. \end{aligned}$$

Consequently, along the lines of the proof of Theorem 1.1.1 one obtains the convergence of an analogue of $(S^{(n)})$ towards fractional Brownian motion also in the long-range seedbank model of [BGKS13].

Acknowledgment. We thank Matthias Birkner, Florin Boenkost, Adrian González Casanova, Alan Hammond and Nicola Kistler for stimulating discussions and valuable hints. We also thank the *Allianz für Hochleistungsrechnen Rheinland-Pfalz* for granting us access to the High Performance Computing Elwetritsch, on which simulations have been performed which inspired results in this work. We are also grateful to two anonymous referees whose careful reading and thoughtful comments led to a substantial improvement of the presentation.

Chapter 2

Branching fractional Brownian motion: discrete approximations and maximal displacement[1]

[1]On arXiv as [GI23]

Abstract

We construct and study branching fractional Brownian motion with Hurst parameter $H \in (1/2, 1)$. The construction relies on a generalization of the discrete approximation of fractional Brownian motion (Hammond and Sheffield [HS13]) to power law Pólya urns indexed by trees. We show that the first order of the speed of branching fractional Brownian motion with Hurst parameter H is $ct^{H+1/2}$ where c is explicit and only depends on the Hurst parameter. A notion of "branching property" for processes with memory emerges naturally from our construction.

2.1 Introduction

Fractional Brownian Motion with Hurst parameter H, also known as H-fractional Brownian motion, is the unique centered Gaussian process B^H with stationary increments and variance $\mathbf{Var}\left[B_t^H\right] = t^{2H}$. We will restrict to the case $H \in (\frac{1}{2}, 1)$, for which fractional Brownian Motion has positively correlated increments (and lacks the Markov property). By a well-known result of Mandelbrot and van Ness [MVN68], fractional Brownian motion (FBM) has a kernel representation in terms of Wiener increments $\mathrm{d}W(s)$:

$$B^H(t) = \int_{(-\infty,t]} K(s,t)\,\mathrm{d}W(s), \quad t \in \mathbb{R}. \tag{2.1.1}$$

The definition of the kernel $K = K^H$ will be recalled in Section 2.3.1. In the following we will construct and analyse a *branching* version of fractional Brownian motion, more specifically *fractional Brownian motion indexed by a Yule tree* $\mathfrak{Y}$. This construction works in two steps: the first is for a given realisation $\mathfrak{y}$ of $\mathfrak{Y}$, the second is a randomisation over $\mathfrak{y}$.

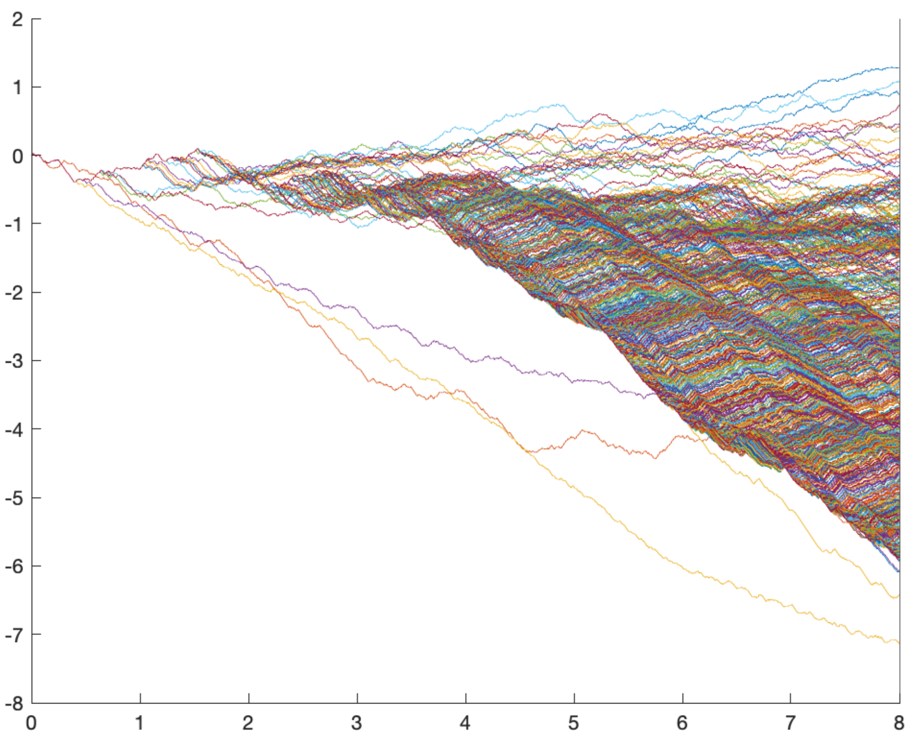


Figure 2.1: This shows a simulation of the branching fractional Brownian motion with Hurst parameter $H = 0.85$. The y-axis is measured in units of $\left(\sum_{\ell \geq 0} q_\ell^2\right)^{-\frac{1}{2}}$ for q_ℓ defined by (2.2.10) and $\alpha = H - \frac{1}{2}$. The simulation is based on the discrete approximation described in Section 2.2.2

The representation (2.1.1) lends itself for a construction of the $\mathfrak{y}$-indexed FBM. To this purpose, think of $\mathfrak{y}$ as the disjoint union of countably many edges $\{h\} \times (\ell_h, r_h]$, where

$h \in \bigcup_{n\in\mathbb{N}_0}\{0,1\}^n$ is the Ulam-Harris name (see [Har63, Section 2.1] and Section 2.2.2) of the edge and $(\ell_h, r_h]$ is the life time interval of the edge. For each $v = (h,t) \in \mathfrak{y}$, let $t(v) := t$. In this way, the *ancestral lineage* $\mathfrak{a}(v) \subset \mathfrak{y}$ of each $v \in \mathfrak{y}$ is in isometric correspondence with $(-\infty, t(v)]$. Let $dW_{\mathfrak{y}}(v)$, $v \in \mathfrak{y}$, be a standard Gaussian white noise on $\mathfrak{y}$. Then the $\mathfrak{y}$*-indexed FBM* has the representation

$$B^H_{\mathfrak{y}}(v) := \int_{\mathfrak{a}(v)} K(t(u), t(v))\, \mathrm{d}W_{\mathfrak{y}}(u), \quad v \in \mathfrak{y}. \tag{2.1.2}$$

This construction is implicit in work by Adler and Samorodnitsky, see [AS95], who introduced branching fractional Brownian as a starting point for their construction of super fractional Brownian motion as a process of measures on the path space that can be seen as a high density limit of branching fractional Brownian motion. For related approaches to tree-indexed processes with memory see [KLS19], who studied the long-term behaviour of their cloud of particles with a focus on the bulk.

Our paper has two main contributions. Firstly, in Theorem 2.3.2 we will construct discrete approximations of $\mathfrak{y}$-indexed FBM $B^H_{\mathfrak{y}}$ (and thus also the branching fractional Brownian motion) for $H > \frac{1}{2}$ via a discrete approximation. This approach is based on the power law Pólya's urn of Hammond and Sheffield [HS13] and a recent analysis of the random genealogy that underlies this urn, see [IW23].

In a nutshell, the approximation of B^H works as follows. For $\alpha := H - \frac{1}{2}$, consider a family $(R_i)_{i\in\mathbb{Z}}$ of independent $\mathbb{N}$-valued random variables with distribution

$$\mu\left(\{n, n+1, \dots\}\right) = n^{-\alpha}, \qquad n \in \mathbb{N}. \tag{2.1.3}$$

The types (say -1 or $+1$) of the individuals $i \in \mathbb{Z}$ are determined recursively: Each individual i selects $i - R_i$ as its parent and inherits the parent's type. As it turns out, this leads to a forest of countably many trees, see Section 2.2.1. If each of these trees is assigned the type -1 or $+1$ by a fair coin tossing, it further turns out that the rescaled sums over the types of the individuals $1, \dots, \lfloor tn \rfloor$, $t \geq 0$, converge to fractional Brownian motion with Hurst parameter H as $n \to \infty$.

The proof of Theorem 2.3.2, which will be carried out in Section 2.6, reveals interesting analytic identities. While the representation (2.1.1) (as well as (2.1.2)) has the flavour of a moving average, the Hammond-Sheffield approximations of FBM is more in the spirit of an autoregression. Indeed, as an aside of our proof of Theorem 2.3.2, we will obtain a "microscopic" interpretation of the prediction formula of Gripenberg and Norros [GN96].

The simulation depicted in Figure 2.1 is based on the discrete approximation provided by Theorem 2.3.2.

Our second main contribution in this paper is the analysis of the speed of the rightmost particle of branching fractional Brownian motion, in the spirit of McKean's celebrated work [McK75] about branching Brownian motion, which was later substantially refined by Bramson [Bra78] and Lalley and Sellke [LS87]. Specifically we will show that the maximum $M^H(t)$ of a BFBM with Hurst parameter $H > \frac{1}{2}$ behaves asymptotically like

$$m(t) := t^{H+\frac{1}{2}} \sqrt{\frac{\sqrt{\pi}2^{2H+1}H}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)\left(H+\frac{1}{2}\right)^2}} \tag{2.1.4}$$

in the sense that for all $\varepsilon > 0$

$$\mathcal{P}\left(\left|\frac{M^H(T)}{m(t)} - 1\right| > \varepsilon\right) \to 0 \quad \text{for } t \to \infty. \tag{2.1.5}$$

See Figure 2.2 for an illustration of $m(t)$ for t fixed and varying H. Note that for $H = \frac{1}{2}$ $m(t)$ equals $\sqrt{2}t$, which is consistent with the results about branching Brownian motion.

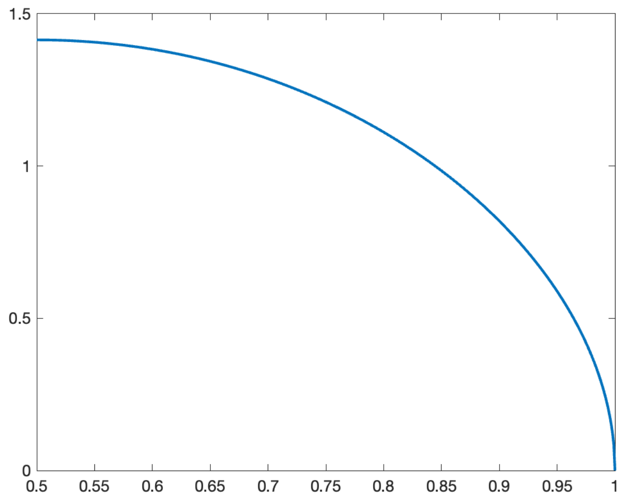


Figure 2.2: This is an illustration of the behaviour of $m(t)/t^{H+\frac{1}{2}}$ as a function of H.

Up to the constant factor in (2.1.4), the leading order $t^{H+\frac{1}{2}}$ of the maximum can be easily explained. Indeed, for a number $\lfloor e^t \rfloor$ of independent normally distributed random variables with variance t^{2H} the leading order of the maximum would be given by

$$\mathfrak{m}(t) := \sqrt{2}\; t^{H+\frac{1}{2}}$$

as suggested by the estimate

$$e^t\, \mathbf{P}\left(\mathcal{N}(0, t^{2H}) \geq \mathfrak{m}(t)\right) \approx e^t \exp\left(-\frac{\mathfrak{m}(t)^2}{2t^{2H}}\right) = 1.$$

Due to correlations we obtain (2.1.2) which only differs by the constant prefactor depending on H. The statement on $M^H(t)$ is proved in Theorem 2.3.8, which in turn is strongly connected to Theorem 2.3.6, see Remark 2.3.9. The proof techniques of these theorems rely heavily on exploring the connection between branching random walks and the so-called generalised random energy models (GREM). Those models were introduced and discussed by Derrida in the 1980's in [Der80], [Der85] and [Der81]. An introduction to those can be found in Lecture Notes by Kistler [Kis15].

The connection between GREMs and branching random walks has first been explored by Arguin, Bovier and Kistler in [ABK11] for branching Brownian motion. In contrast to Brownian motion fractional Brownian motion is a process with memory. The corresponding GREM is then a continuous GREM with decreasing variance. Of such a GREM, Bovier and Kurkova analyse the leading order of the maximum in [BK04b]. Their results lead to Theorem 2.3.6, where the underlying tree is a deterministic binary branching tree, see Section 2.7. For subleading orders work by Maillard and Zeitouni [MZ16] allows some conjectures in our regime, see Remark 2.3.12. For the proof of Theorem 2.3.8, where the underlying tree is a Yule tree, we rely on methods developed first in [ABK11] and later in [KS15] by Kistler and Schmidt. In Remark 2.3.13 we give a more detailed account of (some of) the literature on the GREM, with an emphasis on work that is relevant for the ideas that are used in the present paper.

Our proof of Theorem 2.3.8 is conceptual in the sense that it makes use of the connection between GREMs and branching random walks, Section 2.8. As pointed out to us by an anonymous reviewer of a previous version of this paper, Theorem 2.3.8 can also be deduced in an expedited manner from known results on branching Brownian motion by using the representation (2.1.2) and the Payley-Wiener partial integration formula, see Remark 2.3.11.

2.2 Preliminaries

2.2.1 The Hammond-Sheffield random walk

We start by briefly recalling the urn model of [HS13] along the lines of [IW23].

For $\alpha \in \left(0, \frac{1}{2}\right)$ and a slowly varying function $L : \mathbb{R} \to \mathbb{R}^+$ let $\mu := \mu_{\alpha,L}$ be a probability measure on $\mathbb{N}$ having regularly varying tails

$$\mu(\{n, n+1, \dots\}) \sim n^{-\alpha} L(n), \tag{2.2.1}$$

and let R be an $\mathbb{N}$-valued random variable with distribution μ.

Let $\mathcal{G}_\mu$ be a random directed graph with vertex set $\mathbb{Z}$ and edge set $E\left(\mathcal{G}_\mu\right)$ 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

$$E\left(\mathcal{G}_\mu\right) := \left\{(i, i - R_i) : i \in \mathbb{Z}\right\}. \tag{2.2.2}$$

This induces the (random) equivalence relation

$$i \sim j :\Longleftrightarrow i \text{ and } j \text{ belong to the same connected component of } \mathcal{G}_\mu. \tag{2.2.3}$$

If $\alpha > 1/2$ there is only one connected component and this relation is trivial.

The individuals' types arise as follows: Assume that each component of $\mathcal{G}_\mu$ gets its type by an independent copy of a real-valued random variable Y with

$$\mathbf{E}[Y] = 0 \text{ and } 0 < \mathbf{E}[Y^4] < \infty. \tag{2.2.4}$$

In the situation of [HS13], Y is a centered Rademacher(p) variable, i.e.

$$Y = \xi - (2p - 1) \text{ with } \mathbf{P}(\xi = +1) = p,\ \mathbf{P}(\xi = -1) = 1 - p. \tag{2.2.5}$$

Denote by $\mathscr{C}_i$ the component which contains i (note that $\mathscr{C}_i = \mathscr{C}_j$ if $i \sim j$). For $i \in \mathbb{Z}$ the type of the component $\mathscr{C}_i$ will be denoted by Y_i. Define the "random walk" (with dependent increments)

$$S_n := \sum_{i=1}^{n} Y_i, \quad n = 0, 1, \dots. \tag{2.2.6}$$

Let us analyze the covariance structure of this random walk to motivate its relation to fractional Brownian motion: By construction,

$$\sigma_n^2 := \mathbf{Var}[S_n] = \sum_{i,j \in [n]} \mathbf{Cov}[Y_i, Y_j] = \mathbf{E}[Y^2] \sum_{i,j \in [n]} \mathbf{P}(i \sim j). \tag{2.2.7}$$

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

$$\sum_{i,j \in [n]} \mathbf{P}(i \sim j) \sim \frac{C_1}{\alpha(2\alpha+1)} \cdot \frac{n^{2\alpha+1}}{L(n)^2} \quad \text{as } n \to \infty; \tag{2.2.8}$$

with C_1 as in (2.C.2), see [IW23, (1.8)].

For $\frac{i-1}{n} \le t \le \frac{i}{n}$, $i, n \in \mathbb{N}$, let

$$S^{(n)}(t) \text{ be the linear interpolation of } S_i/\sigma_n \text{ and } S_{i+1}/\sigma_n. \tag{2.2.9}$$

Because (S_n) has stationary increments by construction, it follows from (2.2.7) and (2.2.8) that $S^{(n)}$ has, asymptotically as $n \to \infty$, the covariance structure of fractional Brownian motion with Hurst parameter $H := \frac{1}{2} + \alpha$. In order to prove that $S^{(n)}$ converges (in the sense of finite dimensional distributions) to fractional Brownian motion, it is shown in [IW23] that the finite dimensional distributions of $S^{(n)}$ are asymptotically Gaussian. This is provided by [IW23, Theorem 1.1], see also [HS13, Theorem 1.1].

The renewal function,

$$q_n := \mathbf{P}\left(\tilde{R}_1 + \ldots + \tilde{R}_j = n \text{ for some } j \ge 0\right) \tag{2.2.10}$$

with $\tilde{R}_1, \tilde{R}_2, \dots$ being independent copies of R, an $\mathbb{N}$-valued random variable with distribution μ as in (2.2.1), is the main ingredient to describe the random graph G_μ. As in [IW23] we will work under the condition

$$q_n \sim \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \frac{n^{\alpha-1}}{L(n)} \quad \text{as } n \to \infty, \tag{2.2.11}$$

see [IW23, Theorem 1.1(B)]. This is equivalent to the validity of the Strong Renewal Theorem for the renewal process with increment distribution (2.2.1), see Caravenna and Doney [CD19], whose Theorem 1.4 gives necessary and sufficient conditions in terms of μ for the validity of (2.2.11). A well-known sufficient criterion for (2.2.11) is Doney's [Don97] criterion

$$\sup_{n \ge 1} \frac{n\mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty. \tag{2.2.12}$$

We will work under the more restrictive assumption of $L \equiv 1$ and (2.2.5) with $p = \frac{1}{2}$ to simplify notation. This means that we assign a type to each component of the random graph G_μ either $+1$ or -1 with probability $\frac{1}{2}$.

Another key ingredient to analyse this model is the pair coalescence probability $\mathbf{P}(0 \sim n)$. Under the above assumption we get

$$\mathbf{P}(0 \sim n) \sim \frac{1}{\sum_{\ell \geq 0} q_\ell^2} \cdot \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3} \cdot n^{2\alpha-1} \tag{2.2.13}$$

for $n \to \infty$, see [IW23, Proposition 2.1].

2.2.2 Tree-indexed Hammond-Sheffield-urns

Now we introduce a branching version of the HS-model. Denote by $\mathcal{Y}$ the set of binary branching $\mathbb{R}$-trees with countably many branches. Let $\mathfrak{y} \in \mathcal{Y}$ (for example a realisation of Yule tree, in which every individual splits into two after an Exp(1)-distributed time), see Figure 2.4. We now recall some notation introduced in Section 2.1 and add some more needed: We denote by $\mathcal{B}$ the collection of branches of the tree, such that $\mathfrak{y} := \bigcup_{b \in \mathcal{B}} b$. Assume that at a branch $b \in \mathcal{B}$ there is a branching event at time s, then we name this new branch (b, s). For the sake of notational ease we shorten this to bs. The main branch is named 0, see Figure 2.4. For two branches $b, \tilde{b}$ we denote by $b \wedge \tilde{b}$ the time when they split, for example $0 \wedge 0rs = r$ and $0r \wedge 0rs = s$, as depicted in Figure 2.4. The point on the branch b at time t on the tree is denoted by $(b, t)_\mathfrak{y}$. If bs and b are two branches, we say that $bs \backslash s = b$. We say that the branch b is older than the branch $\tilde{b}$ if the last digit ($\in \mathbb{R}$) of $\tilde{b}$ is bigger than the last digit of b. For example, 0 is the oldest branch and b is older than bs.

Remark 2.2.1. Let us connect our notation (introduced in the paragraph above) with the classic Ulam Harris notation, which we discussed before Equation (2.1.2): The branch $0s_1$ in Figure 2.3 consists of the edges with Ulam-Harris names $f = \emptyset, g = 1, h = (h_1, h_2) = (1, 0)$ (which we write as 10). It branches of from the main branch at time $s_1 = \ell_{h_1} = \ell_1$.
Generally speaking: A branch containing the edge with Ulam-Harris name $h = (h_1, h_2, \dots,)$ splits-off at times

$$\left\{\ell_{h_1, h_2, \dots, h_k} : h_k = 1\right\}.$$

Our notation consists of the branch times at which a 1 is added to the Ulam-Harris names of an edge. A branch b consists then of the edges between branching events. See Figure 2.3 for a translation between those notations in a concrete example.

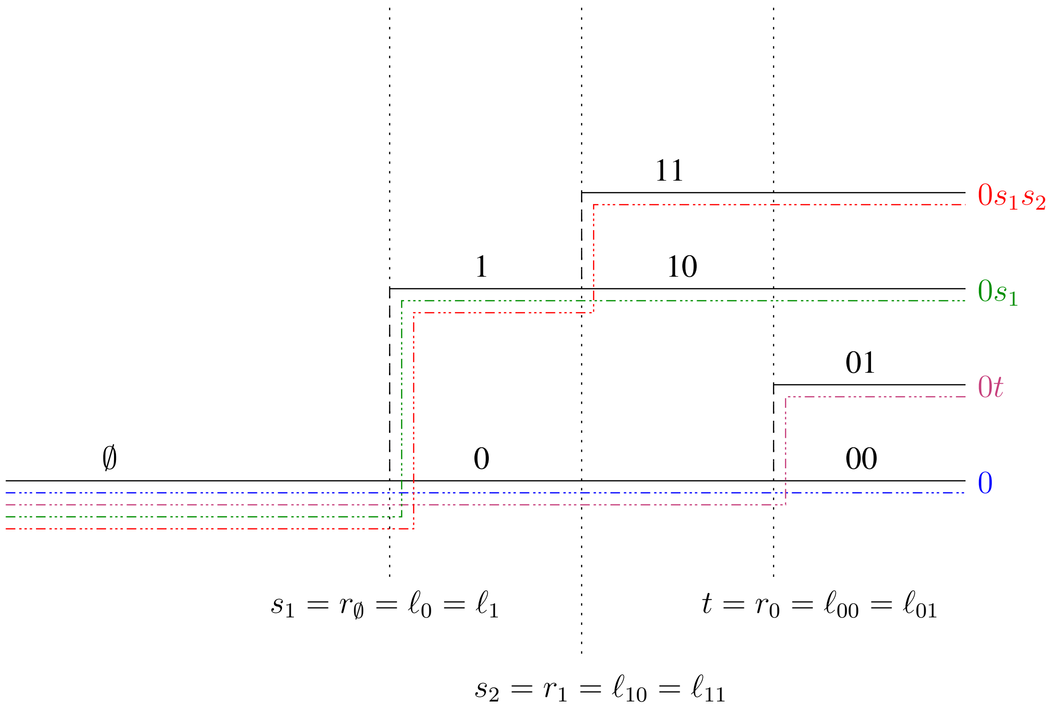


Figure 2.3: This figure contains a translation between the notations introduced in Section 2.2.2 and in the Introduction. The Ulam-Harris names of the edges are written in black. The birth and death times are s_1, s_2 and t. The names of the branches in our notation and the corresponding ancestral lines are displayed in multiple colours. See Remark 2.2.1 for a detailed explanation of the translation between the two notations.

We are now ready to construct the n-th discrete approximation of branching H-fractional Brownian motion:

1. Sample a Yule tree $\mathfrak{y}$.

2. Sample a HS-model over the discretisation of the branch 0. More specifically, sample an HS-model over $\mathbb{Z}$, and identify the integer $i \in \mathbb{Z}$ with the point in the tree $(0, i/n)_{\mathfrak{y}}$. This is done as described in Section 2.2.1, using the branch 0 instead of the real numbers.

3. We proceed recursively. Let the $R_i^{(b)}$ be independent and have distribution μ. Note that each individual $(b, i)_{(n)}$ has an ancestor at distance $R_i^{(b)}$ to the left to which it is connected regardless of the branch it lies in.
Formally, assume that the HS model has been sampled on the oldest m branches. Furthermore, assume that $b = s_1 \dots s_{m+1}$ is the $(m+1)$ oldest branch and that its last

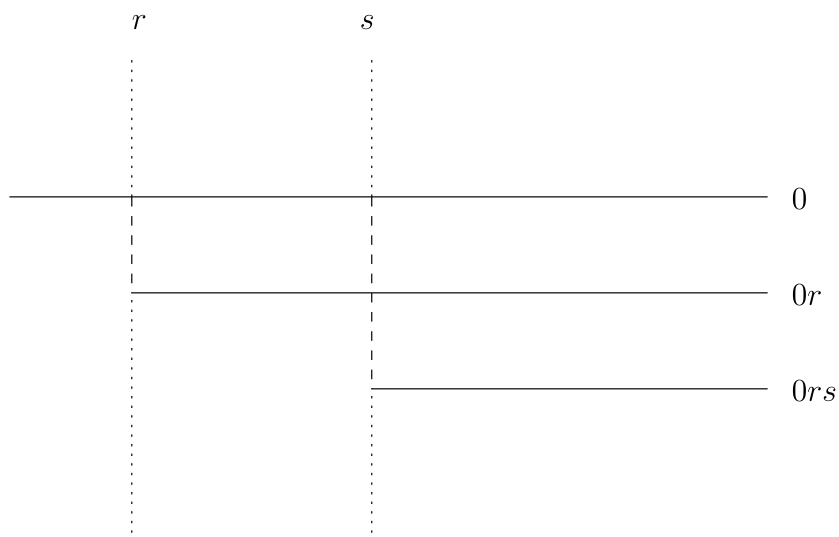


Figure 2.4: Three branches of the tree $\mathfrak{y}$ are shown. The ancestral branch is denoted by 0. The branch splitting off from the ancestral branch at time r is denoted by $0r$, and the branch splitting off from the branch $0r$ at time s is denoted by $0rs$. Time runs from left to right, and distances in $\mathfrak{y}$ are measured horizontally. Each branch b is conceived as a copy of $\mathbb{R}$, with common ancestries being glued together.

digit is $s_{m+1} > 0$. For $i \in n\mathbb{Z}, i \geq \lfloor s_{m+1}n \rfloor$ set

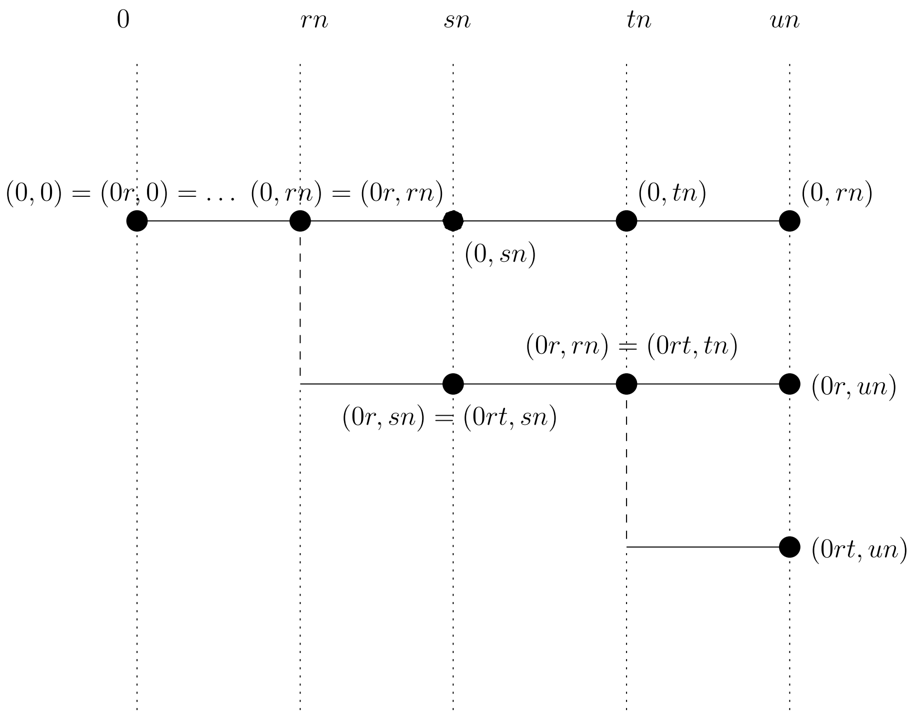
$$r := \max \Big\{ r \leq m+1 : (i - R_i^{(b)} - s_{m+1}n)/n + \lfloor n(s_m - s_r) \rfloor / n > 0 \Big\} \vee 0.$$

The point $\big(b, +i/n\big)_{\mathfrak{y}}$ is connected to $\Big(s_1 \ldots s_r, (i - R_i^{(b)})/n\Big)_{\mathfrak{y}}$. See Figure 2.5.

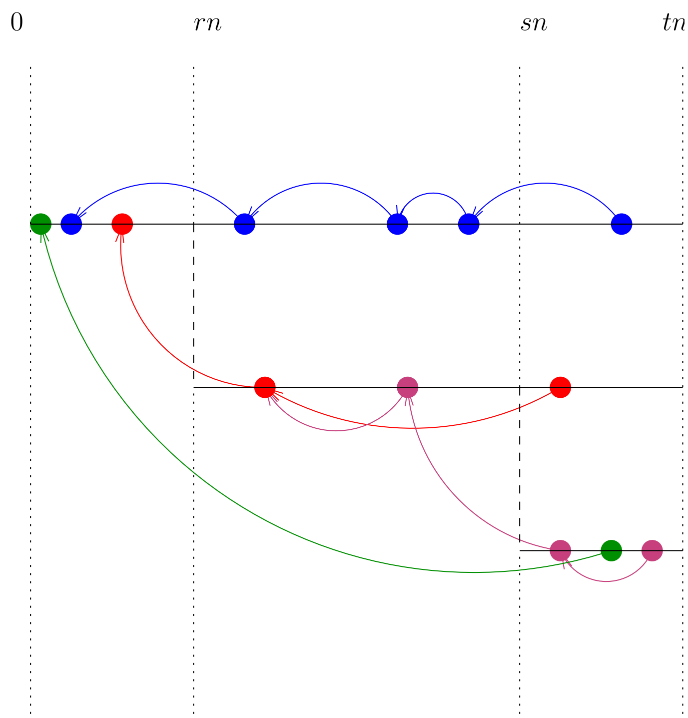
Note that the above procedure produces a random graph with almost surely infinitely many connected components, see Figure 2.5B. We assign a type ± 1 to each component independently with probability $\frac{1}{2}$. By $(bs, k)_{(n)} = (bs, k/n)_{\mathfrak{y}}$ we denote individual number $k \in \mathbb{Z}$ in the branch bs, which branched off from branch b at time s, in the n-th discrete approximation. We will mostly omit the subscripts and just write (bs, k). If $k \leq \lfloor sn \rfloor$ this is the corresponding individual in branch b, if $k > \lfloor sn \rfloor$ it is the $\big[k - \lfloor sn \rfloor\big]$-th individual after the branch point at which bs branched off b. Especially this means that $(b, 0)$ corresponds to the same individual for all branches $b \in \mathcal{B}$, see Figure 2.5A. For an individual (b, k) we denote by $Y_{(b,k)}$ the type of its component, where we often will omit the b if $b = 0$.

The random walk for the n-th discrete approximation along the main branch 0 is defined by

$$S_0^{(n)}(t) := \frac{1}{c(n)} \left[\sum_{\ell=1}^{\lfloor tn \rfloor} Y_\ell + \big[tn - \lfloor tn \rfloor\big] Y_{\lfloor tn \rfloor} + \big[\lceil tn \rceil - tn\big] Y_{\lceil tn \rceil} \right], \qquad t \geq 0 \qquad (2.2.14)$$



(A) An illustration of a branching HS-model and the multiple names an individual can have.



(B) An illustration of a branching HS-model in which we follow three ancestral lines. Two of them coalesce. Red and purple belong to the same component.

Figure 2.5

for the scaling function

$$c(n) := \left(n^{2\alpha+1} \frac{1}{\sum_{\ell \geq 0} q_\ell^2} \cdot \frac{1}{\alpha(2\alpha+1)} \cdot \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3} \right)^{\frac{1}{2}}. \tag{2.2.15}$$

Now define the random walk along a branch bs inductively by

$$S_{bs}^{(n)}(t) := \mathbf{1}_{t \leq s} S_b^{(n)}(t) + \mathbf{1}_{t>s} \left[S_b^{(n)}(s) + \frac{\sum_{\ell=\lfloor ns \rfloor+1}^{\lfloor tn \rfloor} Y_{(bs,\ell)} + \left[tn - \lfloor tn \rfloor\right] Y_{(bs,\lfloor tn \rfloor)} + \left[\lceil tn \rceil - tn\right] Y_{(bs,\lceil tn \rceil)}}{c(n)} \right]. \tag{2.2.16}$$

Observe that this means that for two branches b and $\tilde{b}$ the processes $S_b^{(n)}$ and $S_{\tilde{b}}^{(n)}$ are equal till $b \wedge \tilde{b}$ and share some common memory afterwards. See Figure 2.6 for a simulation using different values of n.

From now on always denote by $\mathbf{P}_\mathfrak{y}, \mathbf{E}_\mathfrak{y}$ the law of the $\mathfrak{y}$-indexed random walks given the tree $\mathfrak{y}$. Thus we immediately get the following:

Proposition 2.2.2. *Let $\mathfrak{y} \in \mathcal{Y}$. Then the sequence of processes $\left(S_b^{(n)}\right)_{n \geq 1}$ converges for all branches $b \in \mathcal{B}$ in distribution to fractional Brownian motion with Hurst-parameter $H = \frac{1}{2} + \alpha$ as $n \to \infty$.*

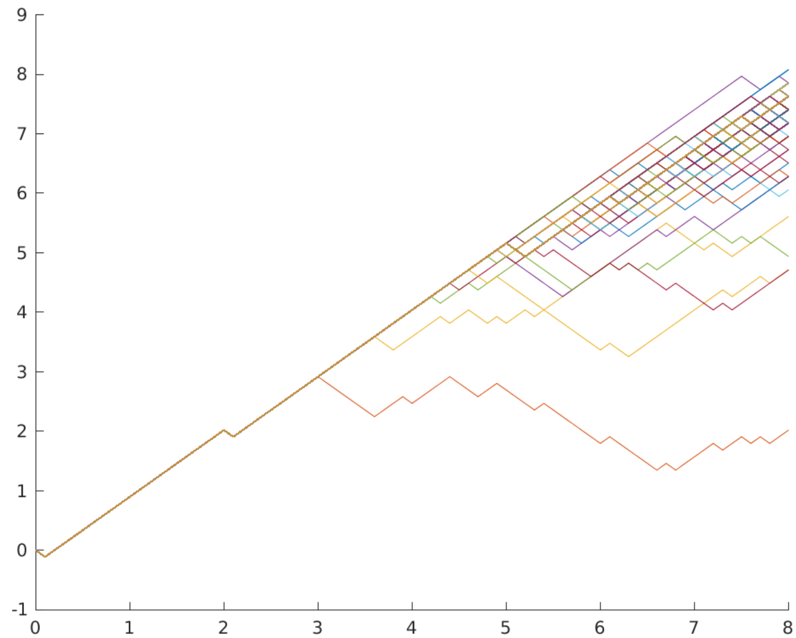
This follows directly by [IW23, Theorem 1.1] since each branch itself is a HS-model. A first step in showing the validity of our discrete construction is the following proposition.

Proposition 2.2.3. *Let $\mathfrak{y} \in \mathcal{Y}$. For not necessarily different branches $b_1, \ldots, b_m \in \mathcal{B}$ and $t_1, \ldots, t_m \in \mathbb{R}$ the vector*

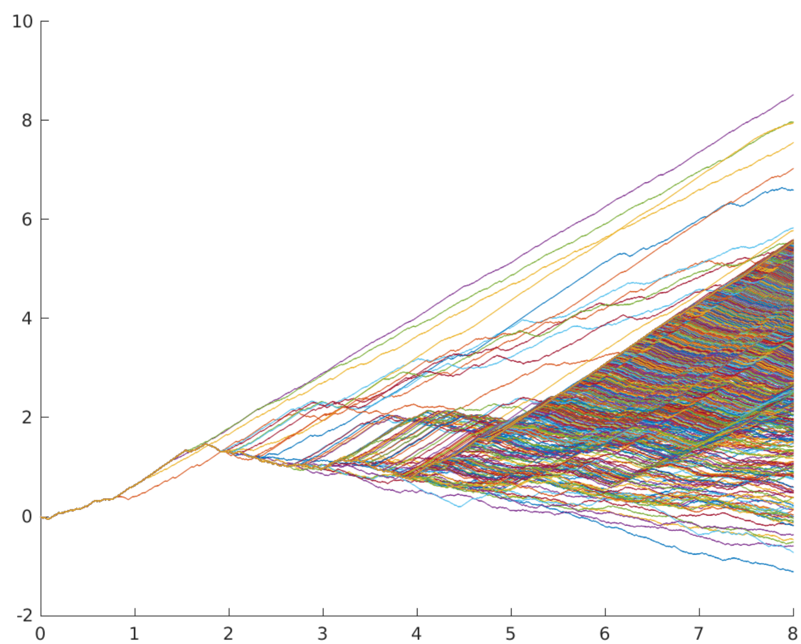
$$\left(S_{b_i}^{(n)}(t_i) \right)_{i=1,\ldots,m} \tag{2.2.17}$$

is asymptotically jointly Gaussian as $n \to \infty$.

Proposition 2.2.3 will be shown by the results of Section 2.5 and 2.6, which contain a notion of "branching property" for processes with memory.



(A) $n = 10$



(B) $n = 300$

Figure 2.6: This is a simulation of $S_{\mathfrak{y}}^{(n)} := \left(\left(S_b(t) \right)_{0 \le t \le 8} \right)_{b \in \mathcal{B}}$ for $\alpha = 0.45$. The y-axis is measured in units of $\left(\sum_{\ell \ge 0} q_\ell^2 \right)^{-\frac{1}{2}}$ for q_ℓ defined by (2.2.10) and $\alpha = H - \frac{1}{2}$.

2.3 Main results

In this section we will describe the main contributions of this paper, which include a discrete approximation of branching fractional Brownian motion and the derivation of the speed of its maximum.

2.3.1 Tree-indexed fractional Brownian motion

Remember the classical construction of fractional Brownian motion already stated in the Introduction, see (2.1.1): Given a Brownian motion W, Gaussian white noise on $\mathbb{R}$, we construct a fractional Brownian Motion with Hurst parameter H by setting

$$B(t) := \int_{\mathbb{R}} K(s,t)\,\mathrm{d}W(s) \tag{2.3.1}$$

with

$$K(s,t) := K^H(s,t) := \frac{1}{C_H}\left[\mathbf{1}_{s\le 0}\left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right) + \mathbf{1}_{t\ge s>0}(t-s)^{H-\frac{1}{2}}\right], \tag{2.3.2}$$

and

$$C_H := \left(-\frac{2^{-2H}\Gamma(-H)\Gamma\left(H+\frac{1}{2}\right)}{\sqrt{\pi}}\right)^{\frac{1}{2}}, \tag{2.3.3}$$

see [MVN68, Definition 2.1, Corollary 3.4], where this process is called *reduced fractional Brownian motion*. Note, that the normalization corresponds to the one used by Samorodnitsky and Taqqu in [ST17, Chapter 7.2] and differs from the one used in [MVN68]. In the same way we can proceed and construct on top of branching Brownian motion $\left(\left(W_b(t)\right)_{[0,T]}\right)_{b\in\mathcal{B}}$, Gaussian white noise on an $\mathbb{R}$-tree $\mathfrak{y}\in\mathcal{T}$, fractional branching Brownian motion:

Definition 2.3.1. Let $\left(\left(W_b(t)\right)_{[0,T]}\right)_{b\in\mathcal{B}}$ be Gaussian white noise on an $\mathbb{R}$-tree $\mathfrak{y}\in\mathcal{Y}$. The process $\left(\left(B_b(t)\right)_{[0,T]}\right)_{b\in\mathcal{B}}$ defined by

$$B_b(t) := \int_{\mathbb{R}} K(s,t)\,\mathrm{d}W_b(s) \tag{2.3.4}$$

is called $\mathfrak{y}$-*indexed branching fractional Brownian motion* (BFBM).

Note that the above definition corresponds to the process defined in (2.1.2) and the construction of [KLS19, Section 6]. If one now computes the covariance

$$\mathbf{Cov}_{\mathfrak{y}}\left[B_b\left(t_1\right), B_{\tilde{b}}\left(t_2\right)\right] \tag{2.3.5}$$

one obtains for branches b and $\tilde{b}$ with $b\wedge\tilde{b} = s$

$$\begin{aligned} & \rho^{\mathrm{K}}\left(t_1,t_2,s\right) \\ := \;& \mathbf{E}_{\mathfrak{y}}\left[B_b(t_1)B_{\tilde{b}}(t_2)\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{C_H^2}\mathbf{E}_{\mathfrak{y}}\Bigg[\left(\int_{-\infty}^{0}(t_1-\xi)^{H-\frac12}-(-\xi)^{H-\frac12}\,\mathrm{d}B_b(\xi)+\int_0^{t_1}(t_1-\xi)^{H-\frac12}\,\mathrm{d}B_b(\xi)\right)\\
&\quad\cdot\left(\int_{-\infty}^{0}(t_2-\xi)^{H-\frac12}-(-\xi)^{H-\frac12}\,\mathrm{d}B_{\tilde b}(\xi)+\int_0^{t_2}(t_2-\xi)^{H-\frac12}\,\mathrm{d}B_{\tilde b}(\xi)\right)\Bigg]\\
&= \frac{1}{C_H^2}\left(\int_{-\infty}^{0}\left((t_1-\xi)^{H-\frac12}-(-\xi)^{H-\frac12}\right)\left((t_2-\xi)^{H-\frac12}-(-\xi)^{H-\frac12}\right)\mathrm{d}\xi\right)\\
&\quad+\frac{1}{C_H^2}\left(\int_0^{s}(t_1-\xi)^{H-\frac12}(t_2-\xi)^{H-\frac12}\,\mathrm{d}\xi\right). \qquad (2.3.6)
\end{aligned}
$$

For $t_1=t_2\equiv t>s$ this gives

$$
\begin{aligned}
&\rho^{\mathrm{K}}(t,t,s)\\
&= \frac{1}{C_H^2}\left(\int_{-\infty}^{0}\left((t-\xi)^{H-\frac12}-(-\xi)^{H-\frac12}\right)^2\mathrm{d}\xi\right)+\frac{1}{C_H^2}\left(\int_0^{s}(t-\xi)^{2H-1}\,\mathrm{d}\xi\right)\\
&= t^{2H}-(t-s)^{2H}\frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac12\right)} =: t^{2H}-C_\rho(t-s)^{2H} \qquad (2.3.7)
\end{aligned}
$$

Observe that the prefactor one of t^{2H} in (2.3.7) is due to the normalization used in (2.3.2).

2.3.2 A discrete approximation via tree-indexed Hammond-Sheffield random walks

Let us now discuss the discrete approximation (2.2.16) to branching fractional Brownian motion:

Theorem 2.3.2. *Let $T>0$. Let $\mathfrak{y}\in\mathcal{Y}$. Then*

$$
S_{\mathfrak{y}}^{(n)} := \left(\left(S_b^{(n)}(t)\right)_{[0,T]}\right)_{b\in\mathcal{B}} \qquad (2.3.8)
$$

converges in distribution to BFBM.

Note that $S_{\mathfrak{y}}^{(n)}$ is a random function from the $\mathbb{R}$-tree $\mathfrak{y}$ into the real numbers $\mathbb{R}$. The above convergence in distribution occurs with respect to the topology induced by uniform convergence.

The proof can be found in Section 2.6.

Since again pair coalescence probabilities are a main ingredient to understand the structure of the model and its limit we state a simple proposition similar to [IW23, Proposition 2.1]:

Proposition 2.3.3. *For branches b and $\tilde{b}$ with $b \wedge \tilde{b} = s$ the probability that (b,i) and $(\tilde{b},j)$ (which for $i,j > sn$ lie in the two different branches) lie in the same component of the* branching *HS-model is given by*

$$\mathbf{P}_{\mathfrak{y}}\left((b,i) \sim (\tilde{b},j)\right) = \frac{1}{\sum_{\ell\geq 0} q_\ell^2} \sum_{r \geq (i\wedge j)-sn} q_r q_{r+|j-i|} = \frac{1}{\sum_{\ell\geq 0} q_\ell^2} \sum_{r\geq 0} q_{i-sn+r} q_{j-sn+r} \tag{2.3.9}$$

for $i,j > sn$.

The proof can be found in Appendix 2.D. This can now be used to compute the covariance structure of the discrete approximations in Theorem 2.3.2: For branches $b, \tilde{b}$ with $b \wedge \tilde{b} = s$ we have

$$\begin{aligned}\mathbf{Cov}_{\mathfrak{y}}\left[S_b^{(n)}(t_1), S_{\tilde{b}}^{(n)}(t_2)\right] &\sim \frac{1}{2}\left[t_1^{2\alpha+1} - (t_1-s)^{2\alpha+1} + t_2^{2\alpha+1} - (t_2-s)^{2\alpha+1}\right] \\ &\quad + \frac{\int_0^{t_1-s}\int_0^{t_2-s}\int_0^{\infty} (y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1}{\Gamma(1-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^{-1}(\alpha(2\alpha+1))^{-1}}. \\ &=: \rho^{\mathrm{HS}}(t_1,t_2,s) \end{aligned} \tag{2.3.10}$$

This formula and its proof, which can be found in Appendix 2.C, help to identify the sources of positive correlation. We will show that the covariance structure obtained with the discrete approximations and the kernel construction are indeed the same, see Corollary 2.5.8. This implies that the two constructions of BFBM lead to the same object.

In particular it turns out that for $H \equiv \alpha + \frac{1}{2}$

$$\rho^{\mathrm{K}}(t_1,t_2,s) = \rho^{\mathrm{HS}}(t_1,t_2,s) =: \rho(t_1,t_2,s) \tag{2.3.11}$$

for $\rho^{\mathrm{K}}(t_1,t_2,s)$ defined by (2.3.6) and $\rho^{\mathrm{HS}}(t_1,t_2,s)$ defined by (2.3.10). This equality is elusive to us without the stochastic interpretation presented here. Interesting analytic identities emerge of this and are the content of Appendix 2.E.

2.3.3 A prediction formula for fractional Brownian motion

To better understand the connection between the representations (2.2.16) and (2.3.21) let us go back to fractional Brownian motion and provide additional insight on the following result by Gripenberg and Norros:

Proposition 2.3.4 ([GN96, Theorem 3.1])**.** *For a fractional Brownian motion B with Hurst parameter $\frac{1}{2} < H < 1$ for all $t > 0$*

$$\mathbf{E}\left[B_t | \sigma(B_s, s \leq 0)\right] = \int_{-\infty}^{0} g(t,s)\,\mathrm{d}B_s \tag{2.3.12}$$

with

$$g(t,-s) := \frac{\sin\left(\pi\left(H-\frac{1}{2}\right)\right)}{\pi} t^{-H+\frac{1}{2}} \int_0^t \frac{\xi^{H-\frac{1}{2}}}{\xi+s}\,\mathrm{d}\xi. \tag{2.3.13}$$

Remark 2.3.5. Note, that the integral in (2.3.13) has no singularity at 0 since since $H > \frac{1}{2}$. For the same reason we have that for all $t \geq 0$ the integrals $\int_{-\infty}^0 |g(t,s)| \, \mathrm{d}s$ and $\int_{-\infty}^0 |g(t,s)^2| \, \mathrm{d}s$ are finite. By [GN96, p. 404] this is enough for the integral in (2.3.12) to exist.

The proof technique in [GN96] consists in using the representation formulae (2.3.21), see [GN96, Proof of Theorem 3.1]. The HS-model allows to understand the kernel g in (2.3.13) as the probability that a certain *increment from the past will get copied into the present*. See Section 2.4 for a proof of Proposition 2.3.4 which makes use of this.

2.3.4 The maximal displacement of branching fractional Brownian motion

In the following we identify the leading order of $\max_{b \in \mathcal{B}} B_b(t)$.

Theorem 2.3.6. *Let $\mathfrak{y}_{\text{bin}}$ be a deterministic binary branching tree (every branch branches into two after time 1, see Figure 2.7), and recall the definition of C_H from (2.3.3). The leading order of the maximum of a BFBM*

$$\left(\left(B_b(t) \right)_{t \geq 0} \right)_{b \in \mathcal{B}} \tag{2.3.14}$$

with Hurst parameter $H \in \left(\frac{1}{2}, 1\right)$ is

$$m(t) := t^{H+\frac{1}{2}} \sqrt{\frac{\log(2)\sqrt{\pi} 2^{2H+1} H}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)\left(H+\frac{1}{2}\right)^2}} = t^{H+\frac{1}{2}} \cdot \frac{\sqrt{2\log(2)}}{C_H} \cdot \frac{1}{\left(H+\frac{1}{2}\right)} \tag{2.3.15}$$

in the sense that

$$\mathbf{E}_{\mathfrak{y}_{\text{bin}}} \left[\frac{\max_{b \in \mathcal{B}} B_b(t)}{m(t)} \right] \to 1 \quad \textit{for } t \to \infty. \tag{2.3.16}$$

See Section 2.7 for a proof.

Remark 2.3.7. Arguments in the flavour of Arguin, Bovier and Kistler [ABK11], which were continued in work by Kistler and Schmidt [KS15], see Section 2.8, give us a similar statement for branching fractional Brownian motion on Yule trees, see the next theorem. The idea of the proof is to first show that the maximum of a BFBM can only be attained by a trajectory staying very close to the maximum all along the way. For this purpose we will show exactly this for a series of Generalized random energy models (GREMs) approximating BFBM. Finally we observe that trajectories staying close to the maximum are indeed of the desired order.
In Remark 2.3.11 we give another proof of Theorem 2.3.8 which combines results of [ABK11] with the Mandelbrot-van-Ness-representation of fractional Brownian motion in terms of Wiener integrals and the Payley-Wiener partial integration formula.

Theorem 2.3.8. *Let $\mathfrak{Y}$ be a (binary branching) Yule tree with branching rate* 1 *and let* η *denote its distribution. Then the leading order of the maximum of a BFBM*

$$\left(\left(B_b(t)\right)_{t\geq 0}\right)_{b\in\mathcal{B}} \tag{2.3.17}$$

with Hurst parameter $H \in \left(\frac{1}{2}, 1\right)$ *is*

$$m(t) := t^{H+\frac{1}{2}} \sqrt{\frac{\sqrt{\pi}2^{2H+1}H}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)\left(H+\frac{1}{2}\right)^2}} = t^{H+\frac{1}{2}} \cdot \frac{\sqrt{2}}{C_H} \cdot \frac{1}{\left(H+\frac{1}{2}\right)} \tag{2.3.18}$$

for C_H *defined by* (2.3.3) *in the sense that*

$$\int_{\mathbb{T}} \mathbf{P}_{\mathfrak{y}} \left(\left| \frac{\max_{b\in\mathcal{B}} B_b(t)}{m(t)} - 1 \right| > \varepsilon \right) \eta\left(\mathrm{d}\mathfrak{y}\right) \to 0 \quad \text{for } t\to\infty \tag{2.3.19}$$

and all $\varepsilon > 0$.

Note that this result means

$$\mathcal{P}\left(\left| \frac{\max_{b\in\mathcal{B}} B_b(t)}{m(t)} - 1 \right| > \varepsilon \right) \to 0, \tag{2.3.20}$$

where $\mathcal{P}$ averages over the Yule tree and the fBM increments along the branches.

Remark 2.3.9. Observe that the only difference between the formulas (2.3.18) and (2.3.18) for the leading order of the maximum $m(t)$ is the factor $\sqrt{\log 2}$. This comes from the fact that the Yule tree has of order e^t many leaves at time t, while the corresponding quantity for the deterministic binary branching tree is 2^t. While this explanation can be made rigorous using the ideas by Kistler and Schmidt [KS15], a proof of Theorem 2.3.8 can be given along the following lines: In the situation of the Theorem, the distribution of $\max_{b\in\mathcal{B}} B_b(t)$ can be approximated by a *generalized random energy model* (GREM, see [Der80]) in which the variance along all branches is decaying. The latter fact together with the Gaussian increments along all branches gives that the *optimal strategy* is to always follow the maximum path. This allows to compute the order of the maximal displacement directly. For details we refer to Section 2.8

Remark 2.3.10. We conjecture that the above results hold almost surely.

Remark 2.3.11. The proof of Theorem 2.3.8, which we will give in Section 2.8, is conceptual in the sense that it relates the fractional branching Brownian motion rather directly with a generalised random energy model, see Remark 2.3.7. Here we give an alternative proof, which makes use of the Mandelbrot-van-Ness-representation of fractional Brownian motion in terms of Wiener integrals and transforms these via the Payley-Wiener partial integration formula into weighted Riemann-Integrals of the paths of branching Brownian motion and controls their asymptotics using results from [ABK11]: Remember that we can represent branching fractional Brownian motion by the Wiener integral

$$B_b(t) := \int_{\mathbb{R}} \frac{1}{C_H} \left[\mathbf{1}_{s\le 0} \left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) + \mathbf{1}_{t\ge s>0} (t-s)^{H-\frac{1}{2}} \right] \mathrm{d}W_b(s). \tag{2.3.21}$$

Now note that all branches have the integral over $\mathbb{R}_{\le 0}$ in common. Writing

$$Z_t := \int_{\mathbb{R}_-} \frac{1}{C_H} \left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) \mathrm{d}W_0(s) \tag{2.3.22}$$

we thus obtain

$$B_b(t) = Z_t + \int_0^t \frac{1}{C_H} (t-s)^{H-\frac{1}{2}} \,\mathrm{d}W_b(s). \tag{2.3.23}$$

By integration by parts the latter integral is equal to

$$\int_0^t \frac{H-\frac{1}{2}}{C_H} (t-s)^{H-\frac{3}{2}} W_b(s) \,\mathrm{d}s, \tag{2.3.24}$$

and similarly

$$Z_t = \int_{\mathbb{R}_-} \frac{H-\frac{1}{2}}{C_H} \left((t-s)^{H-\frac{3}{2}} - (-s)^{H-\frac{3}{2}} \right) W_0(s) \,\mathrm{d}s. \tag{2.3.25}$$

Since $W_0(s)$ is just an ordinary Brownian motion Z_t is almost surely at most of order t^H. Let $\varepsilon > 0$, then by [ABK11, Corollary 2.6] (see also [Bov15, Theorem 3.8] and [ABBS13, Proposition 2.5]) the integral in (2.3.23) fulfills

$$\mathcal{P} \left(\left| \max_{b\in\mathcal{B}} \frac{\int_0^t \frac{H-\frac{1}{2}}{C_H} (t-s)^{H-\frac{3}{2}} W_b(s) \,\mathrm{d}s}{\int_0^t \frac{H-\frac{1}{2}}{C_H} (t-s)^{H-\frac{3}{2}} \sqrt{2} s \,\mathrm{d}s} - 1 \right| > \varepsilon \right) \to 0 \text{ as } t \to \infty. \tag{2.3.26}$$

Substituting $u = \frac{s}{t}$ gives

$$\begin{aligned}
& \frac{H-\frac{1}{2}}{C_H} \int_0^t (t-s)^{H-\frac{3}{2}} \sqrt{2} s \,\mathrm{d}s \\
=\; & t^{H-\frac{1}{2}} \cdot \sqrt{2} \cdot \frac{H-\frac{1}{2}}{C_H} \int_0^1 u(1-u)^{H-3/2} \,\mathrm{d}u \\
=\; & t^{H+\frac{1}{2}} \cdot \sqrt{2} \cdot \frac{H-\frac{1}{2}}{C_H} \int_0^1 (1-u) u^{H-3/2} \,\mathrm{d}u \\
=\; & t^{H+\frac{1}{2}} \cdot \sqrt{2} \cdot \frac{H-\frac{1}{2}}{C_H} \left[\int_0^1 u^{H-3/2} \,\mathrm{d}u - \int_0^1 u^{H-1/2} \,\mathrm{d}u \right] \\
=\; & t^{H+\frac{1}{2}} \cdot \frac{\sqrt{2}}{C_H} \cdot \frac{1}{(H+\frac{1}{2})},
\end{aligned}$$

which gives the desired result.

[ABK11, Corollary 2.6] can also be used for a weak result in the second order term since it also gives that the maximum path of branching Brownian motion at time t is of order $\sqrt{s} \wedge \sqrt{t-s}$ smaller than the maximum path at time s. So since

$$\int_0^t (t-s)^{H-\frac{3}{2}} \left(\sqrt{s} \wedge \sqrt{t-s}\right) \mathrm{d}s \tag{2.3.27}$$

is of order t^H this entails that

$$\limsup_{t\to\infty} \mathcal{P}\left(\frac{\max_{b\in\mathcal{B}} B_b(t) - t^{H+\frac{1}{2}} \cdot \frac{\sqrt{2}}{C_H} \cdot \frac{1}{(H+\frac{1}{2})}}{t^H} > -K\right) \to 1 \text{ as } K \to \infty. \tag{2.3.28}$$

The next remark suggests that this bound is presumably not sharp.

Remark 2.3.12. In [MZ16] Maillard and Zeitouni continue the study on the distribution of the maximum of branching Brownian motion with time-inhomogeneous variance, which corresponds to the study of the time-inhomogeneous F-KPP equation. (See also [FZ12a] and [FZ12b] by Fang and Zeitouni as discussed in Remark 2.3.13.) They prove that the maximum particle of a branching Brownian motion $W^{\sigma^2,T}$ on $[0,T]$ with time inhomogeneous strictly decreasing variance $\sigma^2\left(\frac{s}{T}\right)$ satisfying $\sigma \in C^2, \inf_t |\sigma'(t)| > 0, \sigma(1) > 0$, which can be written as

$$W^{\sigma^2,T}(t) = \int_0^t \sigma\left(\frac{s}{T}\right) \mathrm{d}W_s \text{ for standard Wiener noise } W \text{ and } 0 \le t \le T, \tag{2.3.29}$$

behaves like

$$T \cdot \int_0^1 \sigma(s)\,\mathrm{d}s - T^{\frac{1}{3}} \cdot 2^{-\frac{1}{3}}\alpha_1 \int_0^1 \sigma(s)^{\frac{1}{3}} \left|\sigma'(s)\right|^{\frac{2}{3}} \mathrm{d}s - \log(T)\cdot\sigma(1), \tag{2.3.30}$$

where α_1 is the largest zero of the Airy function of the first kind.

We can naively fit branching fractional Brownian motion into their setting to obtain a candidate for the order of the second order term of the maximum of fractional branching Brownian motion: Using Equation (2.3.7) we can write

$$\begin{aligned}
\rho(s,t,t) &= t^{2H}\left[1 - C_\rho\left(1-\frac{s}{t}\right)^{2H}\right] \\
&= t^{2H} C_\rho \left[1 - \left(1-\frac{s}{t}\right)^{2H}\right] + \left[1-C_\rho\right] t^{2H} \\
&= t^{2H} C_\rho \cdot \left[\int_0^{s/t} \sigma^2(x)\,\mathrm{d}x\right] + \left[1-C_\rho\right] t^{2H}
\end{aligned}$$

for

$$\sigma^2(x) = 2H(1-x)^{2H-1}. \tag{2.3.31}$$

Let now $W^{\sigma^2,T}$ be a Brownian motion on $[0,T]$ with time inhomogeneous variance $\sigma^2\left(\frac{s}{T}\right)$ and let Z be an independent normally distributed random variable. Then the above representation of ρ gives the equality in distribution

$$W^{\sigma^2,T}(T) + \sqrt{1-C_\rho}\, t^H Z \overset{(d)}{=} B^H(t) \tag{2.3.32}$$

for t fixed and $T = t^{2H} C_\rho$. This also holds for the branching systems, in the sense that for branching Brownian motion $\left(W_b^{\sigma^2,T}\right)_{b\in\mathcal{B}}$ with time inhomogeneous variance $\sigma^2\left(\frac{s}{T}\right)$ we obtain

$$\max_{b\in\mathcal{B}} W_b^{\sigma^2,T}(T) + \sqrt{1-C_\rho}\, t^H Z \overset{(d)}{=} \max_{b\in\mathcal{B}} B^H(t)_b. \tag{2.3.33}$$

This is due to the fact that two branches in [MZ16] which have split at time $xT, x \in (0,1)$ have covariance $T\int_0^x \sigma^2(u)\, \mathrm{d}u$. We conjecture that the results by Maillard and Zeitouni, see [MZ16, Theorem 1.1], give the second order term of the speed of the maximum of a BFBM even if the function σ is not fulfilling the conditions $\inf_t |\sigma'(t)| > 0$, $\sigma(1) > 0$ and $\sigma \in C^2$. By (2.3.33) This suggests that the second order term is of order $t^{\frac{2}{3}H}$. We do not believe that the third order term will be the same as in [MZ16, Theorem 1.1], where it is of logarithmic type.

In the following we give a few remarks concerning the connection between branching random walks and the GREM; for more background see the Lecture Notes by Kistler [Kis15].

Remark 2.3.13. A precursor (and a special case) of the GREM is the REM, Random Energy Model, introduced by Derrida in [Der80], continued in [Der81]. This is the Gaussian random field $(X_\sigma)_{\sigma=1,\dots,2^N}$ for independent X_σ. An overview and analysis of this can be found in lecture notes by Bolthausen and Sznitman, see [BS02].
In [Der85] Derrida then introduced the GREM, a REM with multiple levels, which introduced a hierarchical structure. For example a GREM with two levels is a correlated random field

$$(X_\sigma)_{\sigma\in\left\{1,\dots,2^{N/2}\right\}\times\left\{1,\dots,2^{N/2}\right\}}$$

with

$$X_\sigma \equiv Y_{\sigma_1}^{(1)} + Y_{\sigma_1,\sigma_2}^{(2)}$$

for independent collections of independent random variables

$$\left(Y_{\sigma_1}^{(1)}\right)_{\sigma_1\in\left\{1,\dots,2^{N/2}\right\}} \qquad \text{and} \qquad \left(Y_{\sigma_1,\sigma_2}^{(2)}\right)_{\sigma_1\in\left\{1,\dots,2^{N/2}\right\},\sigma_2\in\left\{1,\dots,2^{N/2}\right\}}.$$

The covariance of the random field then depends on the so called overlap. See [Kis15, Section 2.2] fore a more formal introduction. An analysis of the leading order of the maximum and a general overview can be found in [BK09b].
In [BK04a] Bovier and Kurkova give a detailed analysis of GREMs with finitely many hierarchies, which is then extended to continuous hierarchies by [BK04b]. The latter is used by us to prove Theorem 2.3.6. The analysis of [BK04a] and [BK04b] gives much finer results on the leading order than what one gets by just applying Slepian's Lemma [Sle62, Section 2.10], see Lemma 2.7.1.

In [BK06] Bolthausen and Kistler studied a non-hierarchical version of the GREM and showed that the GREM is in some sense able to overcome correlations in certain regimes since this behaves as a suitable constructed hierarchical GREM. In [BK09a] they studied the corresponding Gibbs measures. This analysis has been extended in recent work of Kistler and Sebastiani [KS23].

The connection of the GREM to branching random walks, especially branching Brownian motion, got first explored by Arguin, Bovier and Kistler in [ABK11]. They were able to analyse the full extremal process, including the maximal particle, the second maximal, and so on. They showed that extremal particles descend from ancestors that split either shortly after zero or just before the observed time. In [ABK13b] they continued the study and showed that the extremal process converges in law to a Poisson cluster point process. In [ABK13a] the authors analysed the empirical distribution of the maximal displacement and showed that a Gumbel distribution with a random shift occurs as a limit thus proving a conjecture of Lalley and Sellke.

After this ground breaking work variations of BBM have been studied, for example by Bovier and Hartung in [BH14] and [BH15]. Fang and Zeitouni [FZ12a] studied the lower orders of the maximum of a branching random walk with time-inhomogeneous variance on a deterministic binary branching tree, which is extended to the study of branching Brownian motion with time-inhomogenous variance in [FZ12b] by Fang and Zeitouni. In more recent work [MZ16] Maillard and Zeitouni then studied the maximum of BBM with decaying variance which is close to our setup. See Remark 2.3.12 for a more detailed explanation.

In Theorem 2.3.8 we consider an object that is closely related to a GREM, with the difference being that underlying tree in this object is not a deterministic binary tree but a Yule tree like in the BBM case.

The GREM and its refinements have also found applications in topics beyond branching random walks, see [Kis15, Section 2] for an outline of applications which include branching diffusions [Bov15]. A recent application is one to TAP equations by Kistler, Schmidt and Sebastiani [KSS23].

2.4 Proof of Proposition 2.3.4: Conditioning fractional Brownian motion on its past

We now give a conceptually new proof of Proposition 2.3.4 making use of the discrete approximation of fractional Brownian motion introduced by [HS13].

Proof. We set $\alpha = H - \frac{1}{2}$. Let $S^{(n)}$ be the discrete approximation to fractional Brownian motion defined by (2.2.9). By $\mathcal{F}_{\leq 0}$ we denote the σ-algebra containing any information about $G_\mu \cap \mathbb{Z}_{\leq 0}$ and the colours of the individuals $i \in \mathbb{Z}_{\leq 0}$. First observe that the family $\left(S^{(n)}(t)\right)_n$ for t fixed is uniformly integrable since $\left(S^{(n)}(t)\right)_n$ is bounded in L_2.

Write

$$E_n := \bigcap_{m=1}^{k} \left\{ S^{(n)}(s_m) \in \mathfrak{B}_m \right\} \tag{2.4.1}$$

for $\mathfrak{B}_1, \mathfrak{B}_2, \ldots, \mathfrak{B}_k \in \mathcal{B}(\mathbb{R}), 0 > s_1 > s_2 > \ldots > s_k$, then for $t > 0$

$$\lim_{n\to\infty} \mathbf{E}\left[S^{(n)}(t)\mathbf{1}_{E_n}\right] = \mathbf{E}\left[B(t)\mathbf{1}_{B_{s_1}\in\mathfrak{B}_1,\ldots,B_{s_k}\in\mathfrak{B}_k}\right] \tag{2.4.2}$$

by [IW23, Corollary 1.2] and uniform integrability. Now set

$$b_{n,-k} := \mathbf{P}\Big(\max\left(A_n \cap \{-\infty, \ldots, -2, -1\}\right) = -k\Big). \tag{2.4.3}$$

The event

$$\Big\{\max\left(A_n \cap \{-\infty, \ldots, -2, -1\}\right) = -k\Big\} \tag{2.4.4}$$

can be decomposed with respect to the last individual right of zero which still belongs to A_n, so

$$b_{n,-k} = \sum_{\ell=0}^{n-1} \mathbf{P}\left(n - \ell \in A_n\right)\mu\left(k + n - \ell\right) = \sum_{\ell=1}^{n} \mathbf{P}\left(\ell \in A_n\right)\mu\left(k + \ell\right). \tag{2.4.5}$$

Using

$$\mathbf{P}\left(0 \in A_n\right) \overset{n\to\infty}{\sim} \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \cdot n^{\alpha-1} \tag{2.4.6}$$

and choosing a μ satisfying

$$\frac{\mu(n)}{n^{-\alpha-1}} \sim \alpha \qquad \text{for } n \to \infty \tag{2.4.7}$$

we obtain for $n \to \infty$

$$b_{n,-k} = \sum_{\ell=1}^{n} \mathbf{P}\left(\ell \in A_n\right)\mu\left(k+\ell\right) \sim \frac{\alpha}{\Gamma(\alpha)\Gamma(1-\alpha)} \sum_{\ell=1}^{n} \left(n-\ell\right)^{\alpha-1}\left(k+\ell\right)^{-\alpha-1}. \tag{2.4.8}$$

For $k = \xi n, \xi > 0$ we further get

$$\begin{aligned} b_{n,-k} \quad &\overset{n\to\infty}{\sim} \quad \frac{\alpha}{\Gamma(\alpha)\Gamma(1-\alpha)} \sum_{\ell=1}^{n} \left(n-\ell\right)^{\alpha-1}\left(k+\ell\right)^{-\alpha-1} \\ &= \quad \frac{\alpha}{\Gamma(\alpha)\Gamma(1-\alpha)} n^{-1} \cdot \frac{1}{n} \sum_{\ell=1}^{n} \left(1 - \frac{\ell}{n}\right)^{\alpha-1} \left(\xi + \frac{\ell}{n}\right)^{-\alpha-1} \\ &\overset{n\to\infty}{\sim} \quad \frac{\alpha}{\Gamma(\alpha)\Gamma(1-\alpha)\,n} \int_0^1 \left(1-x\right)^{\alpha-1}\left(\xi+x\right)^{-\alpha-1} \mathrm{d}x. \end{aligned}$$

Applying the substitution $u = \frac{1}{(x+\xi)^\alpha}$ we get the identity

$$\int_0^1 \left(1-x\right)^{\alpha-1}\left(\xi+x\right)^{-\alpha-1} \mathrm{d}x = \frac{\xi^{-\alpha}}{\alpha+\alpha\xi}, \tag{2.4.9}$$

such that in total

$$b_{n,-k} \overset{n\to\infty}{\sim} \frac{\alpha}{\Gamma(\alpha)\Gamma(1-\alpha)\,n} \cdot \frac{\xi^{-\alpha}}{\alpha+\alpha\xi}. \tag{2.4.10}$$

Using (2.2.14) we have

$$c(n)S^{(n)}(t) = \sum_{\ell=1}^{\lfloor nt \rfloor} Y_\ell + \left[tn - \lfloor tn \rfloor\right] Y_{\lfloor tn \rfloor} + \left[\lceil tn \rceil - tn\right] Y_{\lceil tn \rceil}. \tag{2.4.11}$$

If we now define

$$\sum_{\ell=1}^{nt} a_\ell := \sum_{\ell=1}^{\lfloor nt \rfloor} a_\ell + \left[tn - \lfloor tn \rfloor\right] a_{\lfloor tn \rfloor} + \left[\lceil tn \rceil - tn\right] a_{\lceil tn \rceil} \tag{2.4.12}$$

for a sequence $(a_\ell)_\ell$ we can write

$$c(n)\,\mathbf{E}\left[S^{(n)}(t)\mathbf{1}_{E_n}\right] = \mathbf{E}\left[\sum_{\ell=1}^{nt} \mathbf{E}\left[Y_\ell | \mathcal{F}_{\leq 0}\right] \mathbf{1}_{E_n}\right]. \tag{2.4.13}$$

Observing

$$\mathbf{E}\left[Y_\ell | \mathcal{F}_{\leq 0}\right] = \sum_{s \geq 0} Y_{-s} b_{\ell, -\frac{s}{\ell}\cdot\ell} \tag{2.4.14}$$

gives us

$$c(n)\,\mathbf{E}\left[S^{(n)}(t)\mathbf{1}_{E_n}\right] = \mathbf{E}\left[\mathbf{1}_{E_n} \sum_{s \geq 0} \sum_{\ell=1}^{nt} Y_{-s} b_{\ell, -\frac{s}{\ell}\cdot\ell}\right]. \tag{2.4.15}$$

If we now plug in (2.4.10), we obtain

$$\mathbf{E}\left[S^{(n)}(t)\mathbf{1}_{E_n}\right] \overset{n\to\infty}{\sim} \frac{1}{c(n)}\,\mathbf{E}\left[\mathbf{1}_{E_n} \sum_{s \geq 0} Y_{-s} \sum_{\ell=1}^{nt} \ell^{-1} \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \cdot \frac{\left(\frac{s}{\ell}\right)^{-\alpha}}{1 + \left(\frac{s}{\ell}\right)}\right]. \tag{2.4.16}$$

Setting $\Delta S^{(n)}\left(-\frac{s}{n}\right) := S^{(n)}\left(-\frac{s}{n}\right) - S^{(n)}\left(-\frac{s-1}{n}\right) = \frac{1}{c(n)} Y_{-s}$ the right hand side can be rewritten as

$$\begin{aligned} &\frac{1}{c(n)}\,\mathbf{E}\left[\mathbf{1}_{E_n} \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \sum_{s \geq 0} Y_{-s} \sum_{\ell=1}^{nt} \ell^\alpha s^{-\alpha} (\ell + s)^{-1}\right] \\ = \;& \mathbf{E}\left[\mathbf{1}_{E_n} \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \sum_{s \geq 0} \left(\frac{s}{n}\right)^{-\alpha} \Delta S^{(n)}\left(-\frac{s}{n}\right) \frac{1}{n} \sum_{\ell=1}^{nt} \left(\frac{\ell}{n}\right)^\alpha \left(\frac{\ell+s}{n}\right)^{-1}\right]. \end{aligned}$$

Write

$$f^{(n)}\left(\frac{s}{n}, t\right) := \left(\frac{s}{n}\right)^{-\alpha} \frac{1}{n} \sum_{\ell=1}^{nt} \left(\frac{\ell}{n}\right)^\alpha \left(\frac{\ell+s}{n}\right)^{-1}, \qquad f\left(\frac{s}{n}, t\right) := \left(\frac{s}{n}\right)^{-\alpha} \int_0^t y^\alpha \left(y + \frac{s}{n}\right)^{-1} \mathrm{d}y, \tag{2.4.17}$$

then

$$\sum_{s\geq 0}\left(\frac{s}{n}\right)^{-\alpha}\Delta S^{(n)}\left(-\frac{s}{n}\right)\frac{1}{n}\sum_{\ell=1}^{nt}\left(\frac{\ell}{n}\right)^{\alpha}\left(\frac{\ell+s}{n}\right)^{-1}$$
$$= \sum_{s\geq 0}\Delta S^{(n)}\left(-\frac{s}{n}\right)\left[f\left(\frac{s}{n},t\right)+f^{(n)}\left(\frac{s}{n},t\right)-f\left(\frac{s}{n},t\right)\right].$$

Since $\left|\Delta S^{(n)}\left(-\frac{s}{n}\right)\right| \leq \text{const}\cdot n^{-\alpha-\frac{1}{2}}$ and

$$\left|f^{(n)}\left(\frac{s}{n},t\right)-f\left(\frac{s}{n},t\right)\right| \leq \text{const}\cdot\left(\frac{s}{n}\right)^{-\alpha}\cdot\frac{1}{n}\cdot\int_0^t y^{\alpha}\left(y+\frac{s}{n}\right)^{-1}\mathrm{d}y \tag{2.4.18}$$

we obtain

$$\left|\sum_{s\geq 0}\Delta S^{(n)}\left(-\frac{s}{n}\right)\left[f^{(n)}\left(\frac{s}{n},t\right)-f\left(\frac{s}{n},t\right)\right] \leq \text{const}\cdot\sum_{s\geq 0}n^{-\frac{3}{2}}s^{-\alpha}\int_0^t y^{\alpha}\left(y+\frac{s}{n}\right)^{-1}\mathrm{d}y\right|. \tag{2.4.19}$$

We now split the sum: For $\varepsilon > 0$ small enough we have

$$\text{const}\cdot\sum_{s=0}^{n^{1+\varepsilon}}n^{-\frac{3}{2}}s^{-\alpha}\int_0^t y^{\alpha}\left(y+\frac{s}{n}\right)^{-1}\mathrm{d}y \leq \text{const}\cdot n^{-\frac{3}{2}}\left(n^{1+\varepsilon}\right)^{1-\alpha}\to 0 \tag{2.4.20}$$

as $n\to\infty$, as well as

$$\text{const}\cdot\sum_{s\geq n^{1+\varepsilon}}n^{-\frac{3}{2}}s^{-\alpha}\int_0^t y^{\alpha}\left(y+\frac{s}{n}\right)^{-1}\mathrm{d}y \leq \text{const}\cdot\sum_{s\geq n^{1+\varepsilon}}n^{-\frac{3}{2}}s^{-\alpha}\cdot\frac{n}{s}\to 0 \tag{2.4.21}$$

as $n\to\infty$. So

$$\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\sum_{s\geq 0}\left(\frac{s}{n}\right)^{-\alpha}\Delta S^{(n)}\left(-\frac{s}{n}\right)\frac{1}{n}\sum_{\ell=1}^{nt}\left(\frac{\ell}{n}\right)^{\alpha}\left(\frac{\ell+s}{n}\right)^{-1} \tag{2.4.22}$$

converges in distribution to

$$\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^{\infty}s^{-\alpha}\int_0^t y^{\alpha}(y+s)^{-1}\mathrm{d}y\ \mathrm{d}B_{-s}. \tag{2.4.23}$$

This gives the desired result since by Euler's reflection formula

$$\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} = \frac{\sin\left(\pi\left(H-\frac{1}{2}\right)\right)}{\pi}. \tag{2.4.24}$$

□

2.5 $\mathfrak{y}$-indexed processes (with memory)

In this section we want to put the branching HS-model in a wider context. Using this general theory we will then be able to conclude the validity of Proposition 2.2.3 directly by Theorem 2.2.2. To this purpose we start with the following definition which generalizes the class of branching processes beyond the Markov property.

Definition 2.5.1. Let $\mathfrak{y}$ be a binary branching tree as before and $X = \big(X(t)\big)_{t\in\mathbb{R}}$ a real-valued stochastic process with distribution ν and càdlàg-paths. Given the tree $\mathfrak{y}$, assume that attached to each branch b of the random tree $\mathfrak{y}$ there is a process $\big(X_b(t)\big)_{t\in\mathbb{R}}$ such that for all branches $b, \tilde{b} \in \mathcal{B}$ with $b \wedge \tilde{b} = s$

$$X_b(\xi) = X_{\tilde{b}}(\xi) \qquad \forall \xi \le s \tag{2.5.1}$$

and for all $b \in \mathcal{B}$

$$X_b \overset{(d)}{=} X. \tag{2.5.2}$$

The process

$$X_{\mathfrak{y}} := \Big(\big(X_b(t)\big)_{t\in\mathbb{R}}\Big)_{b\in\mathcal{B}} \tag{2.5.3}$$

will be called a $\mathfrak{y}$-*indexed process (with memory) with basic law* ν.

Such processes can be constructed inductively for all basic laws as long as the tree $\mathfrak{y}$ has not uncountably many branches.

Definition 2.5.2. $X_{\mathfrak{y}}$ has the *pairwise conditional independence property* (PCI) iff for all branches $b, \tilde{b} \in \mathcal{B}$

$$\mathcal{L}\left(\big(X_b(t)\big)_{t\ge b\wedge\tilde{b}} \Big| \big(X_{\tilde{b}}(t)\big)_{t\in\mathbb{R}} \right) = \mathcal{L}\left(\big(X_b(t)\big)_{t\ge b\wedge\tilde{b}} \Big| \big(X_{\tilde{b}}(t)\big)_{t\le b\wedge\tilde{b}} \right) \tag{PCI}$$

holds.

Note that (PCI) is equivalent to $\big(X_b(t)\big)_{t\ge b\wedge\tilde{b}}$ being conditionally independent of $\big(X_{\tilde{b}}(t)\big)_{t\ge b\wedge\tilde{b}}$ given $\big(X_b(t)\big)_{t\le b\wedge\tilde{b}}$ for all pairs of branches $b, \tilde{b}$.
See also [KLS19, Section 2] for a similar approach.
Assume that the process $(X_t)_{t\in\mathbb{R}}$ has second moments. Now indeed for two branches the covariance $\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t_1), X_{\tilde{b}}(t_2)\right]$ of a $\mathfrak{y}$-*indexed processes with basic law* $\mathcal{L}\big((X_t)_{t\in\mathbb{R}}\big)$ having the *pairwise conditional independence property* (PCI) is uniquely determined:

Proposition 2.5.3. *Let*

$$\Big(\big(X_b(t)\big)_{t\in\mathbb{R}}\Big)_{b\in\mathcal{B}} \tag{2.5.4}$$

be a $\mathfrak{y}$-indexed process with basic law $\mathcal{L}\big((X_t)_{t\in\mathbb{R}}\big)$ *and let* (PCI) *be fulfilled. Assume that* $(X_t)_{t\in\mathbb{R}}$ *has finite second moments. For branches* $b, \tilde{b} \in \mathfrak{y}$ *with* $b \wedge \tilde{b} = s$ *and* $t_1, t_2 \ge 0$

$$\mathcal{L}\Big(\big(X_b(t_1), X_{\tilde{b}}(t_2)\big)\Big) \tag{2.5.5}$$

is uniquely determined by the distribution of the process X *along one branch, i.e. by* $\mathcal{L}\left(\big(X_b(t)\big)_{t\in\mathbb{R}}\right)$ *for any branch* $b \in \mathcal{B}$.

Corollary 2.5.4. *Let the basic law $\mathcal{L}\left((X_t)_{t\in\mathbb{R}}\right)$ have finite variance for all t. Then in the setting of the above proposition*

$$\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t_1), X_{\tilde{b}}(t_2)\right] \tag{2.5.6}$$

is uniquely determined by the distribution of the process X along one branch, i.e. by $\mathcal{L}\left(\left(X_b(t)\right)_{t\in\mathbb{R}}\right)$ for any branch $b \in \mathcal{B}$.

Keeping in mind Definition 2.5.1 for branches $b, b_1, \dots, b_m \in \mathfrak{y}$ and

$$\mathcal{S} := \bigcup_{i=1}^{m} b_i \tag{2.5.7}$$

we introduce the notations

$$X_b := \left(X_b(t)\right)_{t\in\mathbb{R}}, \qquad X_{\mathcal{S}} := \left(\left(X_{b_i}(t)\right)_{t\in\mathbb{R}}\right)_{b_i\in\mathcal{S}}, \qquad X_{b\setminus\mathcal{S}} := \left(X_b(t)\right)_{t\geq \max\limits_{i=1,\dots,m} b\wedge b_i}, \tag{2.5.8}$$

as well as

$$X_{\mathcal{S}\setminus b} := \left(\left(X_{b_i}(t)\right)_{t\geq b\wedge b_i}\right)_{b_i\in\mathcal{S}}, \qquad X_{\mathcal{S}\cap b} := \left(\left(X_{b_i}(t)\right)_{t\leq b\wedge b_i}\right)_{b_i\in\mathcal{S}}. \tag{2.5.9}$$

Having defined this we can specify a property stronger than (PCI):

Definition 2.5.5. Let X be a $\mathfrak{y}$*-indexed process with basic law* $\mathcal{L}\left((X_t)_{t\in\mathbb{R}}\right)$. We say that X has the *conditional independence property* (CI) iff

$$\mathcal{L}\left(X_{b\setminus\mathcal{S}}\middle|\, X_{\mathcal{S}}\right) = \mathcal{L}\left(X_{b\setminus\mathcal{S}}\middle|\, X_{\mathcal{S}\cap b}\right) \tag{CI}$$

for all pairwise distinct branches $b, b_1, \dots, b_m \in \mathfrak{y}$ with

$$\mathcal{S} := \bigcup_{i=1}^{m} b_i. \tag{2.5.10}$$

Let us now give a definition of Gaussian $\mathfrak{y}$-indexed processes:

Definition 2.5.6. Let ν be the law of a Gaussian process $\left(X(t)\right)_{t\in\mathbb{R}}$. We call the corresponding $\mathfrak{y}$-indexed process $X_{\mathfrak{y}}$ a *Gaussian* $\mathfrak{y}$*-indexed process* if and only if for all branches $b_1, \dots, b_m \in \mathfrak{y}$ and $t_1, \dots, t_m$

$$\left(X_{b_1}(t_1), \dots, X_{b_m}(t_m)\right) \tag{2.5.11}$$

is multivariate Gaussian distributed.

The following Lemma is proved in Appendix 2.B.

Lemma 2.5.7. *Every* ***Gaussian*** $\mathfrak{y}$-indexed process *fulfilling* (PCI) *has the* conditional independence property (CI).

Proposition 2.5.3 implies another nice property of Gaussian $\mathfrak{y}$*-indexed processes* having the *pairwise conditional independence property* (PCI) namely that their distribution is uniquely determined by the distribution along one branch:

Corollary 2.5.8. *The distribution of a **Gaussian*** $\mathfrak{y}$-indexed process *having the* pairwise conditional independence property (PCI) *is uniquely determined by its basic law.*

Remark 2.5.9. Let $\mathfrak{y}$ be as before and let B be a fractional Brownian motion with Hurst parameter H. By Corollary 2.5.8 $\mathfrak{y}$*-indexed fractional Brownian motion* is the (unique in distribution) *Gaussian* $\mathfrak{y}$*-indexed process* with basic law $\mathcal{L}(B)$ having the *pairwise conditional independence property* (PCI).

Furthermore, (CI) implies the following two corollaries for Gaussian processes:

Corollary 2.5.10. *Let* $X_{\mathfrak{y}}$ *be a* $\mathfrak{y}$-indexed processes. *Assume that the basic law of* $X_{\mathfrak{y}}$ *is Gaussian and that* $X_{\mathfrak{y}}$ *has the property* (CI). *Then* $X_{\mathfrak{y}}$ *is a* Gaussian $\mathfrak{y}$-indexed process .

More generally:

Corollary 2.5.11. *Let* $X_{\mathfrak{y}}^{(n)}$ *be a sequence of* $\mathfrak{y}$-indexed processes. *Assume that the basic law of* $X_{\mathfrak{y}}^{(n)}$ *is asymptotically Gaussian for* $n \to \infty$ *and that the* conditional independence property (CI) *holds for each* $n \in \mathbb{N}$. *Then* $X_{\mathfrak{y}}^{(n)}$ *is asymptotically a* Gaussian $\mathfrak{y}$-indexed process .

The following can be seen as a standard example for the fact, that bivariate Gaussianity is not sufficient for multivariate Gaussianity.

Example 2.5.12. We give an example why in Corollary 2.5.10 the property (CI) instead of (PCI) is needed. This example also shows why we need the multivariate Gaussianity in Lemma 2.5.7. Let $\mathfrak{y}$ have the branches $0, 0r, 0rs$ and let Z_i be independent standard normal random variables. Set

$$\begin{aligned} X_0(t) &:= \mathbf{1}_{t\leq r} Z_0 + \mathbf{1}_{r\leq t\leq s} Z_1 + \mathbf{1}_{s\leq t} Z_2, \\ X_{0r}(t) &:= \mathbf{1}_{t\leq r} Z_0 + \mathbf{1}_{r\leq t\leq s} Z_3 + \mathbf{1}_{s\leq t} Z_4, \\ X_{0rs}(t) &:= \mathbf{1}_{t\leq r} Z_0 + \mathbf{1}_{r\leq t\leq s} Z_3 + \mathbf{1}_{s\leq t} \mathrm{sign}\left(Z_2 Z_4\right) |Z_5|. \end{aligned}$$

Observe that $\left(X_0(t), X_{0r}(t), X_{0rs}(t)\right), t \geq s$ is not jointly Gaussian, but the (PCI) holds while (CI) does not hold since the sign of $X_{0rs}(t), t \geq s$ is determined by $X_{0r}(t), X_0(t)$. So in general (PCI) does not imply (CI) and the joint Gaussianity of $X_{\mathfrak{y}}$ is not implied by X being a Gaussian process.

The following example is particularly relevant in the context of this paper.

Example 2.5.13. One easily verifies that the process

$$S_{\mathfrak{y}}^{(n)} := \left(\left(S_b^{(n)}(t)\right)_{t\in\mathbb{R}}\right)_{b\in\mathcal{B}} \tag{2.5.12}$$

defined through (2.2.16) is a $\mathfrak{y}$*-indexed process with basic law*

$$\mathcal{L}\left(\left(S_0^{(n)}(t)\right)_{t\in\mathbb{R}}\right), \tag{2.5.13}$$

see (2.2.14) for the definition of $S_0^{(n)}$. Note that $S_{\mathfrak{y}}^{(n)}$ has the *conditional independence property* (CI). By construction the HS-random walk $S_{\mathfrak{y}}^{(n)}$ has the property (CI). So we obtain that the branching HS-random walk is also asymptotically jointly Gaussian.

2.6 Proof of Theorem 2.3.2 and Proposition 2.2.3

We will now make use of the general theory in the above section to prove Theorem 2.3.2 and Proposition 2.2.3.

Proof of Proposition 2.2.3. Since $\left(\left(S_0^{(n)}(t)\right)_t\right)_{n\in\mathbb{N}}$ is a sequence of Gaussian processes and $S_\mathfrak{y}$ obeys the *conditional independence property* (CI) we can apply Corollary 2.5.11 and get the assertion of Proposition 2.2.3. □

Now Theorem 2.3.2 is a direct consequence of Proposition 2.2.3:

Proof of Theorem 2.3.2. Combining Corollary 2.5.8 and Proposition 2.2.3 gives the desired result since the property (CI) holds. □

2.7 Proof of Theorem 2.3.6

We start with a slightly weaker result than Theorem 2.3.6: Let η denote the distribution of a (binary branching) Yule tree $\mathfrak{y}$ (or even a deterministic binary branching tree like in Theorem 2.3.6), then

$$\liminf_{t\to\infty} \int_{\mathbb{T}} \mathbf{P}_\mathfrak{y}\left(\frac{\max_{b\in\mathcal{B}} B_b(t)}{t^{H+\frac{1}{2}}} > a\right)\eta\left(\mathrm{d}\mathfrak{y}\right) > 0 \quad \text{for } a\in\mathbb{R}_+ \tag{2.7.1}$$

and for all $f(t) \gg t^{H+\frac{1}{2}}$

$$\limsup_{t\to\infty} \int_{\mathbb{T}} \mathbf{P}_\mathfrak{y}\left(\frac{\max_{b\in\mathcal{B}} B_b(t)}{f(t)} > a\right)\eta\left(\mathrm{d}\mathfrak{y}\right) = 0 \quad \text{for } a\in\mathbb{R}_+ \tag{2.7.2}$$

can be obtained by the following well known lemma. Details can be found in Appendix 2.A.

Lemma 2.7.1 (Slepian inequality, [Sle62, Section 2.10]). *Let $\sigma^2 > 0$ and G be a Gaussian vector in $\mathbb{R}^\ell$ with covariance matrix*

$$\Sigma_G(i,j)\begin{cases} = \sigma^2 > 0, & i = j \\ \geq 0, & i \neq j, \end{cases} \tag{2.7.3}$$

and G' a Gaussian vector in $\mathbb{R}^\ell$ with covariance matrix

$$\Sigma_{G'}(i,j)\begin{cases} = \sigma^2 > 0, & i = j \\ \geq \Sigma_G(i,j), & i \neq j. \end{cases} \tag{2.7.4}$$

then for all $a = (a_1, \dots a_\ell) \in \mathbb{R}^\ell$

$$\mathbf{P}\left(G_i < a_i \quad \forall i\right) \leq \mathbf{P}\left(G'_i < a_i \quad \forall i\right). \tag{2.7.5}$$

This gives also for $a \in \mathbb{R}$

$$\mathbf{P}\left(\max_i G_i > a\right) \geq \mathbf{P}\left(\max_i G'_i > a\right). \tag{2.7.6}$$

Essentially the above lemma states that increasing already positive correlations decreases the maximum of a Gaussian vector.

We now prove Theorem 2.3.6 by using results of [BK04b, Theorem 3.1] and connecting the maximum of branching fractional Brownian motion to a *generalized random energy model* (GREM). In order to connect our setting with [BK04b, Theorem 3.1] we briefly describe our problem in their notation.

Remark 2.7.2. Bovier and Kurkova [BK04b] consider a Gaussian process X_σ on the hypercube $\{-1,1\}^N$ whose covariance is given by

$$\mathbf{E}\left[X_\sigma X_{\sigma'}\right] = A\left(d_N(\sigma,\sigma')\right) \text{ for } d_N(\sigma,\sigma') := \frac{1}{N}\left(\min\left\{i : \sigma_i \neq \sigma'_i\right\} - 1\right) \tag{2.7.7}$$

and some probability density function A on $[0,1]$. In our setting the X_σ will correspond to the $B_b(t)$, and the covariance will be given by $\rho\left(t,t,b\wedge\tilde{b}\right)$ for two branches b and $\tilde{b}$, see (2.3.6), (2.3.10) and (2.3.11) for the definition of ρ.

Proof of Theorem 2.3.6. To prove Theorem 2.3.6 we will apply [BK04b, Theorem 3.1]. We think of a tree $\mathfrak{y}_{\text{bin}}$ like the one in figure 2.7, a tree in which every branch branches into two after time 1. If one now assigns to each part of each branch random variables $\Delta Z_{(b),r}$ like it is done in figure 2.8 for $t=3$, such that

$$\Delta Z_{b,r} \sim \mathcal{N}\left(0, \rho\left(t,t,r\right) - \rho\left(t,t,r-1\right)\right), \qquad r > 0 \tag{2.7.8}$$

and

$$\Delta Z_{b,0} \sim \mathcal{N}\left(0, \rho\left(t,t,0\right)\right), \tag{2.7.9}$$

we get that if we define $Z_{(b)}$ as the sum of the random variables along branch b, i.e. for $t=3$

$$\begin{aligned} Z_{013} &= \Delta Z_{013,0} + \Delta Z_{013,1} + \Delta Z_{013,2} + \Delta Z_{013,3} \\ &= \Delta Z_{0,0} + \Delta Z_{01,1} + \Delta Z_{01,2} + \Delta Z_{013,3}, \end{aligned}$$

the following equality in distribution:

$$\left(B_b(t)\right)_{b\in\mathcal{B}} \overset{(d)}{=} \left(Z_b\right)_{b\in\mathcal{B}}. \tag{2.7.10}$$

Note that $\rho(t,t,s)$ is increasing as a function of s, see end of Section 2.C. Now we are in the world of a classical *generalized random energy model* (GREM), see e.g. [BK04b]. So in essence we represented the values of the fractional Brownian motions B_b at time t via a GREM. Since this is now precisely the setting of [BK04b, Theorem 3.1], we get that for a setting like this with

$$\mathbf{Cov}_{\mathfrak{y}_{\text{bin}}}\left[Z_b, Z_{\tilde{b}}\right] = t^{2H}\left[1 - \frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}\left(1-\frac{s}{t}\right)^{2H}\right], \tag{2.7.11}$$

see (2.3.7), for two branches b and $\tilde{b}$ with $b\wedge\tilde{b} = s$ we get that the leading order of the maximum for $t\to\infty$ is

$$t^{H+\frac{1}{2}}\sqrt{2}\sqrt{\log(2)}\int_0^1 \sqrt{\bar{a}(x)}\,\mathrm{d}x, \tag{2.7.12}$$

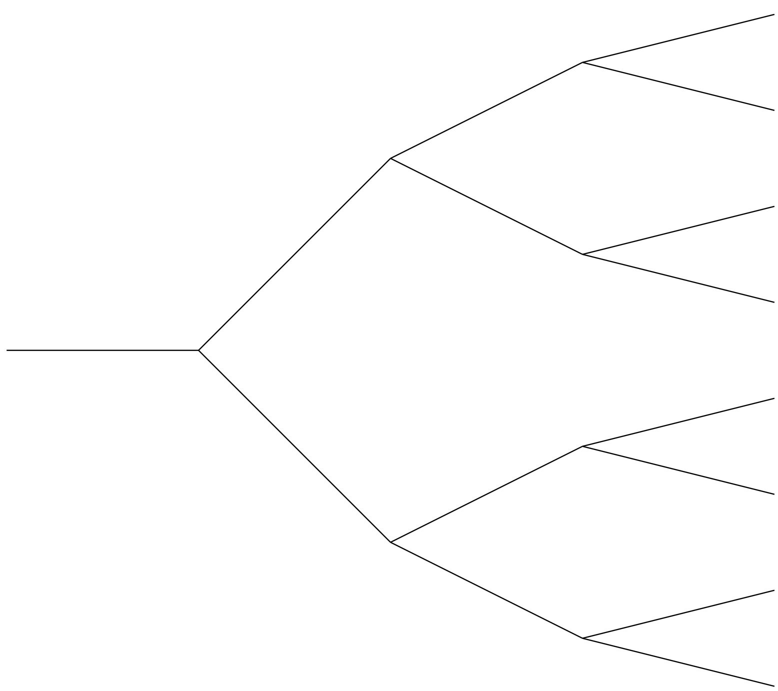


Figure 2.7: Binary Branching GREM

where

$$\bar{a}(x) := \frac{\mathrm{d}}{\mathrm{d}x}\mathbf{Cov}_{\mathfrak{y}_{\mathrm{bin}}}\left[Z_{(b)}, Z_{(bxt)}\right]. \tag{2.7.13}$$

So

$$\bar{a}(x) = \frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}2H\left(1-x\right)^{2H-1}, \tag{2.7.14}$$

which gives

$$\begin{aligned}
& \sqrt{2\log 2}\int_0^1 \sqrt{\bar{a}(x)}\,\mathrm{d}x \\
= \quad & \sqrt{2\log 2\frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}2H}\int_0^1 (1-x)^{H-\frac{1}{2}}\,\mathrm{d}x \\
= \quad & \sqrt{2\log 2\frac{\sqrt{\pi}2^{2H}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}H}\left[-\frac{1}{H+\frac{1}{2}}(1-x)^{H+\frac{1}{2}}\right]_{x=0}^{1} \\
= \quad & \frac{\sqrt{2\log 2\frac{\sqrt{\pi}2^{2H}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}H}}{H+\frac{1}{2}}.
\end{aligned}$$

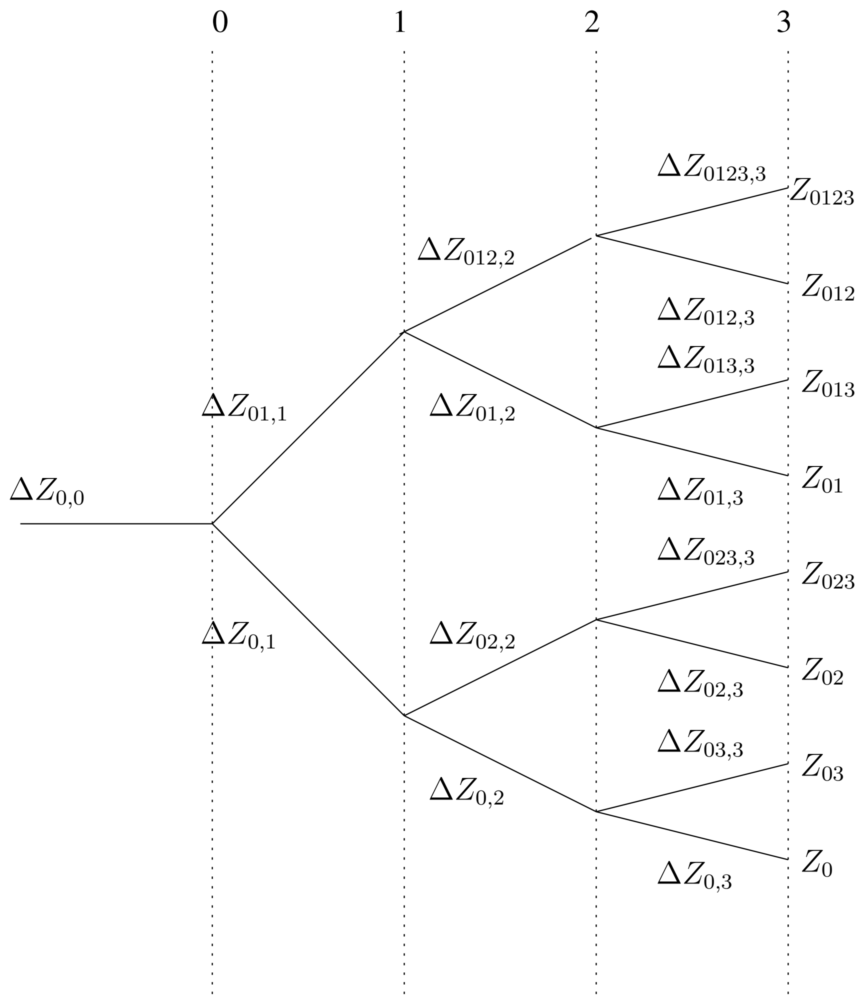


Figure 2.8: Binary Branching GREM

Since $\bar{a}(x)' < 0$ it follows by [BK04b, Theorem 3.1] that as stated

$$\mathbf{E}_{\mathfrak{y}_{\text{bin}}}\left(\frac{\max_{b\in\mathcal{B}} B_b(t)}{m(t)}\right) \to 1 \quad \text{for } t \to \infty \tag{2.7.15}$$

for $m(t)$ defined by (2.3.15). □

2.8 Proof of Theorem 2.3.8

Proof. We will approximate $\left(B_b(t)\right)_{b\in\mathcal{B}}$ by a sequence of GREMs, cf. Section 2.7. For this purpose we discretize time: Let $K \in \mathbb{N}$. For all $i = 1, \ldots, K-1$ we shift all branching events in $\left[\frac{i}{K}t, \frac{i+1}{K}t\right]$ of the Yule tree $\mathfrak{y}$ to $\frac{i}{K}t$, such that we can only have branching events at $0, \frac{t}{K}, \frac{2t}{K}, \ldots, \frac{(K-1)t}{K}$. This approximation to the Yule Tree $\mathfrak{y}$ will be denoted by $\mathfrak{y}_{K,t}$, and its collection of branches by $\mathcal{B}_{K,t}$. In the GREM we set

$$Z_b^{(i)} := \sum_{\ell=0}^{i} \Delta Z_b^{(\ell)}, \tag{2.8.1}$$

where the $\Delta Z_b^{(\ell)}$ are the increments along the branch b, see Figure 2.9. We set $Z_b \equiv Z_b^{(K)}$. For two branches b and $\tilde{b}$ with $b \wedge \tilde{b} = s$ by (2.3.6) we have

$$\mathbf{Cov}_{\mathfrak{y}}\left[B_b(t), B_{\tilde{b}}(t)\right] = t^{2H}\left[1 - C_\rho\left(1 - \frac{s}{t}\right)^{2H}\right] = \rho(t,t,s), \tag{2.8.2}$$

where

$$C_\rho = \frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}. \tag{2.8.3}$$

Thus, in order to represent the $\mathfrak{y}_{K,t}$-indexed fractional Brownian motion as a GREM, we choose, for each $b \in \mathcal{B}_{K,t}$, the distribution of the increment $\Delta Z_b^{(i)}$ as

$$\mathcal{N}\left(0, \rho\left(t,t,\frac{i}{K}t\right) - \rho\left(t,t,\frac{i-1}{K}t\right)\right) \qquad \text{if } i \geq 0 \tag{2.8.4}$$

and

$$\Delta Z_b^{(0)} \equiv Z_b^{(0)} \sim \mathcal{N}\left(0, \rho\left(0,t,t\right)\right) = \mathcal{N}\left(0, \left(1 - C_\rho\right)t^{2H}\right). \tag{2.8.5}$$

Remember that in a GREM $\Delta Z_b^{(i)}$ and $\Delta Z_{\tilde{b}}^{(i)}$ coincide if the branches b and $\tilde{b}$ did not separate till $\frac{iT}{K}$, meaning $b \wedge \tilde{b} > \frac{iT}{K}$. Since we shifted all branching events in the time intervals of length $\frac{1}{K}$ to the left by Slepian's Lemma, see Lemma 2.7.1, the order the maximum of the approximating GREM gives an upper bound for the order of the maximum of BFBM.

Now denote the rescaled standard variations of $\Delta Z_0^{(i)}, i \geq 1$ as

$$\begin{aligned} \Delta f_i &:= \sqrt{2\frac{t}{K}}\sqrt{\rho\left(t,t,\frac{i}{K}t\right) - \rho\left(t,t,\frac{i-1}{K}t\right)} \\ &= \sqrt{2\frac{t}{K}}t^H\left[-C_\rho\left(1-\frac{i}{K}\right)^{2H} + C_\rho\left(1-\frac{i-1}{K}\right)^{2H}\right]^{\frac{1}{2}} \end{aligned} \tag{2.8.6}$$

and the rescaled standard variation of $\Delta Z_0^{(0)}$

$$\Delta f_0 := \sqrt{\rho\left(t,t,0\right)}. \tag{2.8.7}$$

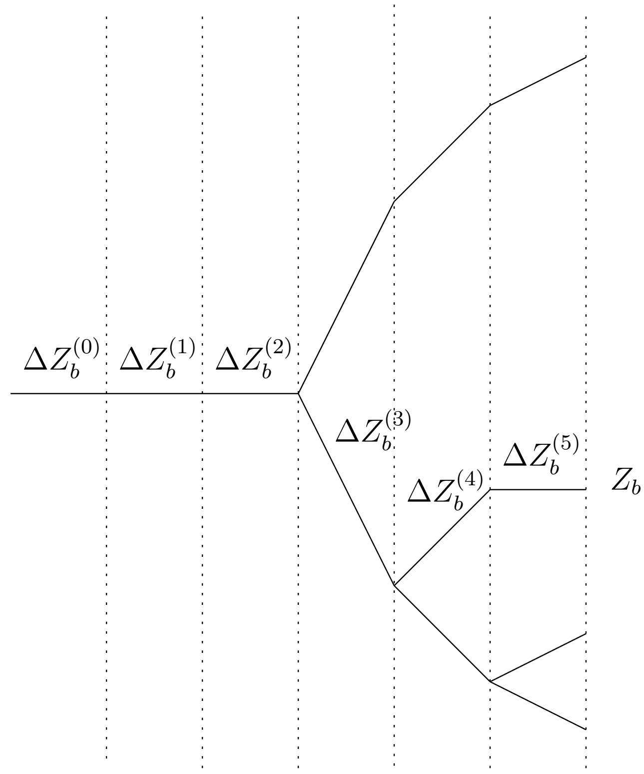


Figure 2.9: A GREM on a discretized Yule tree.

We now start to obtain an upper bound on the order of the maximum. Since Z_b is equal in distribution to the main branch Z_0 we have almost surely

$$\mathbf{P}_{\mathfrak{y}}\left(\exists b \in \mathcal{B}_{K,t} : Z_b > (1+\varepsilon)\sum_{i=0}^{K} \Delta f_i\right) \leq |\mathcal{B}_{K,t}|\, \mathbf{P}_{\mathfrak{y}}\left(Z_0 > (1+\varepsilon)\sum_{i=0}^{K} \Delta f_i\right)$$

$$\leq \quad |\mathcal{B}_{K,t}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \in \left[(\Delta f_i - \ell_i\varepsilon) \pm \varepsilon\right] \quad \forall i\right).$$

Now by independence we get

$$|\mathcal{B}_{K,t}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \in \left[(\Delta f_i - \ell_i\varepsilon) \pm \varepsilon\right] \quad \forall i\right)$$

$$= \quad |\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \prod_{i=1}^{K} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \in \left[(\Delta f_i - \ell_i\varepsilon) \pm \varepsilon\right]\right).$$

By (2.8.4) we obtain

$$|\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \prod_{i=1}^{K} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \in \left[(\Delta f_i - \ell_i\varepsilon) \pm \varepsilon\right]\right).$$

$$\begin{aligned}
&\leq \quad \text{const} \cdot |\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \prod_{i=1}^K \exp\left(-\frac{(\Delta f_i - \ell_i\varepsilon)^2}{2\left(\rho\left(\frac{i}{K}t,t,t\right) - \rho\left(\frac{i-1}{K}t,t,t\right)\right)}\right) \\
&= \quad \text{const} \cdot |\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \prod_{i=1}^K \exp\left(-\frac{t}{K} \cdot \frac{(\Delta f_i - \ell_i\varepsilon)^2}{\Delta f_i^2}\right). \qquad (2.8.8)
\end{aligned}$$

We now examine the product in (2.8.8) in more detail:

$$\prod_{i=1}^K \exp\left(-\frac{t}{K} \cdot \frac{(\Delta f_i - \ell_i\varepsilon)^2}{\Delta f_i^2}\right) = \exp\left(-\frac{t}{K}\sum_{i=1}^K \frac{(\Delta f_i - \ell_i\varepsilon)^2}{\Delta f_i^2}\right).$$

Writing

$$g_i := \sqrt{2C_\rho}\left[-\left(1-\frac{i}{K}\right)^{2H} + \left(1-\frac{i-1}{K}\right)^{2H}\right]^{\frac{1}{2}} \qquad (2.8.9)$$

we can write the exponent as

$$\begin{aligned}
& -\frac{t}{K}\sum_{i=1}^K\left[1+\frac{\ell_i^2\varepsilon^2}{\Delta f_i^2} - \frac{2\ell_i\varepsilon}{\Delta f_i}\right] \\
&= \quad -t - \frac{t\varepsilon^2}{K}\sum_{i=1}^K \frac{\ell_i^2}{\Delta f_i^2} + \frac{2t\varepsilon}{K}\sum_{i=1}^K \frac{\ell_i}{\Delta f_i} \\
&= \quad -t - \frac{t}{K}\varepsilon^2\sum_{i=1}^K \ell_i^2 \frac{K}{g_i^2 t^{2H+1}} + \frac{2t\varepsilon}{K}\sum_{i=1}^K \ell_i \frac{\sqrt{K}}{t^{H+\frac{1}{2}} g_i} \\
&= \quad -t - \frac{\varepsilon^2}{t^{2H}}\sum_{i=1}^K \frac{\ell_i^2}{g_i^2} + \frac{t^{\frac{1}{2}-H}}{K^{\frac{1}{2}}}\varepsilon\sum_{i=1}^K \frac{\ell_i}{g_i}.
\end{aligned}$$

For $\delta > 0$ choose $\varepsilon = \delta t^{H+\frac{1}{2}}$ and $K \equiv t^{1-\xi}$ for some $1 > \xi > 0$ this becomes

$$\begin{aligned}
& -t - t\delta^2\sum_{i=1}^K \frac{\ell_i^2}{g_i^2} + t^{\frac{1}{2}+\frac{\xi}{2}}\delta\sum_{i=1}^K \frac{\ell_i}{g_i} \\
&= \quad -t - K^{\frac{1}{1-\xi}}\delta^2\sum_{i=1}^K \frac{\ell_i^2}{g_i^2} + \delta K^{\frac{1}{2}\frac{1+\xi}{1-\xi}}\sum_{i=1}^K \frac{\ell_i}{g_i}.
\end{aligned}$$

Plugging this into (2.8.8) gives

$$\begin{aligned}
& |\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \exp\left(-t - K^{\frac{1}{1-\xi}}\delta^2\sum_{i=1}^K \frac{\ell_i^2}{g_i^2} + K^{\frac{1}{2}\frac{1+\xi}{1-\xi}}\delta\sum_{i=1}^K \frac{\ell_i}{g_i}\right) \\
&= \quad |\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \exp\left(-t - K^{\frac{1}{1-\xi}}\delta^2\sum_{i=1}^K \frac{\ell_i^2}{g_i^2} + K^{\frac{1}{2}\frac{1+\xi}{1-\xi}}\delta\sum_{i=1}^K \frac{\ell_i}{g_i}\right). \qquad (2.8.10)
\end{aligned}$$

Remembering the form of the g_i and writing

$$h_i := \left(4HC_\rho\left(1-\frac{i}{K}\right)^{2H-1}\right)^{\frac{1}{2}}$$

we get that (2.8.10) is asymptotically equivalent to

$$|\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \exp\left(-t - K^{\frac{2-\xi}{1-\xi}}\delta^2 \sum_{i=1}^K \frac{\ell_i^2}{h_i^2} + K^{\frac{1}{1-\xi}}\delta \sum_{i=1}^K \frac{\ell_i}{h_i}\right). \tag{2.8.11}$$

Since $K^{\frac{2-\xi}{1-\xi}} \gg K^{\frac{1}{1-\xi}}$ and h_i^{-1} is bounded this is logarithmically equivalent to

$$|\mathcal{B}_{t,K}| \sum_{\{(\ell_i):\sum_{i=1}^K \ell_i<0\}} \exp\left(-t - K^{\frac{2-\xi}{1-\xi}}\delta^2 \sum_{i=1}^K \frac{\ell_i^2}{h_i^2}\right). \tag{2.8.12}$$

This we can write as

$$\exp(-t)|\mathcal{B}_{t,K}| \sum_{r=1}^K \sum_{m\geq 1} \sum_{\substack{\{(\ell_i):\sum_{i=1}^K \ell_i<0,\\ \#\{i:\ell_i\neq 0\}=r,\\ \max|\ell_i|=m\}}} \exp\left(-K^{\frac{2-\xi}{1-\xi}}\delta^2 \sum_{i=1}^K \frac{\ell_i^2}{h_i^2}\right). \tag{2.8.13}$$

Now observe that for fixed r and m the third sum in (2.8.13) consists at most of $\binom{K}{r}(2m)^r$ many summands. By Stirling's formula

$$\frac{\#\left\{(\ell_i) : \sum_{i=1}^K \ell_i < 0, \#\{i : \ell_i \neq 0\} = r, \max|\ell_i| = m\right\}}{\exp\big(K\log(K) + \log(2m)r\big)} \to 0$$

as $K \to \infty$. Since $\frac{1}{h_i^2}$ is bounded from below by $\frac{1}{4HC_\rho}$ and at least for one $i \in \{1, \ldots, K\}$ we have $\ell_i^2 = m^2$, the third sum in (2.8.13) is subexponentially small in K and m for all fixed r and m. We conclude that (2.8.13) vanishes for $K \to \infty$.
Finally observe that δ can be taken to zero such that the logarithmic equivalences of (2.8.11) and (2.8.12) still hold. So by

$$\frac{|\mathcal{B}_{t,K}|}{e^t} \Rightarrow \mathrm{Exp}(1)$$

the distribution of $|\mathcal{B}_{t,K}|$ is sufficiently concentrated, such that averaging (2.8.12) over the random tree gives a term of order $o(1)$, such that in total the leading order of the maximum is bounded by

$$\sum_{i=0}^K \Delta f_i \tag{2.8.14}$$

from above. Note that this is of the same order as $\sum_{i=1}^K \Delta f_i$.

To obtain the lower bound we change the above construction, such that we shift all branching events in $\left[\frac{i}{K}t, \frac{i+1}{K}t\right]$ of the Yule tree $\mathfrak{y}$ to $\frac{i+1}{K}t$, such that we can only have branching events at $\frac{t}{K}, \ldots, \frac{Kt}{K}$. By Slepian's Lemma, see Lemma 2.7.1, the order of the maximum of the corresponding GREM gives us a lower bound on the order of the maximum of BFBM. We now aim to apply the Paley–Zygmund inequality to show that for some function $\zeta : \mathbb{R} \to \mathbb{R}$, such that $\zeta(t) \to 1$ as $t \to \infty$, the probability to see a branch b with $\Delta Z_b^{(i)} \geq \Delta f_i \cdot \zeta(t)$ for all $i \geq 1$ is positive. Remember that we are allowed to ignore $\Delta Z_b^{(0)}$ since $\sum_{i=0}^{K} \Delta f_i \sim \sum_{i=1}^{K} \Delta f_i$. Together with the classical concentration inequality by Talagrand, see [Tal95], this gives that as a lower bound on the leading order of the maximum we obtain (2.8.14): Choosing $\xi > \frac{2}{3}, K = t^{1-\xi}$ and

$$\zeta(t) = \left(1 + \frac{1}{2}\log\left(\frac{1}{4\pi}\right) \cdot \frac{K}{t} + \frac{1}{2}\log\left(\frac{K}{t}\right) \cdot \frac{K}{t} + \log(K) \cdot \frac{K}{t}\right)^{\frac{1}{2}},$$

we have $\zeta(t) \downarrow 1$ as $t \to \infty$. Now by a standard Gaussian tail estimate we have

$$\begin{aligned} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \geq \Delta f_i \cdot \zeta(t)\right) &\geq \text{const}\frac{1}{\sqrt{2\pi} \cdot \sqrt{2} \cdot \zeta(t)} \cdot \left(\frac{K}{t}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{t}{K} \cdot \zeta(t)^2\right) \\ &= \exp\left(-\frac{t}{K}\right) \frac{\text{const}}{\zeta(t) \cdot K}. \end{aligned}$$

Since we shifted all branching events a little bit to the right, this implies that the expected number of branches such that $\Delta Z_b^{(i)} \geq \Delta f_i \cdot \zeta(t)$ for all i is bounded from below by const $\cdot$ $\zeta(t)^{-K} K^{-K} \exp\left(-t/K\right)$ for t large enough. The probability that two branches b and $\tilde{b}$ with $b \wedge \tilde{b} = \frac{\ell}{K}t$ fulfill $\Delta Z_b^{(i)}, \Delta Z_{\tilde{b}}^{(i)} \geq \Delta f_i \cdot \zeta(t)$ for all i is given by

$$\left[\prod_{i=1}^{\ell} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \geq \Delta f_i \cdot \zeta(t)\right)\right] \cdot \left[\prod_{i=\ell+1}^{K} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \geq \Delta f_i \cdot \zeta(t)\right)\right]^2.$$

By the corresponding standard Gaussian tail estimate

$$\begin{aligned} \mathbf{P}_{\mathfrak{y}}\left(\Delta Z_0^{(i)} \geq \Delta f_i \cdot \zeta(t)\right) &\leq \text{const}\frac{1}{\sqrt{2\pi} \cdot \sqrt{2} \cdot \zeta(t) \cdot K} \cdot \left(\frac{K}{t}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{t}{K} \cdot \zeta(t)^2\right) \\ &= \exp\left(-\frac{t}{K}\right) \frac{\text{const}}{\zeta(t) \cdot K} \end{aligned}$$

this is bounded from above by

$$\frac{\text{const}}{\zeta(t)^{2K-\ell} \cdot K^{2K-\ell}} \exp\left(-\ell\frac{t}{K} - 2(K-\ell)\frac{t}{K}\right) = \frac{\text{const}}{\zeta(t)^{2K-\ell} \cdot K^{2K-\ell}} \exp\left(-(2K-\ell)\frac{t}{K}\right).$$

Denote by $\mathbf{E}_{\mathcal{P}}$ the expectation with averages with respect to $\mathcal{P}$. Since now in expectation we have of order

$$\exp\left(\frac{t}{K}\left(2K - \ell - 1\right)\right)$$

many choices for two branches that separate at time $\ell \frac{t}{K}$ this directly gives that

$$\mathbf{E}_{\mathcal{P}}\left[\left(\#\left\{b \in \mathcal{B}_{t,K} : \Delta Z_b^{(i)} \geq \Delta f_i \cdot \zeta(t) \qquad \forall i\right\}\right)^2\right]$$

is bounded from above by

$$\begin{aligned} & \text{const} \cdot \sum_{\ell=0}^{K} \exp\left(\frac{t}{K}(2K-\ell-1)\right) \frac{1}{\zeta(t)^{2K-\ell} \cdot K^{2K-\ell}} \exp\left(-(2K-\ell)\frac{t}{K}\right) \\ \leq \; & \text{const} \cdot \zeta(t)^{-K} K^{-K} \exp\left(-t/K\right). \end{aligned}$$

For $1 > \xi > \frac{1}{2}$ the factor $\zeta(t)^{-K}$ converges to 1as $K \to \infty$. So in total the Paley–Zygmund inequality gives that there exists a constant $c(H)$, such that for t large enough

$$\mathcal{P}\left(\#\left\{b \in \mathcal{B}_{t,K} : \Delta Z_b^{(i)} \geq \Delta f_i \qquad \forall i\right\} > 0\right) \geq c(H) > 0.$$

Since $\max_b Z_b$ is a Lipschitz-function of the increments $\Delta Z_b^{(i)}$ we can apply the standard concentration inequality by Talagrand, see [Tal95, (1.14)], such that as a lower bound on the leading order of the maximum we also obtain (2.8.14). Now by (2.8.6)

$$\begin{aligned} \Delta f_i &= \sqrt{2\frac{t}{K}t^H\left[-C_\rho\left(1-\frac{i}{K}\right)^{2H} + C_\rho\left(1-\frac{i-1}{K}\right)^{2H}\right]} \\ &= t^{H+\frac{1}{2}}\sqrt{2C_\rho}\left[\frac{-1}{K}\left[\left(1-\frac{i}{K}\right)^{2H} - \left(1-\frac{i-1}{K}\right)^{2H}\right]\right]^{\frac{1}{2}} \\ &\approx t^{H+\frac{1}{2}}\sqrt{2C_\rho}\sqrt{2H}\left[\frac{1}{K^2}\left(1-\frac{i}{K}\right)^{2H-1}\right]^{\frac{1}{2}} \\ &= 2t^{H+\frac{1}{2}}\sqrt{C_\rho H}\left[\frac{1}{K^2}\left(1-\frac{i}{K}\right)^{2H-1}\right]^{\frac{1}{2}} \\ &= 2t^{H+\frac{1}{2}}\sqrt{C_\rho H}K^{-1}\left(1-\frac{i}{K}\right)^{H-\frac{1}{2}}, \end{aligned}$$

which for $t \equiv K$ large gives

$$\begin{aligned} f\left(\frac{\ell}{K}t\right) &:= \sum_{i=1}^{\ell} \Delta f_i = 2t^{H+\frac{1}{2}}\sqrt{C_\rho H}K^{-1}\sum_{i=1}^{\ell}\left(1-\frac{i}{K}\right)^{H-\frac{1}{2}} \\ &\sim 2t^{H+\frac{1}{2}}\sqrt{C_\rho H}\int_0^{\ell/K}(1-y)^{H-\frac{1}{2}}\,\mathrm{d}y. \end{aligned} \tag{2.8.15}$$

For $\ell = K$ the left hand side of (2.8.15) equals (2.8.14). This gives the desired result. $\square$

2.A Proof of (2.7.1) and (2.7.2)

In this section we give a proof of (2.7.1) and (2.7.2) to show how Slepian's Lemma, Lemma 2.7.1, is already applicable to obtain the order of magnitude of $\max_{b\in\mathcal{B}} B_b(t)$ without the prefactor.

Proof of (2.7.1) *and* (2.7.2). To prove (2.7.1) and (2.7.2) we will make use of Lemma 2.7.1. Given a Yule-Tree with $\mathcal{N}(t) \sim \text{Geo}\left(e^{-t}\right)$ many leaves at time t we choose $s := xt, 0 < x < 1$ arbitrary and couple a process in the following way: All branching events after time s get deleted. All branching events before time s get shifted to exactly s, such that there is one branch till s and at time s exactly $\mathcal{N}(t)$ branch off simultaneously. (So we only consider a modified version of a subtree. To keep notation consistent we do not modify the names of the branches while modifying the tree.) On this tree we build a BFBM $\tilde{B}$ with Hurst parameter H, covariance structure ρ and branching rate 1. Obviously for all branches $\tilde{b}_1, \tilde{b}_2$ in the new tree and all branches b_1, b_2 in $\mathfrak{y}$ which start before time s, we get

$$\rho\,(t,t,s) = \mathbf{Cov}_{\mathfrak{y}}\left[\tilde{B}_{\tilde{b}_1}(t), \tilde{B}_{\tilde{b}_2}(t)\right] \geq \mathbf{Cov}_{\mathfrak{y}}\left[B_{b_1}(t), B_{b_2}(t)\right].$$

By Lemma 2.7.1 the order of the maximum of our new process is lower than the one of the original process, giving us a lower bound on the maximum of our BFBM. Since

$$\tilde{B}_b(t) \overset{(d)}{=} \sqrt{\rho(t,t,s)}Z + \sqrt{t^{2H} - \rho(t,t,s)}Z_b$$

for $Z, (Z_b)_{b\in\mathcal{B}}$ independent and standard Gaussian, we see that the order of the maximum over $\tilde{B}_b(t)$ is of the same order as

$$\max_{b\in\mathcal{B}} \sqrt{t^{2H} - \rho(t,t,s)}Z_b.$$

By a standard tail estimate we get

$$\begin{aligned}
& \mathbf{P}_{\mathfrak{y}}\left(rt^H Z_b > Ct^{H+\frac{1}{2}} + a_t\right) \\
= \;& \mathbf{P}_{\mathfrak{y}}\left(Z_b > \frac{ct^{\frac{1}{2}}}{r} + \frac{a_t}{rt^H}\right) \\
\sim \;& \frac{r}{\sqrt{2\pi}C}\exp\left(-\frac{1}{2}\log(t) - \frac{1}{2}\left(\frac{C^2}{r^2}t + \frac{a_t^2}{r^2t^{2H}} + \frac{2C}{r}t^{\frac{1}{2}-H}a_t\right)\right) \\
= \;& \frac{r}{\sqrt{4\pi}}\exp\left(-\frac{1}{2}\log(t) - \frac{1}{2}\left(2t + o(1) - \log(t)\right)\right) \\
= \;& \frac{1}{\sqrt{4\pi}}\exp\left(-t\right)\left(1 + o(1)\right)
\end{aligned}$$

for

$$C = \sqrt{2}r$$

and

$$a_t = -t^{H-\frac{1}{2}}\log(t)\frac{\rho}{2\sqrt{2}}.$$

So since $e^{-t}\mathcal{N}(t)$ is a martingale with

$$\mathrm{Exp}(1) \sim E_1 = \lim_{t\to\infty} e^{-t}\mathcal{N}(t),$$

we immediately get (2.7.1) . This shows the lower bound. The upper bound follows analogously for $\mathcal{N}(t)$ independent fractional Brownian motions by simply using a first moment bound. □

2.B Proofs of some properties of $\mathfrak{y}$-indexed processes

We start by giving a proof of Lemma 2.5.7:

Proof of Lemma 2.5.7. For all $b' \in \{b_1, \ldots, b_m\}$ and $t, t' > b' \wedge b$ we have

$$\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{b'}(t')\middle|\, X_{b\cap b'}\right] = 0$$

since $X_{\mathfrak{y}}$ has the *pairwise branching property* (PCI), see (PCI). By construction one can choose b' such that

$$X_{\mathcal{S}\cap b} = X_{b'\cap b}.$$

This implies that for all branches $\hat{b} \in \{b_1, \ldots, b_m\}$ and $t, \hat{t} > \hat{b} \wedge b$

$$\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{\hat{b}}(\hat{t})\middle|\, X_{b\cap\mathcal{S}}\right] = \mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{\hat{b}}(\hat{t})\middle|\, X_{b\cap b'}\right].$$

Now

$$\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{\hat{b}}(\hat{t})\middle|\, X_{b\cap b'}\right] = \mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{\hat{b}}(\hat{t})\middle|\, X_{(b\cap b')\setminus\hat{b}}, X_{b\cap\hat{b}}\right].$$

But since

$$\left(X_b(t), X_{(b\cap b')\setminus\hat{b}}\right) \quad \text{and} \quad X_{\hat{b}}(\hat{t})$$

are conditionally independent given

$$X_{b\cap\hat{b}}$$

and

$$X_{\hat{b}}(\hat{t}), \qquad X_{(b\cap b')\setminus\hat{b}}$$

are conditionally independent given

$$X_{b\cap\hat{b}}$$

we obtain

$$\begin{aligned} &\mathcal{L}\left(\left(X_b(t), X_{(b\cap b')\setminus\hat{b}}\right), X_{\hat{b}}(\hat{t})\middle|\, X_{b\cap\hat{b}}\right) \\ = \;&\mathcal{L}\left(\left(X_b(t), X_{(b\cap b')\setminus\hat{b}}\right)\middle|\, X_{b\cap\hat{b}}\right) \otimes \mathcal{L}\left(X_{\hat{b}}(\hat{t})\middle|\, X_{b\cap\hat{b}}\right)\end{aligned}$$

$$
\begin{aligned}
&= \left[\mathcal{L}\left(\left(X_b(t)\right) \Big| \, X_{b\cap\hat{b}} \right) \otimes \mathcal{L}\left(X_{(b\cap b')\setminus\hat{b}} \Big| \, X_{b\cap\hat{b}} \right)\right] \\
&\quad \otimes \mathcal{L}\left(X_{\hat{b}}(\hat{t}) \Big| \, X_{b\cap\hat{b}} \right) \\
&= \left[\mathcal{L}\left(\left(X_b(t)\right) \Big| \, X_{(b\cap b')\setminus\hat{b}}, X_{b\cap\hat{b}} \right) \otimes \mathcal{L}\left(X_{(b\cap b')\setminus\hat{b}} \Big| \, X_{b\cap\hat{b}} \right)\right] \\
&\quad \otimes \mathcal{L}\left(X_{\hat{b}}(\hat{t}) \Big| \, X_{(b\cap b')\setminus\hat{b}}, X_{b\cap\hat{b}} \right)
\end{aligned}
$$

giving us

$$
\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{\hat{b}}(\hat{t}) \Big| \, X_{(b\cap b')\setminus\hat{b}}, X_{b\cap\hat{b}} \right] = \mathbf{Cov}_{\mathfrak{y}}\left[X_b(t), X_{\hat{b}}(\hat{t}) \Big| \, X_{b\cap\hat{b}} \right] = 0.
$$

Now since $X_{\mathfrak{y}}$ is Gaussian this implies that $X_{b\setminus\mathcal{S}}$ and $X_{\mathcal{S}\setminus b}$ are conditionally independent given $X_{\mathcal{S}\cap b}$. This implies that $X_{\mathfrak{y}}$ has the *branching property* (CI). □

Proof of Corollary 2.5.10. Let $t_1 \leq t_2$, then

$$
\left(X_b(t)\right)_{t_1\leq t\leq t_2}
$$

is a Gaussian process for all $b \in \mathcal{B}$. Set $\mathcal{S} := \bigcup_{i=1}^m b_i$ for branches $b_1, \ldots, b_m$.Then by the *branching property* (CI)

$$
\begin{aligned}
\mathcal{L}\left(X_{b\cup\mathcal{S}}\right) &= \mathcal{L}\left(\mathcal{L}\left(X_{b\setminus\mathcal{S}}, X_{\mathcal{S}\setminus b} \Big| \, X_{b\cap\mathcal{S}} \right) \otimes \delta_{X_{b\cap\mathcal{S}}} \right) \\
&= \mathcal{L}\left(\mathcal{L}\left(X_{b\setminus\mathcal{S}} \Big| \, X_{b\cap\mathcal{S}} \right) \otimes \mathcal{L}\left(X_{\mathcal{S}\setminus b} \Big| \, X_{b\cap\mathcal{S}} \right) \otimes \delta_{X_{b\cap\mathcal{S}}} \right).
\end{aligned}
$$

Note that $\mathcal{L}\left(X_{b\setminus\mathcal{S}} \Big| \, X_{b\cap\mathcal{S}} \right)$ is the distribution of a Gaussian process conditioned on its past, as is $\mathcal{L}\left(X_{\mathcal{S}\setminus b} \Big| \, X_{b\cap\mathcal{S}} \right)$ by iteration, such that in total $\mathcal{L}\left(X_{b\cup\mathcal{S}}\right)$ is a Gaussian process. □

The proof of Corollary 2.5.11 is now very similar to the one of Corollary 2.5.10, but Gaussianity gets exchanged with asymptotic Gaussianity.

Proof of Corollary 2.5.11. Let $t_1 \leq t_2$, then

$$
\left(X_b^{(n)}(t)\right)_{t_1\leq t\leq t_2}
$$

is an asymptotically Gaussian process for all $b \in \mathcal{B}$ and by the *branching property* (CI) for branches $b_1, \ldots, b_m$ with $\max_{i=1,\ldots,m} b \wedge b_i = s$

$$
\left(X_b^{(n)}(t)\right)_{t\geq s}, \qquad \left(\left(X_{b_i}^{(n)}(t)\right)_{t\geq b_i\wedge b}\right)_{i=1,\ldots,m}
$$

are conditionally independent given

$$\left(\left(X_{b_i}^{(n)}(t)\right)_{t\le b_i\wedge b}\right)_{i=1,\dots,m}. \tag{2.B.1}$$

By similar arguments as in the proof of Corollary 2.5.10 it follows that the process (2.B.1) is also asymptotically Gaussian. This already gives the assertion. □

Proof of Proposition 2.5.3. First observe that by (2.5.1)

$$\mathcal{L}\left(\left(X_{\tilde b}\right)_{t\ge s}\Big|\left(X_b\right)_{t\in\mathbb{R}}\right)=\mathcal{L}\left(\left(X_b\right)_{t\ge s}\Big|\left(X_b\right)_{t\le s}\right)=\mathcal{L}\left(\left(X_{\tilde b}\right)_{t\ge s}\Big|\left(X_{\tilde b}\right)_{t\le s}\right)$$

is implied by (PCI). Now for branches $b,\tilde b\in\mathfrak{y}$ with $b\wedge\tilde b=s$ by definition we get for $t_1,t_2>s$

$$\begin{aligned}&\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t_1),X_{\tilde b}(t_2)\right]\\ =\;&\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)X_{\tilde b}(t_2)\right]-\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)\right]\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_2)\right].\end{aligned}$$

By (2.5.2) we get

$$\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)\right]\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_2)\right]=\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_1)\right]\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_2)\right]. \tag{2.B.2}$$

Using (PCI) we obtain

$$\begin{aligned}&\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)X_{\tilde b}(t_2)\right]\\ =\;&\mathbf{E}_{\mathfrak{y}}\left[\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)X_{\tilde b}(t_2)\middle|\,\sigma\left(X_{\tilde b}(t),t\le t_2\right)\right]\right]\\ =\;&\mathbf{E}_{\mathfrak{y}}\left[\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)\middle|\,\sigma\left(X_{\tilde b}(t),t\le t_2\right)\right]X_{\tilde b}(t_2)\right]\\ =\;&\mathbf{E}_{\mathfrak{y}}\left[\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)\middle|\,\sigma\left(X_{\tilde b}(t),t\le s\right)\right]X_{\tilde b}(t_2)\right].\end{aligned}$$

And by (2.5.2) we can exchange $X_b(t_1)$ and $X_{\tilde b}(t_1)$, meaning

$$\begin{aligned}&\mathbf{E}_{\mathfrak{y}}\left[\mathbf{E}_{\mathfrak{y}}\left[X_b(t_1)\middle|\,\sigma\left(X_{\tilde b}(t),t\le s\right)\right]X_{\tilde b}(t_2)\right]\\ &=\mathbf{E}_{\mathfrak{y}}\left[\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_1)\middle|\,\sigma\left(X_{\tilde b}(t),t\le s\right)\right]X_{\tilde b}(t_2)\right].\end{aligned} \tag{2.B.3}$$

Together (2.B.3) and (2.B.2) give

$$\begin{aligned}&\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t_1),X_{\tilde b}(t_2)\right]\\ &=\mathbf{E}_{\mathfrak{y}}\left[\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_1)\middle|\,\sigma\left(X_{\tilde b}(t),t\le s\right)\right]X_{\tilde b}(t_2)\right]\\ &-\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_1)\right]\mathbf{E}_{\mathfrak{y}}\left[X_{\tilde b}(t_2)\right].\end{aligned}$$

This gives that for branches $b, \tilde{b} \in \mathfrak{y}$ with $b \wedge \tilde{b} = s$ and $t_1, t_2 > s$

$$\mathbf{Cov}_{\mathfrak{y}}\left[X_b(t_1), X_{\tilde{b}}(t_2)\right]$$

is uniquely determined by the distribution of the process along one branch, i.e. by

$$\mathcal{L}\left(\left(X_b(t)\right)_{t\in\mathbb{R}}\right).$$

If now for all branches $b_1, \ldots b_m$ and $t_1, \ldots t_m \in \mathbb{R}$

$$\left(X_{b_1}(t_1), \ldots, X_{b_m}(t_m)\right)$$

is jointly Gaussian we get that the distribution of the *tree-indexed branching stochastic process*

$$\left(\left(X_b(t)\right)_{t\in\mathbb{R}}\right)_{b\in\mathcal{B}}$$

is uniquely determined by the distribution along one branch, i.e. by

$$\mathcal{L}\left(\left(X_b(t)\right)_{t\in\mathbb{R}}\right)$$

for any branch $b \in \mathcal{B}$. □

2.C Covariances of different branches

To prove Theorem 2.3.2 by Proposition 2.2.3 it is enough to analyze

$$\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{m_1} Y_{b,\ell}, \sum_{\ell=1}^{m_2} Y_{\tilde{b},\ell}\right] \qquad \text{for } m_1 = t_1 n, m_2 = t_2 n, t_1 > t_2 > s, b \in \mathcal{B} \tag{2.C.1}$$

for two branches $b, \tilde{b}$ with $b \wedge \tilde{b} = s$ since by [IW23, Proposition 1.3] tightness is given and asymptotic joint Gaussianity is given by Proposition 2.2.3. We now calculate the covariance between $S_b^{(n)}(t_1)$ and $S_{\tilde{b}}^{(n)}(t_2)$ for $t_1 > t_2 > s$.

First set

$$C_1 := C_2 \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}, \tag{2.C.2}$$

for

$$C_2 := \frac{1}{\sum_{\ell\geq 0} q_\ell^2} > 0, \tag{2.C.3}$$

and

$$C_3 := \frac{1}{\alpha(2\alpha+1)C_1} = \frac{C_2}{\alpha(2\alpha+1)} \cdot \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}. \tag{2.C.4}$$

Recalling (2.2.13) we get

$$C_1 n^{2\alpha-1} \sim \mathbf{P}_{\mathfrak{y}}\left(0 \sim n\right), \tag{2.C.5}$$

see also [IW23, Proposition 2.1]. Denote by A_0 the ancestral line of zero. Set

$$C_q := \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}, \tag{2.C.6}$$

then

$$C_q n^{\alpha-1} \sim q_n := \mathbf{P}_{\mathfrak{y}}\left(-n \in A_0\right), \tag{2.C.7}$$

see (2.2.11) and for further reference [IW23, (1.10)]. Recalling (2.2.8), see also [IW23, (1.8)], we get

$$C_3 n^{2\alpha+1} \sim \mathbf{Var}_{\mathfrak{y}}\left[\sum_{\ell=1}^{n} Y_\ell\right]. \tag{2.C.8}$$

Let's start by using bilinearity and the fact that $Y_{(b,i)} = Y_{(\tilde{b},i)} =: Y_i$ for $i \leq sn$:

$$\begin{aligned}
&\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{m_1} Y_{b,\ell}, \sum_{\ell=1}^{m_2} Y_{\tilde{b},\ell}\right] \\
=\;& \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_{b,\ell} + \sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=1}^{sn} Y_{\tilde{b},\ell} + \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right] \\
=\;& \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_{b,\ell}, \sum_{\ell=1}^{sn} Y_{\tilde{b},\ell}\right] + \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right] \\
&+ \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=1}^{sn} Y_{\tilde{b},\ell}\right] \\
&+ \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right] \\
=\;& \mathbf{Var}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell\right] + \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right] \\
&+ \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=sn+1}^{m_1} Y_{(b,\ell)}, \sum_{\ell=1}^{sn} Y_{b,\ell}\right] \\
&+ \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right] \\
=\;& \mathbf{Var}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell\right] && (2.C.9) \\
&+ \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell, \sum_{\ell=sn+1}^{m_1} Y_{b,\ell}\right] + \mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell, \sum_{\ell=sn+1}^{m_2} Y_{b,\ell}\right] && (2.C.10)
\end{aligned}$$

$$+\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right]. \tag{2.C.11}$$

We now need to check the asymptotics of (2.C.9), (2.C.10) and (2.C.11). These summands can be interpreted in the following way:

(a) The summand in (2.C.9) is the variance of the random walk along one branch.

(b) The first and second summand in (2.C.10) represent the covariance between the increments before and after a branching event.

(c) (2.C.11) represents the covariance between the increments of two branches after a branching event.

Variance of the random walk accumulated before the branching event. For the first summand in (2.C.9) we get

$$\mathbf{Var}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell\right] \sim C_3(sn)^{2\alpha+1},$$

see (2.C.8).

Covariance between the increments of the random walk before and after the branching event. We now analyze the second summand in (2.C.9). By $Y_\ell^2 = 1$ we get

$$\mathbf{Cov}_{\mathfrak{y}}\left[Y_\ell, Y_{b,k}\right] = \mathbf{P}_{\mathfrak{y}}\left(\ell \sim (b,k)\right). \tag{2.C.12}$$

This gives

$$\begin{aligned}&\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{sn} Y_\ell, \sum_{\ell=sn+1}^{m_1} Y_{b,\ell}\right]\\ &= \sum_{\ell=1}^{sn}\sum_{k=sn+1}^{m_1} \mathbf{Cov}_{\mathfrak{y}}\left[Y_\ell, Y_{b,k}\right]\\ &= \sum_{\ell=1}^{sn}\sum_{k=sn+1}^{m_1} \mathbf{P}_{\mathfrak{y}}\left(\ell \sim k\right).\end{aligned}$$

Plugging in the asymptotics (2.C.5) for $n \to \infty$ gives

$$\begin{aligned}&\sum_{\ell=1}^{sn}\sum_{k=sn+1}^{m_1} \mathbf{P}_{\mathfrak{y}}\left(\ell \sim k\right)\\ \sim\ & C_1\sum_{\ell=1}^{sn}\sum_{k=sn+1}^{m_1} (k-\ell)^{2\alpha-1}\end{aligned}$$

$$
\begin{aligned}
&= C_1 n^{2\alpha-1} \cdot n^2 \cdot \frac{1}{n} \sum_{\ell=1}^{sn} \frac{1}{n} \sum_{k=sn+1}^{x_1 n} \left(\frac{k}{n} - \frac{\ell}{n} \right)^{2\alpha-1} \\
&\sim C_1 n^{2\alpha+1} \int_0^s \int_s^{x_1} (y_2 - y_1)^{2\alpha-1} \, \mathrm{d}y_2 \, \mathrm{d}y_1. \qquad (2.C.13)
\end{aligned}
$$

The integral in (2.C.13) can be computed explicitly: The inner integral is given by

$$
\begin{aligned}
&\int_s^{x_1} (y_2 - y_1)^{2\alpha-1} \, \mathrm{d}y_2 \\
&= \left[\frac{1}{2\alpha} (y_2 - y_1)^{2\alpha} \right]_{y_2=s}^{x_1} \\
&= \frac{1}{2\alpha} \left[(x_1 - y_1)^{2\alpha} - (s - y_1)^{2\alpha} \right].
\end{aligned}
$$

For the two summands we compute the outer integral: The first one gives

$$
\begin{aligned}
&\int_0^s (x_1 - y_1)^{2\alpha} \, \mathrm{d}y_1 \\
&= \left[-\frac{1}{2\alpha+1} (x_1 - y_1)^{2\alpha+1} \right]_{y_1=0}^{s} \\
&= \frac{1}{2\alpha+1} \left[{x_1}^{2\alpha+1} - (x_1 - s)^{2\alpha+1} \right].
\end{aligned}
$$

The second gives

$$
\begin{aligned}
&\int_0^s (s - y_1)^{2\alpha} \, \mathrm{d}y_2 \\
&= \left[-\frac{1}{2\alpha+1} (s - y_1)^{2\alpha+1} \right]_{y_1=0}^{s} \\
&= \frac{s^{2\alpha+1}}{2\alpha+1}.
\end{aligned}
$$

So in total we get

$$
\int_0^s \int_s^{x_1} (y_2 - y_1)^{2\alpha-1} \, \mathrm{d}y_2 \, \mathrm{d}y_1 = \frac{1}{2\alpha\,(2\alpha+1)} \left[{x_1}^{2\alpha+1} - (x_1 - s)^{2\alpha+1} - s^{2\alpha+1} \right].
$$

For the second summand in (2.C.9) this gives

$$
\begin{aligned}
&\mathbf{Cov}_{\mathfrak{y}} \left[\sum_{\ell=1}^{sn} Y_\ell, \sum_{\ell=sn+1}^{m_1} Y_{b,\ell} \right] \\
\sim\ & n^{2\alpha+1} C_1 \frac{1}{2\alpha\,(2\alpha+1)} \left[{x_1}^{2\alpha+1} - (x_1 - s)^{2\alpha+1} - s^{2\alpha+1} \right].
\end{aligned}
$$

Analogously to (2.C.9) we get

$$
\mathbf{Cov}_{\mathfrak{y}} \left[\sum_{\ell=1}^{sn} Y_\ell, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell} \right]
$$

$$\sim \quad n^{2\alpha+1} C_1 \frac{1}{2\alpha\,(2\alpha+1)} \left[{x_2}^{2\alpha+1} - (x_2 - s)^{2\alpha+1} - s^{2\alpha+1} \right].$$

Covariance of the increments of the two branches We now aim to compute (2.C.11), the covariance of the increments of two branches, namely

$$\mathbf{Cov}_{\mathfrak{y}} \left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell} \right],$$

by analyzing $\mathbf{P}_{\mathfrak{y}} \left((b,\ell) \sim (\tilde{b}, k) \right)$. Using (2.C.12) we can write (2.C.11) as

$$\begin{aligned} & \mathbf{Cov}_{\mathfrak{y}} \left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell} \right] \\ = & \sum_{\ell=sn+1}^{m_1} \sum_{k=sn+1}^{m_2} \mathbf{Cov}_{\mathfrak{y}} \left[Y_{b,\ell}, Y_{\tilde{b},k} \right] \\ = & \sum_{\ell=sn+1}^{m_1} \sum_{k=sn+1}^{m_2} \mathbf{P}_{\mathfrak{y}} \left((b,\ell) \sim (\tilde{b}, k) \right). \end{aligned}$$

For further analysis we will make use of Proposition 2.3.3.Proposition 2.3.3 now gives

$$\begin{aligned} & \sum_{\ell=sn+1}^{m_1} \sum_{k=sn+1}^{m_2} \mathbf{P}_{\mathfrak{y}} \left((b,\ell) \sim (\tilde{b}, k) \right) \\ = & C_2 \sum_{\ell=sn+1}^{m_1} \sum_{k=sn+1}^{m_2} \sum_{r\geq 0} q_{\ell-sn+r} q_{k-sn+r} \\ = & C_2 \sum_{\ell=1}^{(t_1-s)n} \sum_{k=1}^{(t_2-s)n} \sum_{r\geq 0} q_{\ell+r} q_{k+r}. \end{aligned}$$

By plugging in (2.C.7) we obtain

$$\begin{aligned} & C_2 \sum_{\ell=1}^{(t_1-s)n} \sum_{k=1}^{(t_2-s)n} \sum_{r\geq 0} q_{\ell+r} q_{k+r} \\ \sim \quad & C_2 C_q^2 \sum_{\ell=1}^{(t_1-s)n} \sum_{k=1}^{(t_2-s)n} \sum_{r\geq 0} (\ell+r)^{\alpha-1} (k+r)^{\alpha-1} \\ = \quad & C_2 C_q^2 n^{2\alpha-2} \cdot n^3 \cdot \frac{1}{n} \sum_{\ell=1}^{(t_1-s)n} \frac{1}{n} \sum_{k=1}^{(t_2-s)n} \frac{1}{n} \sum_{r\geq 0} \left(\frac{\ell+r}{n} \right)^{\alpha-1} \left(\frac{k+r}{n} \right)^{\alpha-1} \\ \sim \quad & C_2 C_q^2 n^{2\alpha+1} \int_0^{t_1-s} \int_0^{t_2-s} \int_0^{\infty} (y_3+y_1)^{\alpha-1} (y_3+y_2)^{\alpha-1} \, \mathrm{d}y_3 \, \mathrm{d}y_2 \, \mathrm{d}y_1. \end{aligned}$$

Finally we get

$$\begin{aligned}&\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=sn+1}^{m_1} Y_{b,\ell}, \sum_{\ell=sn+1}^{m_2} Y_{\tilde{b},\ell}\right]\\ \sim\;& C_2 C_q^2 n^{2\alpha+1} \int\limits_0^{t_1-s}\int\limits_0^{t_2-s}\int\limits_0^{\infty} (y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1.\end{aligned}$$

Putting the pieces together Putting it all together gives

$$\begin{aligned}&\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{m_1} Y_{b,\ell}, \sum_{\ell=1}^{m_2} Y_{\tilde{b},\ell}\right]\\ \sim\;& n^{2\alpha+1}\Bigg[C_3 s^{2\alpha+1}\\ &+C_1\frac{1}{2\alpha\,(2\alpha+1)}\left[t_1^{2\alpha+1}-(t_1-s)^{2\alpha+1}-s^{2\alpha+1}\right]\\ &+C_1\frac{1}{2\alpha\,(2\alpha+1)}\left[t_2^{2\alpha+1}-(t_2-s)^{2\alpha+1}-s^{2\alpha+1}\right]\\ &+C_2C_q^2\int_0^{t_1-s}\int_0^{t_2-s}\int_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1\Bigg].\end{aligned}$$

Plugging in (2.C.2), (2.C.3) and (2.C.4), namely

$$C_1 = C_2\frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}, \tag{2.C.14}$$

and

$$C_3 = \frac{1}{\alpha(2\alpha+1)}C_1 = \frac{C_2}{\alpha(2\alpha+1)}\frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3} \tag{2.C.15}$$

and (2.C.6), namely

$$C_q = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}, \tag{2.C.16}$$

we can write

$$\begin{aligned}&\rho^{(n)}(t_1,t_2,s)\\ :=\;&\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{t_1 n} Y_{b,\ell}, \sum_{\ell=1}^{t_2 n} Y_{\tilde{b},\ell}\right]\\ \sim\;& n^{2\alpha+1}C_2\Bigg[s^{2\alpha+1}\frac{1}{\alpha(2\alpha+1)}\frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}\end{aligned}$$

$$+\frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}\frac{1}{2\alpha\,(2\alpha+1)}\left[t_1^{2\alpha+1}-(t_1-s)^{2\alpha+1}-s^{2\alpha+1}\right]$$
$$+\frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3}\frac{1}{2\alpha\,(2\alpha+1)}\left[t_2^{2\alpha+1}-(t_2-s)^{2\alpha+1}-s^{2\alpha+1}\right]$$
$$\left.+\frac{\int_0^{t_1-s}\int_0^{t_2-s}\int_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1}{\Gamma(\alpha)^2\Gamma(1-\alpha)^2}\right].$$

However, writing it this way makes it hard to see that

$$\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{m_1}Y_{b,\ell},\sum_{k=1}^{m_2}Y_{\tilde{b},k}\right] \tag{2.C.17}$$

is indeed increasing in s for fixed m_1, m_2. Writing

$$\mathbf{Cov}_{\mathfrak{y}}\left[\sum_{\ell=1}^{m_1}Y_{b,\ell},\sum_{k=1}^{m_2}Y_{\tilde{b},k}\right]=\sum_{\ell=1}^{m_1}\sum_{k=1}^{m_2}\mathbf{P}_{\mathfrak{y}}\left((b,\ell)\sim(\tilde{b},k)\right) \tag{2.C.18}$$

gives us that (2.C.17) is increasing in s for fixed m_1, m_2 since

$$\mathbf{P}_{\mathfrak{y}}\left((b,\ell)\sim(\tilde{b},k)\right) \tag{2.C.19}$$

is increasing in s. This is all we needed to show for the assertion of Theorem 2.3.2.

2.D Proof of Proposition 2.3.3

The proof of Proposition 2.3.3 is similar to the one in [IW23, Proposition 2.1]. The idea is based on considering two independent ancestral lines, which means lines that do not coalesce even if they meet. For this two lines we estimate the number of common ancestors and decompose it with respect to their most common ancestor.

Proof of Proposition 2.3.3. Assume without loss of generality $j > i$ and let $b, \tilde{b}$ be two branches with $b \wedge \tilde{b} = s$. Denote by $\tilde{A}_{(b,i)}$ and $\tilde{A}_{(\tilde{b},j)}$ two independent (non-coalescing) ancestral lines starting in (b,i) and in $(\tilde{b},j)$ respectively. Since those can not meet before sn we can write

$$\mathbf{E}_{\mathfrak{y}}\left[\left|\tilde{A}_{(b,i)}\cap\tilde{A}_{(\tilde{b},j)}\right|\right]=\sum_{r\geq i-sn}q_r q_{r+j-i}. \tag{2.D.1}$$

By a decomposition with respect to the most recent common ancestor we can write

$$\mathbf{E}_{\mathfrak{y}}\left[\left|\tilde{A}_{(b,i)}\cap\tilde{A}_{(\tilde{b},j)}\right|\right]=\mathbf{P}_{\mathfrak{y}}\left((b,i)\sim(\tilde{b},j)\right)\sum_{r\geq 0}q_r^2. \tag{2.D.2}$$

Giving us

$$\mathbf{P}_{\mathfrak{y}}\left((b,i)\sim(\tilde{b},j)\right)=C_2\sum_{r\geq(i\wedge j)-sn}q_r q_{r+|j-i|}=C_2\sum_{r\geq 0}q_{i-sn+r}q_{j-sn+r}. \tag{2.D.3}$$

□

2.E Some analytic identities

In this section we want to state some analytic identities which are a consequence of (2.3.11). First recall the definitions

$$C_H:=\left(-\frac{2^{-2H}\Gamma(-H)\Gamma\left(H+\frac{1}{2}\right)}{\sqrt{\pi}}\right)^{\frac{1}{2}}, \tag{2.E.1}$$

and

$$\begin{aligned}\rho^{\mathrm{HS}}(t_1,t_2,s) &:= \frac{1}{2}\left[t_1^{2\alpha+1}-(t_1-s)^{2\alpha+1}+t_2^{2\alpha+1}-(t_2-s)^{2\alpha+1}\right]\\ &\quad+\frac{\int\limits_0^{t_1-s}\int\limits_0^{t_2-s}\int\limits_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1}{\Gamma(1-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^{-1}(\alpha(2\alpha+1))^{-1}}.\end{aligned}$$

as well as

$$\begin{aligned}\rho^{\mathrm{K}}(t_1,t_2,s) &:= \frac{1}{C_H^2}\left(\int_{-\infty}^{0}\left((t_1-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)\left((t_2-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)\mathrm{d}\xi\right)\\ &\quad+\frac{1}{C_H^2}\left(\int_0^s(t_1-\xi)^{H-\frac{1}{2}}(t_2-\xi)^{H-\frac{1}{2}}\,\mathrm{d}\xi\right).\end{aligned}$$

We already know that for $H=\alpha+\frac{1}{2}$

$$\rho^{\mathrm{HS}}(t_1,t_2,s)=\rho^{\mathrm{K}}(t_1,t_2,s). \tag{2.E.2}$$

This gives the following analytic identity:

Proposition 2.E.1. *For all $t_1,t_2>s\geq 0$ we have the equality*

$$\begin{aligned}&\frac{1}{2}\left[t_1^{2\alpha+1}-(t_1-s)^{2\alpha+1}+t_2^{2\alpha+1}-(t_2-s)^{2\alpha+1}\right]\\ &+\frac{\int\limits_0^{t_1-s}\int\limits_0^{t_2-s}\int\limits_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1}{\Gamma(1-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^{-1}(\alpha(2\alpha+1))^{-1}}\\ =\;&\frac{1}{C_{\alpha+\frac{1}{2}}^2}\left(\int_{-\infty}^{0}\left((t_1-\xi)^{\alpha}-(-\xi)^{\alpha}\right)\left((t_2-\xi)^{\alpha}-(-\xi)^{\alpha}\right)\mathrm{d}\xi\right)\\ &+\frac{1}{C_{\alpha+\frac{1}{2}}^2}\left(\int_0^s(t_1-\xi)^{\alpha}(t_2-\xi)^{\alpha}\,\mathrm{d}\xi\right).\end{aligned}$$

The special case of $t_1 = t_2 \equiv t > s$ then gives the following identity:

Corollary 2.E.2. *Let $t > s > 0$ then*

$$\begin{aligned} & t^{2\alpha+1} - (t-s)^{2\alpha+1} \frac{\sqrt{\pi} 2^{2\alpha}}{\Gamma(1-H)\Gamma(\alpha)} \\ = \; & t^{2\alpha+1} - (t-s)^{2\alpha+1} \\ & + \frac{\alpha(2\alpha+1)\Gamma(1-\alpha)}{\Gamma(\alpha)\Gamma(1-2\alpha)} \int_0^{t-s} \int_0^{t-s} \int_0^\infty (y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1} \, dy_3 \, dy_2 \, dy_1. \end{aligned}$$

Choosing $t_1 \equiv t > t_2 \equiv 1$ and $s < 1$ in Proposition 2.E.1 shows a beautiful connection to the generalized hypergeometric function, which is defined by

$$ {}_2F_1(a,b;c;z) = \sum_{n \geq 0} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}, \qquad |z| < 1. \tag{2.E.3}$$

Corollary 2.E.3. *For $t > s$*

$$\begin{aligned} & \frac{1}{2}\left[t^{2\alpha+1} - (t-s)^{2\alpha+1} + 1 - (1-s)^{2\alpha+1}\right] \\ & + \frac{\int_0^{t-s}\int_0^{1-s}\int_0^\infty (y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1} \, dy_3 \, dy_2 \, dy_1}{\Gamma(1-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^{-1}(\alpha(2\alpha+1))^{-1}} \\ = \; & \lim_{y\to\infty} \left[\frac{1}{\alpha C_H^2} \left[-\frac{x^{\alpha+1}(t+x)^\alpha \left(\frac{t+x}{t}\right)^{-\alpha} {}_2F_1\left(1+\alpha, \alpha; 2+\alpha; -\frac{x}{t}\right)}{\frac{1}{\alpha}+1} \right.\right. \\ & + \frac{(x+1)^{\alpha+1}(t+x)^{\alpha+1}\left(\frac{t+x}{t-1}\right)^{-\alpha} {}_2F_1\left(1+\alpha,\alpha;2+\alpha;\frac{x+1}{1-t}\right)}{\frac{1}{\alpha}+1} \\ & \left.\left. - \frac{x^{\alpha+1} {}_2F_1(1+\alpha,\alpha;2+\alpha;-x)}{\frac{1}{\alpha}+1} + \frac{x^{1+\alpha}}{\frac{1}{\alpha}+2} \right]\right]_{x=0}^{x=y} \\ & - \frac{1}{1+\alpha}(1-s)^{\alpha+1}(t-1)^\alpha {}_2F_1\left(-\alpha,\alpha+1;\alpha+2;\frac{s-1}{t-1}\right) \\ & + \frac{1}{1+\alpha}(t-1)^{-\alpha} {}_2F_1\left(-\alpha,\alpha+1;\alpha+2;\frac{-1}{t-1}\right). \end{aligned}$$

We start with the proof of Proposition 2.E.1:

Proof of Proposition 2.E.1. We know that (2.E.2) holds. Since

$$\begin{aligned} \rho^{\mathrm{HS}}(t_1,t_2,s) = \; & \frac{1}{2}\left[t_1^{2\alpha+1} - (t_1-s)^{2\alpha+1} + t_2^{2\alpha+1} - (t_2-s)^{2\alpha+1}\right] \\ & + \frac{\int_0^{t_1-s}\int_0^{t_2-s}\int_0^\infty (y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1} \, dy_3 \, dy_2 \, dy_1}{\Gamma(1-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^{-1}(\alpha(2\alpha+1))^{-1}} \end{aligned}$$

and

$$\rho^{\mathrm{K}}(t_1,t_2,s) = \frac{1}{C_H^2}\left(\int_{-\infty}^{0}\left((t_1-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)\left((t_2-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)\mathrm{d}\xi\right) + \frac{1}{C_H^2}\left(\int_0^s (t_1-\xi)^{H-\frac{1}{2}}(t_2-\xi)^{H-\frac{1}{2}}\,\mathrm{d}\xi\right),$$

we have the desired result for $H=\alpha+\frac{1}{2}$. □

Proof of Corollary 2.E.2. Since for $t_1=t_2\equiv t>s$ we have

$$\begin{aligned}
&\rho^{\mathrm{K}}(t,t,s)\\
=\;& \frac{1}{C_H^2}\left(\int_{-\infty}^{0}\left((t-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)^2\mathrm{d}\xi\right)\\
&+\frac{1}{C_H^2}\left(\int_0^s (t-\xi)^{2H-1}\,\mathrm{d}\xi\right)\\
=\;& t^{2H}-(t-s)^{2H}\frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}
\end{aligned}$$

and

$$\begin{aligned}
&\rho^{\mathrm{HS}}(t,t,s)\\
=\;& t^{2\alpha+1}-(t-s)^{2\alpha+1}\\
&+\frac{\alpha(2\alpha+1)\Gamma(1-\alpha)}{\Gamma(\alpha)\Gamma(1-2\alpha)}\int_0^{t-s}\int_0^{t-s}\int_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1.
\end{aligned}$$

this gives the nice identity

$$\begin{aligned}
&\rho^{\mathrm{K}}(t,t,s)\\
=\;& t^{2H}-(t-s)^{2H}\frac{\sqrt{\pi}2^{2H-1}}{\Gamma(1-H)\Gamma\left(H+\frac{1}{2}\right)}\\
=\;& t^{2\alpha+1}-(t-s)^{2\alpha+1}\\
&+\frac{\alpha(2\alpha+1)\Gamma(1-\alpha)}{\Gamma(\alpha)\Gamma(1-2\alpha)}\int_0^{t-s}\int_0^{t-s}\int_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1\\
=\;& \rho^{\mathrm{HS}}(t,t,s).
\end{aligned}$$

Plugging in $H=\alpha+\frac{1}{2}$ gives the desired result. □

Proof of Corollary 2.E.3. For $t_1=t, t_2=1, s<1$ we have for $H=\alpha+\frac{1}{2}$

$$\begin{aligned}
&\rho^{\mathrm{K}}(t_1,t_2,s)\\
=\;& \frac{1}{C_H^2}\left(\int_{-\infty}^{0}\left((t_1-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)\left((t_2-\xi)^{H-\frac{1}{2}}-(-\xi)^{H-\frac{1}{2}}\right)\mathrm{d}\xi\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{C_H^2}\left(\int_0^s (t_1-\xi)^{H-\frac{1}{2}}(t_2-\xi)^{H-\frac{1}{2}}\,\mathrm{d}\xi\right)\\
=\;&\lim_{y\to\infty}\left[\frac{1}{\alpha C_H^2}\left[-\frac{x^{\alpha+1}(t+x)^{\alpha}\left(\frac{t+x}{t}\right)^{-\alpha}{}_2F_1\left(1+\alpha,\alpha;2+\alpha;-\frac{x}{t}\right)}{\frac{1}{\alpha}+1}\right.\right.\\
&+\frac{(x+1)^{\alpha+1}(t+x)^{\alpha+1}\left(\frac{t+x}{t-1}\right)^{-\alpha}{}_2F_1\left(1+\alpha,\alpha;2+\alpha;\frac{x+1}{1-t}\right)}{\frac{1}{\alpha}+1}\\
&-\frac{x^{\alpha+1}{}_2F_1\left(1+\alpha,\alpha;2+\alpha;-x\right)}{\frac{1}{\alpha}+1}\\
&\left.\left.+\frac{x^{1+\alpha}}{\frac{1}{\alpha}+2}\right]\right]_{x=0}^{x=y}\\
&-\frac{1}{1+\alpha}(1-s)^{\alpha+1}(t-1)^{\alpha}{}_2F_1\left(-\alpha,\alpha+1;\alpha+2;\frac{s-1}{t-1}\right)\\
&+(t-1)^{-\alpha}\frac{{}_2F_1\left(-\alpha,\alpha+1;\alpha+2;\frac{-1}{t-1}\right)}{\alpha+1}.
\end{aligned}
$$

To obtain this equality we apply the identity, which can be checked with every well-known CAS, for example `Mathematica`,

$$
\begin{aligned}
&\int\left((x+1)^{\frac{1}{c}}-x^{\frac{1}{c}}\right)\left((b+x)^{\frac{1}{c}}-x^{\frac{1}{c}}\right)\\
=\;&c\left(-\frac{x^{\frac{1}{c}+1}(b+x)^{\frac{1}{c}}\left(\frac{b+x}{b}\right)^{-1/c}{}_2F_1\left(1+\frac{1}{c},-\frac{1}{c};2+\frac{1}{c};-\frac{x}{b}\right)}{c+1}\right.\\
&+\frac{(x+1)^{\frac{1}{c}+1}(b+x)^{\frac{1}{c}}\left(\frac{b+x}{b-1}\right)^{-1/c}{}_2F_1\left(1+\frac{1}{c},-\frac{1}{c};2+\frac{1}{c};\frac{x+1}{1-b}\right)}{c+1}\\
&\left.-\frac{x^{\frac{1}{c}+1}{}_2F_1\left(1+\frac{1}{c},-\frac{1}{c};2+\frac{1}{c};-x\right)}{c+1}+\frac{x^{\frac{c+2}{c}}}{c+2}\right),\qquad b>1,\,c>0
\end{aligned}
$$

with $c=\frac{1}{\alpha}$, $b=t$, and transform the second integral, such that we can make use of [DLMF, (15.6.1)].

By (2.E.2) this is equal to

$$
\begin{aligned}
&\rho^{\mathrm{HS}}(t,1,s)\\
=\;&\frac{1}{2}\left[t^{2\alpha+1}-(t-s)^{2\alpha+1}+1-(1-s)^{2\alpha+1}\right]\\
&+\frac{\int_0^{t-s}\int_0^{1-s}\int_0^{\infty}(y_3+y_1)^{\alpha-1}(y_3+y_2)^{\alpha-1}\,\mathrm{d}y_3\,\mathrm{d}y_2\,\mathrm{d}y_1}{\Gamma(1-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^{-1}(\alpha(2\alpha+1))^{-1}}.
\end{aligned}
$$

This gives the desired identity. □

Acknowledgment We thank Matthias Birkner and Anton Wakolbinger for valuable discussions and suggestions, Nicola Kistler for educating us about generalized random energy models and their connection to branching random walks, Marius Schmidt for the arguments used in Section 2.8, Lisa Hartung for pointing us to [MZ16], and an anonymous reviewer for pointing out an alternative proof of Theorem 2.3.8, see Remark 2.3.11. We also thank the *Allianz für Hochleistungsrechnen Rheinland-Pfalz* for granting us access to the High Performance Computer Elwetritsch, on which the simulations underlying our figures have been performed.

Chapter 3

Depth of most recent common ancestors in the Hammond-Sheffield urn: Proofs and heuristics

We start this chapter with the proofs on the pairwise MRCA-depth results in the regime $\alpha \in (\frac{1}{2}, 1)$ and then follow and continue the approaches outlined in Section S.1.5 about the MRCA-depth of $[n]$ in this regime. An interdependence graph of the proofs and related heuristics is given in Figure 3.1 at the beginning of Section 3.2.

The fact that the HS-trees do not contain a spine was stated in Proposition S.1.5 and proved in Section S.1.4. At the end of this chapter we include a second proof of Proposition S.1.5, which is based on a martingale argument.

3.1 Proofs for the pairwise MRCA results

We will show the assertions of Theorem S.1.11 and S.1.16 by computing the Laplace transform of $\left|\mathrm{MRCA}(0,n)\right|/n$.
In this section we will often use the notation

$$X_n := \left|\mathrm{MRCA}(0,n)\right|.$$

Recall (S.1.10) and set

$$C_q := \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}. \tag{3.1.1}$$

Then we have $q_\ell \sim C_q \cdot \ell^{\alpha-1}$ for $\ell \to \infty$.

3.1.1 Laplace transform of $\left|\mathrm{MRCA}(0,n)\right|$

Lemma 3.1.1. *The Laplace transform f_n of $\left|\mathrm{MRCA}(0,n)\right|$ is given by*

$$f_n(\lambda) = \frac{\sum_{k\geq 0} q_k q_{k+n} e^{-\lambda}}{\sum_{k\geq 0} q_k^2 e^{-\lambda}}.$$

Proof. Let $T \overset{(d)}{=} \text{Exp}(\lambda)$, then we have that for independent (and also independent from T) ancestral lineages A_0 and A_n starting in 0 respectively n

$$\begin{aligned}\mathbf{E}\left[e^{-\lambda X_n}\right] &= \mathbf{E}\left[\mathbf{P}\left(T > X_n | X_n\right)\right] \\ &= \mathbf{E}\left[\mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset | X_n\right)\right] \\ &= \mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset\right)\end{aligned}$$

and

$$\begin{aligned}&\mathbf{E}\left[\left|A_0 \cap A_n \cap [-T, 0]\right|\right] \\ =\ & \sum_{k \geq 0} q_k q_{k+n} \mathbf{P}\left(T \geq k\right) \\ \sim\ & C_q^2 \sum_{k \geq 0} k^{\alpha-1}(k+n)^{\alpha-1} e^{-\lambda k}.\end{aligned}$$

As well as for two independent (and also independent from T) ancestral lineages A_0 and $\tilde{A}_0$ starting in 0

$$\begin{aligned}&\mathbf{E}\left[\left|A_0 \cap A_n \cap [-T, 0]\right|\right] \\ =\ & \mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset\right) \mathbf{E}\left[\left|A_0 \cap \tilde{A}_0 \cap [-T, 0]\right|\right] \\ =\ & \mathbf{P}\left(A_0 \cap A_n \cap [-T, 0] \neq \emptyset\right) \sum_{k \geq 0} q_k^2 e^{-\lambda k}.\end{aligned}$$

In total this gives

$$\mathbf{E}\left[e^{-\lambda X_n}\right] = \frac{\sum_{k \geq 0} q_k q_{k+n} e^{-\lambda k}}{\sum_{k \geq 0} q_k^2 e^{-\lambda k}}.$$

□

3.1.2 Asymptotic distribution of $\left|\text{MRCA}(0, n)\right|/n$: Proof of Theorem S.1.11

Lemma 3.1.2. *As $n \to \infty$, the Laplace transform of $X_n := \left|\text{MRCA}(0, n)\right|$ satisfies*

$$\mathbf{E}\left[e^{-\frac{\theta}{n} X_n}\right] \sim \frac{\theta^{2\alpha-1}}{\Gamma(2\alpha-1)} \int_0^\infty \left(x + x^2\right)^{\alpha-1} e^{-\theta x}\,\mathrm{d}x =: g(\theta). \tag{3.1.2}$$

Proof. With Lemma 3.1.1 we obtain

$$\mathbf{E}\left[e^{-\frac{\theta}{n} X_n}\right] = \frac{\sum_{k \geq 0} q_k q_{k+n} e^{-\frac{\theta}{n} k}}{\sum_{k \geq 0} q_k^2 e^{-\frac{\theta}{n} k}}$$

$$\begin{aligned} &\sim \frac{\sum_{k\geq 0} k^{\alpha-1}(k+n)^{\alpha-1}e^{-\frac{\theta}{n}k}}{\sum_{k\geq 0} k^{2\alpha-2}e^{-\frac{\theta}{n}k}} \\ &\sim \frac{\int_0^\infty x^{\alpha-1}(1+x)^{\alpha-1}e^{-\theta x}\,\mathrm{d}x}{\int_0^\infty x^{2\alpha-2}e^{-\theta x}\,\mathrm{d}x} \end{aligned} \tag{3.1.3}$$

as $n\to\infty$. The substitution $y=\theta x$ gives

$$\begin{aligned} \int_0^\infty x^{2\alpha-2}e^{-\theta x}\,\mathrm{d}x &= \int_0^\infty e^{-y}\left(\frac{y}{\theta}\right)^{2\alpha-2}\frac{1}{\theta}\,\mathrm{d}y \\ &= \theta^{1-2\alpha}\Gamma(2\alpha-1), \end{aligned}$$

such that (3.1.3) is equal to

$$\frac{\theta^{2\alpha-1}}{\Gamma(2\alpha-1)}\int_0^\infty x^{\alpha-1}(1+x)^{\alpha-1}e^{-\theta x}\,\mathrm{d}x = \frac{\theta^{2\alpha-1}}{\Gamma(2\alpha-1)}\int_0^\infty \left(x+x^2\right)^{\alpha-1}e^{-\theta x}\,\mathrm{d}x.$$

□

Lemma 3.1.3. *For the function g defined by (3.1.2) we have*

$$g(\theta)=\mathcal{L}\left(\mathrm{Beta}'\left(1-\alpha,2\alpha-1\right)\right)(\theta),\qquad \theta\geq 0.$$

Proof. We start with an observation on the modified Bessel function of the second kind. By [DLMF, (10.32.8)] we can write the modified Bessel function of the second kind as

$$\begin{aligned} K_\nu(z) &= \frac{\sqrt{\pi}\left(\frac{1}{2}z\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)}\int_1^\infty e^{-zt}\left(t^2-1\right)^{\nu-\frac{1}{2}}\,\mathrm{d}t \\ &= \frac{\sqrt{\pi}\left(\frac{1}{2}z\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)}\int_0^\infty e^{-z}e^{-zt}\left((t+1)^2-1\right)^{\nu-\frac{1}{2}}\,\mathrm{d}t. \end{aligned}$$

By the substitution $s=\frac{t}{2}$ this is equal to

$$\begin{aligned} &\frac{\sqrt{\pi}\left(\frac{1}{2}z\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)}e^{-z}\int_0^\infty e^{-2zs}\left((2s+1)^2-1\right)^{\nu-\frac{1}{2}}2\,\mathrm{d}s \\ =\ &\frac{2\sqrt{\pi}\left(\frac{1}{2}z\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)}e^{-z}\int_0^\infty e^{-2zs}\left(4s^2+2\cdot 2s+1-1\right)^{\nu-\frac{1}{2}}\,\mathrm{d}s \\ =\ &\frac{2\sqrt{\pi}\left(\frac{1}{2}z\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)}e^{-z}\int_0^\infty e^{-2zs}\left(4s^2+4s\right)^{\nu-\frac{1}{2}}\,\mathrm{d}s \\ =\ &\frac{2\cdot 4^{\nu-\frac{1}{2}}\sqrt{\pi}\left(\frac{1}{2}z\right)^\nu}{\Gamma\left(\nu+\frac{1}{2}\right)}e^{-z}\int_0^\infty e^{-2zs}\left(s^2+s\right)^{\nu-\frac{1}{2}}\,\mathrm{d}s. \end{aligned}$$

Setting

$$\nu-\frac{1}{2}=\alpha-1\Leftrightarrow \nu=\alpha-\frac{1}{2},\qquad z=\frac{\theta}{2}$$

Lemma 3.1.2 gives

$$\begin{aligned} g(\theta) &= \frac{\theta^{2\alpha-1}}{\Gamma(2\alpha-1)} \frac{\Gamma(\alpha)}{2\cdot 4^{\alpha-\frac{1}{2}-\frac{1}{2}}\sqrt{\pi}\left(\frac{1}{2}\cdot\frac{1}{2}\theta\right)^{\alpha-\frac{1}{2}}} e^{\frac{1}{2}\theta} K_{\alpha-\frac{1}{2}}\left(\frac{\theta}{2}\right) \\ &= \frac{\Gamma(\alpha)}{\Gamma(2\alpha-1)\sqrt{\pi}} \theta^{\alpha-\frac{1}{2}} e^{\frac{1}{2}\theta} K_{\alpha-\frac{1}{2}}\left(\frac{\theta}{2}\right). \end{aligned}$$

By $K_\nu(z) = K_{-\nu}(z)$, see [DLMF, (10.27.3)], we then obtain

$$\begin{aligned} & g(\theta) \\ = \ & \frac{\Gamma(\alpha)}{\Gamma(2\alpha-1)\sqrt{\pi}} \theta^{\alpha-\frac{1}{2}} e^{\frac{1}{2}\theta} K_{\frac{1}{2}-\alpha}\left(\frac{\theta}{2}\right) \\ = \ & \frac{\Gamma(\alpha)}{\Gamma(2\alpha-1)\sqrt{\pi}} \theta^{\alpha-\frac{1}{2}} e^{\frac{1}{2}\theta} \frac{2\cdot 4^{\frac{1}{2}-\alpha-\frac{1}{2}}\sqrt{\pi}\left(\frac{1}{2}\cdot\frac{1}{2}\theta\right)^{\frac{1}{2}-\alpha}}{\Gamma\left(\frac{1}{2}-\alpha+\frac{1}{2}\right)} e^{-\frac{1}{2}\theta} \cdot \int_0^\infty e^{-\theta s}\left(s^2+s\right)^{\frac{1}{2}-\alpha-\frac{1}{2}} \mathrm{d}s \\ = \ & \frac{\Gamma(\alpha)}{\Gamma(2\alpha-1)\sqrt{\pi}} e^{\frac{1}{2}\theta} \frac{4^{\frac{1}{2}-\alpha}\sqrt{\pi}\left(\frac{1}{4}\right)^{\frac{1}{2}-\alpha}}{\Gamma\left(\frac{1}{2}-\alpha+\frac{1}{2}\right)} e^{-\frac{\theta}{2}} \cdot \int_0^\infty e^{-\theta s}\left(s^2+s\right)^{-\alpha} \mathrm{d}s \\ = \ & \frac{\Gamma(\alpha)}{\Gamma(2\alpha-1)\Gamma(1-\alpha)} \cdot \int_0^\infty e^{-\theta s}\left(s^2+s\right)^{-\alpha} \mathrm{d}s. \end{aligned}$$

Which by

$$\frac{1}{\mathrm{Beta}(1-\alpha, 2\alpha-1)} = \frac{\Gamma(1-\alpha+2\alpha-1)}{\Gamma(1-\alpha)\Gamma(2\alpha-1)} = \frac{\Gamma(\alpha)}{\Gamma(1-\alpha)\Gamma(2\alpha-1)}$$

corresponds to the Laplace transform of the Beta′-distribution with parameters $(1-\alpha, 2\alpha-1)$. □

Proof of Theorem S.1.11. The statement follows by the continuity of Laplace transforms and from the previous two lemmata. □

3.1.3 The MRCA-depth of two nearby individuals: Proof of Proposition S.1.14

To shorten notation from now on denote the Laplace transform of $X_n := \left|\mathrm{MRCA}(0,n)\right|$ by f_n. By standard arguments, see for example [Fel71, Chapter XIII.5, Theorem 2] or [BGT87, Theorem 1.7.1], Proposition S.1.14 is a direct consequence of the following Lemma.

Lemma 3.1.4. *Let $n \in \mathbb{N}$. Then*

$$\lim_{\theta\to 0} \frac{1}{\theta^{2\alpha-1}}\left(1 - f_n\left(\frac{\theta}{n}\right)\right) = C(n,\alpha) \tag{3.1.4}$$

for a constant $C(n,\alpha) > 0$.

Proof. Since

$$q_k \sim C_q k^{\alpha-1} \text{ as } k \to \infty$$

we get

$$1 - f_n\left(\frac{\theta}{n}\right) \sim \frac{\sum_{\ell\geq 0} q_\ell \left(q_\ell - q_{\ell+n}\right) e^{-\frac{\theta}{n}\ell}}{C_q^2 \sum_{\ell\geq 0} \ell^{2\alpha-2} e^{-\frac{\theta}{n}\ell}} \tag{3.1.5}$$

as $\theta \to 0$. Note that for $C_2(\alpha) = \Gamma(2\alpha - 1)$ we have

$$\sum_{\ell\geq 0} \ell^{2\alpha-2} e^{-\frac{\theta}{n}\ell} \sim C_2(\alpha) \left(\frac{n}{\theta}\right)^{2\alpha-1} \tag{3.1.6}$$

as $\theta \to 0$. Now by monotone convergence the right hand side of (3.1.5) fulfills

$$\begin{aligned}
\lim_{\theta\to 0} \frac{\sum_{\ell\geq 0} q_\ell \left(q_\ell - q_{\ell+n}\right) e^{-\frac{\theta}{n}\ell}}{C_q^2 \theta^{2\alpha-1} C_2(\alpha) \left(\frac{n}{\theta}\right)^{2\alpha-1}} &= \sum_{\ell\geq 0} \lim_{\theta\to 0} \frac{\ell^{\alpha-1}\left(\ell^{\alpha-1} - (\ell+n)^{\alpha-1}\right) e^{-\frac{\theta}{n}\ell}}{C_q^2 \theta^{2\alpha-1} C_2(\alpha) \left(\frac{n}{\theta}\right)^{2\alpha-1}} \\
&= \sum_{\ell\geq 0} \lim_{\theta\to 0} \frac{q_\ell \left(q_\ell - q_{\ell+n}\right) e^{-\frac{\theta}{n}\ell}}{C_q^2 C_2(\alpha) (n)^{2\alpha-1}} \\
&= \sum_{\ell\geq 0} \frac{q_\ell \left(q_\ell - q_{\ell+n}\right) \lim_{\theta\to 0} e^{-\frac{\theta}{n}\ell}}{C_q^2 C_2(\alpha) (n)^{2\alpha-1}} \\
&= \sum_{\ell\geq 0} \frac{q_\ell \left(q_\ell - q_{\ell+n}\right)}{C_q^2 C_2(\alpha) (n)^{2\alpha-1}} \\
&= \frac{1}{C_q^2 C_2(\alpha) (n)^{2\alpha-1}} \sum_{\ell\geq 0} q_\ell \left(q_\ell - q_{\ell+n}\right) \\
&= C_q^{-2} C_2(\alpha)^{-1} n^{1-2\alpha} \sum_{\ell\geq 0} q_\ell \left(q_\ell - q_{\ell+n}\right) \\
&= C(n, \alpha)
\end{aligned}$$

for

$$C(n, \alpha) := C_q^{-2} C_2(\alpha)^{-1} n^{1-2\alpha} \sum_{\ell\geq 0} q_\ell \left(q_\ell - q_{\ell+n}\right). \tag{3.1.7}$$

This proves (3.1.4). □

3.1.4 A moderate deviations result: Proof of Proposition S.1.16

The proof of Proposition S.1.16 will require a more detailed analysis of the Laplace transform f_n of $\left|\mathrm{MRCA}(0, n)\right|$.

Lemma 3.1.5. *For every sequence $(\theta_n)_{n\in\mathbb{N}}$ converging to zero such that*

$$\frac{1}{C_q^2} \sum_{\ell\geq 0} q_\ell \left(q_\ell - q_{\ell+n}\right) e^{-\frac{\theta_n \ell}{n}} \sim n^{2\alpha-1} \int_0^\infty x^{\alpha-1} \left(x^{\alpha-1} - (x+1)^{\alpha-1}\right) \mathrm{d}x$$

we have

$$\begin{aligned} C(\alpha) &:= \lim_{\theta\to 0}\lim_{n\to\infty}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right) \\ &= \lim_{n\to\infty}\lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right) \\ &= \lim_{n\to\infty}\frac{1}{\theta_n^{2\alpha-1}}\left(1-f_n\left(\frac{\theta_n}{n}\right)\right) \\ &= \frac{\Gamma(\alpha)\Gamma(3-2\alpha)}{(2\alpha-1)\Gamma(1-\alpha)\Gamma(2\alpha-1)} \end{aligned} \tag{3.1.8}$$

Proof. We already showed that the Laplace transform of $\left|\mathrm{MRCA}(0,n)\right|$ is given by

$$f_n(\lambda)=\frac{\sum_{\ell\geq 0}q_\ell q_{\ell+n}e^{-\lambda\ell}}{\sum_{\ell\geq 0}q_\ell^2 e^{-\lambda\ell}},$$

see Lemma 3.1.1. Since this converges to the Laplace transform of a Beta′-distribution we get that for $\lambda=\frac{\theta}{n}$ and a constant $C(\alpha)$

$$\lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\lim_{n\to\infty}\left(1-f_n\left(\frac{\theta}{n}\right)\right)=C(\alpha), \tag{3.1.9}$$

see Theorem S.1.11. Since the Beta'-distribution with parameters $(1-\alpha,2\alpha-1)$ has tails asymptotically as $y\to\infty$ equal to

$$\frac{y^{1-2\alpha}}{(2\alpha-1)\cdot\mathrm{Beta}(1-\alpha,2\alpha-1)}=\frac{y^{1-2\alpha}}{(2\alpha-1)}\cdot\frac{\Gamma(\alpha)}{\Gamma(1-\alpha)\cdot\Gamma(2\alpha-1)} \tag{3.1.10}$$

[Fel71, Chapter XIII.5, Theorem 2] or [BGT87, Theorem 1.7.1] allow to specify the constant as

$$C(\alpha)=\frac{\Gamma(3-2\alpha)}{(2\alpha-1)}\cdot\frac{\Gamma(\alpha)}{\Gamma(1-\alpha)\cdot\Gamma(2\alpha-1)}. \tag{3.1.11}$$

First note that by (3.1.4) we get that for $n\in\mathbb{N}$ and $C(n,\alpha)$ defined by (3.1.7) we have

$$\lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right)=C(n,\alpha). \tag{3.1.12}$$

We start by showing that for a constant $C_1(\alpha)$

$$\lim_{n\to\infty}\lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right)=C_1(\alpha). \tag{3.1.13}$$

First note that by definition of $C_2(\alpha)$ and $C(n,\alpha)$, see (3.1.7) and (3.1.6), for $K>0$ we have

$$C_2(\alpha)C(n,\alpha) \quad\sim\quad n^{1-2\alpha}\sum_{\ell\geq 0}\ell^{\alpha-1}\left(\ell^{\alpha-1}-(\ell+n)^{\alpha-1}\right)$$

$$
\begin{aligned}
&\sim \quad n^{1-2\alpha} n^{2\alpha-1} \frac{1}{n} \sum_{\ell \geq 0} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) \\
&\sim \quad \frac{1}{n} \sum_{\ell \geq 0} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) \\
&= \quad \frac{1}{n} \sum_{\ell = 0}^{Kn} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) \qquad (3.1.14) \\
&\quad + \frac{1}{n} \sum_{\ell = Kn+1}^{\infty} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) \qquad (3.1.15)
\end{aligned}
$$

as $n \to \infty$. We start with analysing the second sum, see (3.1.15). We have

$$
\begin{aligned}
&\lim_{K\to\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{\ell = Kn+1}^{\infty} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) \\
= &\lim_{K\to\infty} \int_K^\infty x^{\alpha-1} \left(x^{\alpha-1} - (x+1)^{\alpha-1} \right) \mathrm{d}x.
\end{aligned}
$$

Since

$$x^{\alpha-1} - (x+1)^{\alpha-1} \geq 0$$

and

$$x^{\alpha-1} - (x+1)^{\alpha-1} \leq x^{\alpha-2}$$

for $x > 0$, we get

$$\int_K^\infty x^{\alpha-1} \left(x^{\alpha-1} - (x+1)^{\alpha-1} \right) \mathrm{d}x \leq \int_K^\infty x^{2\alpha-3} \,\mathrm{d}x \leq 3K^{2\alpha-2},$$

which converges to zero as $K \to \infty$. So in total

$$\lim_{K\to\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{\ell = Kn+1}^{\infty} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) = 0.$$

We now look at the first sum, see (3.1.14),

$$
\begin{aligned}
&\lim_{K\to\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{\ell = 0}^{Kn} \left(\frac{\ell}{n}\right)^{\alpha-1} \left(\left(\frac{\ell}{n}\right)^{\alpha-1} - \left(\frac{\ell}{n} - 1\right)^{\alpha-1} \right) \\
= &\lim_{K\to\infty} \int_0^K x^{\alpha-1} \left(x^{\alpha-1} - (x+1)^{\alpha-1} \right) \mathrm{d}x \\
=: \;& C_3(\alpha) < \infty.
\end{aligned}
$$

By (3.1.7) we have the asymptotics

$$C(n,\alpha) \sim C_2(\alpha)^{-1} n^{1-2\alpha} \sum_{\ell \geq 0} \ell^{\alpha-1} \left(\ell^{\alpha-1} - (\ell+n)^{\alpha-1} \right)$$

as $n \to \infty$, which then implies (3.1.13).
Now

$$\begin{aligned}
& \lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\lim_{n\to\infty}\left(1-f_n\left(\frac{\theta}{n}\right)\right) \\
= & \lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\lim_{n\to\infty}\frac{\sum_{k\geq 0} q_k\left(q_k-q_{k+n}\right)e^{-\frac{\theta}{n}k}}{\sum_{k\geq 0} q_k^2 e^{-\frac{\theta}{n}k}} \\
= & \lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\frac{\theta^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(1+x)^{\alpha-1}\right)e^{-\theta x}\,\mathrm{d}x}{\int_0^\infty x^{2\alpha-2}e^{-\theta x}\,\mathrm{d}x} \\
= & \lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\frac{\theta^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(1+x)^{\alpha-1}\right)e^{-\theta x}\,\mathrm{d}x}{\Gamma(2\alpha-1)} \\
= & \frac{\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(1+x)^{\alpha-1}\right)\mathrm{d}x}{\Gamma(2\alpha-1)},
\end{aligned}$$

which by $C_2(\alpha)=\Gamma(2\alpha-1)$, see (3.1.6), is equal to $C_1(\alpha)$. This implies

$$\lim_{\theta\to 0}\lim_{n\to\infty}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right)=\lim_{n\to\infty}\lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right). \tag{3.1.16}$$

Now note that by (1.4.8) $C_1(\alpha)$ is equal to $C(\alpha)$. So for θ_n fulfilling

$$\frac{1}{C_q^2}\sum_{\ell\geq 0} q_\ell\left(q_\ell-q_{\ell+n}\right)e^{-\frac{\theta_n}{n}\ell}\sim n^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(x+1)^{\alpha-1}\right)\mathrm{d}x,$$

(3.1.16) and (3.1.9) together imply

$$\begin{aligned}
C(\alpha) &= \lim_{\theta\to 0}\lim_{n\to\infty}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right) \\
&= \lim_{n\to\infty}\lim_{\theta\to 0}\frac{1}{\theta^{2\alpha-1}}\left(1-f_n\left(\frac{\theta}{n}\right)\right) \\
&= \lim_{n\to\infty}\frac{1}{\theta_n^{2\alpha-1}}\left(1-f_n\left(\frac{\theta_n}{n}\right)\right).
\end{aligned}$$

□

Lemma 3.1.6. *For* $X_n=\left|\mathrm{MRCA}(0,n)\right|$ *let*

$$m_n(x):=\mathbf{P}\left(\frac{X_n}{n}>x\right).$$

For every sequence $(\theta_n)_{n\in\mathbb{N}}$ *converging to zero, such that*

$$\frac{1}{C_q^2}\sum_{\ell\geq 0} q_\ell\left(q_\ell-q_{\ell+n}\right)e^{-\frac{\theta_n\ell}{n}}\sim n^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(x+1)^{\alpha-1}\right)\mathrm{d}x,$$

we have

$$\int_0^{\frac{1}{\theta_n}} m_n(z)\,\mathrm{d}z \sim \frac{C(\alpha)}{\Gamma(3-2\alpha)}\theta_n^{2\alpha-2}.$$

Proof. Remember that the random variables X_n have Laplace transforms f_n and that for $X \overset{(d)}{=} \mathrm{Beta}'(1-\alpha, 2\alpha-1)$ we have the convergence $\frac{X_n}{n} \Rightarrow X$. We have already shown

$$C(\alpha) = \lim_{n\to\infty}\lim_{\theta\to 0}\frac{1-f_n\left(\frac{\theta}{n}\right)}{\theta^{2\alpha-1}} = \lim_{\theta\to 0}\lim_{n\to\infty}\frac{1-f_n\left(\frac{\theta}{n}\right)}{\theta^{2\alpha-1}} = \lim_{n\to\infty}\frac{1-f_n\left(\frac{\theta_n}{n}\right)}{\theta_n^{2\alpha-1}} \tag{3.1.17}$$

for $\theta_n = n^{-\delta}$. Note that

$$\tilde{f}_n(\theta) := f_n\left(\frac{\theta}{n}\right) \tag{3.1.18}$$

is the Laplace transform of $\frac{X_n}{n}$. By definition we have

$$\begin{aligned}
\mathcal{L}(m_n)(s) &= \int_0^\infty e^{-xs}\int_x^\infty \tilde{f}_n(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_0^\infty\int_0^\infty e^{-xs}\tilde{f}_n(y)\mathbf{1}_{y\geq x}\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_0^\infty\int_0^\infty e^{-xs}\tilde{f}_n f(y)\mathbf{1}_{y\geq x}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^\infty\int_0^y e^{-xs}\tilde{f}_n(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^\infty \tilde{f}_n(y)\int_0^y e^{-xs}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^\infty \tilde{f}_n(y)\left[-\frac{1}{s}e^{-xs}\right]_{x=0}^{x=y}\mathrm{d}y \\
&= \int_0^\infty \tilde{f}_n(y)\frac{1}{s}\left[1-e^{-ys}\right]\mathrm{d}y \\
&= \frac{1}{s}-\frac{1}{s}\int_0^\infty e^{-ys}\tilde{f}_n(s)\,\mathrm{d}y \\
&= \frac{1}{s}-\frac{1}{s}\mathcal{L}(\tilde{f}_n)(s).
\end{aligned}$$

So we obtain

$$\mathcal{L}(m_n)(\theta) = \frac{1}{\theta}-\frac{\tilde{f}_n(\theta)}{\theta}, \tag{3.1.19}$$

which by (3.1.18) and (3.1.17) gives

$$\mathcal{L}(m_n)(\theta_n) \sim C(\alpha)\theta_n^{2\alpha-2}$$

as $n\to\infty$. We will use this to compute the asymptotics of

$$\int_0^{\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x$$

as $n \to \infty$. For this purpose set

$$h_n(\theta) := \mathcal{L}(m_n)(\theta),$$

then

$$h_n(\theta_n) = \mathcal{L}(m_n)(\theta_n) \sim C(\alpha)\theta_n^{2\alpha-2}$$

as $n \to \infty$, which obviously gives

$$h_n(x\theta_n) = \mathcal{L}(m_n)(x\theta_n) \sim C(\alpha)(x\theta_n)^{2\alpha-2},$$

so

$$\frac{h_n(x\theta_n)}{\theta_n^{2\alpha-2}} \to C(\alpha)x^{2\alpha-2} \qquad \text{as } n \to \infty. \tag{3.1.20}$$

Now note that the right hand side of (3.1.20) is the Laplace transform of the measure

$$\nu([0,y]) = \frac{C(\alpha)}{\Gamma(3-2\alpha)} y^{2-2\alpha},$$

see [BGT87, Proof of Theorem 1.7.1]. While for the left hand side we have

$$\frac{h_n(\theta_n x)}{\theta_n^{2\alpha-2}} = \int_0^\infty e^{-x\theta_n z} \frac{m_n(z)}{\theta_n^{2\alpha-2}} \,\mathrm{d}z = \int_0^\infty e^{-xy} \frac{m_n\left(\frac{y}{\theta_n}\right)}{\theta_n^{2\alpha-2}} \cdot \frac{1}{\theta_n} \,\mathrm{d}y.$$

So in total the left hand side of (3.1.20) is the Laplace transform of

$$\zeta_n(y) := \frac{m_n\left(\frac{y}{\theta_n}\right)}{\theta_n^{2\alpha-2}} \cdot \frac{1}{\theta_n}.$$

By the continuity theorem for Laplace transforms we obtain

$$\int_0^y \zeta_n(x) \,\mathrm{d}x \to \frac{C(\alpha)}{\Gamma(3-2\alpha)} y^{2-2\alpha}$$

as $n \to \infty$. So by substituting $\mathrm{d}x = \theta_n \,\mathrm{d}z$

$$\int_0^y \zeta_n(x) \,\mathrm{d}x = \int_0^y \frac{m_n\left(\frac{x}{\theta_n}\right)}{\theta_n^{2\alpha-2}} \frac{1}{\theta_n} \,\mathrm{d}x = \int_0^{\frac{y}{\theta_n}} \frac{m_n(z)}{\theta_n^{2\alpha-2}} \,\mathrm{d}z$$

we get for $y = 1$

$$\int_0^{\frac{1}{\theta_n}} m_n(z) \,\mathrm{d}z \sim \frac{C(\alpha)}{\Gamma(3-2\alpha)} \theta_n^{2\alpha-2}$$

as $n \to \infty$. □

Lemma 3.1.7. *Let $(\theta_n)_{n\in\mathbb{N}}$ be a sequence converging to zero, such that*

$$\frac{1}{C_q^2}\sum_{\ell\geq 0} q_\ell\left(q_\ell - q_{\ell+n}\right)e^{-\frac{\theta_n \ell}{n}} \sim n^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1}-(x+1)^{\alpha-1}\right)\mathrm{d}x.$$

Then for $m_n(x) := \mathbf{P}\left(\frac{X_n}{n} > x\right)$ and $\tilde{C}(\alpha) := \frac{C(\alpha)}{\Gamma(3-2\alpha)}$ we have

$$m_n\left(x\frac{1}{\theta_n}\right) \sim \tilde{C}(\alpha)(2-2\alpha)x^{1-2\alpha}\theta_n^{2\alpha-1},$$

as $n\to\infty$.

Proof. Set

$$C(n,\theta,a,\alpha) := \frac{\int_0^{a\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x}{a^{2-2\alpha}\theta_n^{2\alpha-2}}.$$

Then by Lemma 3.1.6

$$\lim_{n\to\infty} C(n,\theta,a,\alpha) = \frac{C(\alpha)}{\Gamma\left(3-2\alpha\right)} = \tilde{C}(\alpha), \tag{3.1.21}$$

which by (3.1.11) is equal to

$$\frac{\Gamma(2\alpha-1)\Gamma(\alpha)\Gamma(1-2\alpha)}{\Gamma(1-\alpha)\Gamma\left(3-2\alpha\right)}$$

as well as

$$\int_0^{a\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x = C(n,\theta,a,\alpha)a^{2-2\alpha}\theta_n^{2\alpha-2}.$$

So for $a<b$ we get

$$\begin{aligned}\frac{\int_{a\frac{1}{\theta_n}}^{b\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x}{\theta_n^{2\alpha-2}} &= C(n,\theta,b,\alpha)b^{2-2\alpha} - C(n,\theta,a,\alpha)a^{2-2\alpha} \\ &\overset{n\to\infty}{\to} \tilde{C}(\alpha)\left[b^{2-2\alpha}-a^{2-2\alpha}\right].\end{aligned} \tag{3.1.22}$$

Monotonicity of $m_n(x)$ in x then gives

$$(b-a)\frac{1}{\theta_n}m_n\left(a\frac{1}{\theta_n}\right) \leq \int_{a\frac{1}{\theta_n}}^{b\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x \leq (b-a)\frac{1}{\theta_n}m_n\left(b\frac{1}{\theta_n}\right).$$

So

$$\frac{1}{\theta_n}m_n\left(a\frac{1}{\theta_n}\right) \leq \frac{\int_{a\frac{1}{\theta_n}}^{b\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x}{b-a} \leq \frac{1}{\theta_n}m_n\left(b\frac{1}{\theta_n}\right).$$

Which by division through $\theta_n^{2\alpha-2}$ gives

$$\frac{1}{\theta_n^{2\alpha-1}}m_n\left(a\frac{1}{\theta_n}\right) \leq \frac{\int_{a\frac{1}{\theta_n}}^{b\frac{1}{\theta_n}} m_n(x)\,\mathrm{d}x}{(b-a)\,\theta_n^{2\alpha-2}} \leq \frac{1}{\theta_n^{2\alpha-1}}m_n\left(b\frac{1}{\theta_n}\right).$$

Together with (3.1.22) we obtain

$$\frac{1}{\theta_n^{2\alpha-1}} m_n\left(a\frac{1}{\theta_n}\right) \leq \frac{C(n,\theta,b,\alpha)b^{2-2\alpha} - C(n,\theta,a,\alpha)a^{2-2\alpha}}{(b-a)} \leq \frac{1}{\theta_n^{2\alpha-1}} m_n\left(b\frac{1}{\theta_n}\right).$$

Which together with (3.1.21) gives

$$\limsup_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}} m_n\left(a\frac{1}{\theta_n}\right) \leq \tilde{C}(\alpha)\frac{b^{2-2\alpha} - a^{2-2\alpha}}{(b-a)} \leq \liminf_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}} m_n\left(b\frac{1}{\theta_n}\right).$$

Plugging in $b = a + h$ this gives

$$\limsup_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}} m_n\left(a\frac{1}{\theta_n}\right) \leq \tilde{C}(\alpha)\frac{(a+h)^{2-2\alpha} - a^{2-2\alpha}}{h}.$$

Set

$$\xi(x) := \xi^{2-2\alpha},$$

then

$$\xi'(x) = (2-2\alpha)x^{1-2\alpha},$$

so

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}} m_n\left(a\frac{1}{\theta_n}\right) &\leq \lim_{h\to 0} \tilde{C}(\alpha)\frac{(a+h)^{2-2\alpha} - a^{2-2\alpha}}{h} \\
&= \tilde{C}(\alpha)\xi'(a) \\
&= \tilde{C}(\alpha)(2-2\alpha)a^{1-2\alpha},
\end{aligned}$$

giving

$$\limsup_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}\tilde{C}(\alpha)(2-2\alpha)a^{1-2\alpha}} m_n\left(a\frac{1}{\theta_n}\right) \leq 1.$$

For $a = b - h$ we then obtain

$$\tilde{C}(\alpha)\frac{(b-h)^{2-2\alpha} - b^{2-2\alpha}}{-h} \leq \liminf_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}} m_n\left(b\frac{1}{\theta_n}\right)$$

so as $h \to 0$

$$\begin{aligned}
\lim_{h\to 0} \tilde{C}(\alpha)\frac{(b-h)^{2-2\alpha} - b^{2-2\alpha}}{-h} &= \tilde{C}(\alpha)\xi'(b) \\
&= \tilde{C}(\alpha)(2-2\alpha)b^{1-2\alpha} \\
&\leq \liminf_{n\to\infty} \frac{1}{\theta_n^{2\alpha-1}} m_n\left(b\frac{1}{\theta_n}\right),
\end{aligned}$$

giving

$$1 \leq \liminf_{n\to\infty} \frac{1}{\tilde{C}(\alpha)(2-2\alpha)b^{1-2\alpha}\theta_n^{2\alpha-1}} m_n\left(b\frac{1}{\theta_n}\right).$$

In total for all $x > 0$

$$\lim_{n\to\infty} \frac{1}{\tilde{C}(\alpha)(2-2\alpha)x^{1-2\alpha}\theta_n^{2\alpha-1}} m_n\left(x\frac{1}{\theta_n}\right) = 1,$$

so

$$m_n\left(x\frac{1}{\theta_n}\right) \sim \tilde{C}(\alpha)(2-2\alpha)x^{1-2\alpha}\theta_n^{2\alpha-1},$$

giving

$$\mathbf{P}\left(\frac{X_n}{n} > \frac{x}{\theta_n}\right) \sim \tilde{C}(\alpha)(2-2\alpha)x^{1-2\alpha}\theta_n^{2\alpha-1} = \frac{2-2\alpha}{(2\alpha-1)} \cdot \frac{\Gamma(\alpha)}{\Gamma(1-\alpha)\cdot\Gamma(2\alpha-1)} x^{1-2\alpha}\theta_n^{2\alpha-1}$$

□

Proof of Proposition S.1.16. The general statement of the Proposition is a direct consequence of the previous lemma. We now prove (S.1.33). The sequence $\theta_n := n^{-\delta}$ converges to zero and

$$\frac{1}{C_q^2}\sum_{\ell\geq 0} q_\ell\left(q_\ell - q_{\ell+n}\right)e^{-\frac{\theta_n \ell}{n}} \sim n^{2\alpha-1}\int_0^\infty x^{\alpha-1}\left(x^{\alpha-1} - (x+1)^{\alpha-1}\right)\mathrm{d}x. \tag{3.1.23}$$

So by using

$$q_\ell \sim C_q \ell^{\alpha-1}$$

we only have to check

$$\frac{1}{n^{2\alpha-1}}\sum_{\ell\geq 0}\ell^{\alpha-1}\left(\ell^{\alpha-1} - (\ell+n)^{\alpha-1}\right)\left(1-\exp\left(-\frac{\theta_n}{n}\ell\right)\right) \to 0. \tag{3.1.24}$$

For the end of the proof let C be varying from line to line. Then we have

$$\left(\ell^{\alpha-1} - (\ell+n)^{\alpha-1}\right) \leq C\ell^{\alpha-2}n$$

and

$$\left(1-\exp\left(-\frac{\theta_n}{n}\ell\right)\right) \leq n^{-\delta-1}\ell,$$

so

$$\frac{1}{n^{2\alpha-1}}\sum_{\ell\geq 0}\ell^{\alpha-1}\left(\ell^{\alpha-1} - (\ell+n)^{\alpha-1}\right)\left(1-\exp\left(-\frac{\theta_n}{n}\ell\right)\right) \leq n^{-\delta-2\alpha+1}C\sum_{\ell\geq 0}\ell^{\alpha-1}\ell^{\alpha-2}. \tag{3.1.25}$$

Since

$$\sum_{\ell\geq 0}\ell^{\alpha-1}\ell^{\alpha-2} \leq C,$$

the right hand side of (3.1.25) is in total bounded form above by $Cn^{-\delta+1-2\alpha}$, which converges to 0 as $n \to \infty$. This shows (3.1.24) and thus (S.1.33). □

3.2 Various routes to the scale of the most recent common ancestor of $[n]$

We now turn our focus back on the more complicated question discussed in Section S.1.5.2: How fast does a bulk of individuals $[n]$ find their most recent common ancestor?

We start with the proof of Proposition S.1.23 in Section 3.2.1 and formulate the conjecture that the assumption (S.1.42) made in Proposition S.1.23 is true. Evidence for this is given by simulations in Section 3.2.6. Evidence for the validity of (S.1.44), which implies the assumption of Proposition S.1.23, see Remark S.1.25, is given in Section 3.2.5 by the construction and rigorous analysis of a critical branching random walk to dominate the probability that an individual in the HS-tree belongs to the pruned ancestry of $\mathbb{N}$. Furthermore, we will relate the assertion of Proposition S.1.5 (which says that the HS-tree has no spine) with the conjecture that (S.1.44) is valid. The latter essentially says that the density of the ancestry of $\mathbb{N}$ not belonging to the ancestral lineage of 0 has a density decaying like $|x|^{-\frac{\alpha}{2}}$. For $\alpha > \frac{2}{3}$ this means that the density of the ancestry of $[n]$ is of the same order as the density of one typical ancestral lineage, which is $|x|^{\alpha-1}$. Thus this might be interpreted in the sense that for $\alpha > \frac{2}{3}$ the ancestry of $\mathbb{N}$ consists of kind of a backbone with trees growing out of it.

Proposition S.1.27 is then proved in Section 3.2.2.

In Section 3.2.3 we point out another approach towards a proof of the validity of Statement S.1.21. This approach is based on an assumption on the behaviour of many disjoint renewal chains. In Section 3.2.4 we point out the connection of this approach to the validity of (S.1.42). Figure 3.1 relates the proved implications and shows the connections of the various approaches towards the validity of Statement S.1.21.

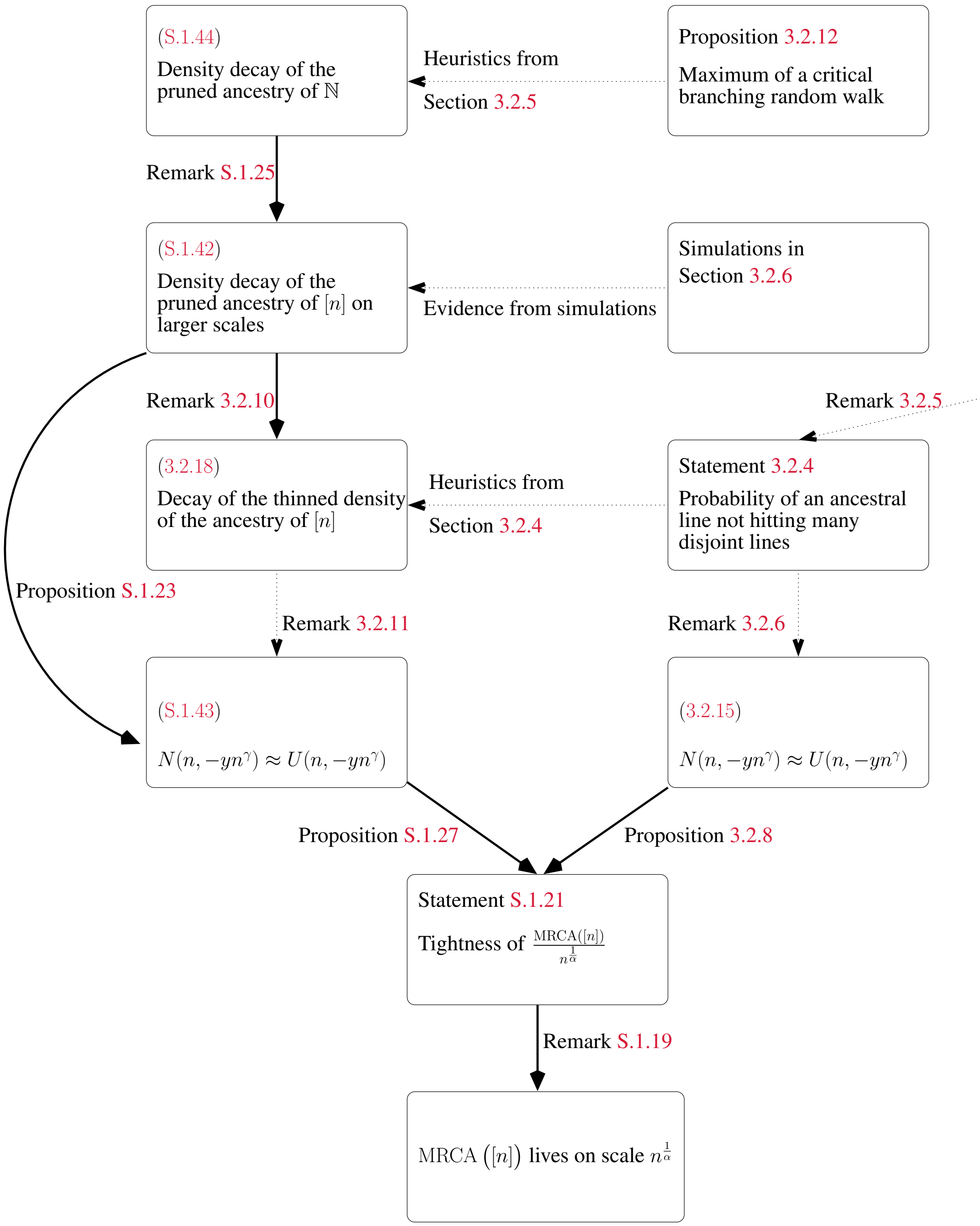


Figure 3.1: Solid arrows represent proved implications, dotted arrows mark evidence or conjectured implications.

3.2.1 Proof of Proposition S.1.23

Proof of Proposition S.1.23. Write J for the set of integers in $\{-(1-\varepsilon)a_n n, \dots, -1\}$ being a parent of an individual $j \in [n]$, that is

$$J := \Big\{ i \in \{-(1-\varepsilon)a_n n, \dots, -1\} : \exists j \in [n] \text{ such that } j - R_j = i \Big\}. \tag{3.2.1}$$

As already discussed in Remark S.1.26, the size of the set J concentrates around its expectation, which is computed in (S.1.46). We will bound

$$\mathbf{E}\left[\sum_{z \in J} \mathbf{1}_{A_z \cap A_{z'} \cap [-a_n n, 0] = \emptyset, \forall z' \geq z,\, z' \in J \cup \{0\}} \right] \tag{3.2.2}$$

from above in order to prove the assertion by Markov's inequality. We start with the conditional expectation

$$\mathbf{E}\left[\sum_{z \in J} \mathbf{1}_{A_z \cap A_{z'} \cap [-a_n n, 0] = \emptyset, \forall z' \geq z,\, z' \in J \cup \{0\}} \middle| J \right]. \tag{3.2.3}$$

We have

$$\begin{aligned}
& \{A_z \cap A_{z'} \cap [-a_n n, 0] = \emptyset, \forall z' \geq z,\, z' \in J \cup \{0\}\} \\
= \; & \Big[\{A_z \cap A_{z'} \cap [-a_n n, 0] = \emptyset, \forall z' \geq z,\, z' \in J \cup \{0\}\} \\
& \qquad \cap \Big\{ \# A_z \cap \{z, \dots, -a_n n\} > (a_n n)^{\alpha - \xi} \Big\} \Big] \\
& \cup \Big[\{A_z \cap A_{z'} \cap [-a_n n, 0] = \emptyset, \forall z' \geq z,\, z' \in J \cup \{0\}\} \\
& \qquad \cap \Big\{ \# A_z \cap \{z, \dots, -a_n n\} \leq (a_n n)^{\alpha - \xi} \Big\} \Big] \\
\subset \; & \Big[\Big\{ \# \{ i : -a_n n \leq i < 0, i \text{ ancestor of } [1,n], i \notin A_0 \} > (a_n n)^{\alpha - \xi} \Big\} \\
& \cup \Big\{ \# A_z \cap \{z, \dots, -a_n n\} \leq (a_n n)^{\alpha - \xi} \Big\} \Big] \\
=: \; & B_1 \cup B_2(z).
\end{aligned}$$

Note that the event B_1 does not depend on z or J at all. By (S.1.42) and Markov's inequality

$$\mathbf{P}(B_1) \leq \text{const} \cdot \frac{(a_n n)^{1 - \frac{\alpha}{2}}}{(a_n n)^{\alpha - \xi}} = (a_n n)^{1 - \frac{3}{2}\alpha + \xi}.$$

Note, that this bound is uniform in J. For B_2 and $\eta > 0$ we have

$$\mathbf{P}(B_2) \;\leq\; \mathbf{P}\left(\Big\{ \# A_{-(1-\varepsilon)a_n n} \cap \{-(1-\varepsilon)a_n n, \dots, -a_n n\} \leq (a_n n)^{\alpha - \xi} \Big\} \right)$$

$$\leq \quad \text{const} \cdot (a_n n)^{-\xi+\eta}.$$

Again, this is a bound uniform in z (by the definition of the set J) and thus in J.

Since $\mathbf{E}[\#J] \leq \text{const} \cdot n^{1-\alpha}$ and $\#J$ concentrates around its expectation we obtain

$$\frac{1}{n(a_n n)^{-\alpha}} \mathbf{E}\left[\sum_{z \in J} \mathbf{1}_{A_z \cap A_{z'} \cap [-a_n n, 0] = \emptyset, \forall z' \neq z,\, z' \in J \cup \{0\}}\right]$$
$$\leq \quad \text{const} \cdot \frac{n^{1-\alpha}}{n(a_n n)^{-\alpha}} \left[(a_n n)^{1-\frac{3}{2}\alpha+\xi} + (a_n n)^{-\xi+\eta}\right],$$

which by

$$\frac{n^{1-\alpha}}{n(a_n n)^{-\alpha}} (a_n n)^{1-\frac{3}{2}\alpha+\xi} = a_n^{1-\frac{\alpha}{2}+\xi} n^{1-\frac{3}{2}\alpha+\xi}$$

and

$$\frac{n^{1-\alpha}}{n(a_n n)^{-\alpha}} (a_n n)^{-\xi+\eta} = a_n^{\alpha-\xi+\eta} n^{-\xi+\eta}$$

gives the desired result since due to our assumptions on the sequence $(a_n)_n$ we have

$$\begin{aligned}
\frac{1}{1-\alpha} \cdot (a_n n)^{1-\alpha} \cdot \left[\left(\frac{1}{a_n}+1\right)^{1-\alpha} - 1\right] &= \frac{1}{1-\alpha} \cdot n^{1-\alpha} \cdot \left[(a_n+1)^{1-\alpha} - a_n^{1-\alpha}\right] \\
&\geq n^{1-\alpha} \cdot (a_n+1)^{-\alpha} \\
&\geq \text{const} \cdot n(a_n n)^{-\alpha}.
\end{aligned}$$

□

3.2.2 Proof of Proposition S.1.27

First we bound the scale of the most recent common ancestor of the individuals having their parent far away by using Proposition S.1.16, which gives good control over the depth of the most recent common ancestor of two individuals.

Lemma 3.2.1. *Let $\alpha \in (\frac{1}{2}, 1)$ and $(b_n)_n$ be a divergent non-decreasing sequence with $b_n = o\big(n^{1/\alpha-1}\big)$. The most recent common ancestor of the individuals $i \in [n]$ having a parent below $-n^{1/\alpha}/b_n$ is then a random variable Y_n with the property that*

$$\frac{Y_n}{n^{1/\alpha} b_n^{\frac{2}{2\alpha-1}}} \to 0 \text{ in probability as } n \to \infty. \tag{3.2.4}$$

Proof of Lemma 3.2.1. Because of our assumption on $(b_n)_n$ the number of individuals $i \in [n]$ having a parent below $-n^{1/\alpha}/b_n$ is divergent as $n \to \infty$. Since $n^{1/\alpha}/b_n \ll n^{1/\alpha}$ we can apply Proposition S.1.23 with $a_n = n^{\frac{1}{\alpha}-1}/b_n$ and get that with probability converging to one not more than of order b_n^α many lines are below $-n^{1/\alpha}/b_n$. Since now the pairwise distances of the fathers of the individuals having a parent below $-n^{1/\alpha}/b_n$ are of order $b_n n^{1/\alpha}$ (this is the order of the maximum of b_n^α many random variables with distribution μ conditioned on being larger than $n^{1/\alpha}$), we can apply Proposition S.1.16 and obtain that they find their MRCA on a scale not larger than the maximum of b_n^α many variables with tails with exponent $2\alpha - 1$ each conditioned on being larger than $b_n n^{\frac{1}{\alpha}}$, that is

$$b_n n^{\frac{1}{\alpha}} \left(b_n^\alpha\right)^{\frac{1}{2\alpha-1}} = n^{\frac{1}{\alpha}} b_n^{1+\frac{\alpha}{2\alpha-1}} \ll n^{\frac{1}{\alpha}} b_n^{\frac{2}{2\alpha-1}}.$$

Note that we only need to take b_n^α many random variables with tails with exponent $2\alpha - 1$ into account since for a set $a_1, \dots a_k \in \mathbb{Z}$

$$\mathrm{MRCA}\left(\{a_1, \dots, a_k\}\right) = \min_{\ell=1,\dots,k} \left\{\mathrm{MRCA}(a_1, a_\ell)\right\}. \tag{3.2.5}$$

□

Lemma 3.2.1 is key for bounding the scale of the most recent common ancestor of $[0, n]$ in a suitable way:

Proposition 3.2.2. *Let $\alpha \in \left(\frac{1}{8}\left(1 + \sqrt{33}\right), 1\right)$ and $\varepsilon > 0$. Then under the same assumptions of Proposition S.1.27 for all non-decreasing diverging sequences $(b_n)_n$*

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\left|\mathrm{MRCA}([0, n])\right|}{b_n n^{\frac{1}{\alpha}}} > \varepsilon\right) = 0.$$

A proof of Proposition 3.2.2 is given below. By the following elementary lemma we are then able to complete the proof of Proposition S.1.27.

Lemma 3.2.3. *Let $(Y_n)_n$ be a sequence of positive random variables. $(Y_n)_n$ is tight if and only if for all diverging positive sequences $(b_n)_n$*

$$\frac{Y_n}{b_n} \to 0 \text{ in probability as } n \to \infty. \tag{3.2.6}$$

Proof of Lemma 3.2.3. If $(Y_n)_n$ is tight, the statement follows by the definition of tightness. Assume that the sequence $(Y_n)_n$ is not tight. Then for all $\varepsilon > 0$ there exists a diverging non-decreasing sequence $(b_n)_n$ and an increasing sequence $(m_n)_n$, such that

$$\mathbf{P}\left(Y_{m_n} > b_n\right) > \varepsilon \qquad \text{for all } n \in \mathbb{N}. \tag{3.2.7}$$

This contradicts the assumption. □

Proposition S.1.27 is then a consequence of Proposition 3.2.2 and Lemma 3.2.3.
We are left with the proof of Proposition 3.2.2.

Proof of Proposition 3.2.2 . First we need to check that the right hand side of (S.1.43) in Proposition S.1.23 is actually decreasing in n for $a_n \ll n^{\frac{1}{\alpha}-1}$. We can focus on $a_n = n^{\delta+1}$, since then for $\gamma = \delta + 1$ the right hand side of (S.1.43) is monotone increasing in γ. So we just analyse $\gamma = 1/\alpha$. We want to choose $\xi > 0$ and $\eta > 0$, such that

$$-\alpha - \frac{1}{\alpha}\left(\frac{1}{2}\alpha - 1 - \xi\right) < 0 \tag{3.2.8}$$

as well as

$$1-\alpha-\frac{\xi}{\alpha}+\eta<0. \tag{3.2.9}$$

(3.2.8) gives the following condition on ξ.

$$\begin{aligned} & -\alpha-\frac{1}{\alpha}\left(\frac{1}{2}\alpha-1-\xi\right)<0 \\ \Leftrightarrow\quad & -\alpha<\frac{1}{\alpha}\left(\frac{1}{2}\alpha-1-\xi\right) \\ \Leftrightarrow\quad & \alpha^2>1+\xi-\frac{1}{2}\alpha \\ \Leftrightarrow\quad & \alpha^2+\frac{1}{2}\alpha-1>\xi. \\ \Leftrightarrow\quad & \left(\alpha+\frac{1}{4}\right)^2-\frac{17}{16}>\xi. \end{aligned}$$

Since the left hand side is >0 for

$$\alpha>\frac{1}{4}\left[\sqrt{17}-1\right] \tag{3.2.10}$$

we can indeed choose $\xi>0$, such that (3.2.8) is satisfied. Now (3.2.9) is monotone decreasing in ξ, so we plug in $\xi=\alpha^2+\frac{1}{2}\alpha-1<\alpha$ and obtain

$$\begin{aligned} & 1-\alpha-\frac{\xi}{\alpha}+\eta<0 \\ \Leftrightarrow\quad & 1-\alpha-\frac{1}{\alpha}\left(\alpha^2+\frac{1}{2}\alpha-1\right)+\eta<0 \\ \Leftrightarrow\quad & 1-\alpha-\left(\alpha+\frac{1}{2}-\frac{1}{\alpha}\right)+\eta<0 \\ \Leftrightarrow\quad & 1-\alpha-\alpha-\frac{1}{2}+\frac{1}{\alpha}+\eta<0. \end{aligned}$$

Since η is arbitrary small the last line is equivalent to

$$1-\alpha-\alpha-\frac{1}{2}+\frac{1}{\alpha}<0\Leftrightarrow\left(\alpha-\frac{1}{8}\right)^2-\frac{33}{64}>0\Leftrightarrow\alpha>\frac{1}{8}\left(1+\sqrt{33}\right). \tag{3.2.11}$$

So we can choose ξ and η such that the right hand side of (S.1.43) is decreasing in n for all $\gamma\leq\frac{1}{\alpha}$ if

$$\alpha>\max\left\{\frac{1}{8}\left(1+\sqrt{33}\right),\frac{1}{4}\left[\sqrt{17}-1\right]\right\}=\frac{1}{8}\left(1+\sqrt{33}\right). \tag{3.2.12}$$

Now let a_n be a divergent non-decreasing sequence. Lemma 3.2.1 then gives us, that the most recent common ancestor of the individuals $i\in[n]$ having a parent below $-n^{1/\alpha}/a_n$ is then a random variable of scale not larger than $n^{1/\alpha}a_n^{\frac{2}{2\alpha-1}}$. Since the only requirement on a_n was that it does not decrease, we can choose $a_n^{\frac{2}{2\alpha-1}}\ll b_n$. $\square$

3.2.3 A conjecture on collision probabilities among many renewal chains

We now point out another approach to a proof of Statement S.1.21. This one is based on a conjectured statement about the behaviour of *many* renewal chains.

Statement 3.2.4 (Probability of an ancestral line not hitting many disjoint lines)**.** *Let $\varepsilon > 0$, $0 < x < y < z$, $f(n) \ll n$ increasing and divergent , and $\gamma \geq 1$. Select randomly an individual V out of those in $[n]$ having their parent in the interval $[-y \cdot n^\gamma, -x \cdot n^\gamma]$. Conditional on the existence of a collection A of ancestral lineages of $[n]$ with $\#A = f(n)$, each having their parent in the interval $[-y \cdot n^\gamma, -x \cdot n^\gamma]$ and all being pairwise disjoint till $-z \cdot n^\gamma$, the probability that the ancestral lineage of V does not coalesce with A till $-z \cdot n^\gamma$ is at most of order $n^{\gamma(1-2\alpha)}$.*

Remark 3.2.5. Denote the ancestral lineage of V by L, and let $\mathcal{A}$ be the union of the ancestral lineages of the collection A. We consider the random set $M := \mathcal{A} \cap [-zn^\gamma, -xn^\gamma]$. The above statement then describes the "non-coalescence event", that L does not hit the set M. We claim that, roughly spoken, from the perspective of L the set M looks more or less like a set that is uniformly spread over $[-zn^\gamma, -xn^\gamma]$. Statement 3.2.4 can then be understood in the following way: An individual $i \in [n]$ with its parent on scale n^γ will at most have a probability of $n^{\gamma(1-2\alpha)}$ to get to the next scale without meeting another ancestral lineage from $[n]$ as long as there are still *many* ancestral lineage on this scale avoiding each other.
We now give a heuristic explanation, why Statement 3.2.4 should be true. Let $\gamma \geq 1$. Assume that $f(n) \gg 1$ many ancestral lineages do not meet over an interval J of length n^γ. Then the intersection of these $f(n)$ many ancestral lineages with J consists of order $f(n)n^{\gamma\alpha}$ many individuals. The relevant question is: what is the probability that another ancestral lineage passes this interval without coalescing with one of those $f(n)$ many? Under the assumption that these $f(n)$ lineages are sufficiently uniformly distributed on the interval J, the *one big jump principle* predicts that with of order n^s many jumps it has a probability of order

$$n^{s-s\alpha} \cdot \left(1 - \frac{f(n)n^{\gamma\alpha}}{n^\gamma}\right)^{n^s} \tag{3.2.13}$$

to pass the interval without hitting one of the other $f(n)$ many ancestral lineages in the interval J. For all divergent $f(n)$ (3.2.13) this is at most increasing in γ till it becomes exponentially small for $s \geq \gamma(1-\alpha)$. So this probability is at most of order $n^{(1-2\alpha)\gamma}$.

In Figure 3.2 we give evidence for the validity of Statement 3.2.4 for $\gamma = 1$.

The validity of Statement S.1.21 then suggests that of the of order $n^{1-\gamma\alpha}$ many individuals having their parents on scale n^γ only of order

$$n^{1-\gamma\alpha} \cdot n^{(1-2\alpha)\gamma} = n^{1+(1-3\alpha)\gamma} \tag{3.2.14}$$

will not coalesce with the $f(n)$ ancestral lineages till $n^{\gamma+\varepsilon}$.

Remark 3.2.6. We now relate the above with (S.1.38). The of order $n^{1+(1-3\alpha)\gamma}$ many individuals not coalescing with the $f(n)$ ancestral lineages till $n^{\gamma+\varepsilon}$ are much less than the $n^{1-(\gamma+\varepsilon)\alpha}$ many individuals of $[n]$ having their parents on scale $n^{\gamma+\varepsilon}$ if

$$1 + (1-3\alpha)\gamma \leq 1 - (\gamma+\varepsilon)\alpha$$

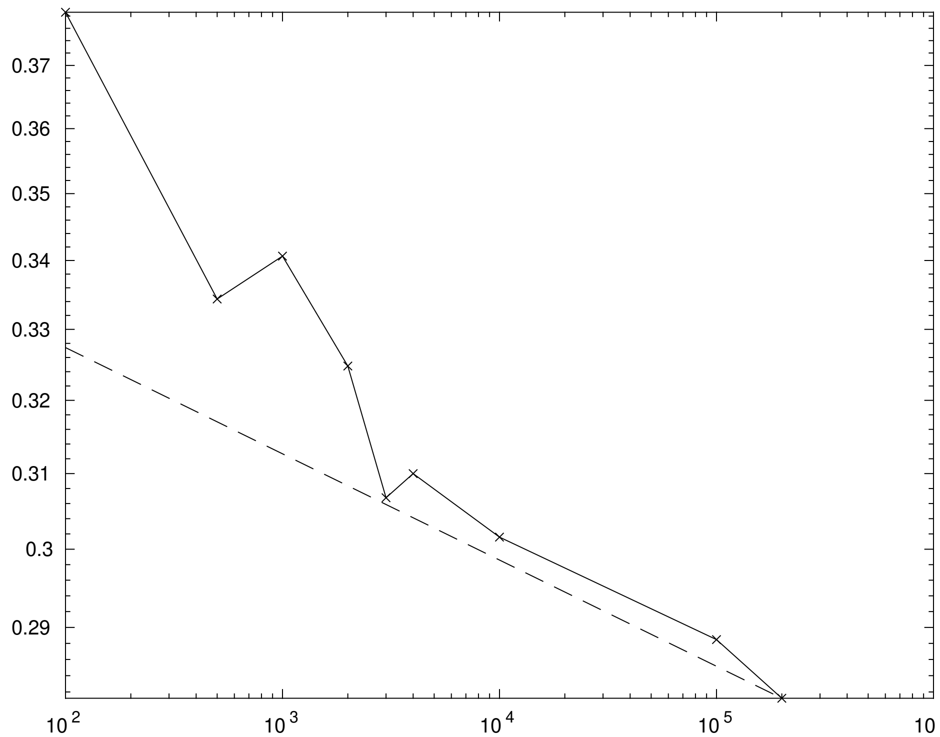


Figure 3.2: Via acceptance rejection we simulate $f(n) = \log(n)$ many ancestral lines of individuals in $[n]$ having their parent in $\{-2n, -2n+1, \dots, -n\}$ and not coalescing till $-5n$. For such a realisation we simulate another line of an individual in $[n]$ having their parent in $\{-2n, -2n+1, \dots, -n\}$ and numerically estimate the probability that this line does not intersect with the ancestry of the other $f(n)$ lines till $-5n$. We average over 10.000 realisations for the $f(n)$ lines and over 100 for the next line in each of those. In the above panels we compare $n^{(1-2\alpha)}$ (dashed line) with the estimated probability for $\alpha = 0.51$

$$\Leftrightarrow \quad (1-2\alpha)\gamma \leq -\varepsilon\alpha$$
$$\Leftrightarrow \quad \frac{\gamma(2\alpha-1)}{\alpha} > \varepsilon.$$

For all $\alpha > \frac{2}{3}$, $\gamma \geq 1$ the last inequality is satisfied for $\varepsilon < \frac{1}{\alpha} - 1$, meaning that for all $\varepsilon < \frac{1}{\alpha} - 1$ the number of ancestral lineages of individuals having a parent on scale n^γ not being coalesced with one of the $f(n)$ many lines till $-n^{\gamma+\varepsilon}$ is of smaller order than the number of individuals of $[n]$ having their parents on scale $n^{\gamma+\varepsilon}$. So let $\alpha > \frac{2}{3}$ and $1/\alpha > \gamma \geq 1$. Then the total number of remaining ancestral lineages of $[n]$ at n^γ, $N(n, -n^\gamma)$, should be of the same order as the number of individuals of $[n]$ having their parents on scale n^γ, which is $U(n, -n^\gamma)$. This motivates the following conjecture.

Conjecture 3.2.7. *For $\alpha > \frac{2}{3}$ the validity of Statement 3.2.4 implies that for all $\varepsilon > 0$, $\gamma \geq 1$, $y \geq 0$*

$$\mathbf{P}\left(N\left(n, -yn^\gamma\right) \geq (1+\varepsilon)U\left(n, -yn^\gamma\right)\right) \to 0 \text{ as } n \to \infty. \tag{3.2.15}$$

In words, this means that $N(n, -xn^\gamma)$ is indeed of the same order as $U(n, -xn^\gamma)$ for $\gamma \geq 1$ (see (S.1.37) and (S.1.5.2) for the definitions of N, respectively U). Even further all these ideas suggest that in fact only of order 1 many individuals *pass from one scale to the next*, since we can choose $f(n)$ arbitrarily slowly growing. Figure 3.3 supports at least the claim that $N(n, -x \cdot n^\gamma)$ is of the same order as $U(n, -x \cdot n^\gamma)$.
The asymptotic bound (3.2.15) bears similarities to (S.1.47). Indeed, we can then prove the following Proposition, which is in the style of to Proposition S.1.27.

Proposition 3.2.8. *The conjectured bound (3.2.15) implies that the sequence*

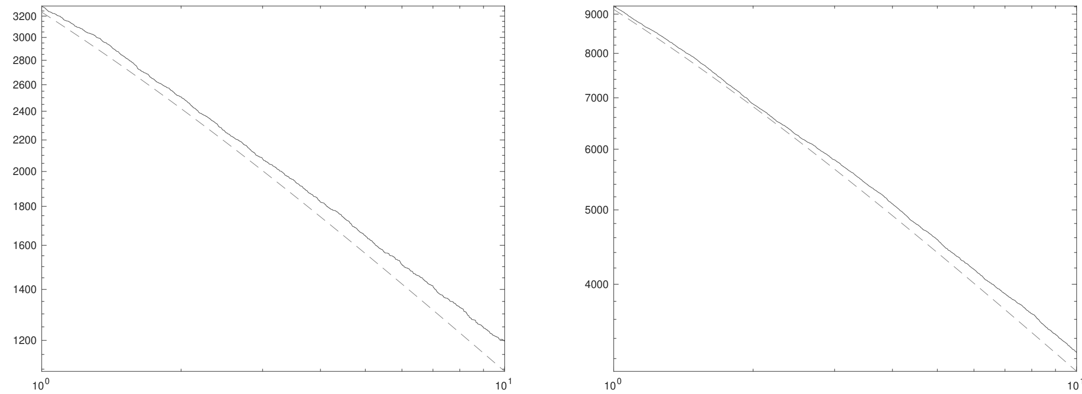
$$\left(\frac{\mathrm{MRCA}([0,n])}{n^{\frac{1}{\alpha}}}\right)_n$$

of random variables is tight.

Proof. We proceed as in the end of the proof of Proposition 3.2.2, see Section 3.2.2. Let a_n be a divergent non-decreasing sequence. Lemma 3.2.1 gives us, that the most recent common ancestor of the individuals $i \in [n]$ having a parent below $-n^{1/\alpha}/a_n$ is then a random variable of scale not larger than $n^{1/\alpha}a_n^{\frac{2}{2\alpha-1}}$. Since the only requirement on a_n was that it does not decrease, we can choose $a_n^{\frac{2}{2\alpha-1}} \ll b_n$. By Lemma 3.2.3 we then obtain the desired result. $\square$

3.2.4 Heuristics for the decay of the thinned density of the ancestry of $[n]$

The heuristic arguments from the previous section also give more evidence for the validity of the MRCA $([n])$-scale Conjecture S.1.18. Observing that with probability $n^{(1-2\alpha)\gamma}$ each of the $n^{1-\alpha\gamma}$ ancestral lineages of $[n]$ jumping directly onto scale n^γ does not coalesce with



(A) $n = 10^8$ **(B)** $n = 10^9$

Figure 3.3: For $n = 10^8, 10^9$ and $\alpha = 0.55$ we simulate $N(n, -x \cdot n)$ for $-x \in [-10, -1]$. The dotted line shows $\frac{n^{1-\alpha}}{1-\alpha}\left[(1+x)^{1-\alpha} - x^{1-\alpha}\right]$ as a proxy on the expected number of individuals having their parent below $-x \cdot n$.

another one out of $[n]$ and then has of order $n^{(1-\alpha)\gamma}$ many ancestors on scale n^γ, we obtain that we should observe of order

$$n^{1-\alpha\gamma} \cdot n^{(1-\alpha)\gamma} \cdot n^{(1-2\alpha)\gamma} = n^{1-4\alpha\gamma+2\gamma} \tag{3.2.16}$$

many ancestors of $[n]$ in an interval of length n^γ additional to the $f(n)$ many lineages, whose existence was postulated in Statement 3.2.4. Since

$$\begin{aligned} & 1 - 4\alpha\gamma + 2\gamma \leq \gamma - \frac{\alpha}{2}\gamma \\ \Leftrightarrow \quad & 1 - \frac{7}{2}\alpha\gamma + \gamma \leq 0 \\ \Leftrightarrow \quad & 1 \leq \frac{7}{2}\alpha\gamma - \gamma \end{aligned}$$

and $\frac{7}{2}\alpha\gamma - \gamma \geq \frac{3}{2}\alpha\gamma$ for $\alpha \geq \frac{1}{2}$ we get that for γ and α satisfying

$$\frac{3}{2}\alpha\gamma \geq 1 \Leftrightarrow \gamma > \frac{2}{3\alpha} \tag{3.2.17}$$

(3.2.16) gives evidence for

$$\sum_{k=0}^{n^\gamma} \mathbf{P}\left(-k \in A_{[n]}\right) \leq \text{const} \cdot n^{\gamma - \frac{\alpha}{2}\gamma} + f(n) \cdot \text{const} \cdot n^{\gamma\alpha}$$

which in turn gives evidence for

$$\mathbf{P}\left(-n^\gamma \in A_{[n]}\right) \leq \text{const} \cdot n^{-\frac{\alpha}{2}\gamma} + f(n) n^{\gamma\alpha - \gamma}.$$

Since for $\alpha > \frac{2}{3}$ the right hand side of (3.2.17) is satisfied for all $\gamma > 1$, this motivates the following modified version conjecture on the decay of the thinned density of the ancestry of $[n]$.

Conjecture 3.2.9 (Thinned density of the ancestry of $[n]$)**.** *Let $\alpha > \frac{2}{3}$ and $x \gg n$ and let $f(n) \ll n$ be diverging and non-decreasing, then*

$$\hat{t}^{(n)}(x) := \mathbf{P}\left(-x \in A_{[n]}\right) - f(n)q_x = O\left(x^{-\frac{\alpha}{2}}\right). \tag{3.2.18}$$

Remark 3.2.10. Note that (S.1.42) implies (3.2.18), since

$$\begin{aligned} & \mathbf{P}\left(-x \in A_{[n]}\right) - f(n)\,\mathbf{P}\left(-x \in A_0\right) \\ \leq\ & \mathbf{P}\left(-x \in A_{[n]}\right) - \mathbf{P}\left(-x \in A_0\right) \\ \leq\ & \mathbf{P}\left(-x \in \hat{A}_{[n]}\right). \end{aligned}$$

Remark 3.2.11. Compared to $t^{(n)}(\ell)$, which is somehow *counting ancestors of $[n]$ not being ancestors of 0*, this result omits $f(n)$ many ancestral lineages in the counting. However, since $f(n)$ is allowed to grow arbitrarily slowly, this is not a huge loss of ground compared to (S.1.42). And as its application in the proof of Proposition S.1.23 shows, the proof could be modified to use Conjecture 3.2.9 instead without weakening the result at all.

3.2.5 Heuristics for the density decay of the pruned ancestry of $\mathbb{N}$: Maximal displacement of a critical branching random walk

For $x \in \mathbb{N}$ set

$$E_x := \{-x \text{ has a descendant above 0 and is no ancestor of } 0\}.$$

Then the asymptotics (S.1.44) are equivalent to

$$t(x) = \mathbf{P}\left(E_x\right) = O\left(x^{-\frac{\alpha}{2}}\right) \qquad \text{as } x \to \infty.$$

In order to analyse this we want to cultivate the following picture of the HS-tree: Set

$$B_k := \left\{y \in A_{\mathbb{N}} \backslash A_0 : \text{MRCA}(0, y) = k\right\} \qquad k \in A_0. \tag{3.2.19}$$

The set B_k then consists of those descendants of k that belong to $-\mathbb{N}\backslash A_0$ and have descen-

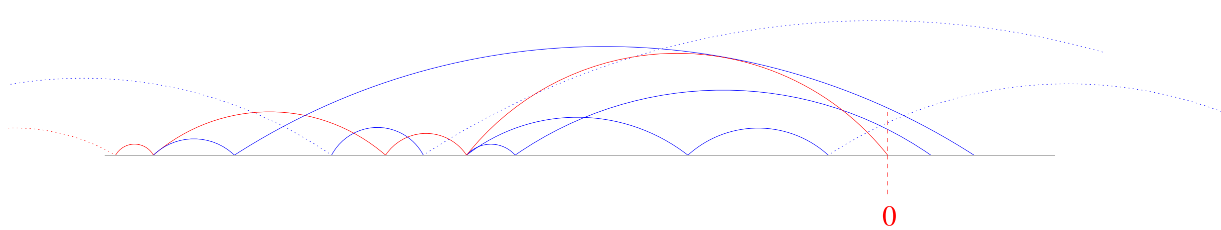


Figure 3.4: This is an illustration of the sets B_k defined by (3.2.19): In red the ancestral lineage of zero, in blue individuals $k \in -\mathbb{N}\backslash A_0$ having descendants in the positive integer numbers.

dants in $\mathbb{N}$. We can then represent $A_{\mathbb{N}_0}$ as the disjoint union

$$\bigcup_{k \in A_0} \left(\{k\} \cup B_k\right)$$

and E_x can then also be expressed as

$$E_x = \{-x \in B_k \text{ for some } k < -x\}.$$

Consider the descendants of an individual $-x \in -\mathbb{N}$ in the HS-tree. The offspring in each generation is then not independent, since for all $k < z$ knowledge of k not being a child of z increases the probability that k is a child of any other individual $j > k$. However, the trees B_{-x} can never hit each other again and can by definition not hit the ancestral lineage of zero and also not the ancestors of any individuals j with $\mathrm{MRCA}(0, j) \geq -x$, see Figure 3.4 for an illustration. Thus we should be able to dominate $\mathbf{P}\left(\{-x \in B_{-x-1}\}\right)$ by independent generations in the following way: Independently for all $\ell \in \mathbb{N}$, individual 0 gets a child with distance ℓ (and hence at position ℓ) with probability $\mu\left(\{\ell\}\right)$. We iterate this procedure generation-wise for all children, thus obtaining a branching random walk on $\mathbb{N}_0$. Since the expected number of children is one, the branching is critical and the associated Galton-Watson tree is almost surely finite.

Denote by M the maximum of the positions of all individuals in this tree. M can thus be understood as the maximum ever attained by in a critical branching random walk that starts at zero. We conjecture, that an upper bound on the order of magnitude of $t(x)$ can be obtained by analysing the asymptotics of $\mathbf{P}(M > x)$ as $x \to \infty$.

Proposition 3.2.12. *Let $\alpha > \frac{1}{2}$. Then for $x \geq 1$*

$$\mathbf{P}(M > x) \leq \text{const} \cdot x^{-\frac{\alpha}{2}}.$$

Proof of Proposition 3.2.12. By construction

$$M \overset{(d)}{=} \max_{i=1,2,\ldots} \left\{B_i(M_i + i)\right\}$$

for independent copies $M_1, M_2, \ldots$ of M and B_i independent with distribution $\mathrm{Ber}\Big(\mu\left(\{i\}\right)\Big)$. So the cumulative distribution function of M fulfills

$$\mathbf{P}(M \leq x) = \prod_{i=1}^{\infty} \Big[\mu(\{i\})\, \mathbf{P}(M \leq x - i) + 1 - \mu\left(\{i\}\right)\Big].$$

We will show

$$\mathbf{P}(M > x) \leq \text{const} \cdot x^{-\frac{\alpha}{2}}$$

which implies the assertion. Since $M \geq 0$ we have

$$\mathbf{P}(M \leq x) = \prod_{i=1}^{x} \Big[1 - \mu\left(\{i\}\right)\left(1 - \mathbf{P}(M \leq x - i)\right)\Big] \cdot \prod_{i \geq x+1} \left[1 - \mu(\{i\})\right]. \tag{3.2.20}$$

In order to ease notation set

$$f(x) := \mathbf{P}(M \leq x) \qquad \text{and} \qquad g(x) := 1 - f(x).$$

We will analyse the log of (3.2.20):

$$\log\left(1-g(x)\right) = \sum_{i=1}^{x} \log\left[1-\mu\left(\{i\}\right)\left(g(x-i)\right)\right] + \sum_{i\geq x+1} \log\left[1-\mu(\{i\})\right].$$

On both sites we replace $\log(1-y)$ by its Taylor series:

$$-\sum_{k\geq 1}\frac{1}{k}g(x)^k = -\sum_{k\geq 1}\frac{1}{k}\sum_{i=1}^{x}\left(\mu\left(\{i\}\right)g(x-i)\right)^k - \sum_{k\geq 1}\frac{1}{k}\sum_{i\geq x+1}\left(\mu\left(\{i\}\right)\right)^k. \qquad (3.2.21)$$

Set

$$\mu^k(x) := \left(\mu\left(\{x\}\right)\right)^k, \quad \overline{\mu^k}(x) := \sum_{i>x}\mu^k(i), \quad g^k(x) := g(x)^k$$

and

$$\overline{\mu}(x) := \overline{\mu^1}(x).$$

By multiplying both sides in (3.2.21) with -1 and writing the first sum on the right hand side as a convolution we obtain

$$\sum_{k\geq 1}\frac{1}{k}g^k(x) = \sum_{k\geq 1}\frac{1}{k}\left(\mu^k * g^k\right)(x) + \sum_{k\geq 1}\frac{1}{k}\overline{\mu^k}(x).$$

Since all summands are positive functions we get the following identity for the Laplace transforms

$$\sum_{k\geq 1}\frac{1}{k}\mathcal{L}\left(g^k\right) = \sum_{k\geq 1}\frac{1}{k}\mathcal{L}\left(g^k\right)\mathcal{L}\left(\mu^k\right) + \sum_{k\geq 1}\frac{1}{k}\mathcal{L}\left(\overline{\mu^k}\right),$$

where we only used linearity and the fact that the Laplace transform of a convolution is the product of the Laplace transforms. Rearranging terms gives the identity

$$\sum_{k\geq 1}\frac{1}{k}\left(1-\mathcal{L}\left(\mu^k\right)\right)\mathcal{L}\left(g^k\right) = \sum_{k\geq 1}\frac{1}{k}\mathcal{L}\left(\overline{\mu^k}\right).$$

Expanding the sums gives

$$\begin{aligned} &\left(1-\mathcal{L}\left(\mu\right)\right)\mathcal{L}\left(g\right) + \frac{1}{2}\left(1-\mathcal{L}\left(\mu^2\right)\right)\mathcal{L}\left(g^2\right) + \sum_{k\geq 3}\frac{1}{k}\left(1-\mathcal{L}\left(\mu^k\right)\right)\mathcal{L}\left(g^k\right) \\ = \;& \mathcal{L}\left(\overline{\mu}\right) + \sum_{k\geq 2}\frac{1}{k}\mathcal{L}\left(\overline{\mu^k}\right). \end{aligned}$$

Now set

$$C_k := \sum_{\ell}\mu^k(\ell), \qquad \overline{C_k} := 1 - C_k,$$

and note

$$C_\ell < 1 \quad \forall \ell \geq 2 \qquad \text{and} \qquad C_1 > C_2 > C_3 > \dots.$$

With this notation we can write

$$\begin{aligned}
&\left(1-\mathcal{L}\left(\mu\right)\right)(\theta)\mathcal{L}\left(g\right)(\theta)+\frac{1}{2}\left(\overline{C_2}+C_2-\mathcal{L}\left(\mu^2\right)(\theta)\right)\mathcal{L}\left(g^2\right)(\theta)\\
&+\sum_{k\geq 3}\frac{1}{k}\left(\overline{C_k}+C_k-\mathcal{L}\left(\mu^k\right)(\theta)\right)\mathcal{L}\left(g^k\right)(\theta)\\
=\;&\mathcal{L}\left(\overline{\mu}\right)+\sum_{k\geq 2}\frac{1}{k}\mathcal{L}\left(\overline{\mu^k}\right)(\theta).
\end{aligned}$$

Since $\mu(x)\sim \text{const}\cdot x^{-\alpha-1}$ the tails of the measures $\overline{\mu^k}$ are integrable for $k\geq 2$. So the right hand side has a singularity of order $\theta^{\alpha-1}$ for $\theta\to 0$ which comes from the first term, while the sum is of lower order.
On the left hand side all terms are positive, so we obtain that $\mathcal{L}\left(g^2\right)(\theta)$ is at most of order $\theta^{\alpha-1}$.
By Lemma 3.2.13 we get that g^2 decreases at least as fast as $\overline{\mu}$. Taking the square root gives the assertion. □

To complete the proof of Proposition 3.2.12 we append the following lemma, which is in the style of Feller's Tauberian results.

Lemma 3.2.13. *Let $g:\mathbb{N}_0\to[0,1]$ be monotone and non-increasing and $\beta>0$ such that*

$$\limsup_{\theta\to 0}\theta^{1-\beta}\mathcal{L}(g)<\infty.$$

Then

$$\limsup_{x\to\infty} g(x)x^{\beta}<\infty.$$

Proof. We prove the statement by contradiction: Assume that there exists a monotone increasing sequence $(x_k)_k$, such that

$$g(x_k)x_k^{\beta}\uparrow\infty,$$

then for $\theta_k:=\frac{1}{x_k}$

$$\begin{aligned}
\mathcal{L}(g)(\theta_k)x_k^{\beta-1} &= \sum_x g(x)e^{-\theta_k x}x_k^{\beta-1}\\
&\geq \sum_{x=1}^{x_k} g(x_k)e^{-\frac{x}{x_k}}x_k^{\beta-1}\\
&\geq \sum_{x=1}^{x_k} g(x_k)e^{-1}x_k^{\beta-1}=e^{-1}x_k^{\beta}g(x_k)\to\infty.
\end{aligned}$$

So

$$\mathcal{L}(g)(\theta_k)x_k^{\beta-1}\to\infty$$

in contradiction to our assumption. □

3.2.6 Simulation-based evidence for the density decay of the pruned ancestry of $[n]$

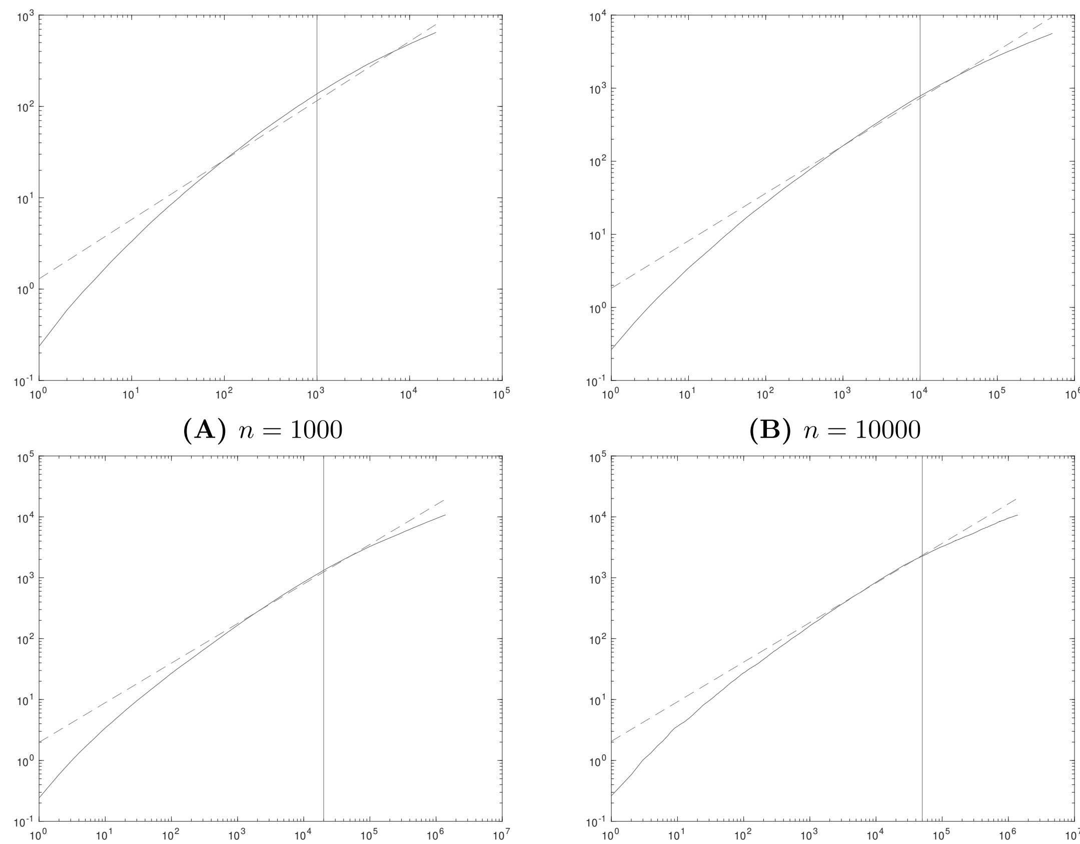
To give further evidence for the validity of (S.1.42) we simulate

$$g(x) := \mathbf{1}_{-x \in \cup_{\ell=1}^{n} A_\ell}, \qquad x = 1, \ldots, n^{\frac{1}{\alpha}}$$

and use

$$G(x) := \sum_{\ell=1}^{x} g(\ell) - \sum_{\ell=1}^{x} q_\ell$$

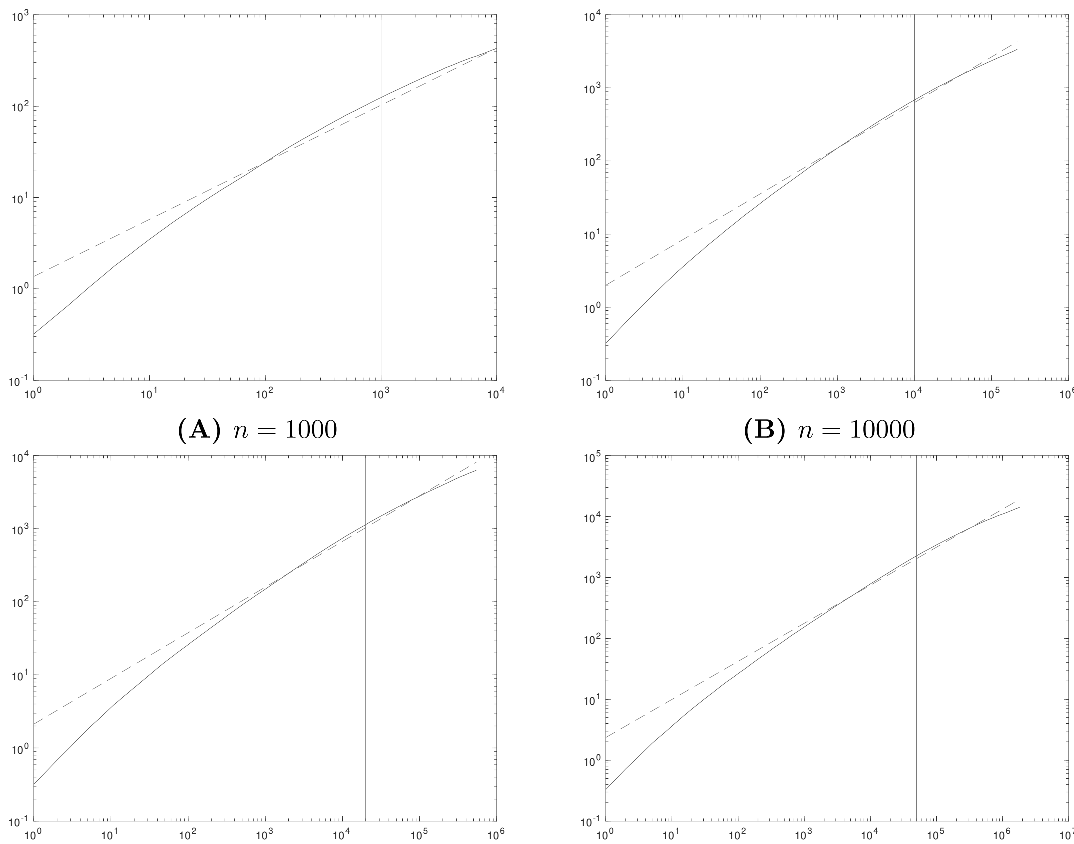
as an estimator for $\sum_{\ell=1}^{x} t^{(n)}(x)$. We performed those simulations for different values of n and α and highlight $x = n$ by a vertical line. The simulations give clear evidence for (S.1.42). We display $\mathrm{const} \cdot x^{1-\frac{\alpha}{2}}$ as a dotted line and an average over 1000 simulations of $G(x)$.



(A) $n = 1000$ **(B)** $n = 10000$

(C) $n = 20000$ **(D)** $n = 50000$

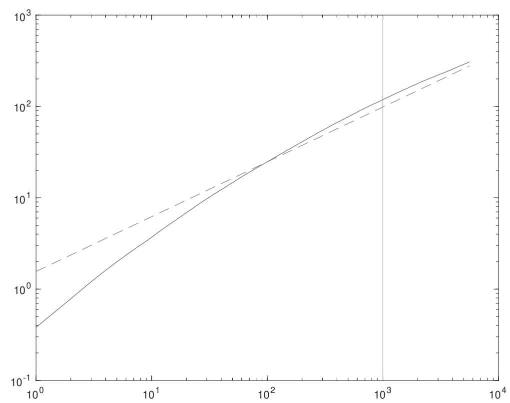
Figure 3.5: $G(x)$ vs. $\text{const} \cdot x^{1-\frac{\alpha}{2}}$ for $\alpha = 0.7$ and different n



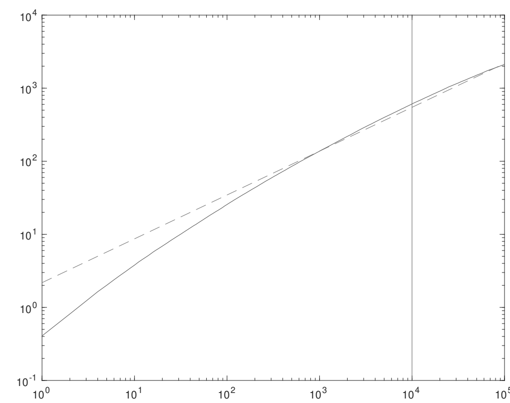
(A) $n = 1000$ **(B)** $n = 10000$

(C) $n = 20000$ **(D)** $n = 50000$

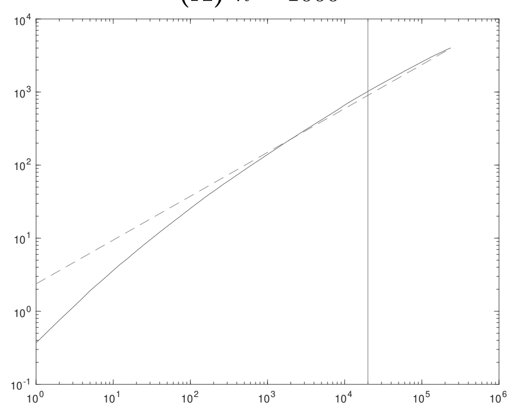
Figure 3.6: $G(x)$ vs. $\text{const} \cdot x^{1-\frac{\alpha}{2}}$ for $\alpha = 0.75$ and different n



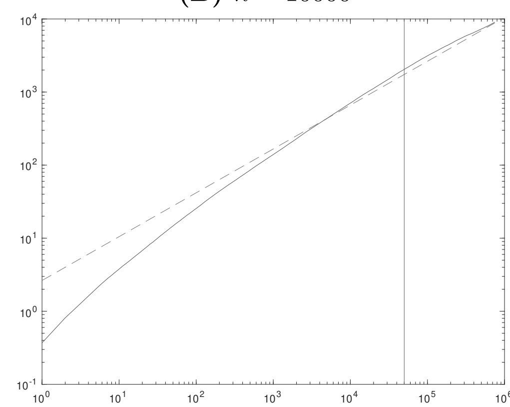
(A) $n = 1000$



(B) $n = 10000$



(C) $n = 20000$



(D) $n = 50000$

Figure 3.7: $G(x)$ vs. const $\cdot\, x^{1-\frac{\alpha}{2}}$ for $\alpha = 0.8$ and different n

3.2.7 A coalescing system of discrete Bessel random walks

We now want to explain a connection between the HS-urn and discrete Bessel random walks. A d-dimensional Bessel-process is described by the stochastic differential equation

$$\mathrm{d}X_t = \mathrm{d}W_t + \frac{d-1}{2}\frac{\mathrm{d}t}{X_t}.$$

The closure of the zero-set of X, is equal in distribution to the range of a $\left(1-\frac{d}{2}\right)$-stable subordinator, see Revuz and Yor [RY99, Chapter XI, Exercise 1.25] and Lawler's lecture notes [Law18] for more details about Bessel processes. In the spirit of Alexander [Ale11] we can look at *discrete Bessel processes* with transition kernel

$$p(x,y) = \begin{cases} \gamma, & y = x \\ \frac{1-\gamma}{2}\left(1-\frac{1-d}{2x}\right), & y = x+1, x>0 \\ \frac{1-\gamma}{2}\left(1+\frac{1-d}{2x}\right), & y = x-1 \geq 0 \\ 1-\gamma, & x=0, y=1 \\ 0, & \text{else} \end{cases} \qquad \text{for } \gamma\in(0,1),\ d>0.$$

Then [Ale11, Thm. 2.1] indeed suggests that for $d = 2(1-\alpha)$ the zero-set of such a random walk is distributed like an ancestral lineage in our setting for a suitable choice of μ fulfilling (2.2.1).

Remark 3.2.14. Note that the mirroring of α around $\frac{1}{2}$ discussed in Remark S.1.12 thus corresponds to a mirroring of the dimension around 1.

Proposition S.1.27 and Proposition S.1.23 can now be understood as a result about discrete Bessel processes coalescing in zero: At times $-n, -n+1, \dots -1$ a random walker with transition kernel p starts in zero and evolves now independent of all the others. But when more than one are in zero at the same time they coalesce. Such a system is similar as the interacting particle systems analysed by Chen and Fan [CF17] with the notable difference, that the density of particles converges to zero as $n\to\infty$.

We now formalise the already verbally described system: The space of particle configurations is given by $\mathbb{N}_0^{\mathbb{N}_0}$ and for $\eta = \big(\eta(i)\big)_{i\in\mathbb{N}_0} \in \mathbb{N}_0^{\mathbb{N}_0}$ we write

$$|\eta| := \sum_{i\in\mathbb{N}_0} \eta(i)$$

for the number of particles in the configuration η. Write $\delta^{(x)}$ for the configuration with only one particle at point $x\in\mathbb{N}_0$, that is $\delta^{(x)}(i) = 0$ for $i\in\mathbb{N}_0\backslash\{x\}$ and $\delta^{(x)}(x) = 1$.

For $\eta,\xi \in \mathbb{N}_0^{\mathbb{N}_0}$ denote by $\mathfrak{p}(t,\eta,\xi)$ the probability that a particle configuration η of independent particles with transition kernel p evolve to the configuration ξ in t steps. In particular this means

$$\mathfrak{p}(t,\eta,\xi) = \begin{cases} 0, & \text{if } t=0 \text{ and } \eta\neq\xi \\ 0, & \text{if } |\eta|\neq|\xi| \\ p(x,y), & \text{if } t=1 \text{ and } \eta=\delta^{(x)}, \xi=\delta^{(y)} \end{cases}.$$

If we introduce the *killing operator*

$$\mathfrak{K}f(\eta) = f\left(\eta - \left(\eta(0)-1\right)_+ \delta_0\right).$$

and set

$$Qf(\eta) := \sum_{\kappa \in \mathbb{N}_0^{\mathbb{N}_0}} \mathfrak{p}(1,\eta,\kappa) f(\kappa)$$

for functions $f : \mathbb{N}_0^{\mathbb{N}_0} \to \mathbb{R}$, such that $Qf(\eta)$ is finite for all $|\eta| < \infty$, the generator of the verbally described system of in zero coalescing particles is given by

$$\mathfrak{K}Qf(\eta) := \sum_{\kappa \in \mathbb{N}_0^{\mathbb{N}_0}} \mathfrak{p}\left(1,\eta,\kappa\right) f\left(\kappa - \left(\kappa(0)-1\right)_+ \delta^{(0)}\right).$$

Meaning that for the system $\hat{\eta} := (\hat{\eta}_t)_{t\geq 0} = \left(\left(\hat{\eta}_t(i)\right)_{i\in\mathbb{N}_0}\right)_{t\geq 0}$ of in zero coalescing random walks with transition kernel p we have

$$\mathbf{E}_{\hat{\eta}_0}\left[f(\hat{\eta}_1)\right] = \mathfrak{K}Qf(\hat{\eta}_0).$$

Now we randomize over $\hat{\eta}(0)$ as sketched above: We start with one particle in zero at time $-n$ evolving according to the transition kernel p. At all time points $-n, -n+1, \ldots -1$ we spawn another particle in zero, however if there is already one or more particle they immediately coalesce. This is equivalent to the following procedure: For all time points $t = -n, \ldots, -1$ generate a $\mathrm{Ber}(\mu(\{t\}))$-distributed random variable B_t. If $B_t = 1$ start a random walk with transition kernel p conditioned on not returning to the origin for at least t steps. This gives a random configuration of particles at time 0. Denote this random configuration by $\mathfrak{e}^{(n)}$. Then (S.1.47) and Conjecture 3.2.7 can be read as

$$\begin{aligned}
&\int_{\mathbb{N}_0^{\mathbb{N}_0}} \mathbf{P}\left(|\hat{\eta}_{a_n n}| \geq (1+\varepsilon)\mathrm{const}(\alpha)\cdot (a_n n)^{1-\alpha}\left[\left(\frac{1}{a_n}+1\right)^{1-\alpha} - 1\right] \middle| \hat{\eta}_0 = \mathfrak{e}^{(n)}\right) \mathrm{d}\mathbf{P}_{\mathfrak{e}^{(n)}} \\
\leq\ & \mathrm{const}\cdot\left[a_n^{1-\frac{\alpha}{2}+\xi} n^{1-\frac{3}{2}\alpha+\xi} + a_n^{\alpha-\xi+\eta} n^{-\xi+\eta}\right].
\end{aligned}$$

Writing $T_{\mathfrak{e}^{(n)}}$ for the minimal random t, such that the in $\mathfrak{e}^{(n)}$ started system $\hat{\eta}(t)$ consists of only one particle, namely $|\hat{\eta}_t| = 1$, Proposition S.1.27 translates to

$$\left(\frac{T_{\mathfrak{e}^{(n)}}}{n^{\frac{1}{\alpha}}}\right)_n \text{ is a tight family of random variables.}$$

Such systems of interacting particles are similar to the ones considered by [CF17]. The methods developed there might additionally give a hydrodynamic description of the above particle system. The notable difference to [CF17] is the density of particles: By (S.1.46) we have that $|\mathfrak{e}^{(n)}|$ is of order $n^{1-\alpha}$ and since the Bessel random walkers are spread out on scale $\sqrt{n}$ we get that the density of particles will be of order $n^{1-\alpha-\frac{1}{2}}$. For $\alpha > \frac{1}{2}$ this is converging to zero. This is the main difficulty in applying the methods of [CF17]. A key tool used in [CF17], see [CF17, Lemma 3.1] (and also Boldrighini, De Masi, Pellegrinotti and Presutti [BDMPP87, (2.1)]), is a time-continuous analogue to the following Lemma. Since we could not find a suitable reference for the time-discrete version state below we give a proof here.

Lemma 3.2.15. *Let S be countable and let $p = (p(x, y))$ be an irreducible stochastic matrix on S with reversible measure $\pi = (\pi(x))$. Let $\eta = (\eta_t)_{t\in\mathbb{Z}_+}$ and $\xi = (\xi_t)_{t\in\mathbb{Z}_+}$ be finite systems of independent random walkers with transition matrix p and fixed starting configurations η_0, ξ_0. Write $(a)_{k\downarrow} := a(a-1)\cdots(a-k+1)$ with $(a)_{0\downarrow} := 1$. Then*

$$\mathbf{E}\left[\prod_{x\in E}\frac{(\eta_t(x))_{\xi_0(x)\downarrow}}{\pi(x)^{\xi_0(x)}}\right] = \mathbf{E}\left[\prod_{x\in E}\frac{(\eta_0(x))_{\xi_t(x)\downarrow}}{\pi(x)^{\xi_t(x)}}\right], \quad t\in\mathbb{Z}_+. \tag{3.2.22}$$

Proof. We will give a combinatoric proof of (3.2.22). Let $|\eta_0| := \sum_x \eta_0(x) = N \in \mathbb{N}$. We describe η by numbered random walkers $X^{(i)} = (X_t^{(i)})_{t\in\mathbb{Z}_+}$, $i \in [N] := \{1, 2, \ldots, N\}$ with starting positions $x_0^{(i)} = X_0^{(i)}$. Then for $k \in \mathbb{N}$

$$(\eta_t(x))_{k\downarrow} = \sum_{\substack{(j_1,\ldots,j_k)\in[N]^k \\ \text{pairwise different}}} \prod_{i=1}^{k} \mathbf{1}_{X_t^{(j_i)}=x}$$

is the number of possible ordered k-tupels of particles at time t at point x. This gives

$$\prod_{x\in S}(\eta_t(x))_{\xi_0(x)\downarrow} = \sum_{(j_{x,\ell})\in J(\xi_0)} \prod_{x\in S}\prod_{\ell=1}^{\xi_0(x)} \mathbf{1}_{X_t^{(j_{x,\ell})}=x} \tag{3.2.23}$$

for

$$\begin{aligned} J(\xi_0) \quad = \quad \big\{ \big(j_{x,\ell}\,; x \in \{y \in S : \xi_0(y) > 0\}, 1 \le \ell \le \xi_0(x)\big) \\ : \text{pairwise different entries in } [N] \ \big\}. \end{aligned}$$

Note that $(j_{x,\ell}) \in J(\xi_0)$ has $|\xi_0|$ many entries, and without loss of generality we can choose $|\xi_0| \le N$, since otherwise (3.2.23) is equal to zero.

Multiplying of (3.2.23) with $\prod_{x\in E}\pi(x)^{-\xi_0(x)}$ and taking the expectation gives

$$\begin{aligned} \mathbf{E}\left[\prod_{x\in E}\frac{(\eta_t(x))_{\xi_0(x)\downarrow}}{\pi(x)^{\xi_0(x)}}\right] &= \sum_{(j_{x,\ell})\in J(\xi_0)} \prod_{x\in S}\prod_{\ell=1}^{\xi_0(x)} \frac{p^t(x_0^{(j_{x,\ell})}, x)}{\pi(x)} \\ &= \sum_{(j_{x,\ell})\in J(\xi_0)} \prod_{x\in S}\prod_{\ell=1}^{\xi_0(x)} \frac{p^t(x, x_0^{(j_{x,\ell})})}{\pi(x_0^{(j_{x,\ell})})}, \end{aligned} \tag{3.2.24}$$

where in the second line we used reversibility.

For $\widetilde{\xi} \in \mathbb{N}_0^S$ with $|\widetilde{\xi}| = |\xi_0|$ set

$$J(\xi_0, \widetilde{\xi}) = \big\{ \big(j_{x,\ell}\big) \in\ J(\xi_0) : \#\{(x,\ell) : x_0^{(j_{x,\ell})} = y\} = \widetilde{\xi}(y) \text{ for all } y \in S \big\}.$$

Note that $J(\xi_0) = \dot\bigcup_{|\widetilde{\xi}|=|\xi_0|} J(\xi_0, \widetilde{\xi})$.

With the representation

$$\xi_t = \sum_{x\in S\,:\,\xi_0(x)>0} \sum_{j=1}^{\xi_0(x)} \delta_{\widetilde{X}_t^{(x,j)}},$$

where the $(\widetilde{X}_t^{(x,j)})_{t\in\mathbb{Z}_+}$, $x \in S, j \in \mathbb{N}$ are independent random walkers with transition kernel p and $\widetilde{X}_0^{(x,j)} = x$, we get that the sum in (3.2.24) restricted on $J(\xi_0, \widetilde{\xi})$ reads as

$$\sum_{(j_{x,\ell})\in J(\xi_0,\widetilde{\xi})} \prod_{x\in S} \prod_{\ell=1}^{\xi_0(x)} \frac{p^t(x, x_0^{(j_{x,\ell})})}{\pi(x_0^{(j_{x,\ell})})} = \mathbf{P}(\xi_t = \widetilde{\xi}) \prod_{y\in S} \frac{(\eta_0(y))_{\widetilde{\xi}(y)\downarrow}}{\pi(y)^{\widetilde{\xi}(y)}}. \tag{3.2.25}$$

Summing (3.2.25) over all $\widetilde{\xi}$ with $|\widetilde{\xi}| = |\xi_0|$ together with (3.2.24) gives (3.2.22). □

3.3 The HS-trees do not contain a spine: A martingale based proof of Proposition S.1.5

We prove Proposition S.1.5 now in a second way via the martingale M_n of the predicted size of the progeny of 0 above n (defined in (3.3.1) below). This proof was suggest by Jan Swart.

Proof of Proposition S.1.5. For $B \subset \mathbb{Z}$ we set

$$C(B) := \{i \in \mathbb{Z} : i - R_i \in B\},$$

so $C(B)$ is the set of children of B. Inductively for $n \geq 0$

$$C^0(B) := B, \qquad C^{n+1}(B) := C\left(C^n(B)\right).$$

And

$$D(A) := \bigcup_{n=0}^{\infty} C^n(B),$$

which is the set of all descendants of B. We look at $B = \{0\}$ and write $D(0) = D\left(\{0\}\right)$. Set

$$D_n(0) := D(0) \cap \{0, \ldots, n\},$$

and

$$C_n(0) := C\left(D_n(0)\right) \cap \{n+1, n+2, \ldots\}.$$

Then $D_n(0)$ is the set of all descendants of 0 between 0 and n and $C_n(0)$ is the set of their children above n. The following properties are equivalent:

1. $\mathbf{P}$ (0 has infinitely many descendants) > 0,
2. $\mathbf{P}$ (There exists at least one spine) > 0,

3. $\mathbf{P}\big(\#D(0)=\infty\big)>0$.

We will now show that

$$M_n := \mathbf{E}\left[\big|C_n(0)\big| \,\big|\, (R_k)_{0<k\leq n}\right], \tag{3.3.1}$$

the conditional expectation of the number of individuals above n that are descendants of 0 and have a parent below $n+1$, given the HS-genealogy up to n. First

$$0 \leq M_n \leq n\,\mathbf{E}\big[\#\text{Children}(0)\big] = n$$

and with the notation

$$\mathcal{C}_n(0) := \big|C_n(0)\big|$$

we have

$$\mathcal{C}_n(0) = \mathcal{C}_{n-1}(0) + \mathbf{1}_{n\in D_n(0)}\big[\#C(n)-1\big],$$

such that

$$\begin{aligned}
&\mathbf{E}\big[M_n|M_{n-1}\big]\\
=\;& \mathbf{E}\left[\mathbf{E}\left[\big|C_n(0)\big| \,\big|\, (R_k)_{0<k\leq n}\right]\big|M_{n-1}\right]\\
=\;& \mathbf{E}\left[\mathbf{E}\left[\mathcal{C}_n(0)\big|\,(R_k)_{0<k\leq n}\right]\big|M_{n-1}\right]\\
=\;& \mathbf{E}\left[\mathbf{E}\left[\mathcal{C}_{n-1}(0)+\mathbf{1}_{n\in D_n(0)}\big[\#C(n)-1\big]\,\big|\,(R_k)_{0<k\leq n}\right]\big|M_{n-1}\right]\\
=\;& M_{n-1}+\mathbf{E}\left[\mathbf{E}\left[\mathbf{1}_{n\in D_n(0)}\big[\#C(n)-1\big]\,\big|\,(R_k)_{0<k\leq n}\right]\big|M_{n-1}\right]\\
=\;& M_{n-1}+\mathbf{E}\left[\mathbf{1}_{n\in D_n(0)}\big[\#C(n)-1\big]\,\big|M_{n-1}\right]\\
=\;& M_{n-1}+\mathbf{E}\left[\mathbf{E}\left[\mathbf{1}_{n\in D_n(0)}\big[\#C(n)-1\big]\,\big|D_{n-1}(0),(R_k)_{0<k\leq n-1}\right]\big|M_{n-1}\right].
\end{aligned} \tag{3.3.2}$$

Using

$$\mathbf{E}\big[\#\text{Children}(0)\big]=1$$

we obtain

$$\mathbf{E}\left[\mathbf{1}_{n\in D_n(0)}\big[\#C(n)-1\big]\,\big|D_{n-1}(0),(R_k)_{0<k\leq n-1}\right]=0,$$

such that (3.3.2) is equal to M_{n-1}. So $(M_n)_n$ is a martingale.

By the martingale convergence theorem the non-negative martingale M_n converges a.s. to a random variable M_∞. Since $\mathbf{E}[R_1]=\infty$ and the gcd. of the support of μ is one, we have that any finite set of individuals has zero descendants with positive probability, so a finite number of lines of descendants die out with positive probability. Since $(M_n)_n$ can not converge to infinity, it thus converges to zero. This gives

$$\#D(0)<\infty \text{ a.s.},$$

which proves the assertion. □

Part II

Ancestral reproductive bias

Chapter 4

Ancestral reproductive bias in continuous time branching trees under various sampling schemes[1]

[1] Appeared as [II24]

Abstract

Cheek and Johnston [CJ23] consider a continuous-time (Bienaymé-)Galton-Watson tree conditioned on being alive at time T. They study the reproduction events along the ancestral lineage of an individual randomly sampled from all those alive at time T. We give a short proof of an extension of their main results [CJ23, Theorems 2.3 and 2.4] to the more general case of Bellman-Harris processes. Our proof also sheds light onto the probabilistic structure of the rate of the reproduction events. A similar method will be applied to explain (i) the different ancestral reproduction bias appearing in work by Geiger [Gei99] and (ii) the fact that the sampling rule considered by Chauvin, Rouault and Wakolbinger in [CRW91, Theorem 1] leads to a time homogeneous process along the ancestral lineage.

4.1 Introduction

Consider a continuous-time branching process with N_t individuals alive at time t, started with one individual at time 0. At the end of its lifetime, an individual is replaced by a random number of independent offspring with distribution $(p_k)_{k\geq 0}$. When lifetimes of the individuals are i.i.d. with an arbitrary distribution μ on $\mathbb{R}_+$, the resulting process is called a *Bellman-Harris* process [BH48]. In the special case of exponentially distributed lifetimes, this process is a continuous-time(Bienaymé-) Galton-Watson process, which is also called *one-dimensional continuous-time Markov branching process*, see [AN72, Chapter 3]. For those processes, Cheek and Johnston [CJ23] study the process of reproduction times and family sizes along the ancestral lineage of an individual sampled from all those alive at a given time $T > 0$, conditioned on the event $\{N_T > 0\}$. We give a short and conceptual probabilistic proof of the main results of [CJ23] in the more general Bellman-Harris setting. The core idea of this proof is as follows:

On the event $\{N_T > 0\}$, we assign to the individuals alive at time T independent random variables, which will be called markers, uniformly distributed on $[0, 1]$. Then the individual whose marker is largest constitutes a uniformly distributed random pick from all the individuals alive at time T. As we will see, the argument s of the generating functions that appear in the analytic arguments of [CJ23] corresponds to the realisation of the largest marker. Sections 4.2 to 4.4 will be devoted to formulating and proving Theorem 4.2.1.

Relating to work of Chauvin, Rouault and Wakolbinger [CRW91], in Section 4.5 we will consider the case of potentially dependent but identically and atomless distributed markers and conditioning on one marker taking the prescribed value s. In contrast to the above, in this case one does not observe a time-inhomogeneity along the sampled ancestral lineage.

In Section 4.6 we will consider a planar embedding of the Bellman-Harris tree conditioned to survive up to time T, and analyse the leftmost ancestral lineage among those surviving until time T. Here we follow Geiger [Gei99], who gave a representation of discrete-time Galton-Watson processes conditioned to survive up to a given number of generations. With this sampling rule we observe a time-inhomogeneity of ancestral reproduction events that is different from the one in [CJ23].

In Section 4.7 we briefly resume the discussion from [CJ23] on a possible relation between the ancestral rate bias and the rate of mutations per cell division in embryogenesis, and illustrate the various sampling schemes from a more biological perspective.

4.2 Sampling an ancestral line at random

Recall that to each individual at time T, we have associated a uniform marker in $[0,1]$. On the event $\{N_T > 0\}$, let the individual V be sampled as described in the Introduction, and let S be its mark. We define the process $(N_t)_{t\geq 0}$ to be right continuous with left limits. As a consequence, if T_1 is the lifetime of the root individual, then N_{T_1} has distribution $(p_k)_{k\geq 0}$. Let J be the random number of reproduction events and $0 < T_1 < T_2 < \cdots < T_J \leq T$ be the random times of reproduction events along the ancestral lineage of V. Let $L_1, \ldots, L_J$ be the offspring sizes in these reproduction events and let $0 < \tau_1 < \tau_2 < \cdots$ be the random arrival times in a renewal process with interarrival time distribution μ. See Figure 5.1 for a sample realisation.

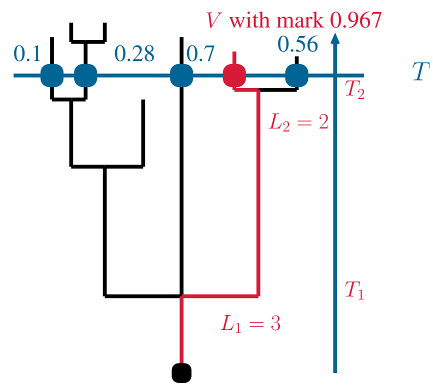


Figure 4.1: An example for a realisation of the random variables S, L_1, L_2, T_1, T_2 in the sampling regime described in Section 4.2.

Denote by $\mathbf{P}$ and $\mathbf{E}$ the probability measure and expectation for $N_0 = 1$.

Theorem 4.2.1. *For $j \geq 0$, $0 < t_1 < \ldots < t_j \leq T \in \mathbb{R}$ and $\ell_1, \ldots, \ell_j \in \mathbb{N}$ we have*

$$\begin{aligned}&\mathbf{P}\left(N_T > 0, J = j, T_1 \in \mathrm{d}t_1, \ldots T_j \in \mathrm{d}t_j, L_1 = \ell_1, \ldots, L_j = \ell_j, S \in \mathrm{d}s\right)\\ &= \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \ldots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} > T\right) \prod_{i=1}^{j} \left(\ell_i p_{\ell_i} \mathbf{E}\left[s^{N_{T-t_i}}\right]^{\ell_i - 1}\right) \mathrm{d}s.\end{aligned} \tag{4.2.1}$$

Corollary 4.2.2. *When integrated over $s \in (0,1)$, (4.2.1) reveals that the process $(T_1, L_1), \ldots, (T_J, L_J)$ of reproduction times and offspring sizes along the ancestral lineage of the uniformly chosen individual (conditioned on $\{N_T > 0\}$) is a mixture of (what could be called) "biased compound renewal processes".*

When the lifetime distribution μ is the exponential distribution with parameter r, then $\tau_1, \tau_2, \ldots$ are the points of a rate r Poisson point process. In this case Corollary 4.2.2 together with (4.2.1) becomes a reformulation of the statements of [CJ23, Theorems 2.3 and 2.4], and at the same time reveals the probabilistic role of the mixing parameter s in the mixture of biased compound Poisson processes that appear in the "Cox process representation" of [CJ23].

Let us write (as in [CJ23]) $F_t(s) := \mathbf{E}[s^{N_t}]$, and abbreviate

$$B(t,T,\ell) := \frac{1}{1 - F_T(0)} \int_0^1 F_{T-t}(s)^{\ell-1} F_T'(s)\,\mathrm{d}s. \tag{4.2.2}$$

[CJ23, Theorem 2.4] (as well as Theorem 4.2.1) says that the rate of size ℓ reproduction along the uniform ancestral lineage at time t is

$$r\ell p_\ell\, B(t,T,\ell).$$

This can be obtained from Corollary 4.2.2 by noting that S has density

$$\frac{F_T'(s)}{1-F_t(0)}.$$

In this sense the factor $B(t,T,\ell)$ can be interpreted as an *(ancestral) rate bias*, on top of the classical term $r\ell p_\ell$. Indeed, the factor $B(t,T,\ell)$ is absent in trees that are biased with respect to their size at time T. Galton-Watson trees of this kind have been investigated (also in the multitype case) by Georgii and Baake [GB03, Section 4]; they are continuous-time analogues of the size-biased trees analysed by Lyons et al. [LPP95] and Kurtz et al. [KLPP97].

In the critical and supercritical case one can check that, for all fixed $u < T$ and $\ell \in \mathbb{N}$ one has the convergence $B(T-u,T,\ell) \to 1$ as $T \to \infty$ because S converges to 1 in probability. In the supercritical case this stabilisation along the sampled ancestral lineage corresponds to the "retrospective viewpoint" that has been taken in [GB03] and, in the more general situation of Crump-Mode-Jagers processes, by Jagers and Nerman [JN96]. The choice $\mu = \delta_1$ renders the case of discrete-time Galton-Watson processes, starting with one individual at time 0 and with reproduction events at times $1, 2, \ldots$. Then, with $T = n \in \mathbb{N}$, and $L_1, \ldots, L_n$ being the family sizes along the ancestral lineage of the sampled individual V, the formula (4.2.1) specialises to

$$\mathbf{P}\left(N_n > 0,\, L_1 = \ell_1, \ldots, L_n = \ell_n,\, S \in \mathrm{d}s\right) = \left(\prod_{i=1}^{n} \ell_i p_{\ell_i}\, \mathbf{E}\left[s^{N_{n-i}}\right]^{\ell_i - 1}\right) \mathrm{d}s. \qquad (4.2.3)$$

4.3 Maxima of i.i.d. random markers

As a preparation for the short probabilistic proof of Theorem 4.2.1 given in the next section, we recall the following well-know fact: Denote by Unif$[0,1]$ the uniform distribution on the interval $[0,1]$. For $\ell \in \mathbb{N}$, let $\widetilde{S}$ be the maximum of ℓ independent Unif$[0,1]$-distributed random variables $U_1, \ldots, U_\ell$. Then the density of $\widetilde{S}$ is

$$\mathbf{P}\left(\widetilde{S} \in \mathrm{d}s\right) = \ell s^{\ell-1}\,\mathrm{d}s, \quad 0 \le s \le 1. \qquad (4.3.1)$$

Indeed, because of exchangeability,

$$\mathbf{P}\left(\widetilde{S} \in \mathrm{d}s\right) = \ell\,\mathbf{P}\left(U_1 \in \mathrm{d}s\right)\mathbf{P}\left(U_2 < s, \ldots, U_\ell < s\right),$$

which equals the r.h.s. of (4.3.1).

The following lemma specialises to (4.3.1) when putting $\tilde{N} \equiv 1$.

Lemma 4.3.1. *Let $\widetilde{N}$ be an $\mathbb{N}_0$-valued random variable, and $\widetilde{N}_1, \widetilde{N}_2, \ldots$ be i.i.d. copies of $\widetilde{N}$. Given $\widetilde{N}_1, \widetilde{N}_2, \ldots$ let $U_{1,1}, \ldots U_{1,\widetilde{N}_1}, U_{2,1}, \ldots U_{2,\widetilde{N}_2}, \ldots$ be independent Unif$[0,1]$-distributed random variables, and write*

$$S_k \;:=\; \max\left\{U_{k,1}, \ldots, U_{k,\widetilde{N}_k}\right\}, \quad k = 1, 2, \ldots$$

$$S^{(\ell)} \quad := \quad \max\{S_1, \ldots, S_\ell\}\,, \quad \ell \in \mathbb{N}$$

where we put $\max(\emptyset) := -\infty$*. Then, for all* $\ell \in \mathbb{N}$*, the density of* $S^{(\ell)}$ *is*

$$\mathbf{P}\left(\widetilde{N}_1 + \ldots + \widetilde{N}_\ell > 0,\, S^{(\ell)} \in \mathrm{d}s\right) = \ell\, \mathbf{E}\left[s^{\widetilde{N}}\right]^{\ell-1} \mathbf{P}\left(\widetilde{N}_1 > 0,\, S_1 \in \mathrm{d}s\right), \quad 0 \le s \le 1. \tag{4.3.2}$$

Proof. Again because of exchangeability, the l.h.s. of (4.3.2) equals

$$\ell\, \mathbf{P}\left(\widetilde{N}_1 > 0,\, S_1 \in \mathrm{d}s\right) \mathbf{P}\left(S_2 < s, \ldots, S_\ell < s\right) \tag{4.3.3}$$

for $s \in [0,1]$. Since by assumption the S_k are i.i.d. copies of S_1, the rightmost factor in (4.3.3) equals

$$\mathbf{P}\left(S_1 < s\right)^{\ell-1} = \mathbf{E}\left[\mathbf{P}\left(S_1 < s \mid \widetilde{N}_1\right)\right]^{\ell-1} = \mathbf{E}\left[s^{\widetilde{N}_1}\right]^{\ell-1} = \mathbf{E}\left[s^{\widetilde{N}}\right]^{\ell-1}.$$

Hence, (4.3.3) equals the r.h.s. of (4.3.2), completing the proof of the lemma. □

The following corollary is immediate.

Corollary 4.3.2. *Let* L *be an* $\mathbb{N}_0$*-valued random variable that is independent of all the random variables appearing in Lemma 4.3.1, with* $\mathbf{P}(L = \ell) = p_\ell$*,* $\ell \in \mathbb{N}_0$*. Then we have for all* $\ell \in \mathbb{N}_0$*,*

$$\mathbf{P}\left(L = \ell,\, \widetilde{N}_1 + \ldots + \widetilde{N}_\ell > 0,\, S^{(\ell)} \in \mathrm{d}s\right) = \ell p_\ell\, \mathbf{E}\left[s^{\widetilde{N}}\right]^{\ell-1} \mathbf{P}\left(\widetilde{N}_1 > 0,\, S_1 \in \mathrm{d}s\right), \quad 0 \le s \le 1. \tag{4.3.4}$$

4.4 Proof of Theorem 4.2.1

We prove the statement (4.2.1) by induction over j, *simultaneously* over all time horizons $T > 0$. We write $\mathbf{P}^T$ for the probability referring to time horizon T; this will be helpful in the induction step where we will encounter two different time horizons.
For $j = 0$, both sides of (4.2.1) are equal to $\mu((T, \infty))\, \mathrm{d}s$.
For $j = 1$, on the event $\{T_1 \in \mathrm{d}t_1\}$, we can directly apply Corollary 4.3.2 to the markers of the L_1 subtrees produced in this event. These subtrees live $T - t_1$ long and thus have sizes distributed as N_{T-t_1}. So the left side of (4.2.1) equals

$$\mathbf{P}\left(\tau_1 \in \mathrm{d}t_1\right) p_{\ell_1} \ell_1\, \mathbf{E}\left[s^{N_{T-t_1}}\right]^{\ell_1 - 1} \mathbf{P}^{T-t_1}\left(T_1 > T - t_1,\, S \in \mathrm{d}s\right),$$

which is using the $j = 0$ case. This is equal to the right hand side of (4.2.1).
Now assume we have proved (4.2.1) for all time horizons T' with $j - 1$ (in place of j), for all times $t'_1, \ldots, t'_{j-1} \le T'$, sizes $\ell'_1, \ldots, \ell'_{j-1} \in \mathbb{N}$ and $s \in [0,1]$. On the event $\{T_1 \in \mathrm{d}t_1, L_1 = \ell_1\}$ the descendants of the ℓ_1 siblings in the first branching event form ℓ_1 independent and

identically distributed trees on the time interval $[t_1, T]$. Thus, using Corollary 4.3.2 and setting $t_1' := t_2 - t_1, \ldots, t_{j-1}' = t_j - t_1$, we obtain that the left hand side of (4.2.1) equals

$$\mathbf{P}(\tau_1 \in \mathrm{d}t_1)\, p_{\ell_1} \ell_1\, \mathbf{E}\left[s^{N_{T-t_1}}\right]^{\ell_1 - 1} \cdot \mathbf{P}^{T-t_1}\left(J = j-1,\, T_1 \in \mathrm{d}t_1{}',\ldots,T_{j-1} \in \mathrm{d}t_{j-1}{}',\, N_{T-t_1} > 0,\, S \in \mathrm{d}s\right). \tag{4.4.1}$$

By the induction assumption, this is equal to

$$\mathbf{P}(\tau_1 \in \mathrm{d}t_1)\, p_{\ell_1} \ell_1\, \mathbf{P}\left(\tau_1' \in \mathrm{d}t_1{}',\ldots,\tau_{j-1}' \in \mathrm{d}t_{j-1}',\, \tau_j' \geq T - t_1\right) \prod_{i=2}^{j} \left(\ell_i p_{\ell_i}\, \mathbf{E}\left[s^{N_{T-t_i}}\right]^{\ell_i - 1}\right), \tag{4.4.2}$$

where $(\tau_1', \tau_2', \ldots)$ have the same distribution as $(\tau_1, \tau_2, \ldots)$. Obviously (4.4.2) equals the r.h.s of (4.2.1). This completes the induction step and concludes the proof. □

4.5 Conditioning on a marker value

Chauvin, Rouault and Wakolbinger [CRW91] consider a Markov process with an atomless transition probability indexed by a continuous-time Galton-Watson-tree and they then condition on an individual at time T to be at a given location.

To relate this to the framework described in the Introduction, we assume that each individual alive at time T in the Bellmann-Harris tree carries a marker in some standard Borel space E and these random marks have the following properties:

(M1) Their marginal distributions (denoted by ν) are identical and do not depend on the reproduction events

(M2) A.s. no pair of marks is equal.

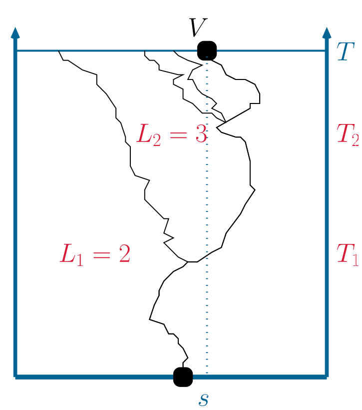


Figure 4.2: An example for a realisation of the random variables L_1, L_2, T_1, T_2 in the sampling regime described in Section 4.5.

Think for example of branching Brownian motion: The positions of the different particles clearly depend on each other via the genealogy, however, at time t the marginal distribution of the position of each particle is a centered Gaussian random variable with variance t, irrespective of its past genealogical events in the underlying continuous-time Galton-Watson tree. Thus (M1), is fulfilled. Since two correlated Gaussian random variables are a.s. not equal if the correlation coefficient is not equal to one, (M2) is also fulfilled.

We now condition on $\{N_T > 0\}$ and, for given $s \in E$, on one of the N_T individuals having marker value s. Remember the previous notation: Denote by V the individual having marker

s. Let J be the random number of reproduction events along the ancestral lineage of V and $0 < T_1 < T_2 < \cdots < T_J < T$ be the random times of these reproduction events. Let $L_1, \dots, L_J$ be the offspring sizes in these reproduction events and let $0 < \tau_1 < \tau_2 < \cdots$ be the random arrival times in a renewal process with interarrival time distribution μ. Figure 5.2 depicts a sample realisation.

The following Theorem generalises (part of) [CRW91, Theorem 2] to general lifetime time distributions.

Theorem 4.5.1. *For $j \geq 0$, $0 < t_1 < \ldots < t_j < T$ and $\ell_1, \dots, \ell_j \in \mathbb{N}$ we have for ν-almost all s*

$$\begin{aligned} &\mathbf{P}\left(J = j,\, T_1 \in \mathrm{d}t_1, \dots, T_j \in \mathrm{d}t_j,\, L_1 = \ell_1, \dots, L_j = \ell_j \middle|\, N_T > 0, \exists\, \text{marker} \in \mathrm{d}s\right) \\ &= \frac{1}{\mathbf{E}[N_T]} \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \dots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} \geq T\right) \prod_{i=1}^{j} \ell_i p_{\ell_i}. \end{aligned} \tag{4.5.1}$$

Proof. Because of properties (M1), (M2) we have

$$\mathbf{P}(N_T > 0, \exists\, \text{marker} \in \mathrm{d}s) = \mathbf{E}[N_T]\nu(\mathrm{d}s), \quad s \in E.$$

Hence (4.5.1) is equivalent to

$$\begin{aligned} &\mathbf{P}\left(J = j,\, T_1 \in \mathrm{d}t_1, \dots, T_j \in \mathrm{d}t_j,\, L_1 = \ell_1, \dots, L_j = \ell_j, N_T > 0, \exists\, \text{marker} \in \mathrm{d}s\right) \\ &= \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \dots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} \geq T\right) \prod_{i=1}^{j} \ell_i p_{\ell_i}\, \nu(\mathrm{d}s). \end{aligned} \tag{4.5.2}$$

As in the proof of Theorem 4.2.1 we prove the statement (4.5.2) by induction over j, *simultaneously* over all time horizons $T > 0$. As before we write $\mathbf{P}^T$ for the probability referring to time horizon T. For $j = 0$ the statement is true, since

$$\mathbf{P}^T(J = 0, N_T > 0, \exists\, \text{marker} \in \mathrm{d}s) = \mathbf{P}\left(\tau_1 \leq T\right)\, \nu(ds).$$

Assume we have proved (4.5.2) for all time horizons T' with $j - 1$ (in place of j), for all times $t'_1, \dots, t'_{j-1} \leq T'$, sizes $\ell'_1, \dots, \ell'_{j-1} \in \mathbb{N}$ and marker distributions with the same marginal ν that satisfy conditions (M1), (M2). Turning to (4.5.2) as it stands, we note that on $\{T_1 = t_1, L_1 = \ell_1\}$, the descendants of the ℓ_1 siblings in the first branching event form ℓ_1 independent and identically distributed trees on the time interval $[t_1, T]$. Let $\mathcal{U}_k$, $k = 1, \dots, \ell_1$, be the set of markers of the individuals at time T that descend from the k-th sibling. By randomly permuting these ℓ_1 siblings, we can assume that the set-valued random variables $\mathcal{U}_k$, $k = 1, \dots, \ell_1$, are exchangeable. Note that the markers in each $\mathcal{U}_k$ satisfy conditions (M1), (M2). Because the markers are a.s. pairwise different by assumption, the marker s belongs to at most one of those $\mathcal{U}_k$, so

$$\mathbf{1}_{\{\exists\, \text{marker} \in \mathrm{d}s\}} = \sum_{k=1}^{\ell_1} \mathbf{1}_{\{\mathcal{U}_k \cap \mathrm{d}s \neq \emptyset\}} \text{ a.s.} \tag{4.5.3}$$

Note that for the sake of intuition we use a differential notation for what formally is an (integral) equality for the distribution of the random point measure formed by the individuals' markers, which by assumption (M2) can be seen as a random set of points.
Putting $t'_1 := t_2 - t_1, \ldots, t'_{j-1} := t_j - t_1$ we thus infer, using the branching property of the Bellman-Harris tree, that the left hand side of (4.5.2) equals

$$\begin{aligned} &\mathbf{P}(\tau_1 \in \mathrm{d}t_1) p_{\ell_1} \ell_1 \\ &\cdot \mathbf{P}^{T-t_1} \Big(J = j-1,\, T_1 \in \mathrm{d}t'_1, \ldots, T_{j-1} \in \mathrm{d}t'_{j-1}, L_1 = \ell_2, \ldots, L_{j-1} = \ell_j, N_{T-t_1} > 0, \exists\, \text{mark} \in \mathrm{d}s \Big). \end{aligned} \tag{4.5.4}$$

By the induction assumption this is equal to

$$\mathbf{P}(\tau_1 \in \mathrm{d}t_1) p_{\ell_1} \ell_1 \, \mathbf{P} \left(\tau'_1 \in \mathrm{d}t'_1, \ldots, \tau'_{j-1} \in \mathrm{d}t'_{j-1}, \tau'_j \geq T - t_1 \right) \prod_{i=2}^{j} \ell_i p_{\ell_i} \, \nu(\mathrm{d}s), \tag{4.5.5}$$

where $(\tau'_1, \tau'_2, \ldots)$ have the same distribution as $(\tau_1, \tau_2, \ldots)$. Obviously (4.5.5) equals the r.h.s. of (4.5.2), which completes the induction step and concludes the proof. □

Remark 4.5.2. If μ is the exponential distribution with parameter r, then $\tau_1, \tau_2, \ldots$ are again the points of a rate r Poisson point process and (4.5.1) implies that reproduction events along the ancestral lineage of V happen according to a time-homogeneous Poisson process with rate $r \sum_\ell \ell p_\ell$. This corresponds to the description of the events along the ancestral line of V given in [CRW91, Theorem 1].

4.6 Sampling the left-most ancestral lineage

We now aim to obtain results about what Geiger [Gei99] calls the leftmost surviving ancestral lineage in a planar embedding of the tree: At any reproduction event we assign independent uniformly on $[0,1]$ distributed markers to all children. An individual can now be uniquely determined by the markers along its ancestral lineage. On the event $\{N_T > 0\}$, let V be the individual whose markers along the entire ancestral lineage comes first in the lexicographic ordering. Let J be the random number of reproduction events and $0 < T_1 < T_2 < \cdots < T_J \leq T$ be the random times of reproduction events along the ancestral lineage of V. Let $L_1, \ldots, L_J$ be the offspring sizes in these reproduction events and let $0 < \tau_1 < \tau_2 < \cdots$ be the random

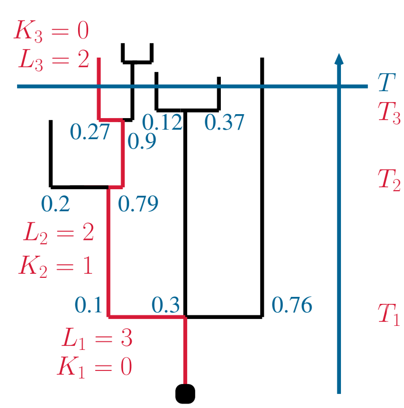


Figure 4.3: An example for a realisation of markers and random variables $L_1, L_2, K_1, k_2, T_1, T_2$ in the sampling regime described in Section 4.6.

arrival times in a renewal process with interarrival time distribution μ. Denote by K_i the number of siblings born at reproduction event number i along the ancestral lineage of V which have a lower lexicographic order than V and whose descendants hence die out before time T. Figure 4.3 shows a realisation for this sampling rule.

Theorem 4.6.1. *For $j \geq 0$, $0 < t_1 < \ldots < t_j < T$, $\ell_1, \ldots, \ell_j \in \mathbb{N}$ and $k_i \in \{1, \ldots, \ell_i - 1\}$ we have*

$$\begin{aligned}&\mathbf{P}\left(N_T > 0, J = j, T_1 \in \mathrm{d}t_1, \ldots, T_j \in \mathrm{d}t_j, L_1 = \ell_1, \ldots, L_j = \ell_j, K_1 = k_1, \ldots, K_j = k_j\right) \\ &= \mathbf{P}\left(\tau_1 \in \mathrm{d}t_1, \ldots, \tau_j \in \mathrm{d}t_j, \tau_{j+1} \geq T\right) \prod_{i=1}^{j} \left(p_{\ell_i} \mathbf{P}\left(N_{T-t_i} = 0\right)^{k_i}\right).\end{aligned} \tag{4.6.1}$$

Proof. The proof of the theorem works in analogy to the one of Theorem 4.2.1, but using following analogue of Lemma 4.3.1. □

Lemma 4.6.2. *Let $\widetilde{N}$ be an $\mathbb{N}_0$-valued random variable, and $\widetilde{N}_1, \widetilde{N}_2, \ldots$ be i.i.d. copies of $\widetilde{N}$. Given $\widetilde{N}_1, \widetilde{N}_2, \ldots$ let $U_1, U_2, \ldots$ be independent Unif$[0,1]$-distributed random variables, and write*

$$\begin{aligned} S^{(\ell)} &:= \min\left\{U_k \mid \widetilde{N}_k \geq 1, k = 1, \ldots, \ell\right\}, \\ K^{(\ell)} &:= \left|\left\{U_k \mid U_k < S^{(\ell)}, k = 1, \ldots, \ell\right\}\right| \end{aligned}$$

where we put $\min(\emptyset) := +\infty$. Then, for all $k < \ell \in \mathbb{N}$ we have

$$\mathbf{P}\left(\widetilde{N}_1 + \ldots + \widetilde{N}_\ell > 0,\ K^{(\ell)} = k\right) = \mathbf{P}\left(\widetilde{N} = 0\right)^k \mathbf{P}\left(\widetilde{N} > 0\right). \tag{4.6.2}$$

Proof. Because $S^{(\ell)}$ and $K^{(\ell)}$ do not depend on the order of $U_1, \ldots, U_\ell$, we can use exchangeability to assume that $U_1 < U_2 < \cdots < U_\ell$. For $K^{(\ell)}$ to be k, $S^{(\ell)}$ has then to be U_{k+1}. This is exactly the case if $\widetilde{N}_1, \ldots, \widetilde{N}_k = 0$ and $\widetilde{N}_{k+1} > 0$. □

4.7 Biological perspectives

Cheek and Johnston [CJ23, Section 5] discuss recent studies ([PMK+21], [CTAS+19]) which suggest that certain mutation rates are elevated for the earliest cell divisions in embryogenesis. Under the assumptions that (1) cell division times vary and (2) mutations arise not only *at* but also *between* cell divisions, Cheek and Johnston argue that this early rate elevation might be parsimoniously explained by their finding that in the supercritical case with no deaths the rate of branching events along a uniformly chosen ancestral lineage is increasing in $t \in [0, T]$ (which is a corollary to their Theorem 2.4).

The two-stage sampling rule

- first sample a random tree ("an adult") that survives up to time T,

- then sample an individual from this tree ("a cell from this adult") at time T

seems adequate for the situation discussed in Cheek and Johnston [CJ23, Section 5]. In other modeling situations, again with a large collection of i.i.d. Galton-Watson trees, one may think of a different sampling rule: Choose individuals at time T uniformly from the union of all time T individuals in all of the trees. This makes it more probable that the sampled individuals belong to larger trees, and in fact corresponds to the size-biasing of the random trees at time T ([GB03, Section 4]). In the two-stage sampling rule we see the different rate bias (4.2.2), discussed at the end of Section 4.2.

As can be seen from [CRW91, Theorem 1] (and Theorem 4.5.1), the rate bias (4.2.2) is also absent along the ancestral lineage of an individual whose marker has a prescribed value s, if one considers a situation in which a neutral marker evolves along the trees in small (continuous) mutation steps, and if one takes, for the prescribed value s, the collection of trees so large that one individual at time T has a marker value close to (ideally: precisely at) s.

The sampling rule that appears in [Gei99] (and Theorem 4.6.1) leads to a rate (and reproduction size) bias along the ancestral lineage that is different from the ones we just discussed. This sampling rule can be defined via i.i.d. real-valued valued neutral markers that are created at each birth and passed to the offspring. The individual sampled at time T (from the tree conditioned to survive up to time T) is the one whose marker sequence is the largest in lexicographic order among the individuals that live in the tree at time T. This interpretation appears of less biological relevance, except in the pure birth (or cell division) case, where one might think of one single marker that is passed on in each generation to a randomly chosen daughter cell.

Acknowledgment We thank Anton Wakolbinger for bringing the work [CJ23] to our attention. We are grateful to him and also to Matthias Birkner, Götz Kersting and Marius Schmidt for stimulating discussions and valuable hints. A substantial part of this work was done during the 2023 seminar week of the Frankfurt probability group in Haus Bergkranz.

Part III

Muller's ratchet

Chapter 5

Muller's ratchet in a near-critical regime: tournament versus fitness proportional selection[1]

[1]Appeared as [IGSW24]

Abstract

Muller's ratchet, in its prototype version, models a haploid, asexual population whose size N is constant over the generations. Slightly deleterious mutations are acquired along the lineages at a constant rate, and individuals carrying less mutations have a selective advantage. The classical variant considers *fitness proportional* selection, but other fitness schemes are conceivable as well. Inspired by the work of Etheridge et al. [EPW09] we propose a parameter scaling which fits well to the "near-critical" regime that was in the focus of [EPW09] (and in which the mutation-selection ratio diverges logarithmically as $N \to \infty$). Using a Moran model, we investigate the "rule of thumb" given in [EPW09] for the click rate of the "classical ratchet" by putting it into the context of new results on the long-time evolution of the size of the best class of the ratchet with (binary) tournament selection. This variant of Muller's ratchet was introduced in [GSW23], and was analysed there in a subcritical parameter regime. Other than that of the classical ratchet, the size of the best class of the tournament ratchet follows an autonomous dynamics up to the time of its extinction. It turns out that, under a suitable correspondence of the model parameters, this dynamics coincides with the so called Poisson profile approximation of the dynamics of the best class of the classcial ratchet.

5.1 Introduction

Muller's ratchet is a prototype model in population genetics. Originally it was conceived to explain the ubiquity of sexual reproduction among eukaryotes despite its many costs [Mul64, Fel74]. In its bare bones version, Muller's ratchet models a haploid, asexual population whose size N is constant over the generations. The neutral part of the random reproduction is given by a Wright-Fisher or a Moran dynamics. Slightly deleterious mutations are acquired along the lineages at a rate m, and individuals carrying less mutations have a selective advantage. The classical variant of Muller's ratchet considers *fitness proportional* selection, where the selective advantage of an individual carrying κ deleterious mutations over a contemporanean that carries a larger number κ' of deleterious mutations is $\frac{\mathfrak{s}}{N}(\kappa' - \kappa)$. Since the mutation mechanism is assumed to be unidirectional, every once in a while the type with the currently smallest number of mutations κ will disappear from the population. As Herbert Muller puts it in his pioneering paper [Mul64], "*an irreversible ratchet mechanism exists in the non-recombining species ... that prevents selection, even if intensified, from reducing the mutational loads below the lightest ..., whereas, contrariwise, 'drift', and what might be called 'selective noise' must allow occasional slips of the lightest loads in the direction of increased weight.*"

It is these "slips of the lightest loads" which are called *clicks of the ratchet*. The question "How often does the ratchet click?" was asked by Etheridge, Pfaffelhuber and one of the present authors in [EPW09], and there it was found that

$$\gamma := \frac{m}{\mathfrak{s} \log(Nm)} \tag{5.1.1}$$

is "*an important factor in determining the rate of the ratchet*". Specifically, under the assumption $1 \ll Nm \ll N$, [EPW09] states the following *Rule of Thumb* for the classical

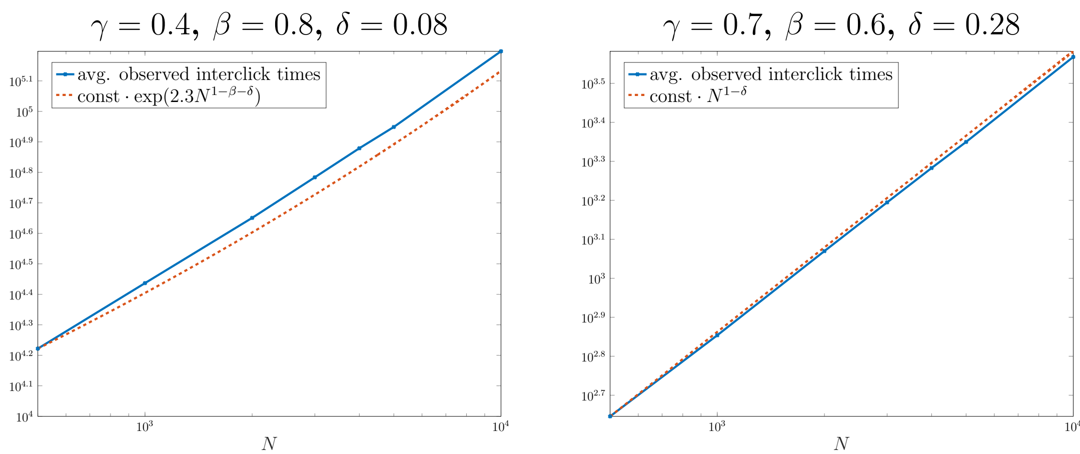
ratchet:
(RTC) *The rate of the (classical) ratchet is of the order* $N^{\gamma-1}m^{\gamma}$ *for* $\gamma \in (\frac{1}{2}, 1)$*, whereas it is exponentially slow in* $(Nm)^{1-\gamma}$ *for* $\gamma < \frac{1}{2}$*.*

With the *mutation-selection ratio*

$$\theta := \frac{m}{\mathfrak{s}},$$

(RTC) predicts the expected interclick time in the case $\gamma \in (\frac{1}{2}, 1)$ as

$$N(Nm)^{-\gamma} = Ne^{-\theta}.$$



(A) **(B)**

Figure 5.1: This is an illustration of the *Rule of Thumb* (RTC) predicting the order of magnitude of the interclick times of the classical ratchet. Each data point was obtained by pooling the interclick times no. 50 to 150 from 100 simulations of the (classical) ratchet for the corresponding parameter configuration (N, β, δ) in the (β, δ)-scaling (5.1.2). In the exponential regime, (RTC) predicts an order of magnitude $\exp(c\,N^{1-\beta-\delta})$ for the interclick times. In panel (A), we see that the constant c is difficult to estimate from simulations up to $N = 10^4$, but $c = 2.3$ as chosen there gives a reasonable fit. For the polynomial regime, (RTC) predicts the order of magnitude $N^{1-\delta}$, which fits very well the data in the situation of panel (B).

As observed by John Haigh ([Hai78]), in the deterministic limit ($N \to \infty$ and $m, \mathfrak{s}$ not depending on N) the type frequency profile in equilibrium becomes Poisson with parameter θ. Consequently, for $\gamma \in (\frac{1}{2}, 1)$ the rule (RTC) goes along with Haigh's prediction that the rate of the ratchet should be proportional to the inverse of the size of the best class.

For a polynomial mutation rate $m = N^{-\beta}$, $0 < \beta < 1$, the condition that γ remains constant (or at least bounded away from 0 and ∞) as $N \to \infty$ amounts to the requirement that the mutation-selection ratio θ is of the order $\log N$ as $N \to \infty$.

For the purpose of illustration we will consider a family of parameter scalings for (m, θ) which we call the (β, δ)-*scaling of the classical ratchet*:

$$m = N^{-\beta}, \qquad \theta = \delta \log N. \tag{5.1.2}$$

This amounts to *moderate mutation-selection*, with the mutation-selection ratio θ diverging logarithmically in N. The factor δ in front of $\log N$ turns out to be critical for the click rate. Indeed, in the (β, δ)-scaling, (5.1.1) takes the form

$$\gamma(\beta, \delta) = \frac{\delta}{1-\beta}.$$

The condition $0 < \gamma < 1$ from (RTC) restricts the pair (β, δ) to the triangle

$$\Delta := \{(\beta, \delta) : 0 < \beta,\ 0 < \delta < 1-\beta\}. \tag{5.1.3}$$

The *polynomial* and the *exponential regime* predicted by (RTC) correspond to

$$\mathcal{P} := \{\tfrac{1}{2} < \gamma(\beta,\delta) < 1\} = \{(\beta,\delta) \in \Delta : \tfrac{1}{2}(1-\beta) < \delta < 1-\beta\},$$

$$\mathcal{E} := \{0 < \gamma(\beta,\delta) < \tfrac{1}{2}\} = \{(\beta,\delta) \in \Delta : 0 < \delta < \tfrac{1}{2}(1-\beta)\},$$

and the predictions for the orders of magnitude of the expected interclick times take the form

$$N(Nm)^{-\gamma} = N^{1-\delta} \qquad \text{for} \quad \gamma \in (\tfrac{1}{2}, 1), \tag{5.1.4}$$

$$\exp\left(\text{const}(Nm)^{1-\gamma}\right) = \exp\left(\text{const}\, N^{1-\beta-\delta}\right) \qquad \text{for} \quad \gamma \in (0, \tfrac{1}{2}). \tag{5.1.5}$$

In view of the predicted transition from polynomial to exponential click rates we refer to $\mathcal{P} \cup \mathcal{E}$ as a *near-critical regime*. See Figure 5.1 for an illustration of (RTC) via simulations.

The evidence for (RTC) that is given in [EPW09] is based on a diffusion approximation for the evolution of the relative size X_0 of the *best class* (which consists of the individuals that carry the least amount of mutations in the current population). Because of the fitness proportional selection, the drift coefficient in this diffusion approximation contains the first moment M of the type frequency configuration $(X_0, X_1, \ldots)$. In order to obtain an approximate autonomous dynamics for X_0, the empirical first moment M has to be predicted based on X_0. A classical way to do this uses the so-called *Poisson profile approximation*, which we will explain in some detail in Section 5.3.

In the present paper we will consider a variant of Muller's ratchet in which fitness proportional selection is replaced by *(binary) tournament selection*. This kind of selection has been studied in the context of evolutionary computation ([BT96, BFM18]) and has found attention also in the biological literature [PBB+15]. In the ratchet's context this means that selective advantage of an individual carrying κ deleterious mutations over a contemporanean that carries a larger number κ' of deleterious mutations is constant (say $\frac{s}{N}$ for some $s = s_N > 0$), irrespective of the value of the difference $\kappa' - \kappa$. For the Moran version of the tournament ratchet, which was introduced in [GSW23] and whose definition we recall in Section 5.2, this means that "pairwise selective fights" are always won by the fitter individual.

Other than in the classical ratchet, the size of the (m, s)-tournament ratchet's best class follows an autonomous dynamics *up to its time of extinction*; at this time the class which was so far the second-best becomes the best one. As we will see in Section 5.3, this dynamics is *equal* to that of the Poisson profile approximation of the size of the classical $(m,\mathfrak{s})$-ratchet's best class, provided that

$$\rho := \frac{m}{s} = 1 - \exp(-m/\mathfrak{s}) = 1 - e^{-\theta}. \tag{5.1.6}$$

We now state a main finding of the present paper.

Rule of thumb for the near-critical tournament ratchet (RTT): *As $N \to \infty$, the expected time between clicks is*

$$\asymp \sqrt{\frac{N}{m}} \qquad \text{if} \quad Nm(1-\rho)^2 \to 0, \tag{5.1.7}$$

$$\asymp \exp\left(Nm(1-\rho)^2\right) \qquad \text{if} \quad Nm(1-\rho)^2 \to \infty. \tag{5.1.8}$$

Here and below, $\asymp$ stands for logarithmic equivalence, i.e. $a_N \asymp b_N$ means $\log a_N \sim \log b_N$, or equivalently $\frac{\log a_N}{\log b_N} \to 1$.

We will not give a complete proof of (RTT) in this work, but will present Theorem 5.3.4 which gives strong evidence for its validity. See Figure 5.2 for an illustration of (RTT) in the light of Theorem 5.3.4. In Remark 5.3.5 we will discuss what are the ingredients missing to go from Theorem 5.3.4 to a proof of (RTT), and we will also indicate a different route to the proof of (RTT), using the technique developed in [GSW23].

We emphasise that, in view of the correspondence (5.1.6), Theorem 5.3.4 also is a result on the asymptotics of the Poisson profile approximation of the classical ratchet, here in terms of Moran processes with mutation and selection. A similar asymptotics was obtained in [EPW09] heuristically by passing right away to the diffusion approximation for logistic branching processes.

In view of (5.1.6) we define, in analogy to (5.1.2), the (β,δ)-scaling for the tournament ratchet as

$$m = N^{-\beta}, \qquad \rho = \frac{m}{s} = 1 - N^{-\delta}.$$

With this scaling, (RTT) takes the following form: *As $N \to \infty$, the expected time between clicks is*

$$\asymp N^{\frac{1+\beta}{2}} \qquad \text{if} \quad (\beta,\delta) \in \mathcal{P}, \tag{5.1.9}$$

$$\asymp \exp\left(N^{1-\beta-2\delta}\right) \qquad \text{if} \quad (\beta,\delta) \in \mathcal{E}. \tag{5.1.10}$$

While both (RTC) and (RTT) state the same boundary ($\gamma = \frac{1}{2}$) between the polynomial and the exponential regime, the exponents differ between (5.1.4) and (5.1.9) as well as between (5.1.5) and (5.1.10). Specifically, in the polynomial regime $\mathcal{P}$ the exponent $\frac{1+\beta}{2}$ for the tournament ratchet is larger than the exponent $1-\delta$ for the classical ratchet.

Here is an explanation for the polynomial regime. The centers of attraction of the equilibrium profile weights of the best and the second best class differ asymptotically by the factor $\sqrt{1-\rho} = N^{\frac{\delta}{2}}$ for the tournament ratchet (see (5.3.5)), while they are given by the Poisson weights $e^{-\theta}$ and $\theta e^{-\theta}$ for the classical ratchet and hence for the latter differ only by the factor $\theta = \delta \log N$ (and thus have the same polynomial order $N^{1-\delta}$). This latter factor is only logarithmic in N; therefore, when starting the "new best class" at the time of a click in its "old" center of attraction, the tournament ratchet has a longer way to go than the classical ratchet. The exponent $\frac{1+\beta}{2}$ in (5.1.7) will be obtained by a Green function analysis in the proof of Theorem 5.3.4. This analysis will also explain the exponent $1-\delta$ in (5.1.4), which corresponds to Haigh's prediction, saying that "the interclick times are of the order of the size of the best class". An intuitive explanation for the appearance of the exponent $1-\beta-2\delta$ in (5.1.8) will be given at the end of Section 5.3.2. The reason why this exponent is different from the one appearing in (5.1.5) is that [EPW09] work here not with the Poisson profile approximation, but with (a rescaling of the diffusion approximation of) the so-called *relaxed Poisson profile approximation*.

Similar as [EPW09], the papers [PSW12, NS12, AP13, MPV20, BS22] used a diffusion approximation for the classical ratchet and modifications thereof. Metzger and Eule [ME13] consider, as a proxy to the classical ratchet, a two-type Moran model with selective advantage s of type 0 over type 1 and mutation rate m from type 0 to type 1. Their formula (8) corresponds to our formula (5.1.6) but their approximations for the classical ratchet concentrate on a regime in which θ remains bounded (see the discussion around [ME13, (23)], and also [WL10, (7),(8)]), whereas we focus here on a regime in which $\theta = \theta_N$ diverges logarithmically in N.

In [GSW23] it was discovered that the tournament ratchet has a dual which consists of a hierarchy of competing logistic processes. The main results of [GSW23] (on the click rate of the tournament ratchet and its type frequency profile between clicks) were obtained for the so-called subcritical regime (see Sec. 5.2.2) and were proved there via duality, with the help of recent results on logistic processes (see [Lam05, SaSha13, CCM16]). This "backward in time" view, which comes on top of an Ancestral Selection Graph decorated with mutation events, opens a route for proving the above stated result (RTT) and for analysing the type frequency profile of the tournament ratchet also in the near-critical regime. This will be pursued in future work.

In [GSW23] the rate of the tournament ratchet was identified in the subcritical regime (i.e. for $\rho = m/s < 1$ and not depending on N) up to logarithmic equivalence. Thus our Theorem 5.3.4 b), which is valid both for the near-critical *and* the subcritical regime, provides an essential step in sharpening the rate asymptotics of [GSW23] from logarithmic equivalence to asymptotic equivalence, see Remark 5.3.5a).

5.2 Muller's ratchet as a Moran process with mutation and selection

5.2.1 Model and basic concepts

In the Moran version of Muller's ratchet, neutral resampling within any ordered pair of individuals happens at rate $\frac{1}{2N}$, and mutation from κ to $\kappa+1$ takes place at rate m/N along each individual lineage. Selective reproduction for an individual i of type $\kappa(i)$ happens at rate $\frac{1}{N}\sum_j \Phi(\kappa(j)-\kappa(i))$, where the sum is taken over all those individuals j whose type $\kappa(j)$ is larger (and therefore "worse") than $\kappa(i)$. Here Φ is the *fitness function*, with $\Phi(0)=0$ and $\Phi(-d) = -\Phi(d)$ for $d \in \mathbb{N}$. For the classical case of *proportional selection*, one has $\Phi(\kappa' - \kappa) = \mathbf{s}\,(\kappa' - \kappa)$, while for the case of *(binary) tournament selection* one has $\Phi(\kappa' - \kappa) = \mathbf{s}\,(\mathbf{1}_{\{\kappa' > \kappa\}} - \mathbf{1}_{\{\kappa' < \kappa\}})$. In the sequel we will refer to these two Moran variants of Muller's ratchet briefly as the *classical ratchet* and the *tournament ratchet*. Both models have $(N, m, \mathbf{s})$ as their parameter triple, and in both models a crucial role is played by the *mutation-selection ratio* $\frac{m}{\mathbf{s}}$. In this section we reserve the symbol $\mathbf{s}$ for the selection parameter. Later, this will be specified as different parameters s and $\mathfrak{s}$ for the tournament and the classical ratchet, respectively. The following definition gives the rates for the type frequencies of the two ratchets.

Definition 5.2.1.

a) Writing N_κ for the current number of individuals of type κ, the jump rates are specified as follows:

   - Resampling: for $\kappa \neq \kappa'$,
     $(N_\kappa, N_{\kappa'})$ jumps to $(N_\kappa + 1, N_{\kappa'} - 1)$ at rate $\frac{1}{2N} N_\kappa N_{\kappa'}$

   - Mutation: for κ,
     $(N_\kappa, N_{\kappa+1})$ jumps to $(N_\kappa - 1, N_{\kappa+1} + 1)$ at rate $m N_\kappa$

   - Selection: for $\kappa < \kappa'$,
     $(N_\kappa, N_{\kappa'})$ jumps to $(N_\kappa + 1, N_{\kappa'} - 1)$ at rate $\begin{cases} \frac{\mathbf{s}}{N} N_\kappa N_{\kappa'} (\kappa' - \kappa) & \text{for the classical ratchet} \\ \frac{\mathbf{s}}{N} N_\kappa N_{\kappa'} & \text{for the tournament ratchet} \end{cases}$

b) The currently best type is

$$K^*(t) := \min\left\{\kappa \in \mathbb{N}_0 : N_\kappa(t) > 0\right\}.$$

c) The click times of the ratchet are the jump times of K^*, i.e. the times at which the currently best type is lost from the population. The type frequency profile seen from the currently best type has the (random) weights

$$X_k^{(N)}(t) := \frac{1}{N} N_{K^*(t)+k}(t), \qquad k = 0, 1, 2 \ldots \tag{5.2.1}$$

We say that a (non-random) type frequency profile $(p_k)_{k \in \mathbb{N}_0}$ obeys the mutation-selection equilibrium conditions (for the parameters m and $\mathbf{s}$) if

$$m(p_k - p_{k-1}) = \mathbf{s}\, p_k \left(\sum_{k' \in \mathbb{N}_0} p_{k'} \Phi(k' - k) \right), \quad k = 0, 1, 2 \ldots, \tag{5.2.2}$$

where we put $p_{-1} := 0$.

For the classical ratchet, (5.2.2) turns into

$$m(p_k - p_{k-1}) = \mathbf{s}\, p_k(\mu - k), \quad k = 0, 1, 2 \ldots, \tag{5.2.3}$$

where $\mu := \sum_\ell \ell p_\ell$ is the first moment of the profile. As already noticed by John Haigh ([Hai78]), (5.2.3) is solved by the *Poisson weights* with first moment $\mu = \frac{m}{\mathbf{s}}$. Indeed, this is the unique solution of (5.2.3) under the condition $p_0 > 0$.

For the tournament ratchet, (5.2.2) turns into

$$m\,(p_k - p_{k-1}) = \mathbf{s}\,p_k \left(\sum_{k' \in \mathbb{N}_0} p_{k'} \left(\mathbf{1}_{\{k'>k\}} - \mathbf{1}_{\{k'<k\}} \right) \right), \quad k = 0, 1, 2 \ldots \tag{5.2.4}$$

Here the condition $p_0 > 0$ leads to the requirement $m < \mathbf{s}$ and yields $p_0 = 1 - \frac{m}{\mathbf{s}}$. Various properties of the solution $(p_{k'})$ of (5.2.4) are stated in [GSW23] Theorem 2.3. The r.h.s. of (5.2.4) equals

$$\mathbf{s}\,p_k \left(1 - p_k - 2 \sum_{k'=0}^{k-1} p_{k'} \right), \quad k = 0, 1, 2 \ldots \tag{5.2.5}$$

A formal analogy between (5.2.3) and (5.2.4) results because (5.2.5) is close to $2\mathbf{s}\,p_k(\frac{1}{2} - g(k))$, where g is the cumulative distribution function of $(p_{k'})$. In this sense the role played by the profile's first moment in (5.2.3) is taken by the profile's median in (5.2.4).

5.2.2 The subcritical regime of the tournament ratchet

We now report briefly on the main results of the recent paper [GSW23]. The parameters of the tournament ratchet will be denoted by (m, s) and its mutation-selection ratio by $\rho := \frac{m}{s}$. In [GSW23], as $N \to \infty$, the mutation-selection ratio $\rho = \frac{m}{s}$ is kept constant and smaller than 1, and it is assumed that $m \to 0$ and $mN \to \infty$. (For technical reasons, mN is assumed to be of larger order of $\log \log N$, which keeps the regime slightly away from that of weak mutation, in which mN would be of order one as $N \to \infty$.) We will refer to this regime as the *subcritical regime* of the tournament ratchet. Informally stated, the main results of [GSW23] are

• *In the subcritical regime the click rate of the tournament ratchet on the* $\frac{1}{m}$*-timescale is, as* $N \to \infty$*, logarithmically equivalent to*

$$e^{-2Nm\left(\frac{1}{\rho} - 1 + \log \rho\right)}. \tag{5.2.6}$$

• *In the subcritical regime and for* N *large, the empirical type frequency profile at generic time points between clicks of the tournament ratchet is with high probability close to the mutation-selection equilibrium system* (p_k) *given by* (5.2.4) *with* $p_0 = 1 - \rho$.

See Theorems 2.2 and 2.3 in [GSW23], which there are proved via a hierarchical duality. As discussed in Remark 5.3.5.a), Theorem 5.3.4 b) can be considered as a significant step in sharpening (5.2.6) to an asymptotic equivalence.

5.3 A synopsis of the classical and the tournament ratchet

5.3.1 The dynamics of the best classes

For $k = 0, 1, \ldots$ let $Y_k^{\mathrm{C}}(t) = N_{K^*+k}^{\mathrm{C}}(t)$ and $Y_k^{\mathrm{T}}(t) = N_{K^*+k}^{\mathrm{T}}(t)$ be the sizes of the

$(k+1)^{\text{st}}$-best class of the classical and the tournament ratchet, where $(N_\kappa^{\mathrm{C}})_{\kappa\in\mathbb{N}_0}$ and $(N_\kappa^{\mathrm{T}})_{\kappa\in\mathbb{N}_0}$ follow the dynamics specified in Definition 5.2.1. Here we assume that the mutation rate m is equal for both ratchets, but the selection coefficients are different:

$$\mathbf{s} = \begin{cases} \frac{m}{\theta} =: \mathfrak{s} & \text{for the classical ratchet} \\ \frac{m}{\rho} =: s & \text{for the tournament ratchet.} \end{cases}$$

The jump rates from n to $n-1$ are given for both Y_0^{C} and Y_0^{T} by

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+m\right), \tag{5.3.1}$$

but the jump rates from n to $n+1$ are different: those of Y_0^{T} are

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+\frac{m}{\rho}\left(1-\frac{n}{N}\right)\right), \tag{5.3.2}$$

while those of Y_0^{C} are

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+\frac{m}{\theta}\sum_{k=1}^{\infty} kX_k\right) \tag{5.3.3}$$

where $(X_k(t))_{k\in\mathbb{N}_0}$ is the type frequency profile as defined in (5.2.1), with (N_κ^{C}) in place of (N_κ). Writing

$$M(t) := \sum_{k=1}^{\infty} kX_k(t)$$

for the first moment of the type frequency profile (X_k), the upward jump rate (5.3.3) takes the form

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+m\frac{M}{\theta}\right). \tag{5.3.4}$$

An inspection of the jump rates in Definition 5.2.1 reveals that for each $k \in \mathbb{N}$ the process $(Y_0^{\mathrm{T}}, \ldots, Y_k^{\mathrm{T}})$ obeys an autonomous dynamics up to the extinction time of Y_0^{T}; for $k = 0$ this is evident from (5.3.1) and (5.3.2). For later reference we note here that $(Y_0^{\mathrm{T}}, Y_1^{\mathrm{T}})$ has, asymptotically as $N \to \infty$, the center of attraction

$$(\mathfrak{a}, \mathfrak{b}) := \left(N(1-\rho), N\sqrt{1-\rho}\right) \tag{5.3.5}$$

provided $Nm \to \infty$ and $\rho \to 1$. To see this, note that the dynamics of $(Y_0^{\mathrm{T}}, Y_1^{\mathrm{T}})$ is autonomous up to the first hitting of $\{0\} \times \{0, \ldots, N\}$, and that the states of $(Y_0^{\mathrm{T}}, Y_1^{\mathrm{T}})$ for which the upward jump rates are asymptotically equal to the downward jump rates have the asymptotic (Np_0, Np_1), with (p_0, p_1) given by (5.2.4) and (5.2.5). In addition to $p_0 = 1-\rho$, this leads to the equation

$$p_1(1-p_1-2(1-\rho)) = \rho(p_1-(1-\rho)),$$

with the solution $p_1 = \sqrt{1-\rho}\left(\sqrt{\rho + \frac{1}{4}(1-\rho)} - \frac{1}{2}\sqrt{1-\rho}\right) \sim \sqrt{1-\rho}$ as $\rho \uparrow 1$.

In contrast to the tournament ratchet, the rates (5.3.4) depend not only on the size of the best class but also on the profile $(X_k(t))_{k\geq 0}$ (via its first moment $M(t)$). There are various ways to predict $M(t)$ on the basis of $Y_0^{\mathrm{C}}(t)$, and thereby to replace (5.3.4) by a rate which is autonomous. One of them will be described in the remainder of this section, a second one will be addressed in Remark 5.3.2. As observed already by John Haigh [Hai78], such a strategy requires a regime in which "genetic drift", i.e. the fluctuations due to neutral reproduction, needs a time to take Y_0^{C} to extinction which is large compared to the time which the noiseless classical ratchet needs to "relax" towards its (new) equilibrium. The dynamics of the latter is

$$\mathrm{d}x_k(t) = \left(\mathfrak{s}\sum_{\ell} x_\ell(\ell - k) + m(x_{k-1}(t) - x_k(t))\right)\mathrm{d}t, \qquad k = 0, 1, \dots \tag{5.3.6}$$

(with $x_{-1} \equiv 0$). As already indicated after (5.2.3), the unique vector of probability weights on $\mathbb{N}_0$ which has a non-vanishing weight at 0 and is a stationary point of (5.3.6) is given by the *Poisson profile*

$$\pi_k = e^{-\theta}\frac{\theta^k}{k!}, \quad k \geq 0. \tag{5.3.7}$$

For the initial profile

$$x(0) := \tfrac{1}{1-\pi_0}(\pi_1, \pi_2, \dots),$$

the *relaxation time* τ which it takes for $x_0(t)$ to come down from $\frac{1}{1-\pi_0}\pi_1$ to $\frac{e}{e-1}\pi_0$ turns out to be

$$\tau = \frac{\log\theta}{\mathfrak{s}},$$

(see [EPW09, Remark 4.3][2]). The time to extinction of a neutral Moran(N)-process starting in $N\pi_0 = Ne^{-\theta}$ is of the order $Ne^{-\theta}$. Haigh's requirement can thus be formulated as

$$Ne^{-\theta} \gg \frac{\log\theta}{\mathfrak{s}},$$

which in the (β, δ)-scaling (5.1.2) just means that $\beta + \delta < 1$.

5.3.2 The Poisson profile approximation for the classical ratchet

Here the idea is to think of the profile $(X_k)_{k\geq 1}$ as (nearly) proportional to the Poisson profile (5.3.7), and as the mass $\pi_0 - X_0$ being distributed proportionally upon this profile. This leads to the so-called *Poisson profile approximation* of $(X_k)_{k\geq 1}$ based on X_0, given by

$$\Pi(X_0) := \left(X_0, \frac{1-X_0}{1-\pi_0}(\pi_1, \pi_2, \dots)\right).$$

[2] In order to ease the look-up we use here and below the numbering of the arxiv version of [EPW09], which otherwise is identical in content with the version published in the LMS Lecture Note Series.

(cf [EPW09, (2.5)]). The first moment of $\Pi(X_0)$ is

$$M(X_0) := (1 - X_0)\,\frac{\theta}{1-\pi_0},$$

in accordance with [EPW09, (5.3a)]. Plugging this into (5.3.4) in place of M leads to the following *Poisson profile approximation* of the upward jump rates (5.3.4):

$$n\left(\frac{1}{2}\left(1-\frac{n}{N}\right)+\frac{m}{1-e^{-\theta}}\left(1-\frac{n}{N}\right)\right). \tag{5.3.8}$$

We denote the birth-and death-process on $\mathbb{N}_0$ with downward jump rates (5.3.1) and upward jump rates (5.3.8) by Y_{PPA}; this process can be seen as an approximation of Y_0^{C}.

Remark 5.3.1. A crucial observation is that the upward jump rates (5.3.8) and (5.3.2) are equal if and only if $\rho = 1 - e^{-\theta}$. In other words, under the "dictionary" (5.1.6), the jump rates (5.3.1) and (5.3.2) of the size of the best class of the (m, s)-tournament ratchet are equal to the jump rates (5.3.1) and (5.3.8) of the Poisson profile approximation for the size of the best class of the classical $(m, \mathfrak{s})$-ratchet.

Remark 5.3.2. Not least to provide a systematic framework for previous approaches ([SCS93, GC00]) to the approximation of the size of the ratchet's best class, [EPW09] embedded the Poisson profile approximation (PPA) into a one-parameter family RPPA(A), $A \geq 0$, the so-called *relaxed Poisson profile approximations*. Roughly, the idea was to take some *delay* into account for the prediction of M based on X_0. For $A = 1$, this results (see [EPW09, (5.3b)]) in

$$M(X_0) := \theta + \frac{1}{e-1}\left(1-\frac{X_0}{\pi_0}\right), \tag{5.3.9}$$

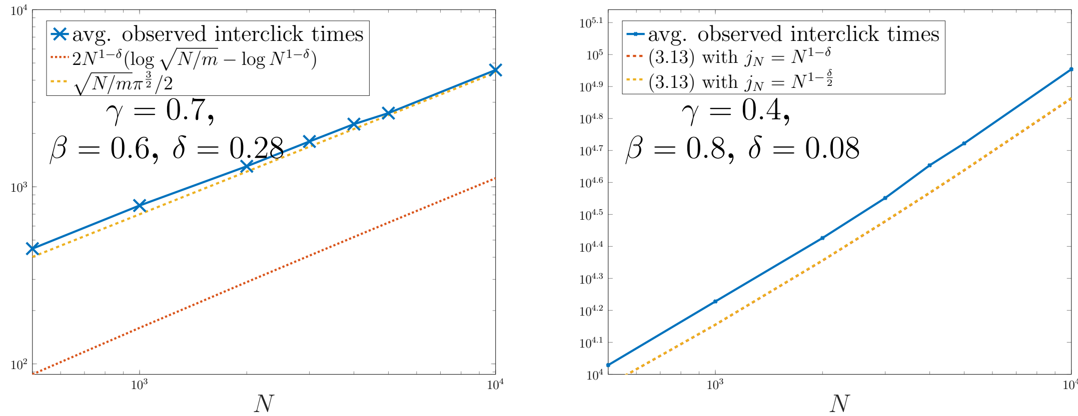
which then is plugged into the upward jump rate (5.3.4) in place of M. In Figure 5.6 we compare the quality of the PPA and RPPA(1) approximations for the rate of the classical ratchet in the light of simulations of our Moran model.

5.3.3 On the expected time to extinction of the best class

In this subsection we focus on the birth-and-death process $Y := Y_0^{\mathrm{T}}$ with jump rates (5.3.1) and (5.3.2). As observed in Remark 5.3.1, this process has the same dynamics as the process Y_{PPA} defined in Section 5.3.2, provided the mutation rates are equal and the selection coefficients are translated through the "dictionary" (5.1.6).

Remark 5.3.3. Before turning to a rigorous analysis, let us give a heuristics for the long-term behaviour of Y, which also points towards (RTT) as well as part of (RTC). The rates (5.3.1) and (5.3.2) display 3 parts: the fluctuation terms $\pm\frac{n}{2}\left(1-\frac{n}{N}\right)$, the net linear birth rate $n\frac{m}{\rho}(1-\rho)$ and the quadratic death rate $\frac{m}{\rho}\frac{n^2}{N}$. The center of attraction of Y (which we encountered already in (5.3.5)) is that (asymptotic) value of n for which the net linear birth rate equals the quadratic death rate and thus equals

$$\mathfrak{a} = N(1-\rho).$$



(A) **(B)**

Figure 5.2: This is an illustration of the *Rule of Thumb for the tournament ratchet* (RTT) in the light of Theorem 5.3.4. Each data point was obtained by pooling the interclick times no. 50 to 150 from 100 simulations of the tournament ratchet for the corresponding value of N. Here, in panel (A) $(\beta, \delta) = (0.6, 0.28)$, which belongs to the polynomial regime $\mathcal{P}$, and in panel (B) $(\beta, \delta) = (0.8, 0.08)$, which belongs to the exponential regime $\mathcal{E}$. Each panel shows two predictions based on the asymptotics of Theorem 5.3.4, using the initial values $\mathfrak{a} = N^{1-\delta}$ and $\mathfrak{b} = N^{1-\delta/2}$, respectively. In the exponential regime the predictions using $\mathfrak{a}$ and $\mathfrak{b}$, respectively, are virtually indistinguishable, while in the polynomial regime the prediction using $\mathfrak{b}$ is by far better than the one using $\mathfrak{a}$.

As long as Y is below $\mathfrak{a}/2$, it is stochastically bounded from below by a binary Galton-Watson process Y^ℓ with supercriticality $m(1-\rho)/2$, and stochastically bounded from above by a binary Galton-Watson process Y^u with supercriticality $m(1-\rho)$. By Haldane's formula (which in this case coincides with the formula for the escape probability of a simple random walk with constant drift), the survival probability of the offspring of one individual in Y^ℓ (resp Y^u) is $\sim N^{-\beta-\delta}$ (resp. $\sim 2N^{-\beta-\delta}$). Hence the probability that Y when starting in $\mathfrak{a}/4$ hits 0 before reaching $\mathfrak{a}/2$, is asymptotically between $\left(1 - 2N^{-\beta-\delta}\right)^{N^{1-\delta}/4}$ and $\left(1 - N^{-\beta-\delta}\right)^{N^{1-\delta}/4}$, which converge to 0 if and only if $1 - \beta - 2\delta > 0$, i.e. $\gamma > \frac{1}{2}$. In this case the number of excursions which Y makes from $\mathfrak{a}/4$ up to $\mathfrak{a}/2$ before going extinct is geometric with expectation asymptotically between $\exp\left(\frac{1}{4}N^{1-\beta-2\delta}\right)$ and $\exp\left(\frac{1}{2}N^{1-\beta-2\delta}\right)$. This gives an intuitive explanation why $\gamma = \frac{1}{2}$ is the bound between the exponential and the polynomial regime, and also sheds light on the result of Theorem 5.3.4

In the case $\gamma > \frac{1}{2}$, the center of attraction plays a negligible role. What becomes relevant then is that threshold for n above which the quadratic death rate $\frac{m}{\rho}\frac{n^2}{N}$ becomes large. Obviously, the order of magnitude of this threshold is $\sqrt{\frac{N}{m}} = N^{\frac{1+\beta}{2}}$. Above this threshold, Y is strongly pushed downwards, making the time spent above the threshold negligible. Below the threshold, Y behaves similar to a (driftless linear) birth-and-death process with upward and downward jump rates (5.3.1). This gives a qualitative explanation of the orders of magnitude of the expected times to extinction that are obtained in Theorem 5.3.4 also for the polynomial regime.

The proof of the following theorem is the content of Section 5.4. This proof relies on an asymptotic analysis of the Green function represented by formula (5.4.1). The fit of a numerical calculation of the Green function based on this formula with the empirical occupation times of the size of the best class of the tournament ratchet is displayed in Figure 5.3. A heuristic explanation of the orders obtained in Theorem 5.3.4 has been given in Remark 5.3.3.

In the following we use the notation

$$f(N) \ll g(N) \iff \lim_{N\to\infty} \frac{f(N)}{g(N)} = 0. \tag{5.3.10}$$

Also, we will usually suppress the N-dependence in the notation, as for example in Y, m and ρ in the following theorem. Note that this theorem comprises a larger regime than the one described by the (β, δ)-scaling for $(\beta, \delta) \in \Delta$, see (5.1.3).

Theorem 5.3.4. *Let T_0 be the extinction time of the birth-and-death process Y with jump rates (5.3.1) and (5.3.2), let $1 \gg m \gg \frac{1}{N}$, and let ρ be a sequence in $[\rho_0, 1)$ for some fixed $\rho_0 \in (0,1)$.*

a) [Polynomial regime] *Assume $Nm(1-\rho)^2 \to 0$ as $N \to \infty$. Let (j_N) be a sequence of natural numbers in $[N]$. If $j_N \ll \sqrt{\frac{N/m}{\log(N/m)}}$, then*

$$\mathbf{E}_{j_N}[T_0] \sim 2j_N \left(\log \sqrt{\frac{N}{m}} - \log j_N \right), \tag{5.3.11}$$

whereas if $j_N \gg \sqrt{\frac{N}{m}}$, then

$$\mathbf{E}_{j_N}[T_0] \sim \frac{\pi^{3/2}}{2} \sqrt{\frac{N}{m}}. \tag{5.3.12}$$

The expected number of returns of the process Y to $\lceil \mathfrak{a} \rceil$, when starting above $\mathfrak{a} = (1-\rho)N$, is asymptotically equivalent to $\frac{1}{m(1-\rho)}$ as $N \to \infty$.

b) [Exponential regime] *Assume $Nm(1-\rho)^2 \to \infty$ and $1 \ll j_N \leq N$ as $N \to \infty$. Then*

$$\mathbf{E}_{j_N}[T_0] \sim \left(1 - \exp\left(-2m \left(\frac{1}{\rho} - 1 \right) j_N \right) \right) \sqrt{\frac{\pi}{mN}}\, v_N, \tag{5.3.13}$$

with

$$v_N := \frac{1}{m\left(\frac{1}{\rho} - 1\right)} \exp\left(2Nm(1-\rho)^2 \eta(m,\rho)\right), \tag{5.3.14}$$

$$\eta(m,\rho) := -\frac{1}{2m} \left[\frac{1}{1-\rho} \log\left(\frac{1+2m}{1+2m/\rho} \right) + \sum_{\ell=1}^{\infty} \left(1 - \frac{1}{(1+2m)^\ell} \right) \frac{(1-\rho)^{\ell-1}}{\ell(\ell+1)} \right].$$

In particular, with

$$e_N := \frac{1}{1-\rho} \sqrt{\frac{\pi}{mN}}\, v_N \tag{5.3.15}$$

one has

$$\mathbf{E}_{j_N}[T_0] \sim \begin{cases} e_N & \textit{if } j_N \gg \frac{1}{m(1-\rho)} \\ e_N(1-\exp(-2C/\rho)) & \textit{if } j_N \sim \frac{C}{m(1-\rho)} \\ e_N\, 2j_N\, m(1/\rho - 1) & \textit{if } j_N \ll \frac{1}{m(1-\rho)} \end{cases} . \tag{5.3.16}$$

The expected number of returns of the process Y to $\lfloor \mathfrak{a} \rfloor$, when starting above $\mathfrak{a} = (1-\rho)N$, is asymptotically equivalent to (5.3.14) *as $N \to \infty$.*

Remark 5.3.5. a) Theorem 5.3.4 constitutes an essential step on the way to a proof of the claim (RTT) formulated in (5.1.7) and (5.1.8). One way to complete this proof could lead via the analysis of the system (Y_0^N, Y_1^N) of the sizes of the best and the second-best class of the tournament ratchet; recall that this system is autonomous up to the time of extinction of its first component. Then, $Y_0^N(0)$ and $Y_1^N(0)$ stand for the (random) sizes of the *new* best and second best class at the time of a click. With T_0^N denoting the extinction time of Y_0^N, we conjecture that both in the polynomial and in the exponential regime $Y_1^N(T_0^N)$ will with high probability be $\gg \sqrt{\frac{N}{m}}$, provided that both $Y_0^N(0)$ and $Y_1^N(0)$ are $\gg \sqrt{\frac{N}{m}}$.

b) While the present work focuses on a forward-in-time approach, an alternative route for proving (RTT) is provided by the backward-in-time approach that was developed in [GSW23] in terms of a hierarchical duality for the tournament ratchet. This requires the extension of the backward-in-time analysis from the subcritical to the near-critical regime, and will be a subject of future research.

Remark 5.3.6. a) Theorem 5.3.4 b) suggests the conjecture that not only in the exponential regime of the near-critical case $\rho \uparrow 1$, but also in the entire subcritical case $\rho < 1$ the rate of the tournament ratchet is asymptotically equivalent to (5.3.15). This would improve the logarithmic equivalence (5.2.6) obtained in [GSW23, Theorem 2.2] to an asymptotic equivalence. Here it is worth noticing that (as we will show at the end of Section 5.4.2) the exponents in (5.2.6) and (5.3.14) obey for all $\rho < 1$

$$(1-\rho)^2\eta(m,\rho) \sim \frac{1}{\rho} - 1 + \log\rho \qquad \text{as } m \to 0. \tag{5.3.17}$$

b) In the light of Remark 5.3.1, Theorem 5.3.4 is relevant not only for the tournament ratchet, but also for the Poisson profile approximation of the classical ratchet. Prominent starting values for Y are

– with regard to the classical ratchet: $n_0^C := N\pi_1 = N\theta e^{-\theta}$, which in the (β,δ)-scaling equals $N^{1-\delta}\delta\log N$,

– with regard to the tournament ratchet: $n_0^T := N\sqrt{1-\rho}$, which according to (5.3.5) is the asymptotic center of attraction of the size of its second best class, and in the (β,δ)-scaling equals $N^{1-\delta/2}$.

Figures 5.4 and 5.5 illustrate that these asymptotics of the starting values can indeed be seen in simulations of the classical and the tournament ratchet. The starting values n_0^C and n_0^T are used in Figures 5.6 and 5.7.

For $(\beta, \delta) \in \mathcal{P}$ we have

$$1 - \delta < \frac{1+\beta}{2} < 1 - \tfrac{\delta}{2}.$$

Hence Theorem 5.3.4.a) gives, in accordance with (5.1.7) and (5.1.4),

$$\mathbf{E}_{n_0^T}[T_0] \asymp N^{\frac{1+\beta}{2}} \qquad \text{and} \qquad \mathbf{E}_{n_0^C}[T_0] \asymp N^{1-\delta}.$$

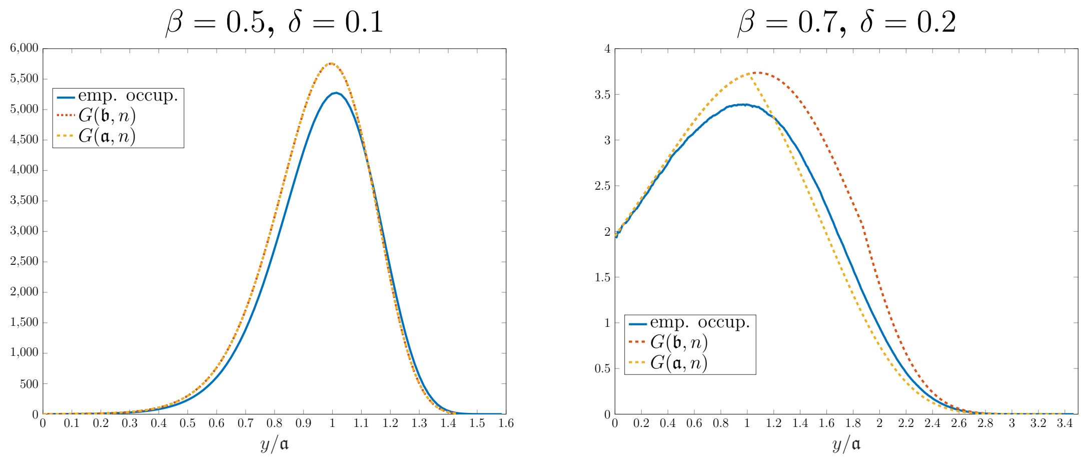
c) Recalling that $m = \rho s$ with $\rho < 1$ (and all these parameters depending on N), the difference of the upward and downward jump rates (5.3.2) and (5.3.1) is

$$\lambda_n - \mu_n = n\left(s - m - s\frac{n}{N}\right)$$

and their sum is $\lambda_n + \mu_n \sim n$ as long as $n \ll N$. Hence the dynamics of Y^N (although its state space is $\{0, 1, \ldots, N\}$ rather than $\mathbb{N}_0$) bears similarities to that of a logistic branching process. Indeed, we conjecture that a logistic branching process $\widehat{Y}^N$ with upward and downward jump rates $\widehat{\lambda}_n$ and $\widehat{\mu}_n$ given by

$$\widehat{\lambda}_n = n\left(\frac{1}{2} + s\right), \qquad \widehat{\mu}_n = n\left(\frac{1}{2} + m + s\frac{n}{N}\right) \tag{5.3.18}$$

will exhibit very similar asymptotics of the expected times to extinction as those obtained for the process Y^N in Theorem 5.3.4. This would complement results of [SaSha13] and [CCM16], both of which do not cover the parameter regime given by (5.3.18). The paper [SaSha13] considers jump rates of the form $\widehat{\lambda}_n = ns$, $\widehat{\mu}_n = nm + n(n-1)\theta$ with constant s, m and small θ; this corresponds to (5.3.18) but without the fluctuation terms $\frac{1}{2}$ which are of dominant order in (5.3.18). (For conceptual clarification we point out that [SaSha13] addresses the case of a constant ratio $s/m > 1$ as supercritical, while in our context this corresponds to a subcritical mutation-selection ratio.) The paper [CCM16] considers quasi-equilibria and extinction times of a class of birth-and-death processes that is more general than logistic branching processes, but imposes a scaling condition of the dynamics which is not fulfilled by (5.3.1) and (5.3.2). Still, both papers point to interesting routes which may offer alternatives to our way of proving Theorem 5.3.4.



(A) **(B)**

Figure 5.3: The empirical occupation times of the size of the best class in a simulation of the tournament ratchet are compared to the Green functions $G(\mathfrak{a}, \cdot)$, $G(\mathfrak{b}, \cdot)$, which are computed numerically using formula (5.4.4). Panels (A) and (B) feature the exponential and the polynomial regime, respectively, with $\gamma = 0.2$ in panel (A) and $\gamma = \frac{2}{3}$ in panel (B). In panel (B) the population size is $N = 500$ and simulations were run up to the first $10^4 + 1$ clicks, where the first click was ignored. In (A), 101 clicks were observed and the first one ignored. Here the population size was $N = 100$. See [EPW09, Figure 5] for similar plots concerning the classical ratchet.

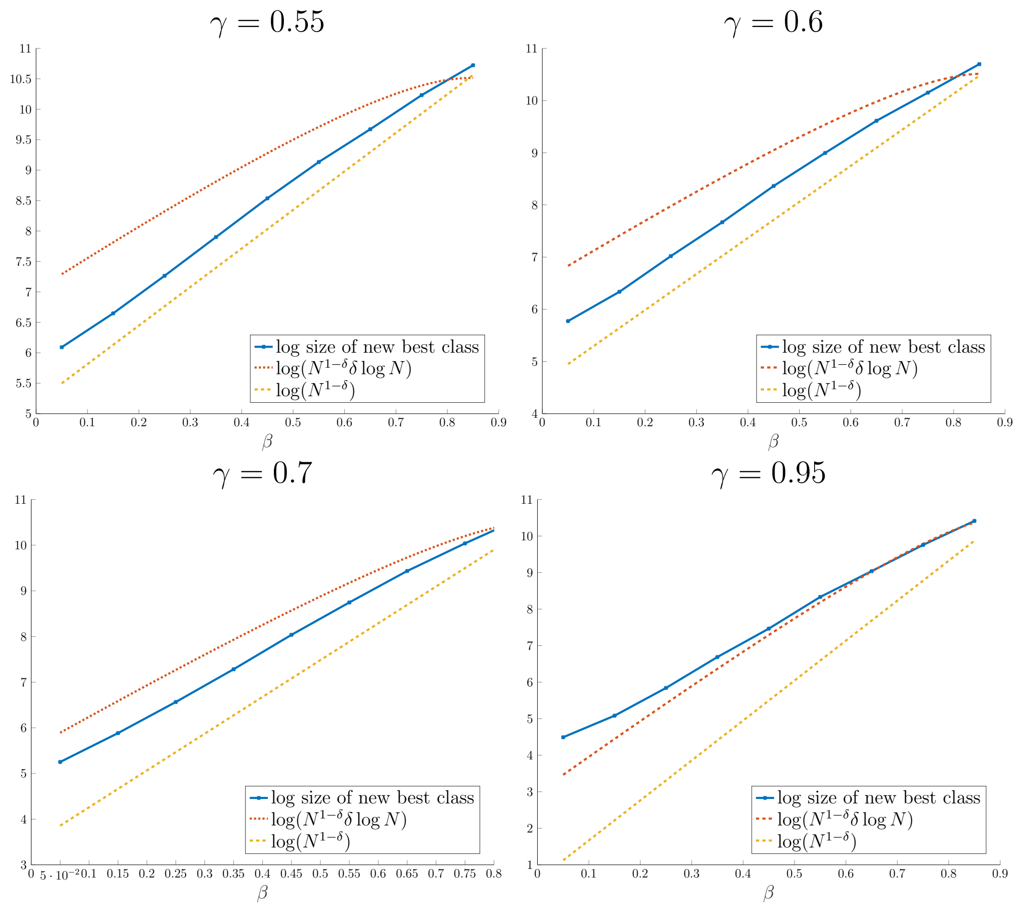


Figure 5.4: For $N = 10^5$ we compare the size of the "new best class" of the classical ratchet immediately after a click (observed in simulations) with the two theoretical predictions $N\pi_0 = N^{1-\delta}$ and $N\pi_1 = N^{1-\delta}\delta\log N$, cf. Remark 5.3.6. b). For various values of $\gamma = \delta/(1-\beta)$, we consider (the logarithms of) these observed and predicted quantities as functions of β. Each data point was obtained by pooling the interclick times no. 5 to 30 from 20 simulations of the classical ratchet for the corresponding parameter configuration. Roughly, the average of the observed logarithmic sizes of the new best class seems to wander away from $N\pi_0$ towards $N\pi_1$ (and beyond) as γ increases.

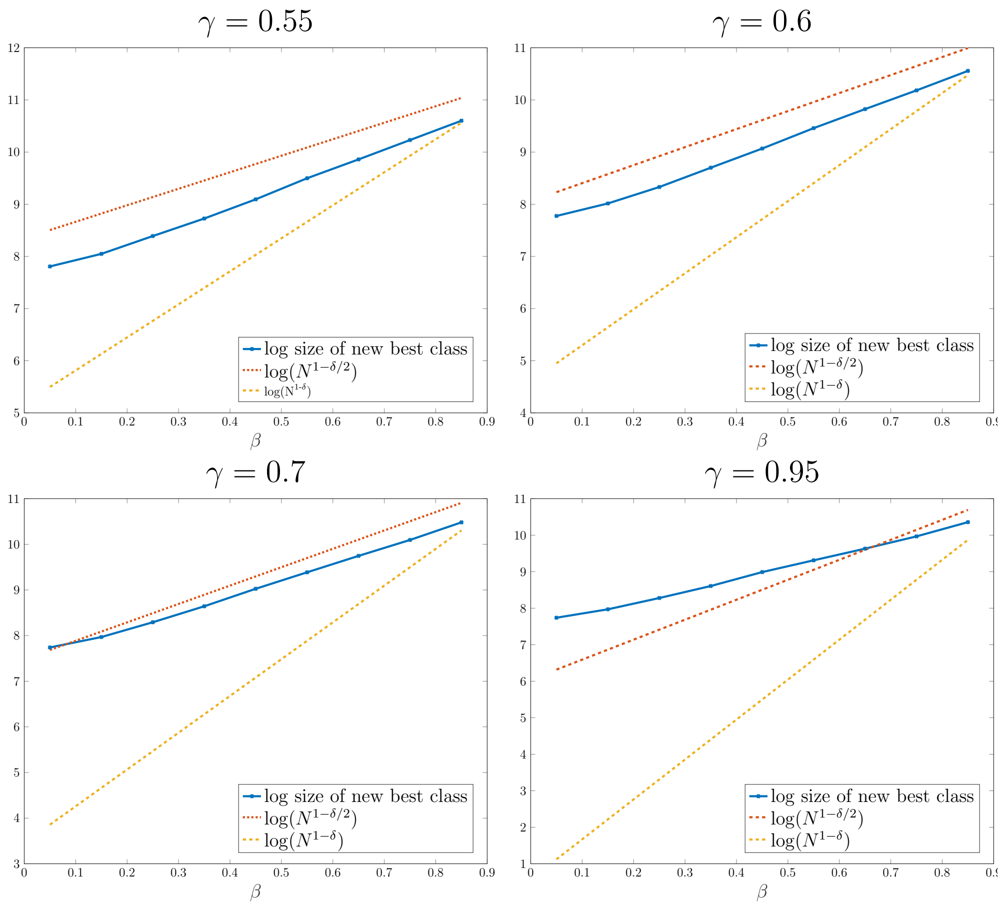


Figure 5.5: For $N = 10^5$ we compare the size of the "new best class" of the tournament ratchet immediately after a click (as observed by simulations) with $\mathfrak{a} = N^{1-\delta}$ and $\mathfrak{b} = N^{1-\delta/2}$, which are the centers of attraction of the best and the second best class of the tournament ratchet(cf. Remark 5.3.6. b)). For various values of $\gamma = \delta/(1-\beta)$, we consider (the logarithms of) these quantities as functions of β. Each data point was obtained by pooling the interclick times no. 5 to 30 from 20 simulations of the tournament ratchet for the corresponding parameter configuration. For a wide range of parameters with γ between 1/2 and 1, $\mathfrak{b}$ is a better fit for the size of the new best class than $\mathfrak{a}$.

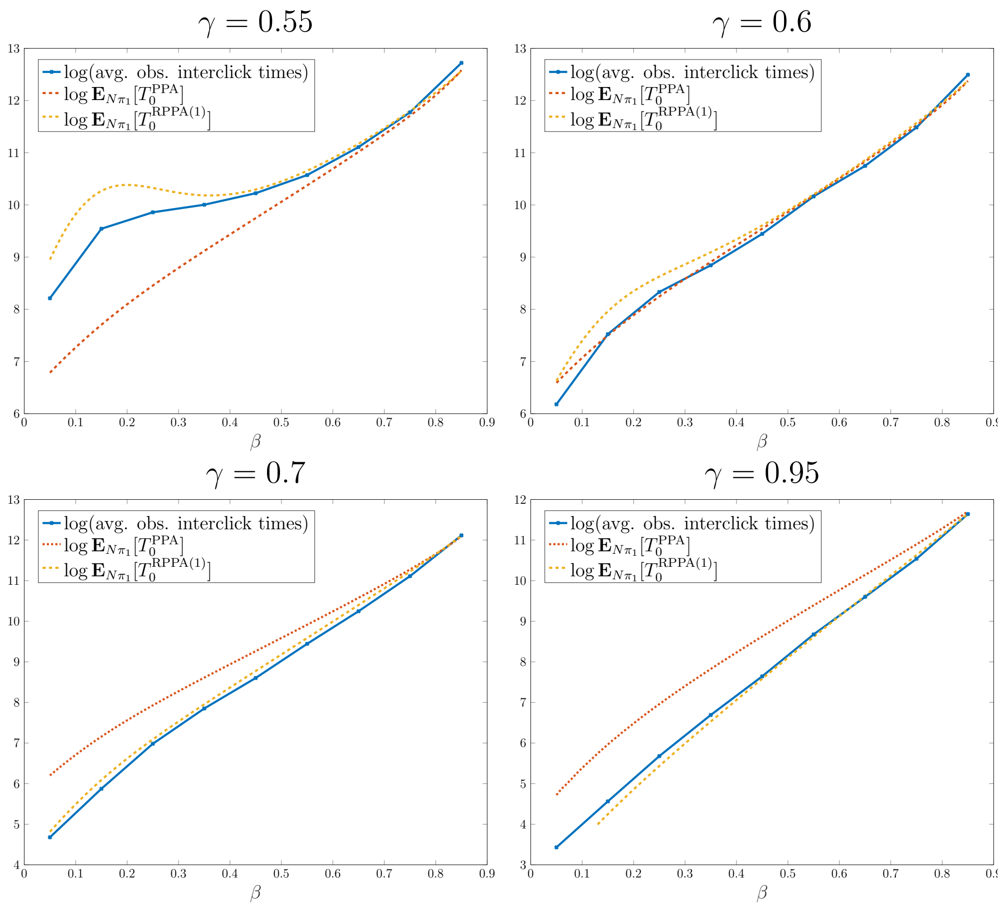


Figure 5.6: For fixed population size $N = 10^5$ the predictions for the expected interclick time of the classical ratchet based on a numerical calculation of the Green function (i) of the PPA and (ii) of the RPPA(1) approximation are compared with simulations. Here, formula (5.4.3) is used (i) for the jump rates (5.3.1) and (5.3.2), and (ii) for the downward jump rate (5.3.1) and the upward jump rate resulting from (5.3.4) and (5.3.9). Each data point was obtained by pooling the interclick times no. 5 to 30 from 20 simulations of the tournament ratchet for the corresponding parameter configuration. Each plot shows this for one fixed value of γ with varying β.

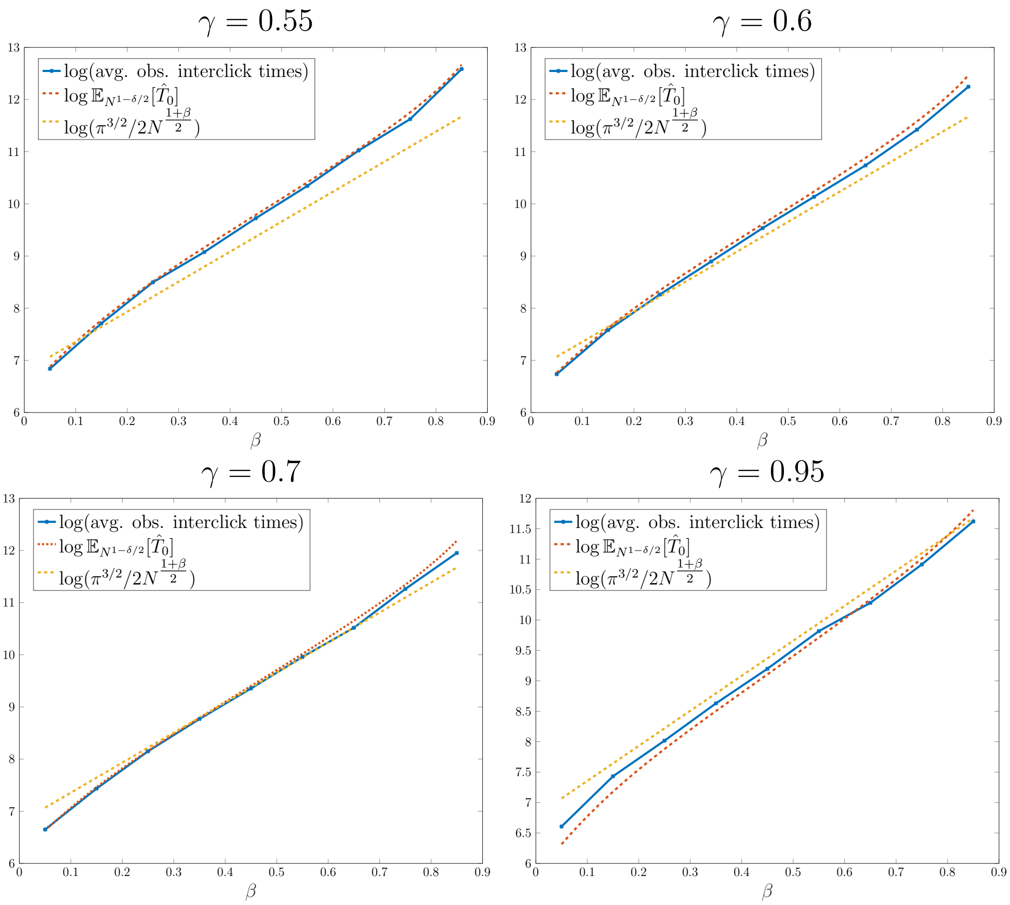


Figure 5.7: For fixed population size $N = 10^5$ the predictions for the expected interclick time of the tournament ratchet based on i) a numerical calculation of the Green function (using formula (5.4.3)) and ii) the asymptotics provided by Theorem 5.3.4 are compared with simulations. Each data point was obtained by pooling the interclick times no. 5 to 30 from 20 simulations of the tournament ratchet for the corresponding parameter configuration. Each plot shows this for one fixed value of γ and for varying β.

5.4 Proof of Theorem 5.3.4

5.4.1 Green function

The proof is based on an asymptotic analysis of the Green function of $Y = Y^N$,

$$G(j,n) := G^N(j,n) = \mathbf{E}_j \left[\int_0^{T_0} I_{\{Y_t^N = n\}} \, \mathrm{d}t \right], \quad 1 \le j,n \le N$$

as $N \to \infty$. By assumption the upward and downward jump rates of Y from n are given by

$$\lambda_n := n \left(\frac{1}{2} \left(1 - \frac{n}{N} \right) + \frac{m}{\rho} \left(1 - \frac{n}{N} \right) \right), \tag{5.4.1}$$

$$\mu_n := n \left(\frac{1}{2} \left(1 - \frac{n}{N} \right) + m \right).$$

Recall that all quantities, including λ_n and μ_n, depend on N, even if we suppress this in the notation for the sake of readability. We express the Green function in terms of the *oddsratio products*

$$r_0 := 1, \quad r_k := \prod_{\ell=1}^{k} \frac{\mu_\ell}{\lambda_\ell}, \qquad k \in \{1, \dots, N-1\}. \tag{5.4.2}$$

The following lemma is well known, see e.g. [SaSha13, (2.4)] for a proof of (5.4.5) via a decomposition with respect to excursions from j. For convenience we include a derivation of (5.4.3) in Section 5.4.5. See also [DSS05, (15)] for a similar representation of $G(j,n)$.

Lemma 5.4.1. *For $1 \le j,n \le N$,*

$$G(j,n) = \frac{1}{\mu_n} \sum_{\ell=0}^{j-1 \wedge n-1} \prod_{k=\ell+1}^{n-1} \frac{\lambda_k}{\mu_k}. \tag{5.4.3}$$

In Figure 5.3, formula (5.4.3) is compared to empirical occupation times from simulations of the process Y.
With

$$R_k := \sum_{i=0}^{k-1} r_i, \qquad k \in \{1, \dots, N\},$$

we obtain from (5.4.3):

$$G(j,n) = \begin{cases} \frac{R_{j \wedge n}}{\lambda_n r_n} \text{ if } n < N, \\ \frac{R_j}{\mu_N r_{N-1}} \text{ if } n = N. \end{cases} \tag{5.4.4}$$

Consequently,

$$\mathbf{E}_j[T_0] = \sum_{n=1}^{N} G(j,n) = \sum_{n=1}^{N-1} \frac{R_{n \wedge j}}{\lambda_n r_n} + \frac{R_j}{\mu_N r_{N-1}}. \tag{5.4.5}$$

Note that

$$U(j) := \log r_j \tag{5.4.6}$$

(sometimes also referred to as *potential*, cf. [DSS05, (16)]) is an additive functional, and (5.4.3) translates into

$$G(j,n) = \frac{1}{\mu_n} \sum_{\ell=0}^{j-1 \wedge n-1} e^{-(U(n-1)-U(\ell))}.$$

5.4.2 Asymptotics for the cumulated oddsratio products

In view of (5.4.5) we are going to find asymptotics for the terms r_k and R_k as $N \to \infty$. Our analysis, see Lemmas 5.4.3 and 5.4.7, shows that, as j increases, r_j is essentially constant on a large interval, before it starts to decrease as j approaches the center of attraction $N(1-\rho)$. The asymptotics of the cumulated oddsratio products R_j and of the terms $G(j,n)$ will be analysed, depending on the order of magnitude of j, in Lemmas 5.4.3, 5.4.4 and 5.4.5 for the polynomial regime, and in Lemmas 5.4.7 and 5.4.8 for the exponential regime.

We recall the notation $f(N) \ll g(N)$ from (5.3.10). Also, we recall that we usually suppress the N-dependence in the notation, as for example in m, ρ and j.

We can express $\log r_j$ as

$$\log r_j = \sum_{k=1}^{j} \log\left(\frac{\mu_k}{\lambda_k}\right) = j \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{k=1}^{j} \log\left(\frac{1-k/((1+2m)N)}{1-k/N}\right). \tag{5.4.7}$$

This expression allows us the following asymptotic description which is key in what follows.

Lemma 5.4.2. *Let $\xi = \xi_N$ be a sequence converging to 0 so slowly that $\xi \gg m$. Then for N large enough and $j \le (1-\xi)N$*

$$0 \le \log r_j - j\log\left(\frac{1+2m}{1+2m/\rho}\right) - \sum_{\ell=1}^{\infty}\left(1 - \frac{1}{(1+2m)^\ell}\right)\frac{1}{\ell(\ell+1)}\frac{j^{\ell+1}}{N^\ell} \le \text{const}\cdot\frac{m}{\xi}. \tag{5.4.8}$$

Lemma 5.4.3. *Let $K := K_N > 0$ either be constant or a diverging sequence. Then for all $k \le K\sqrt{N/m}$*

$$e^{k^2 m/((1+2m)N) - 4(1-\rho)K\sqrt{mN}} \le r_k \le e^{k^2 m/N + K^3/(\sqrt{mN}-K)} \tag{5.4.9}$$

and

$$e^{-4(1-\rho)K\sqrt{mN}} \int_0^{k-1} e^{x^2 m/((1+2m)N)}\,\mathrm{d}x \le R_k \le e^{K^3/(\sqrt{mN}-K)} \int_1^k e^{x^2 m/N}\,\mathrm{d}x. \tag{5.4.10}$$

Lemma 5.4.2 and Lemma 5.4.3 will be proved in Section 5.4.5. We conclude this subsection by showing (5.3.17). To this end, note the two asymptotic equivalences

$$\log\left(\frac{1+2m}{1+2m/\rho}\right) = \log\left(1 + \frac{2m(\rho-1)}{\rho+2m}\right) \sim 2m\left(1-\frac{1}{\rho}\right)$$

and

$$\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{(1-\rho)^{\ell-1}}{\ell(\ell+1)}\sim 2m\sum_{\ell=1}^{\infty}\frac{(1-\rho)^{\ell-1}}{\ell+1},$$

which combine to

$$(1-\rho)^2\eta(m,\rho)\sim\frac{1}{\rho}-1-(1-\rho)-\sum_{\ell=2}^{\infty}\frac{(1-\rho)^{\ell}}{\ell}=\frac{1}{\rho}-1+\log\rho.$$

5.4.3 The polynomial regime: Proof of Theorem 5.3.4.a)

Throughout this subsection we assume $Nm(1-\rho)^2\to 0$ as $N\to\infty$.

We start with the expected number of returns to $a:=\lfloor N(1-\rho)\rfloor$ of the process Y when starting in a. (5.4.4) together with Lemma 5.4.3 gives

$$G(a,a)=\frac{R_a}{\lambda_a r_a}\sim\frac{\int_1^a e^{x^2\frac{m}{N}}\,\mathrm{d}x}{\lambda_a r_a}.$$

From [Wei, (1), (9)] we get

$$\int_1^a e^{x^2\frac{m}{N}}\,\mathrm{d}x\sim\frac{e^{a^2m/N}}{2\frac{m}{N}a},$$

and hence by using (5.4.9) as well as $\lambda_a\sim\frac{1}{2}\rho(1-\rho)N$ we get

$$G(a,a)\sim\frac{1}{m\rho(1-\rho)^2N}\qquad\text{as } N\to\infty,$$

which together with the asymptotics $\lambda_a+\mu_a\sim\rho(1-\rho)N$ gives

$$G(a,a)(\lambda_a+\mu_a)\sim\frac{1}{m(1-\rho)}\qquad\text{as } N\to\infty. \tag{5.4.11}$$

In order to prove the rest the following two lemmas will be proved in Section 5.4.6. Here and below, we will omit the Gauss brackets in the summation bounds for better readability.

Lemma 5.4.4. *Let $\zeta:=\zeta_N\to 0$ such that $\zeta\gg\left[N(1-\rho)^2m\right]^{1/4}$, and $K=K_N$ such that $K\to\infty$ and*

$$K\left((1-\rho)\sqrt{Nm}\vee(Nm)^{-1/6}\right)\to 0.$$

Then

$$\sum_{k=\zeta\sqrt{N/m}}^{K\sqrt{N/m}}\frac{R_k}{\lambda_k r_k}=\sqrt{\frac{N}{m}}\left(\frac{\pi^{3/2}}{2}+O(\zeta)+O\left(\frac{1}{K}\right)\right)\ \text{as } N\to\infty.$$

For the sake of readability, let us introduce the function $f(k)$ via

$$\frac{m}{N}k^2f(k):=k\log\left(\frac{1+2m}{1+2m/\rho}\right)+\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{k^{\ell+1}}{\ell(\ell+1)N^{\ell}}. \tag{5.4.12}$$

Lemma 5.4.5. [3] *Let $K = K_N$ and $\xi = \xi_N$ be sequences with $K_N \to \infty$ and $1 \gg \xi \gg m$. Then there exists a constant $C > 0$ such that for all j with $K\sqrt{N/m} \le j \le N(1-\xi)$*

$$\frac{R_j}{r_j} \le \mathrm{const}\frac{1}{m}\frac{N}{jf(j)}. \tag{5.4.13}$$

With these two lemmas we have the tools for proving Theorem 5.3.4.a), which concerns the polynomial regime. We will distinguish between the cases $j \gg \sqrt{N/m}$ and $j \ll \sqrt{N/m}$, since the potential U (given by (5.4.6)) turns out to be essentially flat below $\sqrt{N/m}$.

5.4.3.1 Proof of (5.3.11)

Abbreviating $\gamma := \log(1/m)$ and recalling that we are in the case

$$j \ll \sqrt{\frac{N/m}{\log(N/m)}}, \tag{5.4.14}$$

we decompose the mean extinction time from state j given by (5.4.5) as follows

$$\begin{aligned}
\mathbf{E}_j[T_0] &= \sum_{k=1}^{j-1}\frac{R_{k\wedge j}}{\lambda_k r_k} + \sum_{k=j}^{\gamma\sqrt{N/m}}\frac{R_{k\wedge j}}{\lambda_k r_k} + \sum_{k=\gamma\sqrt{N/m}+1}^{N}\frac{R_{k\wedge j}}{\lambda_k r_k} \\
&= \sum_{k=1}^{j-1}\frac{R_k}{\lambda_k r_k} + R_j\sum_{k=j}^{\gamma\sqrt{N/m}}\frac{1}{\lambda_k r_k} + R_j\sum_{k=\gamma\sqrt{N/m}+1}^{N-1}\frac{1}{\lambda_k r_k} + \frac{R_j}{\mu_N r_{N-1}} \\
&=: E_1(j) + E_2(j) + E_3(j) + E_4(j).
\end{aligned} \tag{5.4.15}$$

The term $E_4(j)$ is bounded by $\frac{R_{N-1}}{\mu_N r_{N-1}}$. Since $R_{N-1}/r_{N-1} \le N$ and $\mu_N = mN$ we have $\frac{R_{N-1}}{\mu_N r_{N-1}} \le 1/m$, which is $o\left(\sqrt{N/m}\right)$ because of the standing assumption that $Nm \to \infty$. In view of the asymptotics

$$R_k \sim k \quad \text{for } k \ll \sqrt{N/m} \tag{5.4.16}$$

and $\lambda_k \sim k/2$ for $k \ll N$, and because of the inequality $\lambda_k r_k = \mu_k r_{k-1} \ge mkr_{k-1}$ for any $k \le N-1$, we have

$$E_1(j) + E_3(j) \le 4\sum_{k=1}^{j-1}\frac{k}{k} + \frac{2j}{m}\sum_{k=\gamma\sqrt{N/m}+1}^{N}\frac{1}{kr_{k-1}}.$$

Recall the definition of f from (5.4.12). We see that $f(k) \ge 1/2$ when $k \gg \sqrt{N/m}$. Hence there exists a finite constant C such that

$$\sum_{k=\gamma\sqrt{N/m}+1}^{N}\frac{1}{kr_{k-1}} \le \sum_{k=\gamma\sqrt{N/m}}^{N}\frac{e^{-\frac{mk^2}{2N}}}{k} \le \int_\gamma^\infty \frac{e^{-x^2/2}}{x}\,\mathrm{d}x \le \frac{e^{-\gamma^2/2}}{\gamma^2}.$$

[3] This Lemma corrects a mistake of [IGSW24]: There the estimates of F_4 and F_5, see Section 5.4.3.2, and especially this Lemma were not entirely correct. Thus there are small differences compared to the original publication [IGSW24].

In order to see the first inequality we argue as follows: From Lemma 5.4.2 we get that for any k,

$$\begin{aligned}
\log r_k &\geq k\log\left(\frac{1+2m}{1+2m/\rho}\right)+\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{1}{\ell(\ell+1)}\frac{k^{\ell+1}}{N^{\ell}}\\
&\geq k\log\left(\frac{1+2m}{1+2m/\rho}\right)+\left(1-\frac{1}{(1+2m)}\right)\frac{1}{2}\frac{k^2}{N}.
\end{aligned}$$

From the observation that

$$\log\left(\frac{1+2m}{1+2m/\rho}\right)\sim-\frac{2m}{\rho}(1-\rho),\quad \left(1-\frac{1}{(1+2m)}\right)\frac{1}{2}\frac{k}{N}\sim m\frac{k}{N}$$

and

$$\frac{k}{N}\gg(1-\rho)$$

for k at least of order $\sqrt{N/m}$ we see that $r_{k-1}\geq e^{mk^2/(2N)}$. From this we get

$$E_1(j)+E_3(j)\leq 4j+2je^{\gamma}\frac{e^{-\gamma^2/2}}{\gamma^2},$$

which is of lower order than the r.h.s of (5.3.11). We will now analyse $E_2(j)$, which turns out to be the dominant term. First, as $j\ll\sqrt{N/m}$, it may be simplified as follows,

$$E_2(j)\sim 2j\sum_{k=j}^{\log(1/m)\sqrt{N/m}}\frac{1}{kr_k}.$$

By sandwiching arguments using (5.4.8) and (5.4.12) we conclude that

$$\sum_{k=j}^{\gamma\sqrt{N/m}}\frac{1}{kr_k}\sim\int_j^{\gamma\sqrt{N/m}}\frac{e^{-\frac{m}{N}x^2f(x)}}{x}\,\mathrm{d}x.$$

Thanks to (5.4.14) there exists a sequence $\xi=\xi_N\to 0$ such that $\xi^2\gg 1/\log\left(\sqrt{N/m}\right)$ and $j_N\leq\xi_N\sqrt{N/m}$. From (5.4.9), if $k\leq\xi\sqrt{N/m}$ and N is large enough, then $|(m/N)k^2f(k)|\leq 2\xi^2$. Hence

$$e^{-2\xi^2}\left(\log\left(\xi\sqrt{N/m}\right)-\log j\right)\leq\int_j^{\xi\sqrt{N/m}}\frac{e^{-\frac{m}{N}x^2f(x)}}{x}\,\mathrm{d}x\leq e^{2\xi^2}\left(\log\left(\xi\sqrt{N/m}\right)-\log j\right).$$

Moreover, $f(k)\geq\xi^2/2$ for $k\geq\xi\sqrt{N/m}$. We deduce

$$\int_{\xi\sqrt{N/m}}^{\gamma\sqrt{N/m}}\frac{e^{-\frac{m}{N}x^2f(x)}}{x}\,\mathrm{d}x\leq\int_{\xi\sqrt{N/m}}^{\gamma\sqrt{N/m}}\frac{e^{-\frac{m}{N}x^2\xi^2/2}}{x}\,\mathrm{d}x\leq\int_{\xi}^{\infty}\frac{e^{-y^2\xi^2/2}}{y}\,\mathrm{d}y.$$

By substituting $t = \xi y$ in the integral, the right hand side can be written as

$$\int_{\xi^2}^{\infty} \frac{e^{-t^2/2}}{t}\,\mathrm{d}t,$$

which is of order ξ^{-2}. Since this is of lower order than $\log\left(\sqrt{N/m}\right)$ we deduce that

$$E_2(j) \sim 2j\left(\log\left(\sqrt{N/m}\right) - \log j\right). \tag{5.4.17}$$

This ends the proof of of (5.3.11). □

5.4.3.2 Proof of (5.3.12)

We recall that this concerns the case $j = j_N \gg \sqrt{N/m}$. Let $K = K_N$ be a sequence which converges to ∞ so slowly that $K\sqrt{N/m} \le j$ and that K satisfies the requirements of Lemma 5.4.3. Moreover, let $\xi = \xi_N$ be a sequence with $\xi \to 0$ and $\xi \gg m$. In the **first part of the proof** we impose the condition

$$j_N \le N(1 - \xi_N). \tag{5.4.18}$$

Let $\zeta = \zeta_N$ be a sequence converging to 0. Using again (5.4.5), we decompose the mean extinction time from state j as follows:

$$\begin{aligned} &\mathbf{E}_j[T_0] \\ =\ & \sum_{k=1}^{\zeta\sqrt{N/m}} \frac{R_k}{\lambda_k r_k} + \sum_{k=\zeta\sqrt{N/m}+1}^{K\sqrt{N/m}} \frac{R_k}{\lambda_k r_k} + \sum_{k=K\sqrt{N/m}+1}^{N(1-\xi)} \frac{R_{k\wedge j}}{\lambda_k r_k} + \sum_{k=N(1-\xi)+1}^{N-1} \frac{R_j}{\lambda_k r_k} + \frac{R_j}{\mu_N r_{N-1}} \\ =:\ & F_1 + F_2 + F_3(j) + F_4(j) + F_5(j). \end{aligned}$$

The asymptotic of the second sum, F_2, has been derived in Lemma 5.4.4 and leads to the r.h.s. of (5.3.12). It thus suffices to show that $F_1 + F_3(j) + F_4(j) + F_5(j) = o\left(\sqrt{N/m}\right)$. To bound F_1 we need the following lemma, see Section 5.4.6 for a proof.

Lemma 5.4.6. [4] *Assume* $1 \ll Nm \ll (1-\rho)^{-2}$. *Then*

$$\lim_{N\to\infty} \sup_{1\le j\le N/2} \frac{R_j}{r_j \lambda_j} \le 4.$$

For any sequence for $\zeta = \zeta_N \to 0$ we have

$$\zeta\sqrt{\frac{N}{m}} \ll \frac{N}{2}.$$

[4]This lemma did not appear in [IGSW24]. With the strategy used to bound F_1 in [IGSW24] the here presented version of Lemma 5.4.5 would only be enough to prove (5.3.12) under the additional condition $1 - \rho = o\left(\frac{1}{(\log(N/m))^2\sqrt{Nm}}\right)$.

and hence by Lemma 5.4.6 we have that

$$F_1 = O\left(\zeta\sqrt{\frac{N}{m}}\right) = o\left(\sqrt{\frac{N}{m}}\right).$$

Like $E_4(j)$ the term $F_5(j)$ is bounded by $1/m$, which is $o\left(\sqrt{N/m}\right)$ because of the standing assumption that $Nm \to \infty$.
Let us now turn to the analysis of $F_3(j)$. For this we have the upper bound

$$\sum_{k=K\sqrt{N/m}+1}^{N(1-\xi)} \frac{R_k}{\lambda_k r_k}. \tag{5.4.19}$$

By using

$$\lambda_k \sim \frac{k}{2}\left(1-\frac{k}{N}\right) = \frac{1}{2}\frac{k}{N}(N-k)$$

and the bound (5.4.13), the term (5.4.19) is asymptotically bounded from above by

$$\text{const} \sum_{k=K\sqrt{N/m}+1}^{N-1} \frac{N}{k(N-k)}\frac{N}{2mk} = \frac{N^2}{m}\sum_{k=K\sqrt{N/m}+1}^{N-1} \frac{1}{k^2(N-k)}.$$

We claim that this is $o\left(\sqrt{N/m}\right)$, which is equivalent to

$$\frac{N^2}{m}\sqrt{\frac{m}{N}}\sum_{k=K\sqrt{N/m}+1}^{N-1} \frac{1}{k^2(N-k)} \tag{5.4.20}$$

converging to zero. This term we approximate by an integral

$$\begin{aligned}
&\frac{N^2}{m}\sqrt{\frac{m}{N}}\sum_{k=K\sqrt{N/m}+1}^{N-1} \frac{1}{k^2(N-k)} \\
= \;& \frac{N^2}{m}\sqrt{\frac{m}{N}}\cdot\frac{1}{N^3}\cdot N\cdot\frac{1}{N}\sum_{k=K\sqrt{N/m}+1}^{N-1} \frac{1}{(k/N)^2(1-k/N)} \\
\sim \;& \frac{N^2}{m}\sqrt{\frac{m}{N}}\cdot\frac{1}{N^3}\cdot N\cdot\int_{K/\sqrt{mN}}^{1-\frac{1}{N}} \frac{1}{x^2(1-x)}\,\mathrm{d}x \\
= \;& \frac{1}{\sqrt{Nm}}\int_{K/\sqrt{mN}}^{1-\frac{1}{N}} \frac{1}{x^2(1-x)}\,\mathrm{d}x.
\end{aligned}$$

The integral is of order

$$\left(K/\sqrt{mN}\right)^{-1} \vee \log N = (\frac{1}{K}\sqrt{mN}) \vee \log N.$$

Thus (5.4.20) is of order $\frac{(\sqrt{Nm}/K)\vee \log N}{\sqrt{Nm}}$, which converges to zero as $N \to \infty$. We are left with the analysis of

$$F_4(j) = \sum_{k=N(1-\xi)+1}^{N-1} \frac{R_{k\wedge j}}{\lambda_k r_k}.$$

This sum is bounded by

$$R_j \sum_{k=N(1-\xi)+1}^{N-1} \frac{1}{\lambda_k r_k}.$$

Since we assumed $j \leq N - \xi N$ this is again bounded from above by

$$\frac{R_{N(1-\xi)}}{r_{N(1-\xi)}} \sum_{k=N(1-\xi)+1}^{N-1} \frac{r_{N(1-\xi)}}{\lambda_k r_k}.$$

Recall from (5.4.7) that for $j, \ell \in \mathbb{N}$,

$$\log r_{j+\ell} - \log r_j = \ell \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{k=j+1}^{j+\ell} \log\left(\frac{1-k/((1+2m)N)}{1-k/N}\right).$$

Noticing that the term

$$\log\left(\frac{1-k/((1+2m)N)}{1-k/N}\right)$$

is increasing with k, and performing a Taylor expansion, we obtain that for

$$k = N(1-\xi) + \ell$$

we have the inequality

$$\log r_k \geq \log r_{N(1-\xi)}. \tag{5.4.21}$$

Together with Lemma 5.4.5 we obtain

$$\begin{aligned} &\frac{R_{N(1-\xi)}}{r_{N(1-\xi)}} \sum_{k=N(1-\xi)+1}^{N-1} \frac{r_{N(1-\xi)}}{\lambda_k r_k} \\ \leq\ & \frac{\text{const}}{(1-\xi)mf(N(1-\xi))} \cdot \sum_{k=N(1-\xi)+1}^{N-1} \frac{r_{N(1-\xi)}}{\lambda_k r_k}. \end{aligned}$$

By (5.4.21) and since for $\ell = 0, \ldots, \xi N$

$$\lambda_{(1-\xi)N+\ell} \quad \geq \quad \frac{1}{2}\left((1-\xi)N+\ell\right)\left(1-(1-\xi)-\frac{\ell}{N}\right)$$

$$= \left((1-\xi)+\frac{\ell}{N}\right)(\xi N-\ell)\sim \xi N-\ell$$

we obtain

$$\sum_{k=N(1-\xi)+1}^{N-1}\frac{r_{N(1-\xi)}}{\lambda_k r_k}=O\left(\log(N\xi)\right). \tag{5.4.22}$$

Now observe that for N large enough

$$\begin{aligned}
f\left((1-\xi)N\right) &\geq \frac{N}{m((1-\xi)N)^2}\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{N^{\ell+1}(1-\xi)^{\ell+1}}{\ell(\ell+1)N^\ell} \\
&\quad +\frac{N}{m(1-\xi)N}\log\left(\frac{1+2m}{1+2m/\rho}\right) \\
&= \frac{1}{m(1-\xi)^2}\sum_{\ell=1}^{\infty}\frac{(1+2m)^\ell-1}{(1+2m)^\ell}\frac{(1-\xi)^{\ell+1}}{\ell(\ell+1)}+\frac{1}{m(1-\xi)}\log\left(\frac{1+2m}{1+2m/\rho}\right) \\
&\geq \frac{1}{m(1-\xi)^2}\sum_{\ell=1}^{\infty}\frac{2m\ell}{(1+2m)^\ell}\frac{(1-\xi)^{\ell+1}}{\ell(\ell+1)}-\text{const} \\
&= \frac{2}{(1-\xi)}\sum_{\ell=1}^{\infty}\frac{(1-\xi)^\ell}{(1+2m)^\ell}\frac{1}{\ell+1}-\text{const} \\
&= \frac{2(1+2m)}{(1-\xi)^2}\sum_{\ell=2}^{\infty}\frac{(1-\xi)^\ell}{(1+2m)^\ell}\frac{1}{\ell}-\text{const}.
\end{aligned}$$

By the identity

$$\sum_{k=2}^{\infty}\frac{z^k}{k}=-z-\log(1-z), \qquad |z|<1$$

and $\xi \gg m$ we get that there exists a constant $c>0$, such that for N large enough

$$f\left((1-\xi)N\right)\geq c|\log\xi|.$$

If $m \ll N^{-\epsilon}$ for some ϵ, then together with (5.4.22) we obtain

$$F_4(j)=o\left(\sqrt{\frac{N}{m}}\right) \tag{5.4.23}$$

for some sequence ξ obeying $m \ll N^{-\epsilon} \ll \xi$. If $m \gg (\log N)^2/N$, then by (5.4.22) we again obtain (5.4.23). If neither $m \ll N^{-\epsilon}$ for some $\epsilon>0$ nor $m \gg (\log N)^2/N$ holds, then both are true along two subsequences, which together cover the entire sequence $(m_N)_N$. This finishes the proof of (5.3.12) in the case $j \leq N(1-\xi)$.

In the **remaining part of the proof** we consider sequences which not necessarily satisfy the restriction (5.4.18). In view of the first part it suffices to show that the expected time which Y needs to come down from N to $N(1-\xi)$ is of lower order than $\sqrt{N/m}$. For this, we impose an additional condition on the sequence ξ, and will show the following claim: Let

$\xi = \xi_N$ be a sequence converging to 0 and obeying $m \ll \xi \ll m(N/m)^{1/4}$ as $N \to \infty$. Then $\mathbb{E}_N[T_{N(1-\xi)}] = o(\sqrt{N/m})$.

To prove this claim, let $\mathcal{Y}$ be the time-discrete birth-and-death process corresponding to Y. By (5.3.1) and (5.3.2) the probability of $\mathcal{Y}$ to go down in the next step when starting in k is given by

$$\frac{\frac{1}{2}\left(1-\frac{k}{N}\right)+m}{\left(1-\frac{k}{N}\right)+m+\frac{m}{\rho}\left(1-\frac{k}{N}\right)}.$$

which for $N(1-\xi) \le k \le N$ is bounded from below by

$$q := \frac{1}{2}\frac{1+\frac{2m}{\xi}}{1+\frac{m}{\xi}+\frac{m}{\rho}}.$$

Let us put $p := 1-q$, and consider the (p,q)-random walk W on $\mathbb{Z}$ as well as the random walk $\widehat{W}$ on $\mathbb{Z} \cap \{\dots, N-2, N-1, N\}$ that is obtained by reflecting W at N, i.e. by putting $\mathbb{P}_N(\widehat{W}_1 = N) := p$, $\mathbb{P}_N(\widehat{W}_1 = N-1) := q$. A suitable coupling of $\mathcal{Y}$ and $\widehat{W}$ (both starting in N) shows that for $N(1-\xi) \le k \le N$ the expected number visits of $\mathcal{Y}$ to k before $\mathcal{Y}$ reaches $N(1-\xi)$ is not larger than the expected number of visits of $\widehat{W}$ to k before $\widehat{W}$ reaches $N(1-\xi)$. The expected number of visits of the transient random walk W to its starting point is $\frac{q}{q-p} \sim \frac{\xi}{m}$, and the same is true for $\widehat{W}$. The jump rates (5.3.1) and (5.3.2) from state $k \ge N(1-\xi)$ add up to

$$k\left[\frac{N-k}{N}+m+\frac{N-k}{N}\frac{m}{\rho}\right] \ge \frac{1}{2}Nm.$$

Altogether, we obtain the estimate

$$\mathbb{E}_N[T_{N(1-\xi)}] \le \frac{2}{Nm}\xi N\frac{q}{p-q} \sim \frac{2\xi^2}{m^2}, \tag{5.4.24}$$

whose r.h.s. is $o(\sqrt{N/m})$ due to our assumption on ξ.

This concludes the proof of Theorem 5.3.4.a). □

5.4.4 The exponential regime: Proof of Theorem 5.3.4.b)

Throughout this section we assume $Nm(1-\rho)^2 \to \infty$. In this regime the process Y should spend a long time around its center of attraction $(1-\rho)N$, which makes the following decomposition of (5.4.5) natural: for a small $\zeta > 0$ write

$$\begin{aligned}\mathbf{E}_j[T_0] &= \sum_{k=1}^{(1-\zeta)(1-\rho)N}\frac{R_{k\wedge j}}{\lambda_k r_k}+\sum_{k=(1-\zeta)(1-\rho)N+1}^{(1+\zeta)(1-\rho)N}\frac{R_{k\wedge j}}{\lambda_k r_k}+\sum_{k=(1+\zeta)(1-\rho)N+1}^{N-1}\frac{R_{k\wedge j}}{\lambda_k r_k}+\frac{R_j}{\mu_N r_{N-1}}\\ &=: A(\zeta)+B(\zeta)+C(\zeta)+\frac{R_j}{\mu_N r_{N-1}}.\end{aligned} \tag{5.4.25}$$

The assertion of Theorem 5.3.4.b) will be derived at the end of this section from Proposition 5.4.8, whose proof, in turn, will rely on the following lemma. The proof of this lemma as well as that of Proposition 5.4.8 will be given in Section 5.4.7.

Lemma 5.4.7. *Let $j = j_N$ be a sequence of natural numbers converging to ∞, and let $\xi < 1/2$. Then*

- *If $j \le \xi(1-\rho)N$, then for sufficiently large N*
$$\frac{1 - e^{-(1+2\xi)2m(1-\rho)j/\rho}}{(1+2\xi)2m(1-\rho)/\rho} \le R_j \le \frac{1 - e^{-(1-2\xi)2m(1-\rho)j/\rho}}{(1+2\xi)2m(1-\rho)/\rho}.$$

- *If $1/(m(1-\rho)) \ll j \le (2-\xi)(1-\rho)N \wedge N(1-\sqrt{m})$, then, under the assumption $\xi \ge 2\log(mN(1-\rho)^2)/(mN(1-\rho)^2)$,*
$$R_j \sim \frac{\rho}{2m(1-\rho)} \qquad \text{as } N \to \infty.$$

- *If $j = C(1-\rho)N \le N(1-\sqrt{m})$, with $\frac{1}{1-\rho} \ge C > 2/\rho$ (implying $\rho > \frac{2}{3}$), then*
$$R_j \sim \rho(1 - C(1-\rho))\frac{\exp\left(-2m(1-\rho)^2 NH(C)\right)}{2m(C-1)(1-\rho)} \qquad \text{as } N \to \infty,$$
where the function $H(.) = H((m,\rho),.)$ on $\mathbb{R}_+$ is defined by
$$H(y) := -\frac{y}{2m}\left[\frac{1}{1-\rho}\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1 - \frac{1}{(1+2m)^\ell}\right)\frac{(1-\rho)^{\ell-1}y^\ell}{\ell(\ell+1)}\right]. \tag{5.4.26}$$

This is the central ingredient for the proof of the following proposition, which, in turn, will be key for the proof of Theorem 5.3.4.b).

Proposition 5.4.8. *Let $A(\zeta)$, $B(\zeta)$ and $C(\zeta)$ be defined by (5.4.25). Then for $\zeta = \zeta_N$ converging to 0 so slowly that $\zeta\sqrt{mN}(1-\rho) \to \infty$, we have*
$$B(\zeta) \sim \left(R_j \wedge \frac{\rho}{2m(1-\rho)}\right)\sqrt{\frac{\pi}{mN}}\frac{2}{1-\rho}\exp\left(2m(1-\rho)^2 NH(1)\right) \qquad \text{as } N \to \infty, \tag{5.4.27}$$
and
$$A(\zeta) + C(\zeta) = o(B(\zeta)) \qquad \text{as } N \to \infty.$$

The proof of this proposition will be given in Section 5.4.7.

Proof of Theorem 5.3.4.b). First note that we can asymptotically neglect $\frac{R_j}{\mu_N r_{N-1}}$ with the same arguments as those used for $E_4(j)$ and $F_5(j)$, see Section 5.4.3.1 and Section 5.4.3.2.
(i) For $j = j_N = O(1/(m(1-\rho)))$ and any sequence $\xi = \xi_N$ converging to zero we have
$$1 - \exp\left(-(1 \pm 2\xi)2m(1-\rho)j/\rho\right) \sim 1 - \exp\left(-2m(1-\rho)j/\rho\right).$$

Hence Proposition 5.4.8 together with the first bullet point of Lemma 5.4.7 gives
$$\mathbf{E}_j[T_0] \sim B(\zeta) \sim \frac{1 - \exp\left(-2m(1-\rho)j/\rho\right)}{2m(1-\rho)}\sqrt{\frac{\pi}{mN}}\frac{2\rho}{1-\rho}\exp\left(2m(1-\rho)^2 NH(1)\right),$$

(ii) For $j \gg 1/(m(1-\rho))$ Proposition 5.4.8 together with the second bullet point of Lemma 5.4.7 gives

$$\mathbf{E}_j[T_0] \sim B(\zeta) \sim \frac{\rho}{m(1-\rho)^2}\sqrt{\frac{\pi}{mN}}\exp\left(2m(1-\rho)^2 NH(1)\right).$$

(iii) To conclude (5.3.13) from (i) and (ii) it suffices to observe that $\eta(m,\rho) = H(1)$, and that the assumption on j in (ii) implies the convergence $1-\exp(-2m(1-\rho)j_N/\rho) \to 1$ as $N \to \infty$. The claimed asymptotics (5.3.16) is an immediate consequence of (5.3.13).
(iv) It remains to prove the claim on the expected number of excursions from $a := \lfloor \mathfrak{a} \rfloor$, with $\mathfrak{a} = (1-\rho)N$ being the asymptotic center of attraction of Y. In view of (5.4.4) this expected number equals

$$(\lambda_a + \mu_a)G(a,a) = (\lambda_a + \mu_a)\frac{R_a}{\lambda_a r_a}. \tag{5.4.28}$$

In order to estimate r_a we observe that (5.4.8), when expressed in terms of the function H (which was defined in (5.4.26)), gives the asymptotics

$$\frac{1}{r_a} \sim \exp\left(2m(1-\rho)^2 NH(1)\right) \qquad \text{as } N \to \infty. \tag{5.4.29}$$

In addition, the second bullet point of Lemma 5.4.7 gives

$$R_a \sim \frac{\rho}{2m(1-\rho)} \qquad \text{as } N \to \infty. \tag{5.4.30}$$

Since $\lambda_a \sim \mu_a$ as $N \to \infty$, the combination of (5.4.29) and (5.4.30) shows that (5.4.28) is asymptotically equivalent to (5.3.14). $\square$

5.4.5 Proofs of Lemmas 5.4.1, 5.4.2 and 5.4.3

Proof of Lemma 5.4.1. We denote the time-discrete embedded process corresponding to Y by $\mathcal{Y}$, and write $\mathcal{G}(m,n)$ for the expected number of visits at n of $\mathcal{Y}$ when starting in m. Let us start with an analysis of $\mathcal{G}(n,n)$. By standard arguments we have

$$\mathcal{G}(n,n) = \frac{1}{\phi(n)}, \tag{5.4.31}$$

where $\phi(n)$ is the escape probability of $\mathcal{Y}$ from the state n, i.e.

$$\phi(n) = \frac{\mu_n}{\mu_n + \lambda_n}(1 - h^{(n)}(n-1)), \tag{5.4.32}$$

where $h^{(n)} : \{0, 1, \ldots, n\} \to [0,1]$ is $\mathcal{Y}$-harmonic on $\{1, \ldots, n-1\}$ and satisfies the boundary conditions $h^{(n)}(0) = 0$, $h^{(n)}(n) = 1$. Hence

$$h^{(n)}(\ell) = \frac{\sum_{j=0}^{\ell-1} r_j}{\sum_{k=0}^{n-1} r_k}, \quad \ell = 0, \ldots, n, \tag{5.4.33}$$

with the oddsratio product r_k as in (5.4.2). From (5.4.33) we obtain

$$1 - h^{(n)}(n-1) = \frac{r_{n-1}}{\sum_{k=0}^{n-1} r_k}. \tag{5.4.34}$$

For $G(n,n)$, the expected time spent by Y in n when starting in n, we thus obtain the relation

$$G(n,n) = \frac{\mathcal{G}(n,n)}{\lambda_n + \mu_n} = \frac{1}{\phi(n)} \cdot \frac{1}{\mu_n + \lambda_n}.$$

Combining this with (5.4.31), (5.4.32) and (5.4.34) we arrive at

$$G(n,n) = \frac{1}{\mu_n} \cdot \sum_{k=0}^{n-1} \frac{r_k}{r_{n-1}} = \frac{1}{\mu_n} \sum_{\ell=0}^{n-1} \prod_{k=\ell+1}^{n-1} \frac{\lambda_k}{\mu_k}. \tag{5.4.35}$$

For $j > n$ we have

$$G(j,n) = G(n,n), \tag{5.4.36}$$

while for $j < n$

$$G(j,n) = h^n(j) G(n,n) = \frac{\sum_{\ell=0}^{j-1} r_\ell}{\sum_{\ell=0}^{n} r_\ell} G(n,n).$$

Together with (5.4.35) this gives for $j < n$

$$\begin{aligned} G(j,n) &= \frac{1}{\mu_n} \sum_{k=0}^{n-1} \frac{r_k}{r_{n-1}} \cdot \frac{\sum_{\ell=0}^{j-1} r_\ell}{\sum_{\ell=0}^{n-1} r_\ell} \\ &= \frac{1}{\mu_n} \sum_{\ell=0}^{j-1} \prod_{k=\ell+1}^{n-1} \frac{\lambda_k}{\mu_k}. \end{aligned} \tag{5.4.37}$$

It remains to observe that the three cases $j > n$, $j = n$, $j < n$ (which are covered by (5.4.36), (5.4.35) (5.4.37)) combine to (5.4.3) □

Proof of Lemma 5.4.2. We have already seen

$$\begin{aligned} \log r_j = \sum_{k=1}^{j} \log\left(\frac{\mu_k}{\lambda_k}\right) &= j \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{k=1}^{j} \log\left(\frac{1 - k/(1+2m)N}{1 - k/N}\right) \\ &= j \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{k=1}^{j} \sum_{\ell=1}^{\infty} \left(1 - \frac{1}{(1+2m)^\ell}\right) \frac{1}{\ell} \left(\frac{k}{N}\right)^\ell. \end{aligned}$$

As $j \leq N-1$, we may apply Fubini's theorem to write

$$\sum_{k=1}^{j} \sum_{\ell=1}^{\infty} \left(1 - \frac{1}{(1+2m)^\ell}\right) \frac{1}{\ell} \left(\frac{k}{N}\right)^\ell = \sum_{\ell=1}^{\infty} \left(1 - \frac{1}{(1+2m)^\ell}\right) \frac{1}{\ell} \sum_{k=1}^{j} \left(\frac{k}{N}\right)^\ell.$$

Now sandwiching arguments yield

$$\frac{j^{\ell+1}}{(\ell+1)N^\ell} = \int_0^j \left(\frac{x}{N}\right)^\ell \mathrm{d}x$$

$$
\begin{aligned}
&\le \sum_{k=1}^{j}\left(\frac{k}{N}\right)^{\ell}\\
&\le \int_1^j \left(\frac{x}{N}\right)^{\ell} \mathrm{d}x + \left(\frac{j}{N}\right)^{\ell} = \frac{1}{(\ell+1)N^{\ell}}(j^{\ell+1}-1) + \left(\frac{j}{N}\right)^{\ell}.
\end{aligned}
$$

This means that if we introduce

$$
\Delta_j := \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{1}{\ell}\sum_{k=1}^{j}\left(\frac{k}{N}\right)^{\ell} - \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{1}{\ell(\ell+1)}\frac{j^{\ell+1}}{N^{\ell}},
$$

we have

$$
0 \le \Delta_j \le \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{1}{\ell}\left(\frac{j}{N}\right)^{\ell}
$$

From the observation that

$$
1-\frac{1}{(1+2m)^{\ell}} \le 2m\ell,
$$

we get

$$
0 \le \Delta_j \le 2m\sum_{\ell=1}^{\infty}\left(\frac{j}{N}\right)^{\ell}.
$$

Now let $\xi \gg m$. Then $j \le (1-2\xi)N$ implies that for N marge enough, $j+1 \le (1-\xi)N$. Hence

$$
\Delta_j \le 2m\sum_{\ell=1}^{\infty}(1-\xi)^{\ell} = 2m\frac{1-\xi}{\xi} = o(1).
$$

This entails that there exists C such that for any $j \le (1-2\xi)N$ and N large enough,

$$
0 \le \log r_j - j\log\left(\frac{1+2m}{1+2m/\rho}\right) - \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{1}{\ell(\ell+1)}\frac{j^{\ell+1}}{N^{\ell}} \le \text{const}\cdot\frac{m}{\xi}. \quad (5.4.38)
$$

□

Proof of Lemma 5.4.3. We set out from (5.4.8). Notice that for any $x, y \ge 0$, $1-(1+x)^{-y} \le xy$. Hence, for $k \le K\sqrt{N/m}$,

$$
\begin{aligned}
\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{k^{\ell+1}}{\ell(\ell+1)N^{\ell}} &\le 2m\sum_{\ell=1}^{\infty}\frac{k^{\ell+1}}{(\ell+1)N^{\ell}} \le m\sum_{\ell=1}^{\infty}\frac{k^{\ell+1}}{N^{\ell}}\\
&= k^2m/N + mN\sum_{\ell=3}^{\infty}\frac{k^{\ell}}{N^{\ell}}\\
&= k^2m/N + mN\frac{k^3}{N^3}\frac{1}{1-k/N}\\
&\le k^2m/N + \frac{K^3}{\sqrt{mN}-K}.
\end{aligned}
$$

Conversely we have

$$\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{k^{\ell+1}}{\ell(\ell+1)N^\ell}\geq\left(1-\frac{1}{(1+2m)}\right)\frac{k^2}{2N}\sim k^2m/N.$$

Finally, for $k\leq K\sqrt{N/m}$ we also have

$$\left|k\log\left(\frac{1+2m}{1+2m/\rho}\right)\right|=k\left|\log\left(1-\frac{2m(1-\rho)}{\rho+2m}\right)\right|\sim 2mk(1-\rho)\leq 2K\sqrt{m(1-\rho)^2N}=o(1).$$

This concludes the proof of (5.4.9). The estimate (5.4.10) then follows by approximating the sum $R_k:=\sum_{\ell=0}^{k-1}r_\ell$ from below by an integral from 0 to $k-1$ and from above by an integral from 1 to k. □

5.4.6 Proofs of Lemmas 5.4.4, 5.4.5 and 5.4.6

Proof of Lemma 5.4.4. Noting that our choices of K and ζ entail

$$\frac{k^2m}{(1+2m)N}\gg 4(1-\rho)\sqrt{mN}\qquad\forall k\geq\zeta\sqrt{N/m}$$

as well as

$$\begin{aligned}4(1-\rho)K\sqrt{mN}&\to 0\\ \frac{K^3}{\sqrt{mN}-1}&\to 0.\end{aligned}$$

We can deduce

$$\sum_{k=\zeta\sqrt{N/m}}^{K\sqrt{N/m}}\frac{R_k}{kr_k}\sim\int_{y=\zeta\sqrt{N/m}}^{K\sqrt{N/m}}e^{-\frac{m}{N}y^2}\frac{\mathrm{d}y}{y}\int_{z=0}^{y}e^{\frac{m}{N}z^2}\,\mathrm{d}z=:I_N$$

from (5.4.9) and (5.4.10). We thus need to find an equivalent of I_N for large N. Three successive changes of variables entail the equalities:

$$\begin{aligned}I_N&=\int_{y=\zeta\sqrt{N/m}}^{K\sqrt{N/m}}\frac{\mathrm{d}y}{y}\int_{z=0}^{y}e^{-\frac{m}{N}(y^2-z^2)}\,\mathrm{d}z\\
&=\int_{y=\zeta\sqrt{N/m}}^{K\sqrt{N/m}}\mathrm{d}y\int_{\lambda=0}^{1}e^{-\frac{m}{N}y^2(1-\lambda^2)}d\lambda=\sqrt{\frac{N}{m}}\int_{\lambda=0}^{1}d\lambda\int_{z=\zeta}^{K}e^{-z^2(1-\lambda^2)}\,\mathrm{d}z\\
&=\sqrt{\frac{N}{m}}\int_{\lambda=0}^{1}\frac{d\lambda}{\sqrt{1-\lambda^2}}\int_{w=\zeta\sqrt{1-\lambda^2}}^{K\sqrt{1-\lambda^2}}e^{-w^2}dw.\end{aligned}$$

Hence

$$I_N \leq \sqrt{\frac{N}{m}} \int\limits_{\lambda=0}^{1} \frac{d\lambda}{\sqrt{1-\lambda^2}} \int\limits_{w=0}^{\infty} e^{-w^2} dw = \sqrt{\frac{N}{m}} \frac{\pi}{2} \frac{\sqrt{\pi}}{2} = \sqrt{\frac{N}{m}} \frac{\pi^{3/2}}{4}.$$

For the lower estimate we proceed as follows:

$$\begin{aligned}
\frac{\pi^{3/2}}{4} - I_N \sqrt{\frac{m}{N}} &= \int\limits_{\lambda=0}^{1} \frac{d\lambda}{\sqrt{1-\lambda^2}} \left(\int\limits_{w=0}^{\zeta\sqrt{1-\lambda^2}} e^{-w^2} dw + \int\limits_{w=K\sqrt{1-\lambda^2}}^{\infty} e^{-w^2} dw \right) \\
&\leq \int\limits_{\lambda=0}^{1} \frac{d\lambda}{\sqrt{1-\lambda^2}} \left(\zeta + \int\limits_{w=K\sqrt{1-\lambda^2}}^{\infty} e^{-w^2} dw \right) \\
&\leq \frac{\pi}{2}\zeta + \int\limits_{\lambda=0}^{1} \frac{d\lambda}{\sqrt{1-\lambda^2}} \left(\int\limits_{w=K\sqrt{1-\lambda^2}}^{\infty} e^{-w^2} dw \right).
\end{aligned}$$

We write the double integral on the r.h.s. as

$$\int_0^1 \int_{z=K}^{\infty} e^{-z^2(1-\lambda^2)} \, \mathrm{d}z \, \mathrm{d}\lambda = \int_{z=K}^{\infty} e^{-z^2} \int_0^1 e^{z^2\lambda^2} \, \mathrm{d}\lambda \, \mathrm{d}z. \tag{5.4.39}$$

By substituting $x = \lambda z$ the inner integral is equal to

$$\frac{1}{z} \int_0^z e^{x^2} \, \mathrm{d}x,$$

which by [Wei, (1),(9)] is bounded from above by const $\cdot \, e^{z^2}/z^2$, such that in total (5.4.39) is bounded from above by

$$\int_K^{\infty} e^{-z^2} \left(\text{const} \cdot \frac{e^{z^2}}{z^2} \right) \mathrm{d}z \leq \text{const} \cdot \frac{1}{K},$$

so in total

$$\frac{\pi^{3/2}}{4} - I_N \sqrt{\frac{m}{N}} = O(\zeta) + O\left(\frac{1}{K}\right).$$

Hence, as $\lambda_k \sim k/2$ for $k \ll N$, we have proved that

$$\sum_{k=\zeta\sqrt{N/m}}^{K\sqrt{N/m}} \frac{R_k}{\lambda_k r_k} \sim \sum_{k=\zeta\sqrt{N/m}}^{K\sqrt{N/m}} 2\frac{R_k}{k r_k} = \sqrt{\frac{N}{m}} \left(\frac{\pi^{3/2}}{2} + O\left(\zeta\right) + O\left(\frac{1}{K}\right) \right).$$

□

Proof of Lemma 5.4.5. Lemma 5.4.2 enables us to write

$$r_j \sim \exp\left(\frac{m}{N} j^2 f(j)\right) \qquad \text{for } j \le N - \xi N.$$

So

$$\begin{aligned} R_j &= \sum_{\ell=0}^{j-1} r_\ell \sim \sum_{\ell=0}^{j-1} \exp\left(\frac{m}{N}\ell^2 f(\ell)\right) \\ &\le \text{const} \cdot \int_0^j \exp\left(\frac{m}{N}x^2 f(x)\right) \mathrm{d}x \end{aligned}$$

and

$$\begin{aligned} \frac{R_j}{r_j} &\le \text{const} \exp\left(-\frac{m}{N} j^2 f(j)\right) \cdot \int_0^j \exp\left(\frac{m}{N} x^2 f(x)\right) \mathrm{d}x \\ &= \text{const} \cdot \int_0^j \exp\left(\frac{m}{N}\left(x^2 f(x) - j^2 f(j)\right)\right) \mathrm{d}x. \end{aligned} \tag{5.4.40}$$

Since f is non-decreasing this is bounded from above by

$$\text{const} \exp\left(-\frac{m}{N} j^2 f(j)\right) \cdot \int_0^j \exp\left(\frac{m}{N} x^2 f(j)\right) \mathrm{d}x. \tag{5.4.41}$$

By substituting $z = \sqrt{\frac{m}{N} f(j)}x$ the integral is equal to

$$\sqrt{\frac{N}{m f(j)}} \int_0^{\sqrt{\frac{m}{N}} j f(j)} e^{z^2} \, \mathrm{d}z.$$

Since $f(x) \ge \frac{1}{2}$ for $x \gg \sqrt{N/m}$, we obtain by applying [Wei, (1),(9)], that this is bounded from above by

$$\text{const} \cdot \frac{e^{j^2 \cdot \frac{m}{N}}}{j} \cdot \frac{N}{m f(j)}.$$

So (5.4.41) - and hence also (5.4.40) - is asymptotically bounded by const $\cdot \frac{N}{m j f(j)}$, which concludes the proof. □

Proof of Lemma 5.4.6. Note, that the mapping $k \mapsto r_k$ is decreasing on $\{0, \dots, \mathfrak{a}\}$ and increasing on $\{\mathfrak{a}+1, \mathfrak{a}+2, \dots\}$. Hence for $0 \le k \le j$,

$$\frac{r_k}{r_j} \le \begin{cases} \frac{r_0}{r_\mathfrak{a}} & \text{if } k \le \mathfrak{a} \\ 1 & \text{if } k > \mathfrak{a}. \end{cases}$$

From Lemma 5.4.2 we get

$$\frac{r_0}{r_\mathfrak{a}} \sim e^{\varphi(0) - \varphi(\mathfrak{a})} = \exp\left(-\int_0^\mathfrak{a} \varphi'(x)\, dx\right), \tag{5.4.42}$$

where for $x \geq 1$

$$\begin{aligned}\varphi(x) &:= \frac{m}{N}x^2 f(x) \\ &= x\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{x^{\ell+1}}{\ell(\ell+1)N^\ell}.\end{aligned}$$

The derivative of φ is

$$\varphi'(x) = \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{x^{\ell}}{(\ell+1)N^\ell}.$$

We observe that for all sufficiently large N and all $x < N$

$$\varphi'(x) \geq \log\left(\frac{1+2m}{1+2m/\rho}\right) \geq -4m(1-\rho).$$

This shows that the exponent in (5.4.42) is bounded from above by

$$4m(1-\rho)\mathfrak{a} = 4Nm(1-\rho)^2,$$

which converges to 0 as $N \to \infty$, thus proving that

$$\frac{r_1}{r_{\mathfrak{a}}} \to 1 \quad \text{as } N \to \infty. \tag{5.4.43}$$

Because of $1 \ll Nm \ll (1-\rho)^{-2}$ we have

$$\frac{R_j}{r_j} = \sum_{k=0}^{j-1}\frac{r_k}{r_j} \leq j\frac{r_0}{r_{\mathfrak{a}}}. \tag{5.4.44}$$

Finally, by definition,

$$\lambda_j \geq \frac{j}{4} \quad \text{for all } j \leq \frac{N}{2}. \tag{5.4.45}$$

Combining (5.4.43), (5.4.44) and (5.4.45) we arrive at (5.4.6). □

5.4.7 Proofs of Lemma 5.4.7 and Proposition 5.4.8

Proof of Lemma 5.4.7. We start by collecting a few properties of the function H defined in (5.4.26). The first two derivatives of H are

$$\begin{aligned}H'(y) &= -\frac{1}{2m}\left[\frac{1}{1-\rho}\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{(1-\rho)^{\ell-1}y^\ell}{\ell}\right] \\ &= -\frac{1}{2m(1-\rho)}\left[\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{(1-\rho)^{\ell}y^\ell}{\ell}\right]\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2m(1-\rho)}\log\left(\frac{1+2m}{1+2m/\rho}\frac{1-(1-\rho)y/(1+2m)}{1-(1-\rho)y}\right)\\
&= -\frac{1}{2m(1-\rho)}\log\left(\frac{1+2m-(1-\rho)y}{(1+2m/\rho)(1-(1-\rho)y)}\right)\\
&= -\frac{1}{2m(1-\rho)}\log\left(1+\frac{2m}{\rho}\frac{(1-\rho)(y-1)}{(1+2m/\rho)(1-(1-\rho)y)}\right). \qquad (5.4.46)
\end{aligned}
$$

Since $\rho \geq \rho_0$ we have that for $y < \frac{1}{2}$ and N large enough

$$
\left|\frac{2m}{\rho}\frac{(1-\rho)(y-1)}{(1+2m/\rho)(1-(1-\rho)y)}\right| \geq \frac{2m}{\rho}\frac{(1-\rho)\frac{1}{2}}{1+2m/\rho},
$$

such that

$$
\left|\log\left(1+\frac{2m}{\rho}\frac{(1-\rho)(y-1)}{(1+2m/\rho)(1-(1-\rho)y)}\right)\right| \geq \frac{1}{2}\frac{2m(1-\rho)}{\rho(1+2m/\rho)},
$$

which gives

$$
H'(y) \geq \frac{1}{2m(1-\rho)}\frac{1}{2}\frac{2m(1-\rho)}{\rho(1+2m/\rho)} \geq \frac{1}{2\rho}\frac{1}{1+2m/\rho_0} \geq \frac{1}{4} \qquad (5.4.47)
$$

for $y \leq \frac{1}{2}$ and N large enough. We continue with the analysis of H'' and obtain

$$
\begin{aligned}
H''(y) &= -\frac{1}{2m(1-\rho)}\left[-\frac{(1-\rho)}{1+2m-(1-\rho)y}+\frac{(1-\rho)}{1-(1-\rho)y}\right]\\
&= \frac{1}{2m}\left[\frac{1}{1+2m-(1-\rho)y}-\frac{1}{1-(1-\rho)y}\right] \leq 0.
\end{aligned}
$$

Hence, $H(0) = 0$, H reaches its maximum at $y = 1$, and then decreases, and as $N \to \infty$

$$
H(1) = -\frac{1}{2m}\left[\frac{1}{1-\rho}\log\left(\frac{1+2m}{1+2m/\rho}\right)+\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{(1-\rho)^{\ell-1}}{\ell(\ell+1)}\right],
$$

and

$$
H''(1) = \frac{1}{2\rho m}\left[\frac{1}{1+2m/\rho}-1\right] \sim -\frac{1}{2\rho m}2m/\rho = -\frac{1}{\rho^2}, \qquad \text{as } N \to \infty.
$$

Moreover, H is non-negative on $[0, y_0]$ and negative on (y_0, ∞), with y_0 satisfying

$$
y_0 \sim \frac{2}{\rho} \qquad \text{as } N \to \infty.
$$

For later use we also notice that from (5.4.46) we get for all $y \in \mathbb{R} \setminus \{1\}$

$$
H'(y) \sim \frac{1-y}{\rho(1-(1-\rho)y)} \qquad \text{as } N \to \infty. \qquad (5.4.48)
$$

We now focus on the **second bullet point** of the lemma. So j is of the form

$$
j = \frac{g(N)}{2m(1-\rho)} \qquad (5.4.49)
$$

with $g(N)$ satisfying

$$g(N) \to \infty \quad \text{and} \quad g(N) \le (2/\rho - \xi)2m(1-\rho)^2 N.$$

In particular, this means

$$\frac{1}{m(1-\rho)} \ll j \le (2/\rho - \xi)(1-\rho)N \wedge N(1-\sqrt{m}).$$

Using (5.4.7) we obtain by a sandwiching argument

$$\begin{aligned}
R_j &\sim \int_0^j \exp\left(x \log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1 - \frac{1}{(1+2m)^\ell}\right)\frac{1}{\ell(\ell+1)}\frac{x^{\ell+1}}{N^\ell}\right)\mathrm{d}x \\
&= j\int_0^1 \exp\left(jy\left[\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1 - \frac{1}{(1+2m)^\ell}\right)\frac{1}{\ell(\ell+1)}\frac{(jy)^\ell}{N^\ell}\right]\right)\mathrm{d}y \\
&= j\int_0^1 \exp\left(-2m(1-\rho)^2 N H\left(\frac{jy}{N(1-\rho)}\right)\right)\mathrm{d}y \\
&= j\left(\int_0^\varepsilon \exp\left(-2m(1-\rho)^2 N H\left(\frac{jy}{N(1-\rho)}\right)\right)\mathrm{d}y\right. \\
&\qquad \left. + \int_\varepsilon^1 \exp\left(-2m(1-\rho)^2 N H\left(\frac{jy}{N(1-\rho)}\right)\right)\mathrm{d}y\right)
\end{aligned}$$

where an adequate choice of ε (independent of N) will be made below. By (5.4.49) and the above stated properties of the function H we get that for all $y \in [\varepsilon, 1]$,

$$H\left(\frac{jy}{N(1-\rho)}\right) \ge H\left(\frac{g(N)\varepsilon}{2m(1-\rho)^2 N}\right) \wedge H\left(2/\rho - \xi\right). \tag{5.4.50}$$

Moreover, because

$$H'(0) \sim \frac{1}{\rho} \quad \text{and} \quad H'(2/\rho) \sim -\frac{1}{\rho} \quad \text{as } N \to \infty,$$

combining (5.4.47) and (5.4.50) and setting $\varepsilon = \frac{1}{4}$ we obtain for all $y \in [\varepsilon, 1]$ and N large enough,

$$\begin{aligned}
H\left(\frac{jy}{N(1-\rho)}\right) &\ge \inf_{0 \le u \le \varepsilon} H'((2/\rho - \xi)u)\frac{g(N)\varepsilon}{2\rho m(1-\rho)^2 N} \wedge \inf_{0 \le u \le y_0 - 2/\rho + \xi} |H'(u)|\xi \\
&\ge \frac{g(N)\varepsilon}{8\rho m(1-\rho)^2 N} \wedge \frac{\xi}{4\rho}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\int_{\varepsilon}^{1} \exp\left(-2m(1-\rho)^2 N H\left(\frac{jy}{N(1-\rho)}\right)\right) \mathrm{d}y \\
\leq\ & \exp\left(-2m(1-\rho)^2 N\left(\frac{g(N)\varepsilon}{8\rho m(1-\rho)^2 N} \wedge \frac{\xi}{4\rho}\right)\right) \\
=\ & \exp\left(-\frac{g(N)\varepsilon}{4\rho}\right) \vee \exp\left(-\frac{1}{2} m(1-\rho)^2 N\xi/\rho\right).
\end{aligned}$$

The equivalence $H'(0) \sim 1/\rho$ also entails that

$$\begin{aligned}
& j\int_{0}^{\varepsilon} \exp\left(-2m(1-\rho)^2 N H\left(\frac{jy}{N(1-\rho)}\right)\right) \mathrm{d}y \\
\sim\ & j\int_{0}^{\varepsilon} \exp\left(-2m(1-\rho)^2 N \frac{jy}{N\rho(1-\rho)}\right) \mathrm{d}y \\
=\ & j\int_{0}^{\varepsilon} \exp\left(-2m(1-\rho)jy/\rho\right) \mathrm{d}y \\
\sim\ & \frac{\rho}{2m(1-\rho)} = \frac{j\rho}{g(N)}.
\end{aligned}$$

In total, as

$$\exp\left(-\frac{g(N)\varepsilon}{4\rho}\right) = o\left(\frac{1}{g(N)}\right)$$

and

$$\exp\left(-\frac{1}{2}m(1-\rho)^2 N\xi/\rho\right) \leq \frac{2}{(Nm(1-\rho))^2} \leq \left(\frac{8}{\rho g(N)}\right)^2 = o\left(\frac{1}{g(N)}\right)$$

this gives $R_j \sim \rho/(2m(1-\rho))$ and thus proves the second bullet point of the lemma.

We now turn to the **third bullet point**. So j is of the form $j = C(1-\rho)N$ with $C > 2/\rho$ and $j \leq N(1-\sqrt{m})$. (Note that $C(1-\rho)N \leq N$ and $C > 2/\rho$ imply $\rho > 2/3$.) Using (5.4.8) and sandwiching arguments yields

$$R_j \sim (1-\rho)N \int_0^C \exp\left(-2m(1-\rho)^2 N H(y)\right) \mathrm{d}y.$$

$H(y)$ is non-negative for $y \leq y_0$ with $y_0 \sim 2/\rho$, and negative for $y > y_0$. Moreover, $H'' < 0$. Consequently we get

$$R_j \sim (1-\rho)N \int_{y_0}^C \exp\left(-2m(1-\rho)^2 N H(y)\right) \mathrm{d}y.$$

By (5.4.48) there exists $\mathfrak{x} \geq 0$ such that

$$H'(y) \in [-\mathfrak{x}\xi - (C-1)/(\rho(1-C(1-\rho))), \mathfrak{x}\xi - (C-1)/(\rho(1-C(1-\rho)))], \quad \forall y \in [C-\xi, C+\xi].$$

This entails

$$\begin{aligned} & \int_{C-\xi}^{C} e^{-2m(1-\rho)^2 N((C-1)/(\rho(1-C(1-\rho)))+\mathfrak{r}\xi)(C-y)}\,\mathrm{d}y \\ \leq & \int_{C-\xi}^{C} e^{-2m(1-\rho)^2 N(H(y)-H(C))}\,\mathrm{d}y \\ \leq & \int_{C-\xi}^{C} e^{-2m(1-\rho)^2 N((C-1)/(\rho(1-C(1-\rho)))-\mathfrak{r}\xi)(C-y)}\,\mathrm{d}y. \end{aligned}$$

Moreover,

$$\begin{aligned} \int_{y_0}^{C-\xi} \exp\left(-2m(1-\rho)^2 NH(y)\right)\mathrm{d}y &\leq C\exp\left(-2m(1-\rho)^2 NH(C-\xi)\right) \\ &= o\left(\exp\left(-2m(1-\rho)^2 NH(C)\right)\right). \end{aligned}$$

We deduce that

$$\begin{aligned} R_j &\sim (1-\rho)N\frac{\rho(1-C(1-\rho))\exp\left(-2m(1-\rho)^2 NH(C)\right)}{2m(C-1)(1-\rho)^2 N} \\ &= \frac{\rho(1-C(1-\rho))}{2m(C-1)(1-\rho)}\exp\left(-2m(1-\rho)^2 NH(C)\right), \end{aligned}$$

which proves the third bullet point of the lemma.

Finally, we focus on the **first bullet point**. Let $j \leq \xi(1-\rho)N$ with $\xi \leq \frac{1}{2}$. From the observation that

$$1 - \frac{1}{(1+2m)^\ell} \leq 2m\ell$$

we get

$$\begin{aligned} \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{1}{\ell(\ell+1)}\frac{j^\ell}{N^\ell} &\leq 2m\sum_{\ell=1}^{\infty}\frac{1}{\ell+1}\frac{j^\ell}{N^\ell} \\ &\leq m\sum_{\ell=1}^{\infty}\frac{j^\ell}{N^\ell} \\ &\leq m\frac{\xi(1-\rho)}{1-\xi(1-\rho)} \leq \frac{m}{\rho}\xi(1-\rho). \end{aligned}$$

As

$$\log\left(\frac{1+2m/\rho}{1+2m}\right) = \log\left(1+\frac{2m}{\rho}\frac{1-\rho}{1+2m}\right) \sim \frac{2m}{\rho}(1-\rho)$$

we deduce that

$$\sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^\ell}\right)\frac{1}{\ell(\ell+1)}\frac{j^\ell}{N^\ell} \leq \xi\log\left(\frac{1+2m/\rho}{1+2m}\right).$$

Hence for $j \le \xi(1-\rho)N \le (1-\rho)N$,

$$\begin{aligned} -(1+\xi)\frac{2m}{\rho}(1-\rho)j &\le j\log\left(\frac{1+2m}{1+2m/\rho}\right) + \sum_{\ell=1}^{\infty}\left(1-\frac{1}{(1+2m)^{\ell}}\right)\frac{j^{\ell+1}}{\ell(\ell+1)N^{\ell}} \\ &\le -(1-\xi)\frac{2m}{\rho}(1-\rho)j. \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{1-e^{-(1+2\xi)(2m/\rho)(1-\rho)j}}{(1+2\xi)(2m/\rho)(1-\rho)} &= \int_0^j e^{-(1+2\xi)(2m/\rho)(1-\rho)x}\,\mathrm{d}x \\ \le R_j &\le \int_0^j e^{-(1-2\xi)(2m/\rho)(1-\rho)x}\,\mathrm{d}x = \frac{1-e^{-(1-2\xi)(2m/\rho)(1-\rho)j}}{(1+2\xi)(2m/\rho)(1-\rho)}. \end{aligned}$$

This ends the proof of Lemma 5.4.7. □

Proof of Proposition 5.4.8. Let us first study the asymptotics of $B(\zeta)$. Thanks to the second bullet point of Lemma 5.4.7 we know that for any $(1-\zeta)(1-\rho)N \le k \le (1+\zeta)(1-\rho)N$ one has $R_k \sim \rho/(2m(1-\rho))$. Hence

$$B(\zeta) \sim \left(R_j \wedge \frac{\rho}{2m(1-\rho)}\right) \sum_{k=(1-\zeta)(1-\rho)N+1}^{(1+\zeta)(1-\rho)N} \frac{1}{\lambda_k r_k}.$$

Moreover, for $k \in \big[(1-\zeta)(1-\rho)N, (1+\zeta)(1-\rho)N\big]$ we have $\lambda_k \sim \rho k/2$. Hence

$$\begin{aligned} \frac{2}{\rho(1+2\zeta)(1-\rho)N} \sum_{k=(1-\zeta)(1-\rho)N+1}^{(1+\zeta)(1-\rho)N} \frac{1}{r_k} &\le \sum_{k=(1-\zeta)(1-\rho)N+1}^{(1+\zeta)(1-\rho)N} \frac{1}{\lambda_k r_k} \\ &\le \frac{2}{\rho(1-2\zeta)(1-\rho)N} \sum_{k=(1-\zeta)(1-\rho)N+1}^{(1+\zeta)(1-\rho)N} \frac{1}{r_k}. \end{aligned}$$

We are left with the study of

$$\sum_{k=(1-\zeta)(1-\rho)N+1}^{(1+\zeta)(1-\rho)N} \frac{1}{r_k} \sim N(1-\rho)\int_{1-\zeta}^{1+\zeta} \exp\big(2m(1-\rho)^2 N H(y)\big)\,\mathrm{d}y,$$

where the equivalence is a consequence of (5.4.8). (Note that the function H, see (5.4.26), appears in (5.4.8).) Since the function H reaches its maximum at 1 and $H''(1) \sim -1/\rho^2$, and since ζ satisfies

$$\zeta\sqrt{mN}(1-\rho) \to \infty \quad \text{as } N \to \infty,$$

an application of the Laplace method yields

$$\int_{1-\zeta}^{1+\zeta} \exp\big(2m(1-\rho)^2 N H(y)\big)\,\mathrm{d}y \sim \sqrt{\frac{2\pi\rho^2}{2m(1-\rho)^2 N}}\exp\big(2m(1-\rho)^2 N H(1)\big)$$

$$= \sqrt{\frac{\pi}{mN}} \frac{\rho}{1-\rho} \exp\left(2m(1-\rho)^2 NH(1)\right).$$

Hence

$$B(\zeta) \sim \left(R_j \wedge \frac{\rho}{2m(1-\rho)}\right) \sqrt{\frac{\pi}{mN}} \frac{2}{1-\rho} \exp\left(2m(1-\rho)^2 NH(1)\right).$$

This completes the proof of (5.4.27).

To bound $A(\zeta)$, it is enough to notice that for any $k \leq (1-\zeta)(1-\rho)N$,

$$\lambda_k \geq \frac{k}{2}(1-(1-\zeta)(1-\rho)), \quad r_k \geq \exp\left(-2m(1-\rho)^2 NH(1-\zeta/2)\right), \quad R_{k\wedge j} \leq \frac{\rho}{2m(1-\rho)}.$$

The first inequality is a direct consequence of the definition of λ_k in (5.4.1), the second one stems from equality (5.4.7), and the last one is a consequence of Lemma 5.4.7. Altogether this yields that

$$\begin{aligned} A(\zeta) &\leq \frac{(1-\zeta)(1-\rho)\rho N}{2m(1-\rho)} \frac{2\exp\left(2m(1-\rho)^2 NH(1-\zeta/2)\right)}{(1-\zeta)(1-\rho)N(1-(1-\zeta)(1-\rho))} \\ &= o\left(B(\zeta)\right). \end{aligned}$$

The term $C(\zeta)$ is more delicate to bound and we have to decompose it into several terms. This decomposition depends on the value of ρ:

Let us begin with the simplest case, that is $\rho \leq 2/3$. In this case $(2/\rho)(1-\rho) \geq 1$ and thus $(2/\rho)(1-\rho)N \geq N$. Recall that the positive root y_0 of H satisfies $y_0 \sim 2/\rho$. We may decompose $C(\zeta)$ as follows:

$$\begin{aligned} C(\zeta) &= \sum_{k=(1+\zeta)(1-\rho)N+1}^{(y_0(1-\rho)\wedge 1)N(1-\sqrt{m})} \frac{R_{k\wedge j}}{\lambda_k r_k} + \sum_{k=(y_0(1-\rho)\wedge 1)N(1-\sqrt{m})+1}^{N} \frac{R_{k\wedge j}}{\lambda_k r_k} \\ &:= C_\alpha(\zeta) + C_\beta(\zeta). \end{aligned}$$

Using that H is non-negative and decreasing on $[1, y_0]$ and $y_0 \sim 2/\rho$ we obtain from equality (5.4.7) that for any $(1+\zeta)(1-\rho)N + 1 \leq k \leq (y_0(1-\rho) \wedge 1)N(1-\sqrt{m})$,

$$\lambda_k r_k \geq km \exp\left(-2m(1-\rho)^2 NH(1+\zeta/2)\right) \quad \text{and} \quad R_{k\wedge j} \leq \frac{\rho}{2m(1-\rho)}.$$

Hence

$$C_\alpha(\zeta) \leq \frac{\rho N}{2m(1-\rho)} \frac{1}{(1-\rho)Nm} \exp\left(2m(1-\rho)^2 NH(1+\zeta/2)\right) = o\left(B(\zeta)\right).$$

To bound $C_\beta(\zeta)$ we apply (5.4.24) with $\xi \sim \sqrt{m}$ satisfying

$$(y_0(1-\rho) \wedge 1)N(1-\sqrt{m}) = N(1-\xi).$$

This shows that $C(\zeta) = o(B(\zeta))$ in the case $\rho \leq 2/3$.

Let us now consider the case $\rho > 2/3$. We decompose $C(\zeta)$ into three terms as follows:

$$\begin{aligned} C(\zeta) &= \sum_{k=(1+\zeta)(1-\rho)N+1}^{(2/\rho-\zeta)(1-\rho)N} \frac{R_{k\wedge j}}{\lambda_k r_k} + \sum_{k=(2/\rho-\zeta)(1-\rho)N+1}^{(2/\rho+\zeta)(1-\rho)N} \frac{R_{k\wedge j}}{\lambda_k r_k} + \sum_{k=(2/\rho+\zeta)(1-\rho)N+1}^{N} \frac{R_{k\wedge j}}{\lambda_k r_k} \\ &:= C_1(\zeta) + C_2(\zeta) + C_3(\zeta). \end{aligned}$$

$C_1(\zeta)$ may be bounded with similar arguments as for $A(\zeta)$: for any $(1+\zeta)(1-\rho)N \le k \le (2/\rho - \zeta)(1-\rho)N$,

$$\lambda_k r_k \ge km \exp\left(-2m(1-\rho)^2 NH(1+\zeta/2)\right) \quad \text{and} \quad R_{k\wedge j} \le \frac{\rho}{2m(1-\rho)}.$$

This entails

$$C_1(\zeta) \le \frac{(1-2\zeta)(1-\rho)N}{2m\rho(1-\rho)} \frac{1}{(1-\rho)Nm} \exp\left(2m(1-\rho)^2 NH(1+\zeta/2)\right) = o\left(B(\zeta)\right).$$

Now, recalling the third bullet point of Lemma 5.4.7, that $R_\cdot$ is increasing and that r_k is increasing with k when k is larger than $(1+\zeta)(1-\rho)N$, we get for $(2/\rho-\zeta)(1-\rho)N \le k \le (2/\rho+\zeta)(1-\rho)N$,

$$\lambda_k r_k \ge km \exp\left(-2m(1-\rho)^2 NH(2/\rho - 2\zeta)\right)$$

and

$$R_k \le R_{(2/\rho+\zeta)(1-\rho)N} \sim \rho(1-(2/\rho+\zeta)(1-\rho)) \frac{\exp\left(-2m(1-\rho)^2 NH(2+\zeta)\right)}{2m(2/\rho+\zeta-1)(1-\rho)}.$$

Using that $H'(2/\rho) \sim -1/\rho$, we deduce that there exists a constant $\mathfrak{K}$ such that

$$\begin{aligned} C_2(\zeta) &\le \frac{\mathfrak{K}\zeta(1-\rho)N}{(1-\rho)^2 Nm^2} \exp\left(2m(1-\rho)^2 N(H(2-2\zeta) - H(2+2\zeta))\right) \\ &\le \frac{\mathfrak{K}\zeta}{(1-\rho)m^2} \exp\left(\mathfrak{K}m(1-\rho)^2 N\zeta\right) = o\left(B(\zeta)\right). \end{aligned}$$

Notice that for $k \ge (2/\rho+\zeta)(1-\rho)N - 1$, $r_k = \sup\{r_\ell,\ \ell \le k\}$. Hence, for any j,

$$\begin{aligned} C_3(\zeta) &= \sum_{k=(2/\rho+\zeta)(1-\rho)N+1}^{N} \frac{R_{k\wedge j}}{kq_k r_{k-1}} \\ &\le \frac{1}{m} \sum_{k=(2/\rho+\zeta)(1-\rho)N+1}^{N} \frac{R_k}{kr_{k-1}} \\ &= \frac{1}{m} \sum_{k=(2/\rho+\zeta)(1-\rho)N+1}^{N} \frac{r_0 + \dots + r_{k-2} + r_{k-1}}{kr_{k-1}} \le \frac{N}{m} = o\left(B(\zeta)\right). \end{aligned}$$

This shows that $C(\zeta) = o(B(\zeta))$ in the case $\rho > 2/3$, and thus concludes the proof of Proposition 5.4.8. $\square$

Acknowledgment We are grateful to an anonymous reviewer for a careful reading of a previous version and for suggestions that led to an improvement of the presentation. We thank the *Allianz für Hochleistungsrechnen Rheinland-Pfalz* for granting us access to the High Performance Computing ELWETRITSCH, on which the simulations underlying our figures have been performed. On a more personal note, one of the authors (A.W.) remembers with pleasure the inspiring collaboration with Alison Etheridge and Peter Pfaffelhuber which led to the paper [EPW09]. Together with [GSW23], [EPW09] constitutes the main root of the results and insights achieved in the present paper.

Funding This work was partially funded by the Chair "Modélisation Mathématique et Biodiversité" of VEOLIA-Ecole Polytechnique-MNHN-F.X. AGC was supported by PAPIIT-UNAM (grant IN101722) and by CONACYT Ciencia Básica (grant A1-S-14615).

Chapter 6

Muller's ratchet in a near-critical regime: type frequency profile and sample genealogy

This chapter presents a brief outlook on current work with Charline Smadi and Anton Wakolbinger. This work elaborates on Sections 4 and 5.2 of the preprint [IGSW23], thus relating the topic of Part III with the principal theme of this thesis.

Consider the tournament ratchet with the dynamics specified in Definition S.3.1. We focus on the near-critical exponential regime, that is $m_N(1-\rho_N)^2N \to \infty$ with $\rho_N \uparrow 1$ as $N \to \infty$. As in Chapter 5 we suppress the subscript N, thus writing $m \equiv m_N$, $\rho \equiv \rho_N$. A main objective is the proof of the following.

Claim 6.0.1. *Assume $m(1-\rho)^2N \to \infty$ as $N \to \infty$ and that all individuals at time 0 are of type 0. Let $t_N \gg 1/s$. Then*

a) the probability that a uniformly chosen individual at time t_N is of the $(k+1)$-th best type is asymptotically equal to p_k as $N \to \infty$, where the weights $(p_k)_k$ of the so called type frequency profile are given by

$$p_{-1} = 0, \qquad p_0 = 1-\rho, \qquad \rho(p_k - p_{k-1}) = p_k\left(1 - p_k - 2\sum_{\ell=0}^{k-1} p_\ell\right), \ k \geq 1, \quad (6.0.1)$$

b) the probability that n sampled individuals at time t_N are of types $k_1, \dots, k_n$ is asymptotically equal to

$$\prod_{\ell=1}^{n} p_{k_\ell - K^*(t)},$$

where $K^*(t)$ denotes the best type at time t (see Definition S.3.1).

This extends [GSW23, Theorem 2.3] to (part of) the near-critical regime. The proof of this claim builds on the so called Ancestral Selection Graph (ASG), which first appeared in the pioneering work of Krone and Neuhauser [KN97, NK97] and has been intensely used in mathematical population genetics since then; for some recent related work see for instance [BW18, GS18, GS20, CHS22]. This approach was used in [GSW23] to give a graphical representation of the Moran tournament ratchet and of its hierarchical dual.

The line counting process of the Ancestral Selection Graph is a birth-and-death process on $\mathbb{N}$ with upward jump rate $ns(1-n/N)$ and downward jump rate $\frac{n(n-1)}{2N}$. For the tournament ratchet the lineages are decorated by a Poisson point process with intensity m; the number of points on a (potential ancestral) lineage is the *(mutational) load* accumulated along that lineage. The *load of a potential ancestor* $\mathcal{I}'$ of an individual $\mathcal{I}$ is the minimum of the loads of the lineages connecting $\mathcal{I}$ and $\mathcal{I}'$. If all the individuals at time 0 are of equal type, then the type of an individual sampled at time t is the minimum of the mutational loads carried by the ancestral lineages that lead back from this individual to time 0 (cf. [GSW23, Remark 4.2]). For an individual $\mathcal{I}$ sampled at time t, and $k \geq 0$, the graph made up by its load k potential ancestors is the *load* k *ASG* of $\mathcal{I}$, denoted by $\mathcal{A}_k = (\mathcal{A}_k(t-r))_{r\geq 0}$, see Figure 6.1 for an illustration. The *minimum load ASG (back from the total population at time* u*)*, denoted by $\bar{\mathcal{A}}_0^u$, consists at any time $t' \leq u$ of all those individuals living at time t' that are a load $k_{\min}$-potential ancestor of some individual living at time u, where $k_{\min}$ is the minimum of all the loads of lineages connecting times t' and u. For $m = 0$ this corresponds to the *equilibrium Ancestral Selection Graph* introduced and investigated by Pokalyuk and Pfaffelhuber in [PP13]. We define the *minimum load equilibrium ASG* $\bar{\mathcal{A}}_0$ as the local limit of $\bar{\mathcal{A}}_0^u$ as $u \to \infty$.

We write $A_k(t-r) := \#\mathcal{A}_k(t-r)$, and $\bar{A}_0(t-r) := \#\bar{\mathcal{A}}_0(t-r)$, $r \geq 0$, for the line counting processes of $\mathcal{A}_k$ and $\bar{\mathcal{A}}_0$.

[GSW23, Lemma 5.1] tells that both $A_0(t-r)$ and $\bar{A}_0(t-r)$, $r \geq 0$, are birth-and-death processes whose jump rates are given by

$$ns\left(1-\tfrac{n}{N}\right) \quad \text{from } n \text{ to } n+1, \tag{6.0.2}$$

$$n\left(m+\tfrac{n-1}{2N}\right) \quad \text{from } n \text{ to } n-1. \tag{6.0.3}$$

More generally, for any $k \in \mathbb{N}_0$ the process $(A_0, \ldots, A_k)$ has the Markovian multivariate birth-and-death dynamics specified in [GSW23] Lemma 5.1.

We now give evidence for the validity of Claim 6.0.1. Consider for $k \in \mathbb{N}_0$ the event

$$\begin{aligned}\mathscr{E}_k :=\{&\text{there is a coalescence event between } \mathcal{A}_k \text{ and } \bar{\mathcal{A}}_0 \\ &\text{but no coalescence event between } \mathcal{A}_j \text{ and } \bar{\mathcal{A}}_0 \text{ for } j<k\}.\end{aligned}$$

In [GSW23] the following was proved for the regime $m < s$ and $\rho = m/s$ not depending on N.

(H) *The probability that the randomly sampled individual* $\mathcal{I}$ *has type* k *is asymptotically equal to* $\mathbf{P}(\mathscr{E}_k)$ *as* $N \to \infty$.

See Figure 6.2 for an illustration. As the current work addressed at the beginning of this section reveals, **(H)** is valid also in the near-critical exponential regime. The link between **(H)** and the prediction (6.0.1) for the empirical type frequency profile is established by the following Statement.

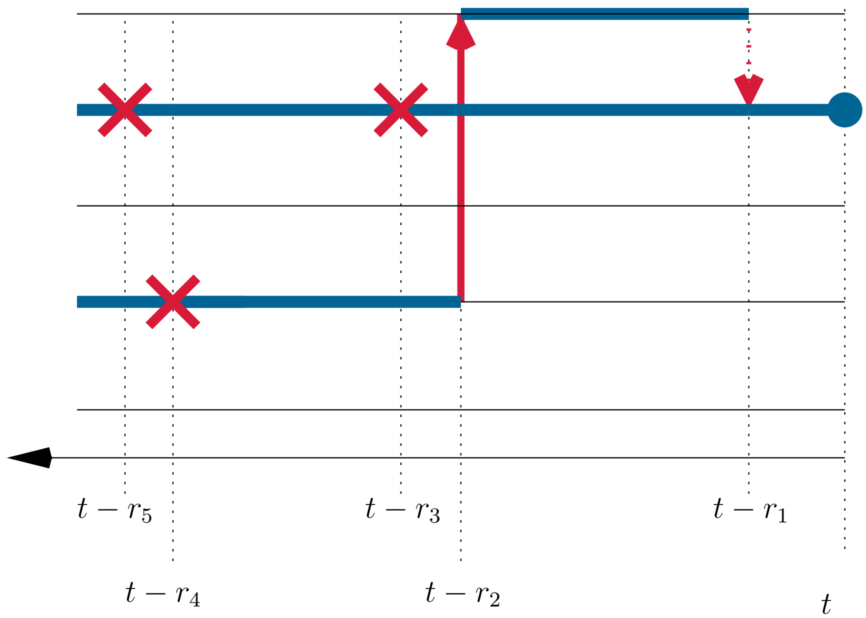


Figure 6.1: For $N = 4$ this is an illustration of the ASG of one individual sampled at time t (and symbolised by a blue dot). The selective arrow (drawn dotted) leads to an increase, and the neutral arrow (drawn bold) leads to a decrease of the line count of the ASG. Between time t and $t - r_1$ the ASG consists of one load 0 potential ancestor, from time $t - r_1$ till $t - r_2$ it consists of two load 0 potential ancestors, from time $t - r_2$ till time $t - r_3$ it consists of one load zero and one load one potential ancestor, from time $t - r_4$ till time $t - r_5$ it consists of two load one potential ancestors, and from time $t - r_5$ on it consists of one load zero and one load one potential ancestor.

Statement 6.0.2. *For $m(1-\rho)^2 N \to \infty$ and for all $n \in \mathbb{N}$, $k_1, \dots, k_n$*

$$\mathbf{P}\left(\mathscr{E}^1_{k_1} \cap \ldots \cap \mathscr{E}^n_{k_n}\right) \sim \prod_{\ell=1}^{n} p_{k_\ell - K^*(t)} \qquad \text{as } N \to \infty, \tag{6.0.4}$$

where the event $\mathscr{E}^{(\ell)}_{k_1}$ refers to the ℓ-th of the n sampled individuals.

In the exponential regime $m(1-\rho)^2 \to \infty$, the quantity $a = 2Nm(1-\rho)$ is the typical (order of) size which the process $\bar{\mathcal{A}}_0$ attains (and keeps) "between clicks". Notably, the condition $m(1-\rho)^2 \to \infty$, ensures that the center of attraction of A_0 is above $1/(1-\rho)$. However, since A_0 starts in 1 and since $\rho \uparrow 1$, it is only slightly supercritical below its center of attraction, such that the probability that A_0 makes it up to its center of attraction is small for large N. Indeed, in view of its dynamics (6.0.2), $(A_0(t-r))_{r \geq 0}$ behaves for sufficiently small r similar to a binary Galton-Watson process, and the entire ASG of the individual $\mathcal{I}$ behaves up to sufficiently small time horizons like a Poisson-decorated Yule-tree. See Figure 6.2 for an illustration. This helps to arrive at (6.0.4): Consider a standard Yule tree $\mathscr{Y}$ carrying a Poisson point process with intensity $\rho = m/s$. Let L be the minimum of the Poisson loads carried by the infinite branches of $\mathscr{Y}$. As an illustration of that, the tree depicted in the red box of Figure 6.2, when viewed as part of a decorated Yule tree, is compatible with the event $\{L = 0\}$. Let $\mathcal{T}$ be a random variable with standard exponential distribution (playing the role of the first splitting time of $\mathscr{Y}$). Given $\mathcal{T}$ let M_0 be a Poisson random variable with parameter $\mathcal{T}\rho$. Decomposing $\mathscr{Y}$ at the time of the first branching event, we obtain that L solves the stochastic fixed point equation

$$L \stackrel{d}{=} M_0 + \min(L_1, L_2), \tag{6.0.5}$$

where, given $\mathcal{T}$, the random variables M_0, L_1 and L_2 are independent. The distribution solving (6.0.5) then has the weights characterised by (6.0.1).

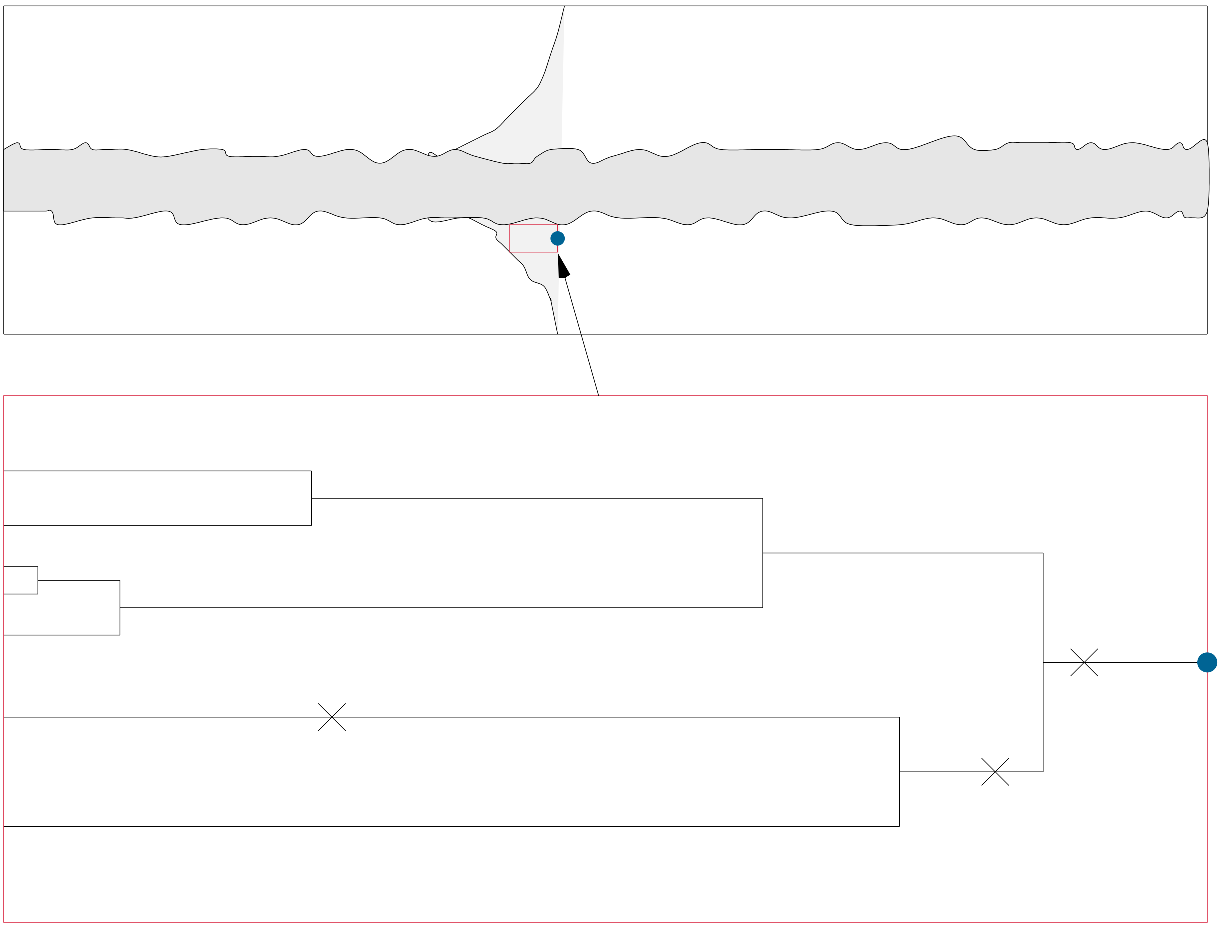


Figure 6.2: This is an illustration of the Poisson-decorated ASG of one individual sampled at time t (and symbolised by the blue dot). Let σ be the time of the occurrence of $\bigcup_{k\in\mathbb{N}_0}\mathscr{E}_k$. Up to the time σ the Poisson-decorated ASG of $\mathcal{I}$ behaves with high probability like a Poisson-decorated Yule tree. The depicted realisation suggests the event $\{L=1\}$, which (for large N) with high probability coincides with the event $\mathscr{E}_1$. According to **(H)** this means that the sampled individual with high probability carries $k_{\min}+1$ mutations, i.e. its type is second best among the individuals at time t.

Bibliography

Bibliography

[ABBS13] E. Aïdékon, J. Berestycki, E. Brunet, and Z. Shi. Branching Brownian motion seen from its tip. *Probab. Theory Related Fields*, 157(1-2):405–451, 2013. doi:10.1007/s00440-012-0461-0.

[ABK11] L.-P. Arguin, A. Bovier, and N. Kistler. Genealogy of extremal particles of branching Brownian motion. *Comm. Pure Appl. Math.*, 64(12):1647–1676, 2011. doi:10.1002/cpa.20387.

[ABK13a] L.-P. Arguin, A. Bovier, and N. Kistler. An ergodic theorem for the frontier of branching Brownian motion. *Electron. J. Probab.*, 18:no. 53, 25, 2013. doi:10.1214/EJP.v18-2082.

[ABK13b] L.-P. Arguin, A. Bovier, and N. Kistler. The extremal process of branching Brownian motion. *Probab. Theory Related Fields*, 157(3-4):535–574, 2013. doi:10.1007/s00440-012-0464-x.

[AKTV14] A. Anderson, R. Kumar, A. Tomkins, and S. Vassilvitskii. The dynamics of repeat consumption. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, page 419–430, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2566486.2568018.

[Ale11] K. S. Alexander. Excursions and local limit theorems for Bessel-like random walks. *Electron. J. Probab.*, 16:no. 1, 1–44, 2011. doi:10.1214/EJP.v16-848.

[AN72] K. B. Athreya and P. E. Ney. *The Galton-Watson Process*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1972. doi:10.1007/978-3-642-65371-1.

[AP13] J. Audiffren and E. Pardoux. Muller's ratchet clicks in finite time. *Stochastic Processes and their Applications*, 123(6):2370–2397, 2013. doi:10.1016/j.spa.2013.02.008.

[AS95] R. J. Adler and G. Samorodnitsky. Super Fractional Brownian Motion, Fractional Super Brownian Motion and Related Self-Similar (Super) Processes. *The Annals of Probability*, 23(2):743 – 766, 1995. doi:10.1214/aop/1176988287.

[BDLO23] N. Broutin, L. Devroye, G. Lugosi, and R. I. Oliveira. Subtractive random forests, 2023, arXiv:2210.10544.

[BDMPP87] C. Boldrighini, A. De Masi, A. Pellegrinotti, and E. Presutti. Collective phenomena in interacting particle systems. *Stochastic Process. Appl.*, 25(1):137–152, 1987. doi:10.1016/0304-4149(87)90194-3.

[Ber99] J. Bertoin. Intersection of independent regenerative sets. *Probab. Theory Related Fields*, 114(1):97–121, 1999. doi:10.1007/s004400050223.

[BFM18] T. Bäck, D. B. Fogel, and Z. Michalewicz. *Evolutionary computation 1: Basic algorithms and operators*. CRC press, 2018.

[BGKS13] J. Blath, A. González Casanova, N. Kurt, and D. Spanò. The ancestral process of long-range seed bank models. *J. Appl. Probab.*, 50(3):741–759, 2013. doi:10.1239/jap/1378401233.

[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.

[BH48] R. Bellman and T. E. Harris. On the theory of age-dependent stochastic branching processes. *Proceedings of the National Academy of Sciences*, 34(12):601–604, 1948. doi:10.1073/pnas.34.12.601.

[BH14] A. Bovier and L. Hartung. The extremal process of two-speed branching Brownian motion. *Electron. J. Probab.*, 19:no. 18, 28, 2014. doi:10.1214/EJP.v19-2982.

[BH15] A. Bovier and L. Hartung. Variable speed branching Brownian motion 1. Extremal processes in the weak correlation regime. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(1):261–291, 2015.

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

[BK89] R. M. Burton and M. Keane. Density and uniqueness in percolation. *Comm. Math. Phys.*, 121(3):501–505, 1989. URL http://projecteuclid.org/euclid.cmp/1104178143.

[BK04a] A. Bovier and I. Kurkova. Derrida's generalised random energy models. I. Models with finitely many hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(4):439–480, 2004. doi:10.1016/j.anihpb.2003.09.002.

[BK04b] A. Bovier and I. Kurkova. Derrida's generalized random energy models 2: models with continuous hierarchies. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 40(4):481–495, 2004. doi:10.1016/j.anihpb.2003.09.003.

[BK06] E. Bolthausen and N. Kistler. On a nonhierarchical version of the generalized random energy model. *Ann. Appl. Probab.*, 16(1):1–14, 2006. doi:10.1214/105051605000000665.

[BK09a] E. Bolthausen and N. Kistler. On a nonhierarchical version of the generalized random energy model. II. Ultrametricity. *Stochastic Process. Appl.*, 119(7):2357–2386, 2009. doi:10.1016/j.spa.2008.12.002.

[BK09b] A. Bovier and I. Kurkova. A short course on mean field spin glasses. In *Spin glasses: statics and dynamics*, volume 62 of *Progr. Probab.*, pages 3–44. Birkhäuser Verlag, Basel, 2009. doi:10.1007/978-3-7643-9891-0_1.

[Bov15] A. Bovier. From spin glasses to branching Brownian motion—and back? In *Random walks, random fields, and disordered systems*, volume 2144 of *Lecture Notes in Math.*, pages 1–64. Springer, Cham, 2015. doi:10.1007/978-3-319-19339-7_1.

[Bra78] M. D. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, 31(5):531–581, 1978. doi:10.1002/cpa.3160310502.

[BS02] E. Bolthausen and A.-S. Sznitman. *Ten lectures on random media*, volume 32 of *DMV Seminar*. Birkhäuser Verlag, Basel, 2002. doi:10.1007/978-3-0348-8159-3.

[BS22] C. Bräutigam and M. Smerlak. Diffusion approximations in population genetics and the rate of Muller's ratchet. *Journal of Theoretical Biology*, 550:111236, 2022. doi:10.1016/j.jtbi.2022.111236.

[BT96] T. Blickle and L. Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394, 1996. doi:10.1162/evco.1996.4.4.361.

[BW18] E. Baake and A. Wakolbinger. Lines of descent under selection. *Journal of Statistical Physics*, 172(1):156–174, 2018. doi:10.1007/s10955-017-1921-9.

[CCM16] J. R. Chazottes, P. Collet, and S. Méléard. Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes. *Probability Theory and Related Fields*, 164(1):285–332, 2016. doi:10.1007/s00440-014-0612-6.

[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.

[CF17] Z.-Q. Chen and W.-T. Fan. Hydrodynamic limits and propagation of chaos for interacting random walks in domains. *Ann. Appl. Probab.*, 27(3):1299–1371, 2017. doi:10.1214/16-AAP1208.

[CHS22] F. Cordero, S. Hummel, and E. Schertzer. General selection models: Bernstein duality and minimal ancestral structures. *The Annals of Applied Probability*, 32(3):1499 – 1556, 2022. doi:10.1214/21-AAP1683.

[CJ23] D. Cheek and S. G. G. Johnston. Ancestral reproductive bias in branching processes. *Journal of Mathematical Biology*, 86(5):70, 2023. doi:10.1007/s00285-023-01907-7.

[CKT20] F. Chierichetti, R. Kumar, and A. Tomkins. Asymptotic behavior of sequence models. In *Proceedings of The Web Conference 2020*, WWW '20, page 2824–2830, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3366423.3380044.

[CR88] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probab. Theory Related Fields*, 80(2):299–314, 1988. doi:10.1007/BF00356108.

[CRW91] B. Chauvin, A. Rouault, and A. Wakolbinger. Growing conditioned trees. *Stochastic Processes and their Applications*, 39(1):117–130, 1991. doi:10.1016/0304-4149(91)90036-C.

[CTAS+19] T. H. Coorens, T. D. Treger, R. Al-Saadi, L. Moore, M. G. Tran, T. J. Mitchell, S. Tugnait, C. Thevanesan, M. D. Young, T. R. Oliver, et al. Embryonal precursors of Wilms tumor. *Science*, 366(6470):1247–1251, 2019. doi:10.1126/science.aax1323.

[Der80] B. Derrida. Random-energy model: Limit of a family of disordered models. *Phys. Rev. Lett.*, 45:79–82, Jul 1980. doi:10.1103/PhysRevLett.45.79.

[Der81] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, Sep 1981. doi:10.1103/PhysRevB.24.2613.

[Der85] B. Derrida. A generalization of the random energy model which includes correlations between energies. *Journal de Physique Lettres*, 46(9):401–407, 1985.

[DLMF] *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.1.12 of 2023-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[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.

[Dro23] R. Drogin. Long range voter models and dynamical fractional brownian motion, 2023, arXiv:2311.03662.

[DSS05] C. R. Doering, K. V. Sargsyan, and L. M. Sander. Extinction times for birth-death processes: exact results, continuum asymptotics, and the failure of the Fokker-Planck approximation. *Multiscale Model. Simul.*, 3(2):283–299, 2005. doi:10.1137/030602800.

[EPW09] A. M. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. *How often does the ratchet click? Facts, heuristics, asymptotics*, page 365–390. London Mathematical Society Lecture Note Series. Cambridge University Press, 2009, arXiv:0709.2775. doi:10.1017/CBO9781139107020.016.

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

[Fel74] J. Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78(2):737–756, 10 1974. doi:10.1093/genetics/78.2.737.

[FZ12a] M. Fang and O. Zeitouni. Branching random walks in time inhomogeneous environments. *Electronic Journal of Probability*, 17(none):1 – 18, 2012. doi:10.1214/EJP.v17-2253.

[FZ12b] M. Fang and O. Zeitouni. Slowdown for time inhomogeneous branching brownian motion. *Journal of Statistical Physics*, 149(1):1–9, 2012. doi:10.1007/s10955-012-0581-z.

[GB03] H.-O. Georgii and E. Baake. Supercritical multitype branching processes: the ancestral types of typical individuals. *Advances in Applied Probability*, 35(4):1090–1110, 2003. doi:10.1239/aap/1067436336.

[GC00] I. Gordo and B. Charlesworth. The Degeneration of Asexual Haploid Populations and the Speed of Muller's Ratchet. *Genetics*, 154(3):1379–1387, 03 2000. doi:10.1093/genetics/154.3.1379.

[Gei99] J. Geiger. Elementary new proofs of classical limit theorems for Galton-Watson processes. *Journal of Applied Probability*, 36(2):301–309, 1999. doi:10.1239/jap/1032374454.

[GI23] A. González Casanova and J. L. Igelbrink. Branching fractional Brownian motion: discrete approximations and maximal displacement, 2023, arXiv:2310.04386.

[GN96] G. Gripenberg and I. Norros. On the prediction of fractional Brownian motion. *J. Appl. Probab.*, 33(2):400–410, 1996. doi:10.2307/3215063.

[GS18] A. González Casanova and D. Spanò. Duality and fixation in Ξ-Wright–Fisher processes with frequency-dependent selection. *The Annals of Applied Probability*, 28(1):250–284, 2018. URL `https://www.jstor.org/stable/26542309`.

[GS20] A. González Casanova and C. Smadi. On Λ-Fleming–Viot processes with general frequency-dependent selection. *Journal of Applied Probability*, 57(4):1162–1197, 2020.

[GSW23] A. González Casanova, C. Smadi, and A. Wakolbinger. Quasi-equilibria and click times for a variant of Muller's ratchet. *Electron. J. Probab.*, 28:Paper No. 159, 37, 2023. doi:10.1214/23-ejp1055.

[Hai78] J. Haigh. The accumulation of deleterious genes in a population—Muller's ratchet. *Theoretical Population Biology*, 14(2):251–267, 1978. doi:10.1016/0040-5809(78)90027-8.

[Har63] T. E. Harris. *The theory of branching processes*, volume Band 119 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin; Prentice Hall, Inc., Englewood Cliffs, NJ, 1963.

[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.

[IGSW23] J. L. Igelbrink, A. González Casanova, C. Smadi, and A. Wakolbinger. Muller's ratchet in a near-critical regime: tournament versus fitness proportional selection, 2023, arXiv:2306.00471v1. URL `https://arxiv.org/abs/2306.00471v1`.

[IGSW24] J. Igelbrink, A. González Casanova, C. Smadi, and A. Wakolbinger. Muller's ratchet in a near-critical regime: Tournament versus fitness proportional selection. *Theoretical Population Biology*, 158:121–138, 2024. doi:10.1016/j.tpb.2024.06.001.

[II24] J. L. Igelbrink and J. Ischebeck. Ancestral reproductive bias in continuous-time branching trees under various sampling schemes. *J. Math. Biol.*, 89(11), 2024. doi:10.1007/s00285-024-02105-9.

[INW68] N. Ikeda, M. Nagasawa, and S. Watanabe. Branching Markov processes. II. *J. Math. Kyoto Univ.*, 8:365–410, 1968. doi:10.1215/kjm/1250524059.

[IW23] J. L. Igelbrink and A. Wakolbinger. Asymptotic Gaussianity via coalescence probabilities in the Hammond-Sheffield urn. *ALEA Lat. Am. J. Probab. Math. Stat.*, 20(1):53–74, 2023. doi:10.30757/alea.v20-04.

[JN96] P. Jagers and O. Nerman. The asymptotic composition of supercritical, multi-type branching populations. *Séminaire de probabilités de Strasbourg*, 30:40–54, 1996. doi:10.1007/BFb0094640.

[Kis15] N. Kistler. Derrida's random energy models. From spin glasses to the extremes of correlated random fields. In *Correlated random systems: five different methods*, volume 2143 of *Lecture Notes in Math.*, pages 71–120. Springer, Cham, 2015. doi:10.1007/978-3-319-17674-1_3.

[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.

[Kle08] A. Klenke. *Wahrscheinlichkeitstheorie*. Masterclass. Springer Berlin Heidelberg, 2008.

[KLPP97] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. *A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes*, pages 181–185. Springer New York, New York, NY, 1997. doi:10.1007/978-1-4612-1862-3_14.

[KLS19] M. A. Kouritzin, K. Lê, and D. Sezer. Laws of large numbers for supercritical branching gaussian processes. *Stochastic Processes and their Applications*, 129(9):3463–3498, 2019. doi:doi.org/10.1016/j.spa.2018.09.011.

[KN97] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51(3):210–237, 1997. doi:https://doi.org/10.1006/tpbi.1997.1299.

[KS15] N. Kistler and M. A. Schmidt. From Derrida's random energy model to branching random walks: from 1 to 3. *Electron. Commun. Probab.*, 20:no. 47, 12, 2015. doi:10.1214/ecp.v20-4189.

[KS23] N. Kistler and G. Sebastiani. On a nonhierarchical generalization of the perceptron GREM. *Stochastic Process. Appl.*, 163:1–24, 2023. doi:10.1016/j.spa.2023.05.008.

[KSS23] N. Kistler, M. A. Schmidt, and G. Sebastiani. On the REM approximation of TAP free energies. *J. Phys. A*, 56(29):Paper No. 294001, 30, 2023. doi:10.1088/1751-8121/acdf30.

[Lam05] A. Lambert. The branching process with logistic growth. *The Annals of Applied Probability*, 15(2):1506–1535, 2005. doi:10.1214/105051605000000098.

[Law18] G. F. Lawler. Notes on the bessel process. 2018. URL `https://api.semanticscholar.org/CorpusID:52200396`.

[LPP95] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of L log L criteria for mean behavior of branching processes. *The Annals of Probability*, 23(3):1125–1138, 1995. doi:10.1214/aop/1176988176.

[LS87] S. P. Lalley and T. Sellke. A Conditional Limit Theorem for the Frontier of a Branching Brownian Motion. *The Annals of Probability*, 15(3):1052 – 1061, 1987. doi:10.1214/aop/1176992080.

[McK75] H. P. McKean. Application of brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Communications on Pure and Applied Mathematics*, 28(3):323–331, 1975. doi:10.1002/cpa.3160280302.

[ME13] J. J. Metzger and S. Eule. Distribution of the fittest individuals and the rate of muller's ratchet in a model with overlapping generations. *PLOS Computational Biology*, 9(11):1–10, 2013. doi:10.1371/journal.pcbi.1003303.

[MPV20] M. Mariani, E. Pardoux, and A. Velleret. Metastability between the clicks of the Muller ratchet, 2020, arXiv:2007.14715.

[Mul32] H. J. Muller. Some genetic aspects of sex. *The American Naturalist*, 66(703):118–138, 1932, arXiv:https://doi.org/10.1086/280418. doi:10.1086/280418.

[Mul64] H. Muller. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9, 1964. doi:10.1016/0027-5107(64)90047-8.

[MVN68] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968. doi:10.1137/1010093.

[MZ16] P. Maillard and O. Zeitouni. Slowdown in branching Brownian motion with inhomogeneous variance. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 52(3):1144 – 1160, 2016. doi:10.1214/15-AIHP675.

[NK97] C. Neuhauser and S. M. Krone. The Genealogy of Samples in Models With Selection. *Genetics*, 145(2):519–534, 02 1997. doi:10.1093/genetics/145.2.519.

[NS12] R. A. Neher and B. I. Shraiman. Fluctuations of fitness distributions and the rate of muller's ratchet. *Genetics*, 191(4):1283–1293, 2012. doi:10.1534/genetics.112.141325.

[PBB+15] T. Paixão, G. Badkobeh, N. Barton, D. Çörüş, D.-C. Dang, T. Friedrich, P. K. Lehre, D. Sudholt, A. M. Sutton, and B. Trubenová. Toward a unifying framework for evolutionary processes. *Journal of Theoretical Biology*, 383:28–43, 2015. doi:0.1016/j.jtbi.2015.07.011.

[PMK+21] S. Park, N. M. Mali, R. Kim, J.-W. Choi, J. Lee, J. Lim, J. M. Park, J. W. Park, D. Kim, T. Kim, et al. Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature*, 597(7876):393–397, 2021. doi:10.1038/s41586-021-03786-8.

[PP13] C. Pokalyuk and P. Pfaffelhuber. The ancestral selection graph under strong directional selection. *Theoretical Population Biology*, 87(C):25–33, 2013. URL `https://EconPapers.repec.org/RePEc:eee:thpobi:v:87:y:2013:i:c:p:25-33`.

[PSW12] P. Pfaffelhuber, P. R. Staab, and A. Wakolbinger. Muller's ratchet with compensatory mutations. *The Annals of Applied Probability*, 22(5):2108 – 2132, 2012. doi:10.1214/11-AAP836.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. doi:10.1007/978-3-662-06400-9.

[SaSha13] S. Sagitov and A. Shaimerdenova. Extinction times for a birth–death process with weak competition. *Lith. Math. J.*, 53(2):220–234, 2013. doi:10.1007/s10986-013-9204-x.

[SCS93] W. Stephan, L. Chao, and J. G. Smale. The advance of muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genetics Research*, 61(3):225–231, 1993. doi:10.1017/S0016672300031384.

[Sle62] D. S. Slepian. The one-sided barrier problem for gaussian noise. *Bell System Technical Journal*, 41:463–501, 1962.

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

[ST17] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Routledge, Nov. 2017. doi:10.1201/9780203738818.

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

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'IHÉS*, 81:73–205, 1995. URL `http://eudml.org/doc/104106`.

[Wei] E. W. Weisstein. Dawson's integral. URL `https://mathworld.wolfram.com/DawsonsIntegral.html`. Visited on February 21, 2024.

[WL10] D. Waxman and L. Loewe. A stochastic model for a single click of muller's ratchet. *Journal of Theoretical Biology*, 264(4):1120–1132, 2010. doi:10.1016/j.jtbi.2010.03.014.