Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2017, Vol. 53, No. 3, 1069–1107
DOI: 10.1214/16-AIHP748


ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

# Horton self-similarity of Kingman's coalescent tree

Yevgeniy Kovchegov[a,1] and Ilya Zaliapin[b,2]

[a] *Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA. E-mail: kovchegy@math.oregonstate.edu*
[b] *Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557-0084, USA. E-mail: zal@unr.edu*

Received 3 January 2015; revised 20 January 2016; accepted 15 February 2016

**Abstract.** The paper establishes Horton self-similarity for a tree representation of Kingman's coalescent process. The proof is based on a Smoluchowski-type system of ordinary differential equations that describes evolution of the number of branches of a given Horton–Strahler order in a tree that represents Kingman's $N$-coalescent, in a hydrodynamic limit. We also demonstrate a close connection between the combinatorial Kingman's tree and the combinatorial level set tree of a white noise, which implies Horton self-similarity for the latter.

**Résumé.** Cet article prouve l'auto-similarité à la Horton pour la représentation par arbres du processus de coalescence de Kingman. La preuve est basée sur un système d'équations différentielles ordinaires de type Smoluchowski décrivant, dans la limite hydrodynamique, l'évolution du nombre de branches d'un ordre de Horton–Strahler donné dans un arbre représentant le $N$-coalescent de Kingman. Nous prouvons aussi un lien étroit entre l'arbre de Kingman combinatoire et l'arbre combinatoire des ensembles de niveaux d'un bruit blanc, ce qui implique l'auto-similarité à la Horton de ce dernier.

*MSC:* Primary 60C05; secondary 82B99

*Keywords:* Coalescent; Kingman's coalescent; Horton–Strahler order; Horton self-similarity

## 1. Introduction

This study focuses on *Horton self-similarity* for binary rooted tree graphs. The concept is related to Horton–Strahler ordering of the tree branches [9,16] that was introduced in hydrology in the mid-20th century to describe the dendritic structure of river networks and has penetrated other areas of sciences since then [4,6,17]. Devroye and Kruszewski [6] assert that "*the Horton–Strahler number occur in almost every field involving some kind of natural branching pattern*". Roughly speaking, the Horton–Strahler order corresponds to the relative importance of a branch in the tree hierarchy. Specifically, each leaf is assigned order $k = 1$; and each internal vertex with offsprings of orders $i$ and $j$ is assigned order $k = \max(i, j) + \delta_{ij}$, where $\delta_{ij}$ is the Kronecker's delta. A *branch* is defined as a sequence of connected vertices with the same order.

Horton self-similarity refers to the geometric decay of the number $N_k$ of branches of order $k$ [9,14]. A trivial example of Horton self-similarity is given by a perfect binary tree (with all leaves having the same depth) for which $N_k/N_{k+1} = 2$ for all $1 \le k < \Omega - 1$, with $\Omega$ being the maximal branch order in the tree. It is easily seen that for any non-perfect binary tree $N_k/N_{k+1} \ge 2$, with the strict inequality holding for at least one value of $k$. A classical model that exhibits non-trivial Horton self-similarity is a tree representation of critical binary Galton–Watson branching processes [4,12,13], also known in hydrology as Shreve's random topology model for river networks [14,15]. Ronald

[1] Supported in part by a grant from the Simons Foundation (#284262) and by the NSF Award DMS 1412557.
[2] Supported by the NSF Awards DMS 0934871 and DMS 1049092.

Shreve [15] has demonstrated that in this model the ratios $N_k/N_{k+1}$ converge to $R = 4$ as $k$ increases. Recently, the authors established Horton self-similarity with the same asymptotic ratio for the level set tree representation of a homogeneous symmetric Markov chain and demonstrated that in general this representation is not equivalent to the critical Galton–Watson tree [18]. Models that obey Horton self-similarity with ratio different from $R = 2, 4$ are still lacking, however, despite their demonstrated practical importance [4,10,12,19].

This study is a first step toward exploring Horton self-similarity with ratio $R \neq 2, 4$. We consider here the tree generated by Kingman's coalescent process with $N$ particles. The main result is a weaker form of Horton self-similarity, called here *root-Horton law*. The Horton ratio is estimated numerically as $R = 3.043827\ldots$. We also establish a close relation between the combinatorial tree representations of Kingman's $N$-coalescent and a combinatorial level set tree for a sequence of i.i.d. random variables (referred to as *discrete white noise*), which implies Horton self-similarity for the latter. These findings add two important classes of processes – Kingman's coalescent and discrete white noise – to the realm of Horton self-similar systems.

The paper is organized as follows. Section 2 describes Horton–Strahler ordering of tree branches and the related concept of Horton self-similarity. Kingman's coalescent process and its tree representation are defined in Section 3. The main results are summarized in Section 4. Section 5 introduces the Smoluchowski–Horton system of equations that describes the dynamics of Horton–Strahler branches in Kingman's coalescent. This section also establishes the validity of the Smoluchowski–Horton equations, as well as the existence of some related quantities, in the hydrodynamic limit. A proof of the existence of root-Horton law for Kingman's coalescent is presented in Section 6. Section 7 demonstrates a connection between the combinatorial tree representation of Kingman's $N$-coalescent process and combinatorial level set tree of a discrete white noise. The Smoluchowski–Horton system for a general coalescent process with collision kernel is written in Section 8. Section 9 concludes. The complete proofs of hydrodynamic limits are given in the Appendices.

## 2. Self-similar trees

This section defines *Horton self-similarity* for rooted binary trees.

### 2.1. *Rooted trees*

A *graph* $\mathcal{G} = (V, E)$ is a collection of vertices $V = \{v_i\}$, $1 \le i \le N_V$ and edges $E = \{e_k\}$, $1 \le k \le N_E$. In a *simple* undirected graph each edge is defined as an unordered pair of distinct vertices: $\forall 1 \le k \le N_E, \exists! 1 \le i, j \le N_V, i \neq j$ such that $e_k = (v_i, v_j)$ and we say that the edge $k$ *connects* vertices $v_i$ and $v_j$. Furthermore, each pair of vertices in a simple graph may have at most one connecting edge. A *tree* is a connected simple graph $T = (V, E)$ without cycles. In a *rooted* tree, one node is designated as a root; this imposes a natural *direction* of edges as well as the parent-child relationship between the vertices. Specifically, of the two connected vertices the one closest to the root is called *parent*, and the other – *child*. Sometimes we consider trees embedded in a plane (*planar trees*), where the children of the same parent are ordered.

A *time oriented tree* $T = (V, E, S)$ assigns time marks $S = \{s_i\}$, $1 \le i \le N_V$ to the tree vertices in such a way that the parent mark is always larger than that of its children. A *combinatorial tree* $\text{SHAPE}(T) \equiv (V, E)$ discards the time marks of a time oriented tree $T$, as well as possible planar embedding, and only preserves its graph-theoretic structure.

We often work with the space $\mathcal{T}_N$ of combinatorial (not labeled, not embedded) rooted binary trees with $N$ leaves, and the space $\mathcal{T}$ of all (finite or infinite) rooted binary trees.

### 2.2. *The Horton–Strahler orders*

The Horton–Strahler ordering of the vertices of a finite rooted binary tree is performed in a hierarchical fashion, from leaves to the root [4,10,12]. Specifically, each leaf has order $k(\text{leaf}) = 1$. An internal vertex $p$ whose children have orders $i$ and $j$ is assigned the order

$$k(p) = \max(i, j) + \delta_{ij}, \tag{1}$$

where $\delta_{ij}$ is the Kronecker's delta. Figure 1 illustrates this definition. A *branch* is defined as a union of connected vertices with the same order.

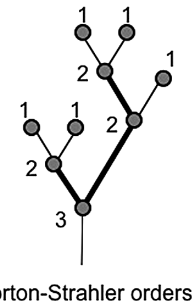


Fig. 1. Example of Horton–Strahler ordering. Two order-2 branches are depicted by heavy lines. The branch to the left from the root consists of one vertex; the branch to the right from the root consists of two vertices.

### 2.3. *Horton self-similarity*

Let $\mathcal{Q}_N$ be a probability measure on $\mathcal{T}_N$ and $N_k^{(\mathcal{Q}_N)}$ be the number of branches of Horton–Strahler order $k$ in a tree generated according to $\mathcal{Q}_N$.

***Definition 1.*** *We say that a sequence of probability laws $\{\mathcal{Q}_N\}_{N\in\mathbb{N}}$ has well-defined asymptotic Horton ratios if for each $k \in \mathbb{N}^+$, random variables $(N_k^{(\mathcal{Q}_N)}/N)$ converge in probability, as $N \to \infty$, to a constant value $\mathcal{N}_k$, called the asymptotic ratio of the branches of order $k$.*

Horton self-similarity implies that the sequence $\mathcal{N}_k$ decreases in a geometric fashion as $k$ goes to infinity. In this work we use a particular form of decay described below.

***Definition 2.*** *A sequence $\{\mathcal{Q}_N\}_{N\in\mathbb{N}}$ of probability laws on $\mathcal{T}$ with well-defined asymptotic Horton ratios is said to obey a root-Horton self-similarity law if and only if the following limit exists and is finite and positive: $\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = R > 0$. The constant $R$ is called the Horton exponent.*

## 3. Coalescent processes, trees

This section reviews Kingman's coalescent process with $N$ particles and introduces its tree representation.

### 3.1. *Kingman's $N$-coalescent process*

We start by considering a general finite coalescent process defined by a collision kernel [2,3,13]. The process begins with $N$ particles (clusters) of mass one. The cluster formation is governed by a symmetric collision rate kernel $K(i, j) = K(j, i) > 0$. Namely, a pair of clusters with masses $i$ and $j$ coalesces at the rate $K(i, j)$, independently of the other pairs, to form a new cluster of mass $i + j$. The process continues until there is a single cluster of mass $N$.

Formally, for a given $N$ consider the space $\mathcal{P}_{[N]}$ of partitions of $[N] = \{1, 2, \dots, N\}$. Let $\Pi_0^{(N)}$ be the initial partition in singletons, and $\Pi_t^{(N)}$ $(t \ge 0)$ be a strong Markov process such that $\Pi_t^{(N)}$ transitions from partition $\pi \in \mathcal{P}_{[N]}$ to $\pi' \in \mathcal{P}_{[N]}$ with rate $K(i, j)$ provided that partition $\pi'$ is obtained from partition $\pi$ by merging two clusters of $\pi$ of masses $i$ and $j$. If $K(i, j) \equiv 1$ for all positive integer masses $i$ and $j$, the process $\Pi_t^{(N)}$ is known as Kingman's $N$-coalescent process.

### 3.2. *Coalescent tree*

A merger history of Kingman's $N$-coalescent process can be naturally described by a time oriented binary tree $T_{\mathrm{K}}^{(N)}$ constructed as follows. Start with $N$ leaves that represent the initial $N$ particles and have time mark $t = 0$. When two clusters coalesce (a transition occurs), merge the corresponding vertices to form an internal vertex with a time mark of the coalescent. The final coalescence forms the tree root. The resulting time oriented binary tree represents the history of the process. We notice that a given unlabeled tree corresponds to multiple coalescent trajectories obtained by relabeling of the initial particles.

Observe that the combinatorial version $\text{SHAPE}(T_{\text{K}}^{(N)})$ of the Kingman's coalescent tree is invariant under time scaling $t_{\text{new}} = Ct_{\text{old}}$, $C > 0$. Thus without loss of generality we let $K(i, j) \equiv 1/N$ in Kingman's $N$-coalescent process. Slowing the process's evolution $N$ times is natural in Smoluchowski coagulation equations that describe the dynamics of the fraction of clusters of different masses.

## 4. Statement of results

The main result of this paper is root-Horton self-similarity for the combinatorial tree $\text{SHAPE}(T_{\text{K}}^{(N)})$ of the Kingman's $N$-coalescent process, as $N$ goes to infinity. Specifically, let $N_k$ denote the number of branches of Horton–Strahler order $k$ in the tree $T_{\text{K}}^{(N)}$ that describes Kingman $N$-coalescent. We show in Section 5, Lemma 3 that for each $k \geq 1$, $N_k/N$ converges in probability to the asymptotic Horton ratio

$$\mathcal{N}_k = \lim_{N\to\infty} N_k/N.$$

Moreover, these $\mathcal{N}_k$ are finite and can be expressed as

$$\mathcal{N}_k = \frac{1}{2}\int_0^\infty g_k^2(x)\,dx,$$

where the sequence $g_k(x)$ solves the following system of ordinary differential equations (ODEs):

$$g'_{k+1}(x) - \frac{g_k^2(x)}{2} + g_k(x)g_{k+1}(x) = 0, \quad x \geq 0$$

with $g_1(x) = 2/(x+2)$, $g_k(0) = 0$ for $k \geq 2$. Equivalently,

$$\mathcal{N}_k = \int_0^1 \bigl(1 - (1-x)h_{k-1}(x)\bigr)^2\,dx,$$

where $h_0 \equiv 0$ and the sequence $h_k(x)$ satisfies the ODE system

$$h'_{k+1}(x) = 2h_k(x)h_{k+1}(x) - h_k^2(x), \quad 0 \leq x \leq 1$$

with the initial conditions $h_k(0) = 1$ for $k \geq 1$.

The root-law Horton self-similarity is proven in Section 6 in the following statement.

**Theorem 1.** *The asymptotic Horton ratios $\mathcal{N}_k$ exist, are finite and satisfy the convergence* $\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = R$ *with* $2 \leq R \leq 4$.

Numerical solution for the sequence $h_k$ provides an estimation of Horton exponent $R = 3.043827\ldots$ and suggests that $\mathcal{N}_k$ also obey a stronger version of Horton self-similarity: $\lim_{k\to\infty}(\mathcal{N}_k R^k) = N_0 > 0$.

Section 7.1 introduces a *level set tree* $\text{LEVEL}(X_i)$ that describes the structure of the level sets of a discrete-time function $X_i$, $i = 1, \ldots, i_{\max}$. In particular, we show that there exists a one-to-one map between finite rooted planar time oriented binary trees and sequences of the local extrema of $X_i$. Let $W = \{W_i\}$ be a *discrete white noise*, that is a process comprised of i.i.d. values with a common atomless distribution. Consider now a process $\tilde{W}_i^{(N)}$ with exactly $N$ local maxima separated by $N-1$ internal local minima such that the latter form a discrete white noise; we call $\tilde{W}_i^{(N)}$ an *extended discrete white noise*.

Let $L_W^{(N)} = \text{LEVEL}(\tilde{W}_i^{(N)})$ be the level set tree of $\tilde{W}_i^{(N)}$ and $\text{SHAPE}(L_W^{(N)})$ be the combinatorial tree that retains the graph-theoretic structure of $L_W^{(N)}$ and drops its planar embedding as well as the time marks of the vertices. Furthermore, let $T_{\text{K}}^{(N)}$ be the tree that corresponds to a Kingman's $N$-coalescent, and let $\text{SHAPE}(T_{\text{K}}^{(N)})$ be its combinatorial version that drops the time marks of the vertices. By construction, both the trees $\text{SHAPE}(L_W^{(N)})$ and $\text{SHAPE}(T_K^{(N)})$, belong to the space $\mathcal{T}_N$ of binary rooted trees with $N$ leaves. Section 7.2 establishes the following equivalence.

**Theorem 2.** *The trees* $\text{SHAPE}(L_W^{(N)})$ *and* $\text{SHAPE}(T_K^{(N)})$ *have the same distribution on* $\mathcal{T}_N$.

The equivalence leads to the Horton self-similarity for discrete white noise.

**Corollary 1.** *The combinatorial level set tree of a discrete white noise is root-Horton self similar with the same Horton exponent* $R$ *as for Kingman's coalescent.*

## 5. Smoluchowski–Horton ODEs for Kingman's coalescent

Consider Kingman's $N$-coalescent process and its tree representation $T_K^{(N)}$. In Section 5.1 we informally write Smoluchowski-type ODEs for the number of Horton–Strahler branches in the coalescent tree $T_K^{(N)}$ and consider the asymptotic version of these equations as $N \to \infty$. Section 5.2 formally establishes the validity of the hydrodynamic limit.

### 5.1. *Main equation*

Recall that we let $K(i,j) \equiv 1/N$ in Kingman's $N$-coalescent process. Let $|\Pi_t^{(N)}|$ denote the total number of clusters at time $t \geq 0$, and let $\eta_{(N)}(t) := |\Pi_t^{(N)}|/N$ be the total number of clusters relative to the system size $N$. Then $\eta_{(N)}(0) = N/N = 1$ and $\eta_{(N)}(t)$ decreases by $1/N$ with each coalescence of clusters with the rate

$$\frac{1}{N}\binom{N\eta_{(N)}(t)}{2} = \frac{\eta_{(N)}^2(t)}{2} \cdot N + o(N), \quad \text{as } N \to \infty,$$

since $1/N$ is the coalescence rate for any pair of clusters regardless of their masses. Informally, this implies that the limit relative number of clusters $\eta(t) = \lim_{N\to\infty} \eta_{(N)}(t)$ satisfies the following ODE:

$$\frac{d}{dt}\eta(t) = -\frac{\eta^2(t)}{2}. \tag{2}$$

The corresponding initial condition $\eta(0) = 1$ implies a unique solution $\eta(t) = 2/(2+t)$.

Next, for any $k \in \mathbb{N}^+$ we define $\eta_{k,N}(t)$ to be the number of clusters that correspond to branches of Horton–Strahler order $k$ at time $t$ relative to the system size $N$. Initially, each particle represents a leaf of Horton–Strahler order 1. Accordingly, the initial conditions are set to be, using Kronecker's delta notation,

$$\eta_{k,N}(0) = \delta_1(k).$$

We describe now the evolution of $\eta_{k,N}(t)$ using the definition of Horton–Strahler orders.

Observe that $\eta_{k,N}(t)$ increases by $1/N$ with each coalescence of clusters of Horton–Strahler order $k-1$ that happens with the rate

$$\frac{1}{N}\binom{N\eta_{k-1,N}(t)}{2} = \frac{\eta_{k-1,N}^2(t)}{2} \cdot N + o(N).$$

Thus $\frac{\eta_{k-1,N}^2(t)}{2} + o(1)$ is the instantaneous rate of increase of $\eta_{k,N}(t)$.

Similarly, $\eta_{k,N}(t)$ decreases by $1/N$ when a cluster of order $k$ coalesces with a cluster of order strictly higher than $k$ with the rate

$$\eta_{k,N}(t)\left(\eta_{(N)}(t) - \sum_{j=1}^{k}\eta_{j,N}(t)\right) \cdot N,$$

and it decreases by $2/N$ when a cluster of order $k$ coalesces with another cluster of order $k$ with the rate

$$\frac{1}{N}\binom{N\eta_{k,N}(t)}{2} = \frac{\eta_{k,N}^2(t)}{2}\cdot N + o(N).$$

Thus the instantaneous rate of decrease of $\eta_{k,N}(t)$ is

$$\eta_{k,N}(t)\left(\eta_{(N)}(t) - \sum_{j=1}^{k}\eta_{j,N}(t)\right) + \eta_{k,N}^2(t) + o(1).$$

Now we can informally write the limit rates-in and the rates-out for the clusters of Horton–Strahler order via the following *Smoluchowski–Horton system* of ODEs:

$$\frac{d}{dt}\eta_k(t) = \frac{\eta_{k-1}^2(t)}{2} - \eta_k(t)\left(\eta(t) - \sum_{j=1}^{k-1}\eta_j(t)\right) \tag{3}$$

with the initial conditions $\eta_k(0) = \delta_1(k)$. Here we define $\eta_k(t) = \lim_{N\to\infty}\eta_{k,N}(t)$, provided it exists, and let $\eta_0 \equiv 0$.

Since $\eta_k(t)$ has the instantaneous rate of increase $\eta_{k-1}^2(t)/2$, the relative total number of clusters corresponding to branches of Horton–Strahler order $k$ is given by

$$\mathcal{N}_k = \delta_1(k) + \int_0^\infty \frac{\eta_{k-1}^2(t)}{2}\,dt. \tag{4}$$

This equation has a simple heuristic interpretation. Namely, according to the Horton–Strahler rule (1), a branch of order $k > 1$ can only be created by merging two branches of order $k-1$. In Kingman's coalescent process these two branches are selected at random from all pairs of branches of order $k-1$ that exist at instant $t$. As $N$ goes to infinity, the asymptotic density of a pair of branches of order $(k-1)$, and hence the instantaneous intensity of newly formed branches of order $k$, is $\eta_{k-1}^2(t)/2$. The integration over time gives the relative total number of order-$k$ branches. The validity of Equation (4) is proven below in Lemma 3.

It is not hard to compute the first three terms of the sequence $\mathcal{N}_k$ by solving Equations (2) and (3) in the first three iterations:

$$\mathcal{N}_1 = 1, \qquad \mathcal{N}_2 = \frac{1}{3}, \quad \text{and} \quad \mathcal{N}_3 = \frac{e^4}{128} - \frac{e^2}{8} + \frac{233}{384} = 0.109686868100941\ldots.$$

Hence, we have $\mathcal{N}_1/\mathcal{N}_2 = 3$ and $\mathcal{N}_2/\mathcal{N}_3 = 3.038953879388\ldots$. Our numerical results yield, moreover,

$$\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty}\frac{\mathcal{N}_k}{\mathcal{N}_{k+1}} = 3.0438279\ldots.$$

### 5.2. *Hydrodynamic limit*

This section establishes the existence of the asymptotic ratios $\mathcal{N}_k$ as well as the validity of Equations (2), (3) and (4) in a hydrodynamic limit. We refer to Darling and Norris [5] for a survey of formal techniques for proving that a Markov chain converges to the solution of a differential equation.

Notice that *quasilinearity* of the system of ODEs in (3) implies the existence and uniqueness. Specifically, if the first $k-1$ functions $\eta_1(t), \ldots, \eta_{k-1}(t)$ are given, then (3) is a linear equation in $\eta_k(t)$. The following argument is different from the one presented by Norris [11] for the Smoluchowski equations.

**Lemma 1.** *Let $\eta_{(N)}(t)$ be the relative total number of clusters and $\eta(t)$ be the solution to Equation (2) with the initial condition $\eta(0) = 1$. Then*

$$\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,\infty)} \to 0$$

*in probability as $N \to \infty$.*

A proof of Lemma 1 is given in Appendix A. The proof is divided into steps that we briefly outline below.

- *Steps I, II.* We start by establishing bounds on the number of coalescences within the time interval $[t, t+\delta]$. Specifically, fix $\epsilon_0 \in (0,1)$ and take $\delta > 0$. Given $y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]$, let $u = \binom{Ny}{2}$ and $v = \binom{Ny - \lceil \frac{\delta y^2}{2} N \rceil}{2}$. We use the exponential Markov inequality to show that for any given $t \geq 0$ and large enough $N$ we have

$$P\left(\frac{\delta}{N^2} v - (1+\delta)N^{-1/3} \leq \eta_{(N)}(t) - \eta_{(N)}(t+\delta) \leq \frac{\delta}{N^2} u + N^{-1/3} \Big| \eta_{(N)}(t) = y\right)$$
$$\geq \left(1 - \exp\left\{-N^{1/6} + \frac{4\delta}{\epsilon_0^2}\right\}\right)^2.$$

- *Step III.* The bounds of steps I, II are applied to show that

$$P\left(\left|\frac{\eta^2_{(N)}(t)}{2} + \Delta_\delta \eta_{(N)}(t)\right| \leq \delta + \left(\delta^{-1}+1\right)N^{-1/3} \Big| \eta_{(N)}(t) = y\right)$$
$$\geq \left(1 - \exp\left\{-N^{1/6} + \frac{4\delta}{\epsilon_0^2}\right\}\right)^2 \tag{5}$$

for $N$ large enough, where $\Delta_\delta f(x) := \frac{f(x+\delta) - f(x)}{\delta}$ denotes the forward difference.
- *Step IV.* For $K > 0$, consider an interval $[0, K]$ partitioned into $M$ subintervals

$$[t_0, t_1], [t_1, t_2], \ldots, [t_{M-1}, t_M]$$

of equal length $\delta = K/M$, where $t_0 = 0$ and $t_M = K$. Let $\epsilon_0 = \eta(K)/2 = 1/(2+K)$, where $\eta(t) = 2/(2+t)$ is the solution to Equation (2) with the initial condition $\eta(0) = 1$. Consider the difference equation

$$\Delta_\delta \psi_{(N)}(t_i) = -\frac{\psi^2_{(N)}(t_i)}{2} + \mathcal{E}'(t_i) \tag{6}$$

with initial condition $\psi_{(N)}(0) = 1$, where the error $|\mathcal{E}'(t_i)|$ satisfies

$$\left|\mathcal{E}'(t_i)\right| \leq \delta + \left(\delta^{-1}+1\right)N^{-1/3}.$$

At this step we prove that if $M$ is large enough and for any natural number $j \leq M$ function $\psi_{(N)}(t_i)$ satisfies (6) for all $i \in \{0, 1, \ldots, j-1\}$, then

$$\psi_{(N)}(t_j) \geq \epsilon_0$$

as we take $N$ large enough. This follows from observing that $\eta(t)$ will satisfy a difference equation similar to (6),

$$\Delta_\delta \eta(t_i) = -\frac{\eta^2(t_i)}{2} + \mathcal{E}(t_i) \tag{7}$$

with $|\mathcal{E}(t_i)| \leq \frac{1}{4}\delta$ for all $i \in \{0, 1, \ldots, M-1\}$.
- *Step V.* Consider events

$$A_i = \left\{\Delta_\delta \eta_{(N)}(t_i) = -\frac{\eta^2_{(N)}(t_i)}{2} + \mathcal{E}'(t_i) \text{ and } \left|\mathcal{E}'(t_i)\right| \leq \delta + \left(\delta^{-1}+1\right)N^{-1/3}\right\} \tag{8}$$

for all $i \in \{0, 1, \ldots, M-1\}$. Here we combine the results of steps III and IV and establish that with probability greater than $P(\bigcap_{i=0}^{M-1} A_i) \to 1$ as $M \to \infty$, $\eta_{(N)}(t_i)$ satisfies the difference Equation (6) with $\psi_{(N)}(t) \equiv \eta_{(N)}(t)$.

- *Step VI.* Taking $\psi_{(N)}(t) \equiv \eta_{(N)}(t)$, we compare the difference Equation (6) with (7), and bound the error $|\eta_{(N)}(t) - \eta(t)|$ for all $t \in [0, K]$. Specifically, we show that with probability greater than $P(\bigcap_{i=0}^{M-1} A_i) \to 1$,

$$\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,K]} \leq \frac{15}{4}K^2/M + 4K/M + 3/M \tag{9}$$

for $M$ large enough and $N \geq M^6$. Therefore, letting $M \to \infty$, we obtain

$$\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,K]} \to 0 \quad \text{in probability.}$$

- *Step VII.* Take $\epsilon \in (0, 1)$ and $\gamma > 1$, and consider $K > \frac{2(1-\epsilon)}{\epsilon}\gamma$. This step uses Markov inequality to show that

$$P\left(\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[K,\infty)} < \epsilon\right) > 1 - 1/\gamma,$$

which, together with the results of step VI, implies

$$\limsup_{N\to\infty} P\left(\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,\infty)} < \epsilon\right) \geq 1 - 1/\gamma.$$

We conclude that

$$\lim_{N\to\infty} P\left(\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,\infty)} < \epsilon\right) = 1.$$

Therefore we have shown that $\|\eta_{(N)}(t) - \eta(t)\|_{L^\infty[0,\infty)} \to 0$ in probability, thus establishing Lemma 1.

We now proceed with establishing a hydrodynamic limit for the Smoluchowski–Horton system of ODEs (3). Let

$$\eta_{k,N}(t) := \frac{N_k(t)}{N} \quad \text{and} \quad g_{k,N}(t) := \eta_{(N)}(t) - \sum_{j:j<k} \eta_{j,N}(t).$$

**Lemma 2.** *Consider the relative numbers $\eta_{k,N}(t)$ of clusters that correspond to branches of Horton–Strahler order $k$ and functions $\eta_k(t)$ that solve the system of equations (3) with the initial conditions $\eta_k(0) = \delta_1(k)$. Then,*

$$\left\|\eta_{k,N}(t) - \eta_k(t)\right\|_{L^\infty[0,\infty)} \to 0, \quad \forall k \geq 1,$$

*in probability, as $N \to \infty$.*

A proof of Lemma 2 is given in Appendix B. Here we summarize the steps used in the proof.

- *Step I.* We use the setting from the proof of Lemma 1. Fix $K > 0$ and consider an interval $[0, K]$ partitioned into $M$ subintervals

$$[t_0, t_1], [t_1, t_2], \dots, [t_{M-1}, t_M]$$

of equal length $\delta = K/M$, where $t_0 = 0$ and $t_M = K$. Let $\epsilon_0 = \eta(K)/2 = 1/(2 + K)$. The total number of coalescences within the interval $[t_i, t_{i+1}]$ equals $N[\eta_{(N)}(t_i) - \eta_{(N)}(t_{i+1})]$.

For any $k \in \mathbb{N}^+$ and any $i = 0, 1, \dots, M - 1$ we represent the relative number of coalescences that involve the clusters of order $k$ within $[t_i, t_{i+1}]$ as

$$\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i) = \xi_1 + \xi_2 + \cdots + \xi_m,$$

where $\xi_1, \xi_2, \dots, \xi_m$ are random variables that correspond to the $m$ coalescences (of any Horton–Strahler order) within $[t_i, t_{i+1}]$ in the order of occurrence. Here, each $\xi_r$ can take values in $\frac{1}{N}\{-2, -1, 0, 1\}$; and their dependence on $k$ is omitted to simplify the notations. By construction, conditioned on the values $\{\eta_{j,N}(t_i)\}_j$, the distribution of $\xi_r$ for $1 \leq r \leq m$ is completely determined by the history $\mathcal{T}_{r-1}$ of the preceding $r - 1$ transitions.

Consider a random variable $\xi$ with the values $\{-2, -1, 0, 1\}$ specified by the probabilities $\{p(-2), p(-1), p(0), p(1)\}$:

$$p(-2) := \eta_{k,N}^2(t_i)/\eta_{(N)}^2(t_i),$$

$$p(1) := \begin{cases} \eta_{k-1,N}^2(t_i)/\eta_{(N)}^2(t_i) & \text{if } k > 1, \\ 0 & \text{if } k = 1, \end{cases}$$

$$p(-1) := 2\eta_{k,N}(t_i)g_{k+1,N}(t_i)/\eta_{(N)}^2(t_i),$$

$$p(0) := 1 - p(-2) - p(-1) - p(1).$$

Recall the events $A_i$ defined in (8). We notice that, conditioned on $\bigcap_{i'=0}^{i} A_{i'}$, the total variation distance between the distribution of $\xi_r$ (for a fixed $1 \le r \le m$) and the distribution of $\xi$ is of order $\mathcal{O}(\delta)$. We use this to show that for each $k \in \mathbb{N}^+$, there is a large enough $c_k > 0$ and $a > 0$ such that

$$\begin{aligned} &P\left(\left|\left[\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i)\right] - E[\xi]\delta\frac{\eta_{(N)}^2(t_i)}{2}\right| < c_k\delta^{4/3} \,\Big|\, \bigcap_{i'=0}^{i} A_{i'}\right) \\ &\quad \ge 1 - \exp\{-aM^4\} \end{aligned} \tag{10}$$

for all $i = 0, 1, \ldots, M-1$, $2M^6 > N > M^6$, and $M$ large enough.

- *Step II.* According to the results of step I, we obtain the following system of difference equations:

$$\begin{aligned} \Delta_\delta\eta_{1,N}(t_i) &= -\eta_{1,N}(t_i)\eta_{(N)}(t_i) + \mathcal{E}_1'(t_i), \\ \Delta_\delta\eta_{k,N}(t_i) &= \frac{\eta_{k-1,N}^2(t_i)}{2} - \eta_{k,N}(t_i)g_{k,N}(t_i) + \mathcal{E}_k'(t_i) \quad \text{for } k \ge 2 \end{aligned} \tag{11}$$

with the initial conditions

$$\big(\eta_{1,N}(0), \eta_{2,N}(0), \ldots, \eta_{k,N}(0), \ldots\big) = (1, 0, 0, \ldots),$$

where for a given $\rho \in \mathbb{N}$ and $c = \max_{1\le k\le\rho}\{c_k\}$ we have $|\mathcal{E}_k'(t_i)| < c\delta^{1/3}$ for each $1 \le k \le \rho$. Each equation in this system holds with the probability that converges to unity as $M$ increases.

We now compare the above difference equations (11) to the following system of difference equations that corresponds to the system of ODEs (3):

$$\begin{aligned} \Delta_\delta\eta_1(t_i) &= -\eta_1(t_i)\eta(t_i) + \mathcal{E}_1(t_i), \\ \Delta_\delta\eta_k(t_i) &= \frac{\eta_{k-1}^2(t_i)}{2} - \eta_k(t_i)g_k(t_i) + \mathcal{E}_k(t_i) \quad \text{for } k \ge 2, \end{aligned} \tag{12}$$

where $g_k(t) := \eta(t) - \sum_{i:i<k}\eta_i(t)$, and the error

$$\mathcal{E}_k(t_i) = \frac{\eta_k''(c_{i,k})}{2}\delta \quad \text{for some } c_{i,k} \in (t_i, t_{i+1}).$$

- *Step III.* We show that, conditioning on the event $\bigcap_{i=0}^{M-1} A_i$, we have the following upper bound for any $k \in \{1, \ldots, \rho\}$, all $i \in \{0, 1, \ldots, M-1\}$, and $t \in (t_i, t_{i+1})$:

$$\begin{aligned} \left|\eta_{k,N}(t) - \eta_k(t)\right| &\le \left|\eta_{k,N}(t) - \eta_{k,N}(t_i)\right| + \left|\eta_{k,N}(t_i) - \eta_k(t_i)\right| + \left|\eta_k(t_i) - \eta_k(t)\right| \\ &\le \big(5K^2 + 4K + 4\big)/M + (c+1)2^k\frac{\delta^{1/3}}{\rho}\left[e^{2K\rho} - 1\right] + 3\delta. \end{aligned}$$

We conclude that, for any $k$,

$$\|\eta_{k,N} - \eta_k\|_{L^\infty[0,K]} \to 0 \quad \text{in probability.}$$

- *Step IV.* Finally, observe that for any $\epsilon > 0$ and for $K > 2$ large enough so that $\eta(K) < \epsilon$,

$$\eta_k(t) \le \eta(t) \le \eta(K) < \epsilon \quad \text{for all } t \ge K$$

and

$$\begin{aligned} P\big(\big\|\eta_{k,N}(t) - \eta_k(t)\big\|_{L^\infty[K,\infty)} > \epsilon\big) &\le P\big(\big\|\eta_{k,N}(t)\big\|_{L^\infty[K,\infty)} > \epsilon\big) \\ &\le P\big(\big\|\eta_{(N)}(t)\big\|_{L^\infty[K,\infty)} > \epsilon\big) \\ &= P\big(\eta_{(N)}(K) > \epsilon\big) \\ &\le \frac{2(1-\epsilon)}{\epsilon K}. \end{aligned}$$

The last bound is obtained from Markov inequality for the random variable $T_m$ that represents the time of the $m$th coalescence. Therefore, together with the result of the previous step, we have shown that for each $k$,

$$\|\eta_{k,N} - \eta_k\|_{L^\infty[0,\infty)} \to 0$$

in probability. This completes the proof.

Finally, the last lemma in this section establishes a hydrodynamic limit for the Horton ratios.

**Lemma 3.** *The Horton ratios $N_k/N$ converge in probability to a finite constant $\mathcal{N}_k$ given by* (4), *as $N \to \infty$.*

A proof of Lemma 3 is given in Appendix C.

## 6. The root-Horton self-similarity and related results

We begin this section with preliminary lemmas and propositions, and then proceed to proving Theorem 1.

Let $g_1(t) = \eta(t)$ and $g_k(t) = \eta(t) - \sum_{j:j<k} \eta_j(t)$ be the asymptotic number of clusters of Horton–Strahler order $k$ or higher at time $t$. We can rewrite (3) via $g_k$ using $\eta_k(t) = g_k(t) - g_{k+1}(t)$:

$$\frac{d}{dt} g_k(t) - \frac{d}{dt} g_{k+1}(t) = \frac{(g_{k-1}(t) - g_k(t))^2}{2} - \big(g_k(t) - g_{k+1}(t)\big) g_k(t).$$

Observe that $g_1(t) \ge g_2(t) \ge g_3(t) \ge \cdots$. We now rearrange the terms, obtaining for all $k \ge 2$,

$$\frac{d}{dt} g_{k+1}(t) - \frac{g_k^2(t)}{2} + g_k(t) g_{k+1}(t) = \frac{d}{dt} g_k(t) - \frac{g_{k-1}^2(t)}{2} + g_{k-1}(t) g_k(t). \tag{13}$$

One can readily check that $\frac{d}{dt} g_2(t) - \frac{g_1^2(t)}{2} + g_1(t) g_2(t) = 0$; the above equations hence simplify as follows

$$\begin{aligned} &g'_{k+1}(t) - \frac{g_k^2(t)}{2} + g_k(t) g_{k+1}(t) = 0 \quad \text{with} \\ &g_1(t) = \frac{2}{t+2}, \quad \text{and} \quad g_k(0) = 0 \text{ for } k \ge 2. \end{aligned} \tag{14}$$

Next, returning to the asymptotic ratios of the number of order-$k$ branches to $N$, we observe that (13) implies that, for $k \ge 2$,

$$\mathcal{N}_k = \int_0^\infty \frac{\eta_{k-1}^2(t)}{2}\, dt = \int_0^\infty \frac{(g_{k-1}(t) - g_k(t))^2}{2}\, dt = \int_0^\infty \frac{g_k^2(t)}{2}\, dt$$

since

$$\frac{(g_{k-1}(t)-g_k(t))^2}{2}=\frac{d}{dt}g_k(t)+\frac{g_k^2(t)}{2},$$

where $0\le g_k(t)\le g_1(t)\to 0$ as $t\to\infty$, and $\int_0^\infty \frac{d}{dt}g_k(t)\,dt=g_k(\infty)-g_k(0)=0$ for $k\ge 2$. Let $n_k$ represent the number of order-$k$ branches relative to the number of order-$(k+1)$ branches:

$$n_k:=\frac{\mathcal{N}_k}{\mathcal{N}_{k+1}}=\frac{\frac{1}{2}\int_0^\infty g_k^2(t)\,dt}{\frac{1}{2}\int_0^\infty g_{k+1}^2(t)\,dt}=\frac{\|g_k\|^2_{L^2[0,\infty)}}{\|g_{k+1}\|^2_{L^2[0,\infty)}}.$$

Consider the following limits that represent respectively the root and the ratio asymptotic Horton laws:

$$\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}}=\lim_{k\to\infty}\left(\prod_{j=1}^k n_j\right)^{-\frac{1}{k}}\quad\text{and}\quad\lim_{k\to\infty}n_k=\lim_{k\to\infty}\frac{\|g_k\|^2_{L^2[0,\infty)}}{\|g_{k+1}\|^2_{L^2[0,\infty)}}.$$

Theorem 1 establishes the existence of the first limit. We expect the second, stronger, limit also to exist and both of them to be equal to 3.043827... according to our numerical results. We now establish some basic facts about $g_k$ and $n_k$.

**Proposition 1.** *Let $g_k(x)$ solve the ODE system (14). Then*

(a) $\frac{1}{2}\int_0^\infty g_k^2(t)\,dt=\int_0^\infty g_k(t)g_{k+1}(t)\,dt$,
(b) $\int_0^\infty g_{k+1}^2(t)\,dt=\int_0^\infty (g_k(t)-g_{k+1}(t))^2\,dt$,
(c) $\lim_{t\to\infty} tg_k(t)=2$,
(d) $n_k=\frac{\|g_k\|^2_{L^2[0,\infty)}}{\|g_{k+1}\|^2_{L^2[0,\infty)}}\ge 2$,
(e) $n_k=\frac{\|g_k\|^2_{L^2[0,\infty)}}{\|g_{k+1}\|^2_{L^2[0,\infty)}}\le 4$.

**Proof.** Part (a) follows from integrating (14), and part (b) follows from part (a). Part (c) is done by induction, using the L'Hôpital's rule as follows. It is obvious that $\lim_{x\to\infty} tg_1(t)=2$. We observed earlier that $g_1(t)\ge g_2(t)\ge g_3(t)\ge\cdots$. Hence, for any $k\ge 1$,

$$tg_k(t)\le tg_1(t)=\frac{2t}{t+2}<2,\quad \forall t\ge 0.$$

Also,

$$[tg_{k+1}]'=\frac{tg_k^2(t)}{2}-tg_k(t)g_{k+1}(t)+g_{k+1}(t)=\frac{(g_k(t)-g_{k+1}(t))tg_k(t)+(2-tg_k(t))g_{k+1}(t)}{2}$$

implying $[tg_{k+1}]'\ge 0$ for all $t\ge 0$ as $g_k(t)-g_{k+1}(t)\ge 0$ and $2-tg_k(t)>0$. Hence, $tg_{k+1}(t)$ is bounded and nondecreasing. Thus, $\lim_{t\to\infty} tg_{k+1}(t)$ exists for all $k\ge 1$.

Next, suppose $\lim_{t\to\infty} tg_k(t)=2$. Then by the Mean Value Theorem, for any $t>0$ and for all $y>t$,

$$\frac{g_{k+1}(t)-g_{k+1}(y)}{t^{-1}-y^{-1}}\le\sup_{z:z\ge t}\frac{g'_{k+1}(z)}{-z^{-2}}.$$

Taking $y\to\infty$, obtain

$$\frac{g_{k+1}(t)}{t^{-1}}\le\sup_{z:z\ge t}\frac{g'_{k+1}(z)}{-z^{-2}}.$$

Therefore

$$\lim_{t\to\infty} t g_{k+1}(t) = \lim_{t\to\infty} \frac{g_{k+1}(t)}{t^{-1}} = \limsup_{z\to\infty} \frac{g'_{k+1}(z)}{-z^{-2}} = \limsup_{z\to\infty} \frac{\frac{g_k^2(z)}{2} - g_k(z)g_{k+1}(z)}{-z^{-2}}$$
$$= \limsup_{z\to\infty}\left[z^2 g_k(z)g_{k+1}(z) - \frac{z^2 g_k^2(z)}{2}\right] = 2\lim_{t\to\infty} t g_{k+1}(t) - 2$$

implying $\lim_{t\to\infty} t g_{k+1}(t) = 2$. The statement (d) follows from the tree construction process. An alternative proof of (d) using differential equations is given in the following subsection. Part (e) follows from part (a) together with Hölder inequality

$$\frac{1}{2}\|g_k\|^2_{L^2[0,\infty)} = \int_0^\infty g_k(t)g_{k+1}(t)\,dt \le \|g_k\|_{L^2[0,\infty)} \cdot \|g_{k+1}\|_{L^2[0,\infty)},$$

which implies $\frac{\|g_k\|^2_{L^2[0,\infty)}}{\|g_{k+1}\|^2_{L^2[0,\infty)}} \le 4$. □

***Remark 1.*** *The statements* (a) *and* (b) *of Proposition* 1 *have a straightforward heuristic interpretation, similar to that of Equation* (4) *above. Specifically,* (a) *claims that the asymptotic relative total number of vertices of order* $k+1$ *and above in the Kingman's tree* (*left-hand side*) *equals twice the asymptotic relative total number of vertices of order* $k+1$ *and above except the vertices parental to two vertices of order* $k$ (*right-hand side*)*. This is nothing but the asymptotic property of a binary tree – the number of leaves equals twice the number of internal nodes. The item* (a) *hence merely claims that the Kingman's tree formed by clusters of order above* $k$ *is binary for any* $k \ge 1$*. Similarly, item* (b) *claims that the asymptotic relative total number of vertices of order* $(k+2)$ *and above* (*left-hand side*) *equals the asymptotic relative total number of vertices of order* $(k+1)$ (*right-hand side*)*. This is yet another way of saying that the Kingman's tree is binary.*

Finally, observe that $g_k(t) \to 0$ as $k \to \infty$. Indeed, Proposition 1 and the Dominated Convergence Theorem imply

$$\int_0^\infty g_{k+1}^2(t)\,dt = \int_0^\infty \big(g_k(t) - g_{k+1}(t)\big)^2\,dt \to 0 \quad \text{as } k \to \infty.$$

Next, following (14),

$$g_{k+1}(t) = \int_0^t g'_{k+1}(y)\,dy = \int_0^t \frac{g_k^2(y)}{2}\,dy - \int_0^t g_k(y)g_{k+1}(y)\,dy \to 0 \quad \text{as } k \to \infty.$$

6.1. *Rescaling to* [0, 1] *interval*

Let

$$h_k(x) = (1-x)^{-1} - (1-x)^{-2} g_{k+1}\left(\frac{2x}{1-x}\right)$$

for $x \in [0,1)$. Then $h_0 \equiv 0$, $h_1 \equiv 1$, and the system of ODEs (14) rewrites as

$$h'_{k+1}(x) = 2h_k(x)h_{k+1}(x) - h_k^2(x) \tag{15}$$

with the initial conditions $h_k(0) = 1$.

Observe that the above quasilinearized system of ODEs (15) has $h_k(x)$ converging to $h(x) = \frac{1}{1-x}$ as $k \to \infty$, where $h(x)$ is the solution to Riccati equation $h'(x) = h^2(x)$ over $[0,1)$, with the initial value $h(0) = 1$. Specifically, we have proven that $g_k(x) \to 0$ as $k \to \infty$. Thus

$$h_k(x) = (1-x)^{-1} - (1-x)^{-2} g_{k+1}\left(\frac{2x}{1-x}\right) \to h(x) = \frac{1}{1-x}.$$

Observe that $h_2(x) = (1+e^{2x})/2$, but for $k \geq 3$ finding a closed form expression becomes increasingly hard. Given $h_k(x)$, Equation (15) is a linear first-order ODE in $h_{k+1}(x)$; its solution is given by $h_{k+1}(x) = \mathcal{H}h_k(x)$ with

$$\mathcal{H}f(x) = \left[1 - \int_0^x f^2(y)e^{-2\int_0^y f(s)\,ds}\,dy\right] \cdot e^{2\int_0^x f(s)\,ds}. \tag{16}$$

Hence, the problem we are dealing with concerns the asymptotic behavior of an iterated non-linear functional. Moreover, since (15) is a quasilinearized system of ODEs, it extends to all of $[0, \infty)$ with the same sequence of solutions $h_k(x)$ obtained from iterating $\mathcal{H}$ in (16). In particular,

$$h_{k+1}(1) = \mathcal{H}h_k(1) = \left[1 - \int_0^1 h_k^2(y)e^{-2\int_0^y h_k(s)\,ds}\,dy\right] \cdot e^{2\int_0^1 h_k(s)\,ds}.$$

Here the quantity $n_k$ rewrites in terms of $h_k$ as follows

$$n_k = \frac{\|1 - h_{k+1}/h\|^2_{L^2[0,1]}}{\|1 - h_k/h\|^2_{L^2[0,1]}}.$$

Using the setting of (15), we give an ODE proof to Proposition 1(d). To do so, we first need to prove the following lemma.

**Lemma 4.**

$$\|1 - h_{k+1}/h\|_{L^2[0,1]} = \|h_{k+1}/h - h_k/h\|_{L^2[0,1]}.$$

**Proof.** Observe that

$$h'_{k+1}(x) + \big(h_{k+1}(x) - h_k(x)\big)^2 = h^2_{k+1}(x).$$

We now use integration by parts to obtain

$$\begin{aligned}
\int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h^2(x)}\,dx &= \int_0^1 \frac{h^2_{k+1}(x)}{h^2(x)}\,dx - \int_0^1 \frac{h'_{k+1}(x)}{h^2(x)}\,dx \\
&= \int_0^1 \frac{h^2_{k+1}(x)}{h^2(x)}\,dx + 1 - 2\int_0^1 \frac{h_{k+1}(x)}{h(x)}\,dx \\
&= \int_0^1 \frac{(1 - h_{k+1}(x))^2}{h^2(x)}\,dx
\end{aligned}$$

since $1/h(x) = 1 - x$. $\square$

**Alternative proof of Proposition 1(d).** Notice that $h \geq \cdots \geq h_{k+1} \geq h_k \geq \cdots \geq h_0 \equiv 0$, which follows from $g_1(t) \geq g_2(t) \geq g_3(t) \geq \cdots$. The Lemma 4 implies

$$\begin{aligned}
\|1 - h_{k+1}/h\|^2_{L^2[0,1]} &= \|h_{k+1}/h - h_k/h\|^2_{L^2[0,1]} \\
&= \int_0^1 \big[(1 - h_k/h) - (1 - h_{k+1}/h)\big]^2\,dx \\
&= \|1 - h_{k+1}/h\|^2_{L^2[0,1]} + \|1 - h_k/h\|^2_{L^2[0,1]} - 2\int_0^1 (1 - h_k/h)(1 - h_{k+1}/h)\,dx
\end{aligned}$$

and therefore

$$\begin{aligned}\|1-h_k/h\|^2_{L^2[0,1]} &= 2\int_0^1 (1-h_k/h)(1-h_{k+1}/h)\,dx \\ &= 2\|1-h_{k+1}/h\|^2_{L^2[0,1]} + 2\int_0^1 (h_{k+1}/h-h_k/h)(1-h_{k+1}/h)\,dx \\ &\geq 2\|1-h_{k+1}/h\|^2_{L^2[0,1]}\end{aligned}$$

yielding $2 \leq \frac{\|1-h_k/h\|^2_{L^2[0,1]}}{\|1-h_{k+1}/h\|^2_{L^2[0,1]}} = n_k$ as in Proposition 1(d). □

It is also true that one can improve Proposition 1(d) to make it a strict inequality since one can check that

$$h(x) > \cdots > h_{k+1}(x) > h_k(x) > \cdots > h_0(x) \equiv 0 \quad \text{for } x \in (0,1).$$

6.2. *Proof of the existence of the root-Horton limit*

Here we present the proof of our main Theorem 1. It is based on Lemma 5 and Lemma 6 that will be proven in the following two subsections.

**Lemma 5.** *If the limit* $\lim_{k\to\infty} \frac{h_{k+1}(1)}{h_k(1)}$ *exists, then* $\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty}(\prod_{j=1}^k n_j)^{-\frac{1}{k}}$ *also exists, and*

$$\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to+\infty}\left(\frac{1}{h_k(1)}\right)^{-\frac{1}{k}} = \lim_{k\to\infty}\frac{h_{k+1}(1)}{h_k(1)}.$$

**Lemma 6.** *The limit* $\lim_{k\to\infty} \frac{h_{k+1}(1)}{h_k(1)} \geq 1$ *exists, and is finite.*

**Theorem 1.** *The limit* $\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty}(\prod_{j=1}^k n_j)^{-\frac{1}{k}} = R$ *exists. Moreover,* $R = \lim_{k\to\infty}\frac{h_{k+1}(1)}{h_k(1)}$, *and* $2 \leq R \leq 4$.

**Proof.** The existence and finiteness of $\lim_{k\to\infty}\frac{h_{k+1}(1)}{h_k(1)}$ established in Lemma 6 is the precondition for Lemma 5 that in turn implies the existence and finiteness of the limit $\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}}$ as needed for the root-Horton law. Finally, $2 \leq R \leq 4$ follows from Proposition 1. □

6.3. *Proof of Lemma 5 and related results*

**Proposition 2.**

$$\left\|1-h_{k+1}(x)/h(x)\right\|^2_{L^2[0,1]} \leq \frac{1}{h_{k+1}(1)} \leq \left\|1-h_k(x)/h(x)\right\|^2_{L^2[0,1]}.$$

**Proof.** Integrating from 0 to 1 both sides of the equation

$$\frac{h'_{k+1}(x)}{h^2_{k+1}(x)} = 1 - \frac{(h_{k+1}(x)-h_k(x))^2}{h^2_{k+1}(x)}$$

we obtain $\frac{1}{h_{k+1}(1)} = \int_0^1 \frac{(h_{k+1}(x)-h_k(x))^2}{h^2_{k+1}(x)}\,dx$ as $h_{k+1}(0)=1$.

Hence,

$$\frac{1}{h_{k+1}(1)} = \int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h_{k+1}^2(x)}\,dx \geq \int_0^1 \frac{(h_{k+1}(x) - h_k(x))^2}{h^2(x)}\,dx = \int_0^1 \left(1 - \frac{h_{k+1}(x)}{h(x)}\right)^2 dx$$

by Lemma 4, proving the first inequality.

Now,

$$\frac{1}{h_{k+1}(1)} = \left\|1 - h_k(x)/h_{k+1}(x)\right\|_{L^2[0,1]}^2 \leq \left\|1 - h_k(x)/h(x)\right\|_{L^2[0,1]}^2$$

thus completing the proof. □

**Proof of Lemma 5.** If the limit $\lim_{k\to\infty} \frac{h_{k+1}(1)}{h_k(1)}$ exists and is finite, then $\lim_{k\to\infty}(\frac{1}{h_k(1)})^{-\frac{1}{k}}$ must also exist and be finite. Hence the existence and finiteness of

$$\lim_{k\to\infty} (\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty} \left(\int_0^1 \left(1 - \frac{h_k(x)}{h(x)}\right)^2 dx\right)^{-\frac{1}{k}}$$

follows from Proposition 2. □

### 6.4. *Proof of Lemma 6 and related results*

In this subsection we use the approach developed by Drmota [8] to prove the existence and finiteness of $\lim_{k\to\infty} \frac{h_{k+1}(1)}{h_k(1)} \geq 1$. As we observed earlier this result is needed to prove the existence, finiteness, and positivity of $\lim_{k\to\infty}(\mathcal{N}_k)^{-\frac{1}{k}} = \lim_{k\to\infty}(\prod_{j=1}^k n_j)^{-\frac{1}{k}}$, the root-Horton law.

***Definition 3.*** *Given $\gamma \in (0,1]$. Let*

$$V_{k,\gamma}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \leq x \leq 1-\gamma, \\ \gamma^{-1} h_k(\frac{x-(1-\gamma)}{\gamma}) & \text{for } 1-\gamma \leq x \leq 1. \end{cases}$$

Note that sequences of functions $h_k(x)$ and $V_{k,\gamma}(x)$ can be extended beyond $x = 1$.

Here are some observations we make about the above defined functions.

**Observation 1.** *$V_{k,\gamma}(x)$ are positive continuous functions satisfying*

$$V'_{k+1,\gamma}(x) = 2V_{k+1,\gamma}(x)V_{k,\gamma}(x) - V_{k,\gamma}^2(x)$$

*for all $x \in [0,1] \setminus (1-\gamma)$, with initial conditions $V_{k,\gamma}(0) = 1$.*

**Observation 2.** *Let $\gamma_k = \frac{h_k(1)}{h_{k+1}(1)}$. Then*

$$V_{k,\gamma_k}(1) = h_{k+1}(1) \tag{17}$$

*and*

$$V_{k,\gamma}(1) = \gamma^{-1} h_k(1) \geq h_{k+1}(1) \quad \text{whenever } \gamma \leq \gamma_k. \tag{18}$$

**Observation 3.**

$$V_{k,\gamma}(x) \leq V_{k+1,\gamma}(x)$$

*for all $x \in [0,1]$ since $h_k(x) \leq h_{k+1}(x)$.*

**Observation 4.** *Since* $h_1(x) \equiv 1$ *and* $\gamma_1 = \frac{h_1(1)}{h_2(1)}$,

$$h_2(x) \le V_{1,\gamma_1}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma_1, \\ \gamma_1^{-1} = h_2(1) & \text{for } 1-\gamma_1 \le x \le 1. \end{cases}$$

The above observation generalizes as follows.

**Proposition 3.**

$$h_{k+1}(x) \le V_{k,\gamma_k}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma_k, \\ \gamma_k^{-1} h_k\big(\frac{x-(1-\gamma_k)}{\gamma_k}\big) & \text{for } 1-\gamma_k \le x \le 1. \end{cases}$$

In order to prove Proposition 3 we will need the following lemma.

**Lemma 7.** *For any* $\gamma \in (0,1)$ *and* $k \ge 1$, *function* $V_{k,\gamma}(x) - h_{k+1}(x)$ *changes its sign at most once as* $x$ *increases from* $1-\gamma$ *to* 1. *Moreover, since* $V_{k,\gamma}(1-\gamma) = h(1-\gamma) > h_{k+1}(1-\gamma)$, *function* $V_{k,\gamma}(x) - h_{k+1}(x)$ *can only change sign from nonnegative to negative.*

**Proof.** This is a proof by induction with base at $k = 1$. Here $V_{1,\gamma}(x) = \frac{1}{\gamma}$ is constant on $[1-\gamma, 1]$, while $h_2(x) = (1 + e^{2x})/2$ is an increasing function, and

$$V_{1,\gamma}(1-\gamma) = h(1-\gamma) > h_2(1-\gamma).$$

For the induction step, we need to show that if $V_{k,\gamma}(x) - h_{k+1}(x)$ changes its sign at most once, then so does $V_{k+1,\gamma}(x) - h_{k+2}(x)$. Since both sequences of functions satisfy the same ODE relation (see Observation 1), we have

$$\frac{d}{dx}\Big[\big(V_{k+1,\gamma}(x) - h_{k+2}(x)\big) \cdot e^{-2\int_{1-\gamma}^{x} h_{k+1}(y)\,dy}\Big]$$
$$= \big(2V_{k+1,\gamma}(x) - V_{k,\gamma}(x) - h_{k+1}(x)\big) \cdot \big(V_{k,\gamma}(x) - h_{k+1}(x)\big) \cdot e^{-2\int_{1-\gamma}^{x} h_{k+1}(y)\,dy},$$

where $h_{k+1}(x) \le V_{k+1,\gamma}(x)$ by definition of $V_{k+1,\gamma}(x)$, and $V_{k,\gamma}(x) \le V_{k+1,\gamma}(x)$ as in Observation 3.

Now, let

$$I(x) := \int_{1-\gamma}^{x} \big(2V_{k+1,\gamma}(s) - V_{k,\gamma}(s) - h_{k+1}(s)\big) \cdot \big(V_{k,\gamma}(s) - h_{k+1}(s)\big) \cdot e^{-2\int_{1-\gamma}^{s} h_{k+1}(y)\,dy}\,ds.$$

Then

$$\big(V_{k+1,\gamma}(x) - h_{k+2}(x)\big) \cdot e^{-2\int_{1-\gamma}^{x} h_{k+1}(y)\,dy} = V_{k+1,\gamma}(1-\gamma) - h_{k+2}(1-\gamma) + I(x).$$

The function $2V_{k+1,\gamma}(x) - V_{k,\gamma}(x) - h_{k+1}(x) \ge 0$, and since $V_{k,\gamma}(x) - h_{k+1}(x)$ changes its sign at most once, then $I(x)$ should change its sign from nonnegative to negative at most once as $x$ increases from $1-\gamma$ to 1. Hence

$$V_{k+1,\gamma}(x) - h_{k+2}(x) = \big(V_{k+1,\gamma}(1-\gamma) - h_{k+2}(1-\gamma) + I(x)\big) \cdot e^{2\int_{1-\gamma}^{x} h_{k+1}(y)\,dy}$$

should change its sign from nonnegative to negative at most once as

$$V_{k+1,\gamma}(1-\gamma) = h(1-\gamma) > h_{k+2}(1-\gamma).$$ □

**Proof of Proposition 3.** Take $\gamma = \gamma_k$ in Lemma 7. Then function $h_{k+1}(x) - V_{k,\gamma_k}(x)$ should change its sign from nonnegative to negative at most once within the interval $[1-\gamma_k, 1]$. Hence, $V_{k,\gamma_k}(1-\gamma_k) > h_{k+1}(1-\gamma_k)$ and $h_{k+1}(1) = V_{k,\gamma_k}(1)$ imply $h_{k+1}(x) \le V_{k,\gamma_k}(x)$ as in the statement of the proposition. □

Now we are ready to prove the monotonicity result.

**Lemma 8.**

$$\gamma_k \le \gamma_{k+1} \quad \textit{for all } k \in \mathbb{N}^+.$$

**Proof.** We prove it by contradiction. Suppose $\gamma_k \ge \gamma_{k+1}$ for some $k \in \mathbb{N}^+$. Then

$$V_{k,\gamma_k}(x) \le V_{k,\gamma_{k+1}}(x) = \begin{cases} \frac{1}{1-x} & \text{for } 0 \le x \le 1-\gamma_{k+1}, \\ \gamma_{k+1}^{-1} h_k\left(\frac{x-(1-\gamma_{k+1})}{\gamma_{k+1}}\right) & \text{for } 1-\gamma_{k+1} \le x \le 1 \end{cases}$$

and therefore

$$h_{k+1}(x) \le V_{k,\gamma_k}(x) \le V_{k,\gamma_{k+1}}(x) \le V_{k+1,\gamma_{k+1}}(x)$$

as $h_{k+1}(x) \le V_{k,\gamma_k}(x)$ by Proposition 3.

Recall that for $x \in [1-\gamma_{k+1}, 1]$,

$$V'_{k+1,\gamma_{k+1}}(x) = 2V_{k,\gamma_{k+1}}(x)V_{k+1,\gamma_{k+1}}(x) - V^2_{k,\gamma_{k+1}},$$

where at $1-\gamma_{k+1}$ we consider only the right-hand derivative. Thus for $x \in [1-\gamma_{k+1}, 1]$,

$$\frac{d}{dx}\big(V_{k+1,\gamma_{k+1}}(x) - h_{k+2}(x)\big) = A(x) + B(x)\big(V_{k+1,\gamma_{k+1}}(x) - h_{k+2}(x)\big),$$

where $A(x) = 2V_{k+1,\gamma_{k+1}}(x) - V_{k,\gamma_{k+1}}(x) - h_{k+1}(x) \ge 0$, $B(x) = 2h_{k+1}(x) > 0$, and

$$V_{k+1,\gamma_{k+1}}(1-\gamma_{k+1}) - h_{k+2}(1-\gamma_{k+1}) = h(1-\gamma_{k+1}) - h_{k+2}(1-\gamma_{k+1}) > 0.$$

Hence

$$V_{k+1,\gamma_{k+1}}(1) - h_{k+2}(1) \ge V_{k+1,\gamma_{k+1}}(1-\gamma_{k+1}) - h_{k+2}(1-\gamma_{k+1}) > 0$$

arriving to a contradiction since $V_{k+1,\gamma_{k+1}}(1) = h_{k+2}(1)$. □

**Corollary.** *Limit* $\lim_{k\to\infty} \gamma_k$ *exists.*

**Proof.** Lemma 8 implies $\gamma_k$ is a monotone increasing sequence, bounded by 1. □

**Proof of Lemma 6.** Lemma 6 follows immediately from an observation that $\frac{h_{k+1}(1)}{h_k(1)} = \frac{1}{\gamma_k}$. □

## 7. Relation to the tree representation of white noise

This section establishes a close connection between the combinatorial tree of Kingman's $N$-coalescent and the combinatorial level set tree of a discrete white noise.

### 7.1. *Level set tree of a discrete-time function*

We start with recalling basic facts about tree representation of a discrete-time function; for details and further results see [18]. Consider a function $X_i$ with discrete time index $i = 0, 1, \dots, i_{\max}$ and values distributed without atoms over $\mathbb{R}$. Let $X_t \equiv X(t)$ be a function of continuous time $t \in [0, i_{\max}]$ obtained from $X_i$ by linear interpolation of its values. The level set $\mathcal{L}_\alpha(X_t)$ is defined as the pre-image of the function values above $\alpha$:

$$\mathcal{L}_\alpha(X_t) = \{t : X_t \ge \alpha\}.$$

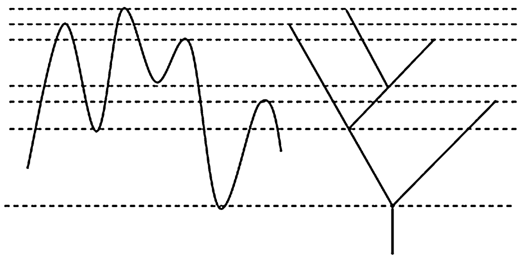


Fig. 2. Function $X_t$ (panel (a)) with a finite number of local extrema and its level set tree $\text{LEVEL}(X)$ (panel (b)).

The level set $\mathcal{L}_\alpha$ for each $\alpha$ is a union of non-overlapping intervals; we write $|\mathcal{L}_\alpha|$ for their number. Notice that $|\mathcal{L}_\alpha| = |\mathcal{L}_\beta|$ as soon as the interval $[\alpha, \beta]$ does not contain a value of local maxima or minima of $X_t$ and $0 \le |\mathcal{L}_\alpha| \le n$, where $n$ is the number of the local maxima of $X_t$.

The *level set tree* $\text{LEVEL}(X_t)$ is a planar time oriented binary tree that describes the topology of the level sets $\mathcal{L}_\alpha$ as a function of threshold $\alpha$, as illustrated in Figure 2. Namely, there are bijections between (i) the leaves of $\text{LEVEL}(X_t)$ and the local maxima of $X_t$, (ii) the internal (parental) vertices of $\text{LEVEL}(X_t)$ and the local minima of $X_t$ (excluding possible local minima at the boundary points), and (iii) the pair of subtrees of $\text{LEVEL}(X_t)$ rooted at a local minima $X(t^*)$ and the first positive excursions (or meanders bounded by $t = 0$ or $t = N$) of $X(t) - X(t^*)$ to right and left of $t^*$. Each vertex in the tree is assigned a mark equal to the value of the local extrema according to the bijections (i) and (ii) above. This makes the tree time oriented according to the threshold $\alpha$. It is readily seen that any function $X_t$ with distinct values of consecutive local minima corresponds to a binary tree $\text{LEVEL}(X_t)$. We refer to [18] for discussion of some subtleties related to this construction as well as for further references.

### 7.2. *Tree representation of white noise*

Let $W_j^{(N)}$, $j = 1, \dots, N-1$, be a *discrete white noise* that is a discrete time process comprised of $N-1$ i.i.d. random variables with a common atomless distribution. Consider now an auxiliary process $\tilde{W}_i^{(N)}$, $i = 1, \dots, 2N-1$ such that it has exactly $N$ local maxima and $N-1$ internal local minima $\tilde{W}_{2j}^{(N)} = W_j^{(N)}$, $j = 1, \dots, N-1$. We call $\tilde{W}_i^{(N)}$ an *extended white noise*; it can be constructed, for example, as follows:

$$\tilde{W}_i^{(N)} = \begin{cases} W_{i/2}^{(N)} & \text{for even } i, \\ \max\left(W_{\max(1, \frac{i-1}{2})}^{(N)}, W_{\min(N-1, \frac{i+1}{2})}^{(N)}\right) + 1 & \text{for odd } i. \end{cases} \tag{19}$$

Let $L_W^{(N)} = \text{LEVEL}(\tilde{W}_i^{(N)})$ be the level set tree of $\tilde{W}_i^{(N)}$ and $\text{SHAPE}(L_W^{(N)})$ be a (random) combinatorial tree that retains the graph-theoretic structure of $L_W^{(N)}$ and drops its planar embedding as well as the vertex marks. By construction, $L_W^{(N)}$ has exactly $N$ leaves.

**Lemma 9.** *The distribution of* $\text{SHAPE}(L_W^{(N)})$ *on* $\mathcal{T}_N$ *is the same for any atomless distribution* $F$ *of the values of the associated white noise* $W_j^{(N)}$.

**Proof.** The condition of atomlessness of $F$ is necessary to ensure that the level set tree is binary with probability 1. By construction, the combinatorial level set tree is completely determined by the ordering of the local minima of the respective trajectory, independently of the particular values of its local maxima and minima. We complete the proof by noticing that the ordering of $W_j^{(N)}$ is the same for any choice of atomless distribution $F$. □

Let $T_{\mathrm{K}}^{(N)}$ be the tree that corresponds to a Kingman's $N$-coalescent, and let $\mathrm{SHAPE}(T_{\mathrm{K}}^{(N)})$ be its combinatorial version that drops the time marks of the vertices. Both the trees $\mathrm{SHAPE}(L_W^{(N)})$ and $\mathrm{SHAPE}(T_{\mathrm{K}}^{(N)})$, belong to the space $\mathcal{T}_N$ of binary rooted trees with $N$ leaves.

**Theorem 2.** *The trees* $\mathrm{SHAPE}(L_W^{(N)})$ *and* $\mathrm{SHAPE}(T_{\mathrm{K}}^{(N)})$ *have the same distribution on* $\mathcal{T}_N$.

The proof below uses the duality between coalescence and fragmentation processes [1]. Recall that a *fragmentation process* starts with a single cluster of mass $N$ at time $t = 0$. Each existing cluster of mass $m$ splits into two clusters of masses $m - x$ and $x$ at the splitting rate $S_t(m, x)$, $1 < m \le N$, $1 \le x < N$. A coalescence process on $N$ particles with time-dependent collision kernel $K_t(x, y)$, $1 \le x, y < N$ is equivalent, upon time reversal, to a discrete-mass fragmentation process of initial mass $N$ with some splitting kernel $S_t(m, x)$. See Aldous [1] for further details and the relationship between the dual collision and splitting kernels in general case.

**Proof of Theorem 2.** We show that both the examined trees have the same distribution as the combinatorial tree of a fragmentation process with mass $N$ and a splitting kernel that is uniform in mass: $S_t(m, x) = S(t)$.

Kingman's $N$-coalescence with kernel $K(x, y) = 1$ is dual to the fragmentation process with splitting kernel [1, Table 3]

$$S_t(m, x) = \frac{2}{t(t+2)}.$$

This kernel is independent of the cluster mass, which means that the splitting of mass $m$ is uniform among the $m - 1$ possible pairs $\{1, m-1\}$, $\{2, m-2\}$, $\ldots$, $\{m-1, 1\}$. The time dependence of the kernel does not affect the combinatorial structure of the fragmentation tree (and can be removed by a deterministic time change).

The level set tree $L_W^{(N)}$ can be viewed as a tree that describes a fragmentation process with the initial mass $N$ equal to the number of local maxima of the trajectory $\tilde{W}_i^{(N)}$. By construction, each subtree of $L_W^{(N)}$ with $n$ leaves corresponds to an excursion (or meander, if we treat one of the boundaries) with $n$ local maxima. This subtree (as well as the corresponding excursion or meander) splits into two by the internal global minimum of $\tilde{W}_i^{(N)}$ at the corresponding time interval.

The global minimum splits the series $\tilde{W}_i^{(N)}$ into two, to the left and right of the minimum, with $M_L$ and $(N - M_L)$ local maxima, respectively. Since the local minima of $\tilde{W}_i^{(N)}$ form a white noise, the distribution of $M_L$ is uniform on $[1, N - 1]$. Next, the internal vertices of the level set tree of the left (or right) time series correspond to its $M_L - 1$ (or $N - M_L - 1$) internal local minima that form a white noise (with the distribution different from that of the initial white noise $W_j^{(N)}$). Hence, the subsequent splits of masses (number of local maxima) continues according to a discrete uniform distribution. And so on down the tree.

Hence, the combinatorial level set tree of $\tilde{W}_i^{(N)}$ has the same distribution as a combinatorial tree of a fragmentation process with uniform mass splitting. This completes the proof. □

***Remark 2.*** *We notice that the dual splitting kernels for multiplicative and additive coalescences* [1, *Table* 3] *only differ by their time dependence, and are equivalent as functions of mass. Hence, the combinatorial structure of the respective trees is the same.*

**Corollary 1.** *The combinatorial level set tree of a discrete white noise* $W^{(N)}$ *is root-Horton self similar with the same Horton exponent* $R$ *as that for Kingman's* $N$*-coalescent.*

**Proof.** Recall the operation of tree *pruning* $\mathcal{R}(T) : \mathcal{T} \to \mathcal{T}$ that cuts the leaves of a finite tree $T$ and removes possible resulting nodes of degree 2 [4,18]. By definition, pruning corresponds to index shift in Horton statistics: $N_k \to N_{k-1}$, $k > 1$. It has been shown in [18] that

$$\mathcal{R}\big[\mathrm{LEVEL}\big(\tilde{W}_i^{(N)}\big)\big] = \mathrm{LEVEL}\big(W_j^{(N)}\big).$$

Hence, Horton self-similarity for one of these processes implies that for the other. The Horton self-similarity for the extended white noise $\tilde{W}^{(N)}$ follows directly from Theorem 2. □

## 8. General coalescent processes

The ODE approach introduced in this paper can be extended to the coalescent kernels other than $K(i,j) \equiv 1$. For that we need to classify the relative number $\eta_k(t)$ of clusters of order $k$ at time $t$ according to the cluster masses. Namely, let $\eta_{k,m}(t)$ be the average number of clusters of order $k$ and mass $m \geq 2^k$ at time $t$. Then

$$\eta_k(t) = \sum_{m=2^k}^{\infty} \eta_{k,m}(t).$$

In the case of a symmetric coalescent kernel $K(i,j) = K(j,i)$ the Smoluchowski–Horton ODEs can be written asymptotically as

$$\begin{aligned} \frac{d}{dt}\eta_{k,m}(t) = &\sum_{i=1}^{k-1}\sum_{\mu=2^k}^{m-2^i} \eta_{k,\mu}(t)\eta_{i,m-\mu}K(\mu, m-\mu) \\ &+ \frac{1}{2}\sum_{\substack{m_1+m_2=m \\ m_1,m_2 \geq 2^{k-1}}} \eta_{k-1,m_1}(t)\eta_{k-1,m_2}(t)K(m_1,m_2) \\ &- \eta_{k,m}(t)\sum_{\widetilde{m}=2^i}^{\infty} K(m,\widetilde{m})\left(\sum_{i=1}^{\infty}\eta_{i,\widetilde{m}}(t)\right) \end{aligned} \tag{20}$$

with the initial conditions $\eta_{1,1}(0) = 1$ and $\eta_{k,m}(0) = 0$ for all $(k,m) \neq (1,1)$.

Observe that when $K(i,j) \equiv 1$, summing the above equations (20) over index $m$ produces the Smoluchowski–Horton ODE (3) for the average relative number of order-$k$ branches $\eta_k(t)$ in Kingman's coalescent process.

## 9. Discussion

This paper establishes the root-Horton self-similarity (Section 6, Theorem 1) for Kingman's $N$-coalescent process, as $N$ goes to infinity. We also demonstrate (Section 7.1, Theorem 2) the distributional equivalence of the combinatorial trees of Kingman's $N$-coalescent to that of a discrete extended white noise with $N$ local maxima, hence extending the self-similarity results to a tree representation of a discrete white noise (Section 7, Corollary 1).

Combining the results of this study with that of Burd et al. [4] and Zaliapin and Kovchegov [18] one observes that Horton self-similarity is a property of tree representation for (i) white noise, (ii) symmetric random walk, (iii) critical binary Galton–Watson branching process, and (iv) Kingman's $N$-coalescent. The listed processes are believed to closely depict physical and biological mechanisms of diverse origin and are commonly used as essential building blocks in scientific modeling. The results of this study and those in [4,18] thus provide at least a partial explanation for the omnipresence of Horton self-similarity in observed and modeled branching structures. This study seems to be the first that rigorously establishes Horton self-similarity with Horton exponent different from $R = 2, 4$.

Our Theorem 1 establishes a weak, root-law, convergence of the asymptotic ratios $\mathcal{N}_k$, while we believe that the stronger (ratio and geometric) forms of convergence are also valid. These stronger Horton laws are usually considered in the literature (e.g., [7,9,12,18]). It seems important to show rigorously at least the ratio-Horton law ($\lim_{k\to\infty} \mathcal{N}_k/\mathcal{N}_{k+1} = R > 0$).

The Smoluchowski–Horton equations (3) that form a core of the presented method and their equivalents (14) and (15) seem to be promising for further more detailed exploration. Indeed, one may hope that the approach that refers explicitly to the Horton–Strahler orders might effectively complement conventional analysis of cluster masses. The analysis of the Smoluchowski–Horton systems can be done within the ODE framework, similarly to the present study, or within the nonlinear iterative system framework (see (16)). The latter approach is still to be explored.

Finally, it is noteworthy that the analysis of multiplicative and additive coalescents according to the general Smoluchowski–Horton system (20) appears, after a certain series of transformations, to follow many of the steps implemented in this paper for Kingman's coalescent, with the ODE system being replaced by a suitable PDE one. These results will be published elsewhere.

## Appendix A: Proof of Lemma 1

We split the proof into smaller steps.

• *Step I.* Fix $\epsilon_0 \in (0,1)$ and take $\delta > 0$. We show below that, given $\eta_{(N)}(t) = y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]$, the number of coalescences during the time interval $[t, t+\delta]$ does not exceed $\frac{\delta}{N}\binom{Ny}{2} + N^{2/3}$ with high probability. Specifically, we use an exponential Markov inequality (Chernoff's bound) with exponent $s > 0$ to bound the probability that a sum of $\frac{\delta}{N}\binom{Ny}{2} + N^{2/3}$ exponential inter-arrival times with the rate not exceeding $\frac{1}{N}\binom{Ny}{2}$ adds up to less than $\delta$. Let $\zeta_i$ be the arrival time of $i$th coalescence and $u = \binom{Ny}{2}$. Then

$$
\begin{aligned}
&P\left(N\left[\eta_{(N)}(t) - \eta_{(N)}(t+\delta)\right] > \frac{\delta}{N}u + N^{2/3} \Big| \eta_{(N)}(t) = y\right)\\
&\quad = P\left(\sum_{i=1}^{\lfloor \frac{\delta}{N}u + N^{2/3}\rfloor} \zeta_i < \delta \Big| \eta_{(N)}(t) = y\right)\\
&\quad \le \frac{e^{s\delta}}{(1+\frac{sN}{u})^{\frac{\delta}{N}u + N^{2/3}}}\\
&\quad \le \exp\left\{s\delta - \left(\frac{\delta}{N}u + N^{2/3}\right)\left(\frac{sN}{u} - \frac{s^2N^2}{u^2}\right)\right\}\\
&\quad = \exp\left\{-\frac{s}{u}N^{5/3} + \frac{\delta s^2}{u}N + \frac{s^2}{u^2}N^{8/3}\right\}
\end{aligned}
$$

as $\ln(1+x) > x - x^2$ for $x > 0$. Taking $s = N^{1/2}$ in the above inequality, we obtain

$$
\begin{aligned}
&P\left(N\left[\eta_{(N)}(t) - \eta_{(N)}(t+\delta)\right] > \frac{\delta}{N}u + N^{2/3} \Big| \eta_{(N)}(t) = y\right)\\
&\quad = \exp\left\{-\frac{1}{u}N^{13/6} + \frac{\delta N^2}{u} + \frac{N^{11/3}}{u^2}\right\}\\
&\quad = \exp\left\{-\frac{2}{Ny(Ny-1)}N^{13/6} + \frac{2\delta N^2}{Ny(Ny-1)} + \frac{4N^{11/3}}{(Ny)^2(Ny-1)^2}\right\}\\
&\quad = \exp\left\{-\frac{2}{y(y-1/N)}N^{1/6} + \frac{2\delta}{y(y-1/N)} + \frac{4N^{-1/3}}{(y)^2(y-1/N)^2}\right\}\\
&\quad \le \exp\left\{-2N^{1/6} + \frac{2\delta}{\epsilon_0(\epsilon_0 - 1/N)} + \frac{4N^{-1/3}}{\epsilon_0^2(\epsilon_0 - 1/N)^2}\right\}\\
&\quad \le \exp\left\{-N^{1/6} + \frac{4\delta}{\epsilon_0^2}\right\} \qquad (21)
\end{aligned}
$$

for $N$ large enough.

• *Step II.* From Step I we know that, given $\eta_{(N)}(t) = y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]$, there are no more than

$$
\begin{aligned}
\frac{\delta}{N}\binom{Ny}{2} + N^{2/3} &= \frac{\delta y^2}{2}N - \frac{\delta y}{2} + N^{2/3}\\
&\le \frac{\delta y^2}{2}N + N^{2/3}
\end{aligned}
$$

coalescing pairs during $[t, t+\delta]$ with probability exceeding $1-\exp\{-N^{1/6}+\frac{4\delta}{\epsilon_0^2}\}$. In this case the exponential rates of inter-arrival times during $[t, t+\delta]$ must be at least

$$\begin{aligned}
&\frac{1}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil-\lceil N^{2/3}\rceil}{2}\\
&=\frac{1}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}-\frac{Ny-\lceil\frac{\delta y^2}{2}N\rceil-1/2-\lceil N^{2/3}\rceil/2}{N}\lceil N^{2/3}\rceil\\
&\geq\frac{1}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}-\left(y-\frac{\delta y^2}{2}\right)\lceil N^{2/3}\rceil\\
&\geq\frac{1}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}-N^{2/3}
\end{aligned}$$

for $N$ large enough. We now use exponential Markov inequality to bound the conditional probability that there are fewer than $\frac{\delta}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}-(1+\delta)N^{2/3}$ coalescents in $[t, t+\delta]$. Specifically, we bound the probability that a sum of $\frac{\delta}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}-(1+\delta)N^{2/3}$ independent exponential random variables of rate not less than $\frac{1}{N}\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}-N^{2/3}$ is greater than $\delta$.

Set $v=\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}$. Since we are interested in the values of $\delta \ll 1$, then

$$(1-\delta)^2\frac{N^2y^2}{2}\leq v=\binom{Ny-\lceil\frac{\delta y^2}{2}N\rceil}{2}\leq u=\binom{Ny}{2}\leq\frac{N^2y^2}{2}. \tag{22}$$

Exponential Markov inequality with exponent $s>0$ implies

$$\begin{aligned}
&P\left(N\big[\eta_{(N)}(t)-\eta_{(N)}(t+\delta)\big]<\frac{\delta}{N}v-(1+\delta)N^{2/3}\,\middle|\,\begin{matrix}N[\eta_{(N)}(t)-\eta_{(N)}(t+\delta)]\leq\frac{\delta}{N}u+N^{2/3}\\ \eta_{(N)}(t)=y\end{matrix}\right)\\
&\leq\frac{e^{-s\delta}}{(1-\frac{sN}{v-N^{5/3}})^{\frac{\delta}{N}v-(1+\delta)N^{2/3}}}\\
&\leq\exp\left\{-s\delta+\left(\frac{\delta}{N}v-(1+\delta)N^{2/3}\right)\left(\frac{sN}{v-N^{5/3}}+\frac{s^2N^2}{(v-N^{5/3})^2}\right)\right\}\\
&\leq\exp\left\{\left(\frac{1}{1-N^{5/3}/v}-1\right)s\delta-\frac{s(1+\delta)N^{5/3}}{v-N^{5/3}}+\left(\frac{\delta}{N}v-(1+\delta)N^{2/3}\right)\frac{s^2N^2}{(v-N^{5/3})^2}\right\}\\
&\leq\exp\left\{\frac{s\delta N^{5/3}/v}{1-N^{5/3}/v}-\frac{s(1+\delta)N^{5/3}}{v}+\frac{\delta v s^2 N}{(v-N^{5/3})^2}\right\}
\end{aligned}$$

as $-x-x^2<\ln(1-x)$ for $x\in(0,\frac{1}{2})$. Take $s=N^{1/2}$ to obtain

$$\begin{aligned}
&P\left(N\big[\eta_{(N)}(t)-\eta_{(N)}(t+\delta)\big]<\frac{\delta}{N}v-(1+\delta)N^{2/3}\,\middle|\,\begin{matrix}N[\eta_{(N)}(t)-\eta_{(N)}(t+\delta)]\leq\frac{\delta}{N}u+N^{2/3}\\ \eta_{(N)}(t)=y\end{matrix}\right)\\
&=\exp\left\{\frac{\delta N^{13/6}/v}{1-N^{5/3}/v}-\frac{(1+\delta)N^{13/6}}{v}+\frac{\delta v N^2}{(v-N^{5/3})^2}\right\}\\
&\leq\exp\left\{\frac{2\delta N^{1/6}}{(1-\delta)^2y^2-2N^{-1/3}}-\frac{2(1+\delta)N^{1/6}}{y^2}+\frac{2\delta y^2}{((1-\delta)^2y^2-2N^{-1/3})^2}\right\}\\
&\leq\exp\left\{\frac{2N^{1/6}}{y^2}\left[\frac{\delta}{(1-\delta)^2-2N^{-1/3}/y^2}-(1+\delta)\right]+\frac{3\delta y^2}{(1-\delta)^4y^4}\right\}
\end{aligned}$$

$$\le \exp\left\{-\frac{N^{1/6}}{y^2}+\frac{3\delta}{(1-\delta)^4 y^2}\right\}$$

$$\le \exp\left\{-N^{1/6}+\frac{4\delta}{\epsilon_0^2}\right\} \tag{23}$$

for $N$ large enough, by using (22).

Thus, multiplying the probabilities of complement events in (21) and (23) we obtain

$$P\left(\frac{\delta}{N^2}v-(1+\delta)N^{-1/3}\le \eta_{(N)}(t)-\eta_{(N)}(t+\delta)\le \frac{\delta}{N^2}u+N^{-1/3}\Big|\eta_{(N)}(t)=y\right)$$

$$\ge \left(1-\exp\left\{-N^{1/6}+\frac{4\delta}{\epsilon_0^2}\right\}\right)^2$$

for any given $t\ge 0$ and $y\in \frac{1}{N}\mathbb{Z}\cap[\epsilon_0,1]$.

- *Step III.* Let $\Delta_\delta f(x):=\frac{f(x+\delta)-f(x)}{\delta}$ denote the forward difference. Now, as we already pointed out in (22),

$$(1-\delta)^2\frac{N^2\eta^2_{(N)}(t)}{2}\le v\le u\le \frac{N^2\eta^2_{(N)}(t)}{2}.$$

Hence,

$$P\left(\left|\frac{\eta^2_{(N)}(t)}{2}+\Delta_\delta\eta_{(N)}(t)\right|\le \delta+\left(\delta^{-1}+1\right)N^{-1/3}\Big|\eta_{(N)}(t)=y\right)$$

$$\ge P\left((1-\delta)^2\frac{\eta^2_{(N)}(t)}{2}-\left(\delta^{-1}+1\right)N^{-1/3}\le -\Delta_\delta\eta_{(N)}(t)\le \frac{\eta^2_{(N)}(t)}{2}+\delta^{-1}N^{-1/3}\Big|\eta_{(N)}(t)=y\right)$$

$$\ge P\left(\frac{\delta}{N^2}v-(1+\delta)N^{-1/3}\le \eta_{(N)}(t)-\eta_{(N)}(t+\delta)\le \frac{\delta}{N^2}u+N^{-1/3}\Big|\eta_{(N)}(t)=y\right)$$

$$\ge \left(1-\exp\left\{-N^{1/6}+\frac{4\delta}{\epsilon_0^2}\right\}\right)^2 \tag{24}$$

for $N$ large enough. The first inequality above uses the fact that

$$(1-\delta)^2\frac{\eta^2_{(N)}(t)}{2}>\frac{\eta^2_{(N)}(t)}{2}-\delta.$$

This is equivalent to

$$(-2+\delta)\frac{\eta^2_{(N)}(t)}{2}>-1,$$

which is always true since $\eta_{(N)}(t)\le 1$ and $\delta>0$.

- *Step IV.* For $K>0$, consider an interval $[0,K]$ partitioned into $M$ subintervals

$$[t_0,t_1],[t_1,t_2],\dots,[t_{M-1},t_M]$$

of equal length $\delta=K/M$, where $t_0=0$ and $t_M=K$. Here $M$ may depend on $N$.

Let $\epsilon_0=\eta(K)/2=1/(2+K)$, where $\eta(t)=2/(2+t)$ is the solution to the equation (2) with the initial condition $\eta(0)=1$. Consider the following difference equation

$$\Delta_\delta\psi_{(N)}(t_i)=-\frac{\psi^2_{(N)}(t_i)}{2}+\mathcal{E}'(t_i) \tag{25}$$

with initial condition $\psi_{(N)}(0) = 1$, where the error $|\mathcal{E}'(t_i)|$ satisfies

$$\left|\mathcal{E}'(t_i)\right| \le \delta + \left(\delta^{-1} + 1\right)N^{-1/3}.$$

**Claim 1.** *If $M$ is large enough, then the following is true as we take $N$ large enough. For any natural number $j \le M$, if function $\psi_{(N)}(t_i)$ satisfies (25) for all $i \in \{0, 1, \dots, j-1\}$, then*

$$\psi_{(N)}(t_j) \ge \epsilon_0.$$

Indeed, if we take $N \ge M^6$, then

$$\left|\mathcal{E}'(t_i)\right| \le \delta + \left(\delta^{-1} + 1\right)N^{-1/3} \le K/M + 1/(KM) + 1/M^2.$$

Now, since $\eta(t) = 2/(2+t)$ is the solution to Equation (2) with the initial condition $\eta(0) = 1$, $\eta(t)$ will satisfy

$$\Delta_\delta \eta(t_i) = -\frac{\eta^2(t_i)}{2} + \mathcal{E}(t_i)$$

for all $i \in \{0, 1, \dots, M-1\}$, where $\mathcal{E}(t_i) = \frac{\eta''(c_i)}{2}\delta = \frac{\eta^3(c_i)}{4}\delta$ for some $c_i \in (t_i, t_{i+1})$. Hence, as $\eta(t) \le 1$ for all $t \ge 0$, $|\mathcal{E}(t_i)| \le \frac{1}{4}\delta$.

Consider the error quantities $\varepsilon_i := \psi_{(N)}(t_i) - \eta(t_i)$. We have

$$\begin{aligned}
\varepsilon_{i+1} &= \psi_{(N)}(t_{i+1}) - \eta(t_{i+1}) \\
&= \left[\psi_{(N)}(t_i) - \frac{\psi^2_{(N)}(t_i)}{2}\delta + \mathcal{E}'(t_i)\delta\right] - \left[\eta(t_i) - \frac{\eta^2(t_i)}{2}\delta + \mathcal{E}(t_i)\delta\right] \\
&= \left[\eta(t_i) + \varepsilon_i - \frac{(\eta(t_i)+\varepsilon_i)^2}{2}\delta + \mathcal{E}'(t_i)\delta\right] - \left[\eta(t_i) - \frac{\eta^2(t_i)}{2}\delta + \mathcal{E}(t_i)\delta\right] \\
&= \left(1 - \eta(t_i)\delta\right)\varepsilon_i - \frac{\varepsilon_i^2}{2}\delta + \delta\left(\mathcal{E}'(t_i) - \mathcal{E}(t_i)\right),
\end{aligned}$$

where $|\mathcal{E}'(t_i) - \mathcal{E}(t_i)| \le \frac{5}{4}K/M + 1/(KM) + 1/M^2 < C_K/M$ if $M > 1$, with $C_K = \frac{5}{4}K + \frac{1}{K} + 1$. Since $\eta(t_i) > \eta(K)$ for all $i \in \{0, 1, \dots, M-1\}$,

$$|\varepsilon_{i+1}| \le \left(1 - \eta(K)K/M\right)|\varepsilon_i| + \frac{\varepsilon_i^2}{2}K/M + KC_K/M^2.$$

Taking $M$ large enough so that $KC_K/M < 2\eta(K)$, we can prove by induction that

$$|\varepsilon_i| \le iKC_K/M^2. \tag{26}$$

Indeed, $\varepsilon_0 = 0$, and if $|\varepsilon_i| \le iKC_K/M^2$, then

$$\begin{aligned}
|\varepsilon_{i+1}| &\le \left(1 - \eta(K)K/M\right)|\varepsilon_i| + \frac{\varepsilon_i^2}{2}K/M + KC_K/M^2 \\
&= |\varepsilon_i| + \left(|\varepsilon_i| - 2\eta(K)\right)|\varepsilon_i|K/(2M) + KC_K/M^2 \\
&\le |\varepsilon_i| + \left(iKC_K/M^2 - 2\eta(K)\right)|\varepsilon_i|K/(2M) + KC_K/M^2 \\
&\le |\varepsilon_i| + KC_K/M^2 \\
&\le (i+1)KC_K/M^2,
\end{aligned}$$

which completes the induction step.

The inequality (26) is therefore valid for all $i \in \{0, \dots, M-1\}$, implying

$$|\varepsilon_i| \le MKC_K/M^2 = \frac{\frac{5}{4}K^2 + K + 1}{M} < \epsilon_0 \tag{27}$$

for $M$ large enough.

Recall that $\epsilon_0 = \eta(K)/2 = 1/(2+K)$. Then, by (27),

$$\psi_{(N)}(t_j) = \eta(t_j) + \varepsilon_j \ge \eta(K) - \epsilon_0 = \epsilon_0$$

for all $j \in \{0, 1, \dots, M-1\}$. This proves the above Claim 1.

• *Step V.* Consider events

$$A_i = \left\{ \Delta_\delta \eta_{(N)}(t_i) = -\frac{\eta^2_{(N)}(t_i)}{2} + \mathcal{E}'(t_i) \text{ and } \left|\mathcal{E}'(t_i)\right| \le \delta + \left(\delta^{-1} + 1\right) N^{-1/3} \right\} \tag{28}$$

for all $i \in \{0, 1, \dots, M-1\}$. Then inequality (24) rewrites as

$$P\left(A_j | \eta_{(N)}(t_j) = y\right) \ge \left(1 - \exp\left\{-N^{1/6} + \frac{4\delta}{\epsilon_0^2}\right\}\right)^2$$

for any $y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]$.

Claim 1 implies that $\bigcap_{i=0}^{j-1} A_i$ is contained in the event $\{\eta_{(N)}(t_j) \in [\epsilon_0, 1]\}$, and therefore, using the Markov property, we obtain

$$\begin{aligned} P\left(A_j \middle| \bigcap_{i=0}^{j-1} A_i\right) &= \sum_{y: y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]} P\left(A_j \middle| \eta_{(N)}(t_j) = y, \bigcap_{i=0}^{j-1} A_i\right) P\left(\eta_{(N)}(t_j) = y \middle| \bigcap_{i=0}^{j-1} A_i\right) \\ &= \sum_{y: y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]} P\left(A_j \middle| \eta_{(N)}(t_j) = y\right) P\left(\eta_{(N)}(t_j) = y \middle| \bigcap_{i=0}^{j-1} A_i\right) \\ &\ge \left(1 - \exp\left\{-N^{1/6} + \frac{4\delta}{\epsilon_0^2}\right\}\right)^2 \end{aligned}$$

as $\sum_{y: y \in \frac{1}{N}\mathbb{Z} \cap [\epsilon_0, 1]} P(\eta_{(N)}(t_j) = y | \bigcap_{i=0}^{j-1} A_i) = P(\eta_{(N)}(t_j) \in [\epsilon_0, 1] | \bigcap_{i=0}^{j-1} A_i) = 1$. Hence, since we have taken $N \ge M^6$,

$$\begin{aligned} P\left(\bigcap_{i=0}^{M-1} A_i\right) &\ge \left(1 - \exp\left\{-N^{1/6} + \frac{4\delta}{\epsilon_0^2}\right\}\right)^{2M} \\ &\ge \left(1 - \exp\left\{-M + \frac{4K}{\epsilon_0^2 M}\right\}\right)^{2M} \\ &\to 1 \quad \text{as } M \to \infty. \end{aligned} \tag{29}$$

We established that with probability greater than $P(\bigcap_{i=0}^{M-1} A_i) \to 1$ as $M \to \infty$, $\eta_{(N)}(t_i)$ satisfies difference equation (25) with $\psi_{(N)}(t) \equiv \eta_{(N)}(t)$.

• *Step VI.* Rewriting (27) for $\psi_{(N)}(t) \equiv \eta_{(N)}(t)$, we see that with probability of at least $P(\bigcap_{i=0}^{M-1} A_i) \to 1$,

$$\left|\eta_{(N)}(t_i) - \eta(t_i)\right| = |\varepsilon_i| < \epsilon_0$$

for all $i \in \{0, 1, \dots, M-1\}$. Now, if $t \in (t_i, t_{i+1})$, then

$$\begin{aligned}\left|\eta_{(N)}(t)-\eta(t)\right| &\le \left|\eta_{(N)}(t)-\eta_{(N)}(t_i)\right| + \left|\eta_{(N)}(t_i)-\eta(t_i)\right| + \left|\eta(t_i)-\eta(t)\right| \\ &\le \big(\eta_{(N)}(t_i)-\eta_{(N)}(t_{i+1})\big) + \left(\frac{5}{4}K^2+K+1\right)/M + \big(\eta(t_i)-\eta(t_{i+1})\big) \\ &= \eta_{(N)}(t_i)-\eta(t_i)+\eta(t_{i+1})-\eta_{(N)}(t_{i+1}) + \left(\frac{5}{4}K^2+K+1\right)/M + 2\big(\eta(t_i)-\eta(t_{i+1})\big) \\ &\le 3\left(\frac{5}{4}K^2+K+1\right)/M + 2\big(\eta(t_i)-\eta(t_{i+1})\big) \\ &\le 3\left(\frac{5}{4}K^2+K+1\right)/M + \delta\end{aligned}$$

as

$$2\big(\eta(t_i)-\eta(t_{i+1})\big) = 2\delta\frac{d}{dt}\eta(c_i) = \delta\eta^2(c_i) \le \delta \quad \text{for some } c_i \in [t_i, t_{i+1}]. \tag{30}$$

Here we used the facts that $\eta_{(N)}(t)$ and $\eta(t)$ are decreasing functions and $\eta(t) = 2/(2+t)$ is the solution to Equation (2). Thus with probability greater than $P(\bigcap_{i=0}^{M-1} A_i) \to 1$,

$$\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[0,K]} \le \left(\frac{15}{4}K^2+3K+3\right)/M + K/M = \frac{15}{4}K^2/M + 4K/M + 3/M \tag{31}$$

for $M$ large enough and $N \ge M^6$.

Therefore, letting $M \to \infty$, we have shown that

$$\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[0,K]} \to 0 \quad \text{in probability.}$$

• *Step VII.* Take $\epsilon \in (0,1)$ and $\gamma > 1$. Let $T_m$ be the time when the first $m = \lfloor (1-\epsilon)N \rfloor$ clusters merge. The expectation for the time $T_m$ is

$$E[T_m] = \frac{N}{\binom{N}{2}} + \frac{N}{\binom{N-1}{2}} + \dots + \frac{N}{\binom{N-m+1}{2}} = \frac{2m}{N-m}.$$

If we take $K > \frac{2(1-\epsilon)}{\epsilon}\gamma$, then $\eta(K) < \eta(\frac{2(1-\epsilon)}{\epsilon}\gamma) < \eta(2(1-\epsilon)/\epsilon) = \epsilon$, and for any $t \ge K$, $|\eta_{(N)}(t)-\eta(t)| > \epsilon$ implies $\eta_{(N)}(t) > \epsilon > \eta(t) > 0$. Thus, by Markov's inequality,

$$\begin{aligned}P\big(\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[K,\infty)} > \epsilon\big) &\le P\big(\eta_{(N)}(K) > \epsilon\big) = P(T_m > K) \\ &\le \frac{2(1-\epsilon)}{\epsilon K} < 1/\gamma.\end{aligned} \tag{32}$$

Now, we take $M > (\frac{15}{4}K^2 + 4K + 3)/\epsilon$. Then, by (31),

$$P\big(\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[0,K]} < \epsilon\big) \ge P\left(\bigcap_{i=0}^{M-1} A_i\right),$$

and

$$\begin{aligned}P\big(\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[0,\infty)} < \epsilon\big) \ \ge\ & P\big(\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[0,K]} < \epsilon\big) \\ &+ P\big(\left\|\eta_{(N)}(t)-\eta(t)\right\|_{L^\infty[K,\infty)} < \epsilon\big) - 1\end{aligned}$$

$$\geq\ P\left(\bigcap_{i=0}^{M-1} A_i\right) + (1-1/\gamma) - 1$$
$$\to 1 - 1/\gamma$$

as we let $M \to \infty$. Hence,

$$\limsup_{N\to\infty} P\left(\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,\infty)} < \epsilon\right) \geq 1 - 1/\gamma$$

for any given $\gamma > 1$. Thus

$$\lim_{N\to\infty} P\left(\left\|\eta_{(N)}(t) - \eta(t)\right\|_{L^\infty[0,\infty)} < \epsilon\right) = 1.$$

Therefore we have shown that $\|\eta_{(N)}(t) - \eta(t)\|_{L^\infty[0,\infty)} \to 0$ in probability. □

## Appendix B: Proof of Lemma 2

• *Step I.* We will use the setting from the proof of Lemma 1. Fix $K > 0$ and consider an interval $[0, K]$ partitioned into $M$ subintervals

$$[t_0, t_1], [t_1, t_2], \ldots, [t_{M-1}, t_M]$$

of equal length $\delta = K/M$, where $t_0 = 0$ and $t_M = K$. Let $\epsilon_0 = \eta(K)/2 = 1/(2+K)$.

Once again, let $\eta_{(N)}(t)$ denote the relative total number of clusters. For $i = 0, 1, \ldots, M-1$, the total number of coalescences within the interval $[t_i, t_{i+1}]$ equals $N[\eta_{(N)}(t_i) - \eta_{(N)}(t_{i+1})]$. Take $N > M^6$. The probability of the event $\bigcap_{i=0}^{M-1} A_i$, where $A_i$ was defined in (28), was bounded below in (29) as follows

$$P\left(\left|N\big[\eta_{(N)}(t_i) - \eta_{(N)}(t_{i+1})\big] - \delta N \frac{\eta_{(N)}^2(t_i)}{2}\right| \leq \delta^2 N + (1+\delta)N^{2/3}\ \forall i = 0, 1, \ldots, M-1\right)$$
$$= P\left(\bigcap_{i=0}^{M-1} A_i\right) \geq \left(1 - \exp\left\{-M + \frac{4K}{\epsilon_0^2 M}\right\}\right)^{2M} \to 1$$

as $M \to \infty$. Recall also that $P(\min_{t\in[0,K]} \eta_{(N)}(t) > \epsilon_0 | \bigcap_{i=0}^{M-1} A_i) = 1$.

Recall that $\eta_{k,N}(t)$ is the number of clusters corresponding to branches of Horton–Strahler order $k$ at time $t$ relative to the system size $N$, and $g_{k,N}(t) := \eta_{(N)}(t) - \sum_{j:j<k} \eta_{j,N}(t)$. Let

$$\mathcal{X}_N := \frac{1}{N}\ell_1(\mathbb{Z}_+) \cap \left\{x \in \ell_1(\mathbb{R}) : \|x\|_1 \leq 1\right\}$$
$$= \left\{x = (x_1, x_2, \ldots) : x_k \in \frac{1}{N}\mathbb{Z}_+\ \forall k,\ \text{and}\ \sum_k x_k \leq 1\right\}.$$

Here,

$$\bar{\eta}_N(t) := \big(\eta_{1,N}(t), \eta_{2,N}(t), \ldots\big) \in \mathcal{X}_N$$

and $\eta_{(N)}(t) = \sum_{k=0}^{\infty} \eta_{k,N}(t) = \|\bar{\eta}_N(t)\|_1$.

For each $m \geq 0$ we define events

$$B_{m,t_i} := \left\{\left|m - \delta N \frac{\eta_{(N)}^2(t_i)}{2}\right| \leq \delta^2 N + (1+\delta)N^{2/3}\right\} \tag{33}$$

and

$$D_{m,t_i} := \{N[\eta_{(N)}(t_i) - \eta_{(N)}(t_{i+1})] = m\}.$$

Now observe that for any integer $i \geq 0$, event $A_i$ defined in (28) can be expanded as follows

$$\begin{aligned}
A_i &= \left\{\Delta_\delta \eta_{(N)}(t_i) = -\frac{\eta_{(N)}^2(t_i)}{2} + \mathcal{E}'(t_i) \text{ and } |\mathcal{E}'(t_i)| \leq \delta + (\delta^{-1}+1)N^{-1/3}\right\} \\
&= \bigcup_{m\geq 0}\left\{N[\eta_{(N)}(t_i) - \eta_{(N)}(t_{i+1})] = m \text{ and } \left|m - \delta N \frac{\eta_{(N)}^2(t_i)}{2}\right| \leq \delta^2 N + (1+\delta)N^{2/3}\right\} \\
&= \bigcup_{m\geq 0}[B_{m,t_i} \cap D_{m,t_i}] \\
&= \bigcup_{x\in\mathcal{X}_N}\bigcup_{m\geq 0}[\{\bar{\eta}_N(t_i) = x\} \cap B_{m,t_i} \cap D_{m,t_i}].
\end{aligned} \tag{34}$$

For integer $i \geq 0$, consider all $x \in \mathcal{X}_N$ such that

$$P\left(\bar{\eta}_N(t_i) = x \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}\right) > 0. \tag{35}$$

Next, for each $x \in \mathcal{X}_N$ satisfying (35), consider all integer $m \geq 0$ such that

$$P(B_{m,t_i} | \bar{\eta}_N(t_i) = x) > 0. \tag{36}$$

Finally for each $x \in \mathcal{X}_N$ satisfying (35) and for each integer $m \geq 0$ satisfying (36), consider the event $D_{m,t_i}$. For any integer $k > 0$ we can represent the coalescences that involve the clusters of order $k$ within $[t_i, t_{i+1}]$ as

$$\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i) = \xi_1 + \xi_2 + \cdots + \xi_m,$$

where $\xi_1, \xi_2, \dots, \xi_m$ are random variables that correspond to the $m$ coalescences (of any Horton–Strahler order) within $[t_i, t_{i+1}]$ in the order of occurrence. Here, each $\xi_r$ can take values in $\frac{1}{N}\{-2, -1, 0, 1\}$; and their dependence on $k$ is omitted to simplify the notations. By construction, conditioning on $\bigcap_{i'=0}^{i-1} A_{i'}$, $\{\bar{\eta}_N(t_i) = x\}$, $B_{m,t_i}$, and $D_{m,t_i}$ for values $x$ and $m$ satisfying (35) and (36), the distribution of $\xi_r$ for $1 \leq r \leq m$ is completely determined by the history $\mathcal{T}_{r-1}$ of the preceding $r-1$ transitions. Also, we have the following bounds:

(1) A transition that decreases $\eta_{k,N}(t)$ by $2/N$ has probability

$$p_l(-2) \leq P\left(\xi_r = -2/N \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}, \mathcal{T}_{r-1}\right) \leq p_u(-2),$$

where

$$p_l(-2) := \begin{cases} \binom{Nx_k - 2m}{2} / \binom{N\|x\|_1}{2} & \text{if } Nx_k - 2m \geq 2, \\ 0 & \text{otherwise,} \end{cases}$$

and $p_u(-2) := \binom{Nx_k + m}{2} / \binom{N\|x\|_1 - m}{2}$.

(2) A transition that increases $\eta_{k,N}(t)$ by $1/N$ has probability

$$p_l(1) \leq P\left(\xi_r = 1/N \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}, \mathcal{T}_{r-1}\right) \leq p_u(1),$$

where

$$p_l(1) := \begin{cases} \binom{Nx_{k-1}-2m}{2} / \binom{N\|x\|_1}{2} & \text{if } Nx_{k-1} - 2m \geq 2, \\ 0 & \text{otherwise,} \end{cases}$$

and $p_u(1) := \binom{Nx_{k-1}+m}{2} / \binom{N\|x\|_1-m}{2}$ if $k > 1$, and if $k = 1$, we let $p_l(1) = p_u(1) = 0$.

(3) A transition that decreases $\eta_{k,N}(t)$ by $1/N$ has probability

$$p_l(-1) \leq P\left(\xi_r = -1/N \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}, \mathcal{T}_{r-1}\right) \leq p_u(-1),$$

where $p_l(-1) := \frac{\max\{(Nx_k - 2m), 0\}(N \sum_{j=k+1}^{\infty} x_j - m)}{\binom{N\|x\|_1}{2}}$ and $p_u(-1) := \frac{(Nx_k + m)(N \sum_{j=k+1}^{\infty} x_j + m)}{\binom{N\|x\|_1 - m}{2}}$.

Next, for $x$ and $m$ satisfying (35) and (36), define probabilities

$$p(-2) := x_k^2 / \|x\|_1^2, \qquad p(1) := \begin{cases} x_{k-1}^2 / \|x\|_1^2 & \text{if } k > 1, \\ 0 & \text{if } k = 1, \end{cases} \qquad p(-1) := \frac{2x_k \sum_{j=k+1}^{\infty} x_j}{\|x\|_1^2},$$

and $p(0) := 1 - p(-2) - p(-1) - p(1)$. Let $\xi$ be a random variable with the values $\{-2, -1, 0, 1\}$ specified by the probabilities $\{p(-2), p(-1), p(0), p(1)\}$. Also let $\xi^+ = \xi \cdot \mathbf{1}_{\xi > 0}$ and $\xi^- = \xi \cdot \mathbf{1}_{\xi < 0}$.

Observe that we have conditioned on a sub-event of $\bigcap_{i'=0}^{i} A_{i'}$. Indeed, by (34)

$$\bigcap_{i'=0}^{i-1} A_{i'} \cap B_{m,t_i} \cap D_{m,t_i} \subseteq \bigcap_{i'=0}^{i} A_{i'}.$$

Here, since $P(\eta_{(N)}(t_i) \in [\epsilon_0, 1] | \bigcap_{i'=0}^{i-1} A_{i'}) = 1$ and $x$ satisfies (35),

$$\eta_{(N)}(t_i) = \|x\|_1 \geq \epsilon_0,$$

and therefore

$$p_l(-2) = p(-2) + \mathcal{O}(\delta) \quad \text{and} \quad p_u(-2) = p(-2) + \mathcal{O}(\delta),$$
$$p_l(1) = p(1) + \mathcal{O}(\delta) \quad \text{and} \quad p_u(1) = p(1) + \mathcal{O}(\delta),$$
$$p_l(-1) = p(-1) + \mathcal{O}(\delta) \quad \text{and} \quad p_u(-1) = p(-1) + \mathcal{O}(\delta).$$

Indeed, since the values $x$ and $m$ satisfy (35) and (36),

$$\left| m - \delta N \frac{\|x\|_1^2}{2} \right| \leq \delta^2 N + (1 + \delta) N^{2/3}$$

as in (33). Hence, $m = \mathcal{O}(\delta N)$. Therefore, conditioning on $\bigcap_{i'=0}^{i-1} A_{i'}$, $\{\bar{\eta}_N(t_i) = x\}$, $B_{m,t_i}$, and $D_{m,t_i}$ for values $x$ and $m$ satisfying (35) and (36), we have

$$p_l(-2) - p(-2) = \frac{\binom{Nx_k - 2m}{2}}{\binom{N\|x\|_1}{2}} - \frac{x_k^2}{\|x\|_1^2}$$
$$= \frac{-4\|x\|_1 x_k \frac{m}{N} + 4\|x\|_1 \frac{m^2}{N^2} + 2\|x\|_1 \frac{m}{N^2} + (x_k - \|x\|_1) \frac{x_k}{N}}{\|x\|_1^2 (\|x\|_1 - 1/N)} = \mathcal{O}(\delta)$$

when $Nx_k - 2m \geq 2$,

$$p_u(-2) - p(-2) = \frac{\binom{Nx_k+m}{2}}{\binom{N\|x\|_1-m}{2}} - \frac{x_k^2}{\|x\|_1^2}$$
$$= \frac{2(x_k^2\|x\|_1 + x_k\|x\|_1^2)\frac{m}{N} + \frac{x_k^2\|x\|_1 - x_k\|x\|_1^2}{N} + \frac{(\|x\|_1^2 - x_k^2)m^2}{N^2} - \frac{(\|x\|_1^2 + x_k^2)m}{N^2}}{(\|x\|_1 - \frac{m}{N})(\|x\|_1 - \frac{m+1}{N})\|x\|_1^2} = \mathcal{O}(\delta),$$

$$p_l(1) - p(1) = \frac{\binom{Nx_{k-1}-2m}{2}}{\binom{N\|x\|_1}{2}} - \frac{x_{k-1}^2}{\|x\|_1^2}$$
$$= \frac{-4\|x\|_1 x_{k-1}\frac{m}{N} + 4\|x\|_1\frac{m^2}{N^2} + 2\|x\|_1\frac{m}{N^2} + (x_{k-1} - \|x\|_1)\frac{x_{k-1}}{N}}{\|x\|_1^2(\|x\|_1 - 1/N)} = \mathcal{O}(\delta)$$

when $Nx_{k-1} - 2m \geq 2$,

$$p_u(1) - p(1) = \frac{\binom{Nx_{k-1}+m}{2}}{\binom{N\|x\|_1-m}{2}} - \frac{x_{k-1}^2}{\|x\|_1^2}$$
$$= \frac{2(x_{k-1}^2\|x\|_1 + x_{k-1}\|x\|_1^2)\frac{m}{N} + \frac{x_{k-1}^2\|x\|_1 - x_{k-1}\|x\|_1^2}{N} + \frac{(\|x\|_1^2 - x_{k-1}^2)m^2}{N^2} - \frac{(\|x\|_1^2 + x_{k-1}^2)m}{N^2}}{(\|x\|_1 - \frac{m}{N})(\|x\|_1 - \frac{m+1}{N})\|x\|_1^2} = \mathcal{O}(\delta),$$

$$p_l(-1) - p(-1) = \frac{(Nx_k - 2m)(N\sum_{j=k+1}^\infty x_j - m)}{\binom{N\|x\|_1}{2}} - \frac{2x_k\sum_{j=k+1}^\infty x_j}{\|x\|_1^2}$$
$$= \frac{-2\|x\|_1(x_k + 2\sum_{j=k+1}^\infty x_j)\frac{m}{N} + 2\frac{x_k}{N}\sum_{j=k+1}^\infty x_j + 4\|x\|_1\frac{m^2}{N^2}}{(\|x\|_1 - \frac{1}{N})\|x\|_1^2} = \mathcal{O}(\delta)$$

when $Nx_k - 2m \geq 0$,

$$p_u(-1) - p(-1) = \frac{(Nx_k + m)(N\sum_{j=k+1}^\infty x_j + m)}{\binom{N\|x\|_1-m}{2}} - \frac{2x_k\sum_{j=k+1}^\infty x_j}{\|x\|_1^2}$$
$$= \frac{2\|x\|_1(2x_k\sum_{j=k+1}^\infty x_j + x_k\|x\|_1 + \|x\|_1\sum_{j=k+1}^\infty x_j)\frac{m}{N}}{(\|x\|_1 - \frac{m}{N})(\|x\|_1 - \frac{m+1}{N})\|x\|_1^2}$$
$$+ \frac{2x_k\|x\|_1\sum_{j=k+1}^\infty x_j\frac{1}{N} + 2(\|x\|_1 - x_k\sum_{j=k+1}^\infty x_j)\frac{m^2}{N^2} - 2x_k\sum_{j=k+1}^\infty x_j\frac{m}{N^2}}{(\|x\|_1 - \frac{m}{N})(\|x\|_1 - \frac{m+1}{N})\|x\|_1^2}$$
$$= \mathcal{O}(\delta).$$

Finally, if $Nx_k - 2m < 2$, then $\frac{x_k^2}{\|x\|_1^2} < \frac{2(m+1)}{N\|x\|_1^2} \leq \frac{2(m+1)}{N\epsilon_0^2} = \mathcal{O}(\delta)$, and if $Nx_k - 2m < 0$, then $\frac{2x_k\sum_{j=k+1}^\infty x_j}{\|x\|_1^2} \leq \frac{4m}{N\|x\|_1} \leq \frac{4m}{N\epsilon_0} = \mathcal{O}(\delta)$.

Next, let $\xi_r^+ = \xi_r \cdot \mathbf{1}_{\xi_r > 0}$ and $\xi_r^- = \xi_r \cdot \mathbf{1}_{\xi_r < 0}$. Then

$$\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i) = X_+ + X_-,$$

where

$$X_+ = \xi_1^+ + \xi_2^+ + \cdots + \xi_m^+$$

and

$$X_- = \xi_1^- + \xi_2^- + \cdots + \xi_m^-.$$

Next, for any $\lambda^+, \lambda^- \geq 0$ and $s \in [0,1]$ consider

$$\begin{aligned}
&E\left[e^{sN[\lambda^+ X_+ + \lambda^- X_-]} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}\right] \\
&\quad = \prod_{r=1}^{m} E\left[e^{sN[\lambda^+ \xi_r^+ + \lambda^- \xi_r^-]} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}, \mathcal{T}_{r-1}\right],
\end{aligned}$$

where for all $r$,

$$\begin{aligned}
&E\left[e^{sN[\lambda^+ \xi_r^+ + \lambda^- \xi_r^-]} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}, \mathcal{T}_{r-1}\right] \\
&\quad \leq e^{-2\lambda^- s} p_u(-2) + e^{-\lambda^- s} p_u(-1) + e^{\lambda^+ s} p_u(1) + \big(1 - p_l(-2) - p_l(-1) - p_l(1)\big) \\
&\quad \leq e^{-2\lambda^- s} p(-2) + e^{-\lambda^- s} p(-1) + e^{\lambda^+ s} p(1) + p(0) + C\delta \\
&\quad = E\big[e^{s[\lambda^+ \xi^+ + \lambda^- \xi^-]}\big] + C\delta
\end{aligned}$$

for a large enough $C > 0$. Hence,

$$E\left[e^{sN[\lambda^+ X_+ + \lambda^- X_-]} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}\right] \leq \big(E\big[e^{s[\lambda^+ \xi^+ + \lambda^- \xi^-]}\big] + C\delta\big)^m.$$

Therefore, by the exponential Markov inequality with the exponent $s$, for all $x$ and $m$ satisfying (35) and (36),

$$\begin{aligned}
&P\left(N\big[\lambda^+ X_+ + \lambda^- X_-\big] \geq E\big[\lambda^+ \xi^+ + \lambda^- \xi^-\big]m + m^{14/15} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}\right) \\
&\quad \leq E\left[e^{sN[\lambda^+ X_+ + \lambda^- X_-]} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}\right] e^{-s(E[\lambda^+ \xi^+ + \lambda^- \xi^-]m + m^{14/15})} \\
&\quad \leq \big(E\big[e^{s[\lambda^+ \xi^+ + \lambda^- \xi^-]}\big] + C\delta\big)^m e^{-s(E[\lambda^+ \xi^+ + \lambda^- \xi^-]m + m^{14/15})} \\
&\quad = \big(E\big[e^{s(\lambda^+[\xi^+ - E[\xi^+]] + \lambda^-[\xi^- - E[\xi^-]])}\big] + e^{-sE[\lambda^+ \xi^+ + \lambda^- \xi^-]} C\delta\big)^m e^{-sm^{14/15}} \\
&\quad = \big(1 + E\big[s\big(\lambda^+\big[\xi^+ - E\big[\xi^+\big]\big] + \lambda^-\big[\xi^- - E\big[\xi^-\big]\big]\big)\big] + C\delta + \mathcal{O}\big(s^2 + s\delta\big)\big)^m e^{-sm^{14/15}} \\
&\quad = \big(1 + C\delta + \mathcal{O}\big(s^2 + s\delta\big)\big)^m e^{-sm^{14/15}} \\
&\quad \leq \exp\big\{m\big[C\delta + \mathcal{O}\big(s^2 + s\delta\big)\big] - sm^{14/15}\big\} \quad \text{as } s, \delta \to 0.
\end{aligned}$$

Next, taking $2M^6 > N > M^6$ and $M$ large enough, and plugging $s = 2C\delta m^{1/15} = \mathcal{O}(M^{-2/3})$ (as $M \to \infty$) into the above exponential Markov inequality, we obtain

$$\begin{aligned}
&P\left(N\big[\lambda^+ X_+ + \lambda^- X_-\big] \geq E\big[\lambda^+ \xi^+ + \lambda^- \xi^-\big]m + m^{14/15} \,\middle|\, \bigcap_{i'=0}^{i-1} A_{i'}, \{\bar{\eta}_N(t_i) = x\}, B_{m,t_i}, D_{m,t_i}\right) \\
&\quad \leq \exp\big\{-C\delta m + \mathcal{O}\big(M^{11/3}\big)\big\} \\
&\quad \leq \exp\big\{-AM^4\big\} \qquad (37)
\end{aligned}$$

for sufficiently small positive $A < CK^2\epsilon_0^2/2 \le CK^2\eta_{(N)}^2(t_i)/2$ and sufficiently large $M$ as we conditioned on $B_{m,t_i}$, e.g. let $A = CK^2\epsilon_0^2/10$.

The exponential in $M^4$ lower bound on

$$P\left(N\left[\lambda^+ X_+ + \lambda^- X_-\right] \le E\left[\lambda^+\xi^+ + \lambda^-\xi^-\right]m - m^{14/15} \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \left\{\bar{\eta}_N(t_i) = x\right\}, B_{m,t_i}, D_{m,t_i}\right)$$

follows via a symmetrical argument. Specifically, for $C > 0$ large enough, and all $s \in [0, 1]$,

$$E\left[e^{-sN[\lambda^+ X_+ + \lambda^- X_-]} \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \left\{\bar{\eta}_N(t_i) = x\right\}, B_{m,t_i}, D_{m,t_i}\right] \le \left(E\left[e^{-s[\lambda^+\xi^+ + \lambda^-\xi^-]}\right] + C\delta\right)^m.$$

Therefore, taking $s = 2C\delta m^{1/15} = \mathcal{O}(M^{-2/3})$, we obtain

$$\begin{aligned}
&P\left(N\left[\lambda^+ X_+ + \lambda^- X_-\right] \le E\left[\lambda^+\xi^+ + \lambda^-\xi^-\right]m - m^{14/15} \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \left\{\bar{\eta}_N(t_i) = x\right\}, B_{m,t_i}, D_{m,t_i}\right) \\
&\quad \le E\left[e^{-sN[\lambda^+ X_+ + \lambda^- X_-]} \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \left\{\bar{\eta}_N(t_i) = x\right\}, B_{m,t_i}, D_{m,t_i}\right] e^{s(E[\lambda^+\xi^+ + \lambda^-\xi^-]m - m^{14/15})} \\
&\quad \le \left(E\left[e^{-s[\lambda^+\xi^+ + \lambda^-\xi^-]}\right] + C\delta\right)^m e^{s(E[\lambda^+\xi^+ + \lambda^-\xi^-]m - m^{14/15})} \\
&\quad = \left(1 + C\delta + \mathcal{O}\left(s^2 + s\delta\right)\right)^m e^{-sm^{14/15}} \\
&\quad \le \exp\left\{m\left(C\delta + \mathcal{O}\left(s^2 + s\delta\right)\right) - sm^{14/15}\right\} \\
&\quad \le \exp\left\{-C\delta m + \mathcal{O}\left(M^{11/3}\right)\right\} \\
&\quad \le \exp\left\{-AM^4\right\}
\end{aligned} \tag{38}$$

for sufficiently small positive $A < CK^2\epsilon_0^2/2 \le CK^2\eta_{(N)}^2(t_i)/2$ and sufficiently large $M$.

Thus, plugging $\lambda^+ = \lambda^- = 1$ into (37) and (38), we obtain the following inequality. For each $k$ and $M$ large enough, there exists $a > 0$ such that

$$\begin{aligned}
&P\left(\left|\left(\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i)\right) - E[\xi]m/N\right| < m^{14/15}/N \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \left\{\bar{\eta}_N(t_i) = x\right\}, B_{m,t_i}, D_{m,t_i}\right) \\
&\quad \ge 1 - \exp\left\{-aM^4\right\}
\end{aligned}$$

for all $i = 0, 1, \ldots, M-1$ and all $x$ and $m$ satisfying (35) and (36).

Now, (34) implies for any event $F$,

$$\begin{aligned}
P\left(F \Big| \bigcap_{i'=0}^{i} A_{i'}\right) = \sum_{x,m} &P\left(F \Big| \bigcap_{i'=0}^{i-1} A_{i'}, \left\{\bar{\eta}_N(t_i) = x\right\}, B_{m,t_i}, D_{m,t_i}\right) \\
&\times P\left(\left\{\bar{\eta}_N(t_i) = x\right\} \cap B_{m,t_i} \cap D_{m,t_i} \Big| \bigcap_{i'=0}^{i} A_{i'}\right),
\end{aligned}$$

where, by (34),

$$\sum_{m,x} P\left(\left\{\bar{\eta}_N(t_i) = x\right\} \cap B_{m,t_i} \cap D_{m,t_i} \Big| \bigcap_{i'=0}^{i} A_{i'}\right) = 1.$$

Therefore, since here $m^{14/15}/N = \mathcal{O}(M^{-4/3})$, $\delta^2 = \mathcal{O}(M^{-2})$, and $(1+\delta)N^{-1/3} = \mathcal{O}(M^{-2})$, there is a large enough $c_k > 0$ such that

$$\begin{aligned}
&P\left(\left|\left[\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i)\right] - E[\xi]\delta\frac{\eta^2_{(N)}(t_i)}{2}\right| < c_k\delta^{4/3} \,\Big|\, \bigcap_{i'=0}^{i} A_{i'}\right) \\
&= \sum_{x,m} P\left(\left|\left[\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i)\right] - E[\xi]\delta\frac{\eta^2_{(N)}(t_i)}{2}\right| < c_k\delta^{4/3} \,\Big|\, \bigcap_{i'=0}^{i-1} A_{i'} \cap \left\{\bar\eta_N(t_i) = x\right\} \cap B_{m,t_i} \cap D_{m,t_i}\right) \\
&\quad \times P\left(\left\{\bar\eta_N(t_i) = x\right\} \cap B_{m,t_i} \cap D_{m,t_i} \,\Big|\, \bigcap_{i'=0}^{i} A_{i'}\right) \\
&\ge \sum_{x,m} P\left(\left|\left[\eta_{k,N}(t_{i+1}) - \eta_{k,N}(t_i)\right] - E[\xi]m/N\right| < m^{14/15}/N \,\Big|\, \bigcap_{i'=0}^{i-1} A_{i'} \cap \left\{\bar\eta_N(t_i) = x\right\} \cap B_{m,t_i} \cap D_{m,t_i}\right) \\
&\quad \times P\left(\left\{\bar\eta_N(t_i) = x\right\} \cap B_{m,t_i} \cap D_{m,t_i} \,\Big|\, \bigcap_{i'=0}^{i} A_{i'}\right) \\
&\ge 1 - \exp\left\{-aM^4\right\}
\end{aligned} \tag{39}$$

for all $i = 0, 1, \dots, M-1$, $2M^6 > N > M^6$, and $M$ large enough, as

$$\bigcap_{i'=0}^{i-1} A_{i'} \cap B_{m,t_i} \cap D_{m,t_i} \subseteq \bigcap_{i'=0}^{i} A_{i'}.$$

- *Step II.* Observe that as we condition on $\{\bar\eta_N(t_i) = x\} \cap B_{m,t_i} \cap D_{m,t_i}$ in (39),

$$\begin{aligned}
E[\xi] &= -2\frac{\eta^2_{k,N}(t_i)}{\eta^2_{(N)}(t_i)} + \frac{\eta^2_{k-1,N}(t_i)}{\eta^2_{(N)}(t_i)}(1-\delta_{1,k}) - \frac{2\eta_{k,N}(t_i)g_{k+1,N}(t_i)}{\eta^2_{(N)}(t_i)} \\
&= \frac{\eta^2_{k-1,N}(t_i)}{\eta^2_{(N)}(t_i)}(1-\delta_{1,k}) - 2\frac{\eta_{k,N}(t_i)g_{k,N}(t_i)}{\eta^2_{(N)}(t_i)}
\end{aligned}$$

and

$$\left|\frac{m}{N} - \delta\frac{\eta^2_{(N)}(t_i)}{2}\right| \le \delta^2 + (1+\delta)N^{-1/3}.$$

Thus, (39) will imply the following system of difference equations with the initial conditions and the error bound as mentioned below.

$$\begin{aligned}
\Delta_\delta\eta_{1,N}(t_i) &= -\eta_{1,N}(t_i)\eta_{(N)}(t_i) + \mathcal{E}'_1(t_i), \\
\Delta_\delta\eta_{k,N}(t_i) &= \frac{\eta^2_{k-1,N}(t_i)}{2} - \eta_{k,N}(t_i)g_{k,N}(t_i) + \mathcal{E}'_k(t_i) \quad \text{for } k \ge 2
\end{aligned} \tag{40}$$

with the initial conditions

$$\left(\eta_{1,N}(0), \eta_{2,N}(0), \dots, \eta_{k,N}(0), \dots\right) = (1, 0, 0, \dots),$$

where for a given $\rho \in \mathbb{N}$ and $c = \max_{1\le k\le\rho}\{c_k\}$ we have $|\mathcal{E}'_k(t_i)| < c\delta^{1/3}$ for each $1 \le k \le \rho$. Here, for each $k$, the $k$th equation holds with probability of at least

$$\begin{aligned}
1 &- \sum_{i=0}^{M}\left[1 - P\left(\bigcap_{i'=0}^{i} A_{i'}\right)\cdot\left(1-\exp\{-aM^{2/3}\}\right)\right] \\
&\ge 1 - M\left[1 - P\left(\bigcap_{i'=0}^{M-1} A_{i'}\right)\cdot\left(1-\exp\{-aM^{2/3}\}\right)\right] \\
&\ge 1 - M\left[1 - \left(1-\exp\left\{-M+\frac{4K}{\epsilon_0^2 M}\right\}\right)^{2M}\left(1-\exp\{-aM^{2/3}\}\right)\right] \\
&\ge 1 + M\exp\{-aM^{2/3}\} - M\left[1 - \left(1-\exp\left\{-M+\frac{4K}{\epsilon_0^2 M}\right\}\right)^{2M}\right] \\
&\to 1 \quad \text{as } M\to\infty.
\end{aligned}$$

Finally, the same error propagation analysis as in Step IV in the proof of Lemma 1 is applied to compare the above difference equations (40) to the difference equations that correspond to the following system of ODEs

$$\begin{aligned}
&\frac{d}{dt}\eta_1(t) = -\eta_1(t)\eta(t), \\
&\frac{d}{dt}\eta_k(t) = \frac{\eta_{k-1}^2(t)}{2} - \eta_k(t)g_k(t) \quad \text{for } k\ge 2
\end{aligned}$$

with the initial conditions

$$\left(\eta_1(0),\eta_2(0),\dots,\eta_k(0),\dots\right) = (1,0,0,\dots),$$

where $g_k(t) := \eta(t) - \sum_{i:i<k}\eta_i(t)$. The above system of ODEs can be converted into the following system of difference equations

$$\begin{aligned}
&\Delta_\delta\eta_1(t_i) = -\eta_1(t_i)\eta(t_i) + \mathcal{E}_1(t_i), \\
&\Delta_\delta\eta_k(t_i) = \frac{\eta_{k-1}^2(t_i)}{2} - \eta_k(t_i)g_k(t_i) + \mathcal{E}_k(t_i) \quad \text{for } k\ge 2
\end{aligned} \tag{41}$$

with the error

$$\mathcal{E}_k(t_i) = \frac{\eta_k''(c_{i,k})}{2}\delta \quad \text{for some } c_{i,k}\in(t_i,t_{i+1}).$$

Here $|\mathcal{E}_1(t_i)| = \frac{|\eta_1''(c_{i,1})|}{2}\delta < \frac{3}{4}\delta$ as $\eta_1''(t) = -[\eta_1(t)\eta(t)]' = \frac{3}{2}\eta_1(t)\eta^2(t)$.
The error for $k>1$ is

$$\left|\mathcal{E}_k(t_i)\right| = \frac{|\eta_k''(c_{i,k})|}{2}\delta \le \frac{k+2}{2}\delta$$

as

$$\begin{aligned}
\eta_k''(t) &= \left[\frac{\eta_{k-1}^2(t)}{2} - \eta_k(t)g_k(t)\right]' \\
&= \eta_{k-1}(t)\eta_{k-1}'(t) - \eta_k'(t)g_k(t) - \eta_k(t)g_k'(t)
\end{aligned}$$

$$= \eta_{k-1}(t)\left(\frac{\eta_{k-2}^2(t)}{2} - \eta_{k-1}(t)g_{k-1}(t)\right) - \left(\frac{\eta_{k-1}^2(t)}{2} - \eta_k(t)g_k(t)\right)g_k(t)$$
$$- \eta_k(t)\left(-\frac{\eta_k^2(t)}{2} - \eta_1(t)\eta(t) + \sum_{i:2\le i<k}\left[\frac{\eta_{i-1}^2(t)}{2} - \eta_i(t)g_i(t)\right]\right)$$

and for each $i$, $|\eta_i(t)| \le 1$ and $|g_i(t)| \le 1$.

• *Step III.* Next, we notice that the error propagates as in (27), iteratively producing for each $k \in \mathbb{N}^+$,

$$\varepsilon_{k,i} := \eta_{k,N}(t_i) - \eta_k(t_i) = \mathcal{O}(M^{-1}).$$

Indeed, if $\varepsilon_i = \eta_{(N)}(t_i) - \eta(t_i)$, then conditioning on the event $\bigcap_{i=0}^{M-1} A_i$, the approximation error $\varepsilon_i$ was shown to satisfy $|\varepsilon_i| \le iKC_K/M^2$.

Let $d_1 := \frac{3}{4}$, and for $k > 1$, $d_k := \frac{k+2}{2}$. Then $|\mathcal{E}_k(t_i)| \le d_k\delta$. Next let $\varepsilon_{0,i} := 0$ for all $i$. Also, we observe that $\varepsilon_{k,0} = 0$ for all $k \ge 0$ because of the same initial conditions in systems (40) and (41).

From the difference equations (40) and (41), we have the error propagating as follows

$$\varepsilon_{k,i+1} = \varepsilon_{k,i} + \delta\left(\frac{\eta_{k-1,N}^2(t_i)}{2} - \frac{\eta_{k-1}^2(t_i)}{2}\right) - \delta\big(\eta_{k,N}(t_i)g_{k,N}(t_i) - \eta_k(t_i)g_k(t_i)\big)$$
$$+ \delta\big(\mathcal{E}_k'(t_i) - \mathcal{E}_k(t_i)\big)$$
$$= \varepsilon_{k,i} + \delta\left(\eta_{k-1}(t_i)\varepsilon_{k-1,i} + \frac{\varepsilon_{k-1,i}^2}{2}\right)$$
$$- \delta\left(\big(\eta_k(t_i) + \varepsilon_{k,i}\big)\left[\varepsilon_i - \sum_{k'=1}^{k-1}\varepsilon_{k',i}\right] + g_k(t_i)\varepsilon_{k,i}\right)$$
$$+ \delta\big(\mathcal{E}_k'(t_i) - \mathcal{E}_k(t_i)\big)$$

and therefore

$$|\varepsilon_{k,i+1}| \le |\varepsilon_{k,i}| + \delta|\varepsilon_{k-1,i}| + \delta\frac{\varepsilon_{k-1,i}^2}{2} + \delta\left[|\varepsilon_i| + \sum_{k'=1}^{k}|\varepsilon_{k',i}|\right]$$
$$+ \delta\varepsilon_{k,i}\left[|\varepsilon_i| + \sum_{k'=1}^{k-1}|\varepsilon_{k',i}|\right] + c\delta^{4/3} + \delta^2 d_k. \tag{42}$$

The inequality (42) is crucial for proving the following statement by induction. We claim that for each integer $\rho > 0$ and $M$ large enough,

$$|\varepsilon_{k,i}| \le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^i - 1\big] \quad \text{for all } k \in \{1,\dots,\rho\} \text{ and } i = 0,1,\dots,M-1.$$

The basis step follows from the initial conditions $\varepsilon_{0,i} = 0$ and $\varepsilon_{k,0} = 0$. The inductive step is obtained from (42) as follows. Suppose for a choice of $k \in \{1,\dots,\rho\}$ and $i$,

$$|\varepsilon_{k',j}| \le (c+1)2^{k'}\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^j - 1\big]$$

for all $j = 0,1,\dots,M-1$ whenever $k' < k$, and

$$|\varepsilon_{k,j}| \le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^j - 1\big]$$

whenever $j \le i$.

Observe that

$$\delta\sum_{k'=1}^{k}|\varepsilon_{k',i}| \le \delta\rho(c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^i-1\big]$$

and hence

$$\begin{aligned}|\varepsilon_{k,i}|+\delta|\varepsilon_{k-1,i}|+\delta\sum_{k'=1}^{k}|\varepsilon_{k',i}|+c\delta^{4/3} &\le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+\delta/2+\delta\rho)(1+2\delta\rho)^i-1\big]-C_1\delta^{4/3}\\ &\le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^{i+1}-1\big]-C_1\delta^{4/3},\end{aligned}$$

with $C_1=(c+1)2^{k-1}\rho^{-1}+(c+1)2^k-c>0$. At the same time, all other terms in (42) are estimated from above by functions that have higher powers of $\delta$:

$$\begin{aligned}&\delta\frac{\varepsilon_{k-1,i}^2}{2}\le (c+1)^2 2^{2k-3}\frac{\delta^{5/3}}{\rho^2}\big[e^{2K\rho}-1\big]^2,\\ &\delta|\varepsilon_i|\le\delta^2 C_K,\\ &\delta\varepsilon_{k,i}|\varepsilon_i|\le C_K(c+1)2^k\frac{\delta^{7/3}}{\rho}\big[e^{2K\rho}-1\big],\\ &\delta\varepsilon_{k,i}\sum_{k'=1}^{k-1}|\varepsilon_{k',i}|\le (c+1)^2 2^{2k}\frac{\delta^{5/3}}{\rho}\big[e^{2K\rho}-1\big]^2,\end{aligned}$$

where we used the observation $(1+2\delta\rho)^i\le(1+2\delta\rho)^M\le e^{2K\rho}$. This implies that

$$|\varepsilon_{k,i+1}|\le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^{i+1}-1\big]$$

for $M$ large enough, and therefore $\delta$ small enough, thus proving the claim. Hence

$$|\varepsilon_{k,i}|\le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[(1+2\delta\rho)^i-1\big]\le (c+1)2^k\frac{\delta^{1/3}}{\rho}\big[e^{2K\rho}-1\big]=\mathcal{O}\big(\delta^{1/3}\big)$$

for any $\rho$ and all $k\in\{1,\dots,\rho\}$.

Therefore, conditioning on the event $\bigcap_{i=0}^{M-1}A_i$, we have the following upper bound for any $k\in\{1,\dots,\rho\}$ and for all $i\in\{0,1,\dots,M-1\}$. If $t\in(t_i,t_{i+1})$, then

$$\begin{aligned}\big|\eta_{k,N}(t)-\eta_k(t)\big| &\le \big|\eta_{k,N}(t)-\eta_{k,N}(t_i)\big|+\big|\eta_{k,N}(t_i)-\eta_k(t_i)\big|+\big|\eta_k(t_i)-\eta_k(t)\big|\\ &\le 2\big(\eta_{(N)}(t_i)-\eta_{(N)}(t)\big)+(c+1)2^k\frac{\delta^{1/3}}{\rho}\big[e^{2K\rho}-1\big]+\big|\eta_k(t_i)-\eta_k(t)\big|\\ &\le 2\big(\eta_{(N)}(t_i)-\eta_{(N)}(t_{i+1})\big)+(c+1)2^k\frac{\delta^{1/3}}{\rho}\big[e^{2K\rho}-1\big]+\big|\eta_k(t_i)-\eta_k(t)\big|\\ &= 2\big(\eta_{(N)}(t_i)-\eta(t_i)\big)+2\big(\eta(t_{i+1})-\eta_{(N)}(t_{i+1})\big)+2\big(\eta(t_i)-\eta(t_{i+1})\big)\\ &\quad+(c+1)2^k\frac{\delta^{1/3}}{\rho}\big[e^{2K\rho}-1\big]+\big|\eta_k(t_i)-\eta_k(t)\big|\\ &\le \big(5K^2+4K+4\big)/M+(c+1)2^k\frac{\delta^{1/3}}{\rho}\big[e^{2K\rho}-1\big]+3\delta\end{aligned}$$

as the net change $|\eta_{k,N}(t) - \eta_{k,N}(t_i)|$ in the number of clusters of order $k$ is dominated by twice the net change $\eta_{(N)}(t_i) - \eta_{(N)}(t)$ in the total number of clusters. We also used

$$\eta_{(N)}(t_{i'}) - \eta(t_{i'}) \le \left(\frac{5}{4}K^2 + K + 1\right)/M \quad \text{for all } i' \in \{0, 1, \dots, M\}$$

shown in (27),

$$2\big(\eta(t_i) - \eta(t_{i+1})\big) \le \delta$$

shown in (30), and that there exists $c_i' \in (t_i, t_{i+1})$ such that

$$\big|\eta_k(t_i) - \eta_k(t)\big| = (t - t_i)\left|\frac{d}{dt}\eta_k\big(c_i'\big)\right| = (t - t_i)\left|\frac{\eta_{k-1}^2(c_i')}{2} - \eta_k\big(c_i'\big)g_k\big(c_i'\big)\right| \le 2\delta.$$

Thus, for any $k$,

$$\|\eta_{k,N} - \eta_k\|_{L^\infty[0,K]} \to 0 \quad \text{in probability.}$$

- *Step IV.* Finally, observe that for any $\epsilon > 0$ and for $K > 2$ large enough so that $\eta(K) < \epsilon$,

$$\eta_k(t) \le \eta(t) \le \eta(K) < \epsilon \quad \text{for all } t \ge K$$

and, by (32),

$$\begin{aligned} P\big(\big\|\eta_{k,N}(t) - \eta_k(t)\big\|_{L^\infty[K,\infty)} > \epsilon\big) &\le P\big(\big\|\eta_{k,N}(t)\big\|_{L^\infty[K,\infty)} > \epsilon\big) \\ &\le P\big(\big\|\eta_{(N)}(t)\big\|_{L^\infty[K,\infty)} > \epsilon\big) \\ &= P\big(\eta_{(N)}(K) > \epsilon\big) \\ &\le \frac{2(1-\epsilon)}{\epsilon K}. \end{aligned}$$

Thus, together with the previous step, we have shown that for each $k$,

$$\|\eta_{k,N} - \eta_k\|_{L^\infty[0,\infty)} \to 0$$

in probability. □

## Appendix C: Proof of Lemma 3

Observe that when we plug in $\lambda^+ = 1$ and $\lambda^- = 0$ into (37) and (38), we obtain that in the difference equations (40), the number of emerging clusters of Horton–Strahler order $j$ within the time interval $[t_i, t_{i+1}]$ divided by $N$ is

$$\frac{p(1)m + \mathcal{O}(m^{14/15})}{N} = \frac{\eta_{j-1,N}^2(t_i)}{2} \cdot \delta + \mathcal{O}\big(\delta^{4/3}\big)$$

for all $i = 0, 1, \dots, M-1$, $\delta = K/M$, and $m$ satisfying

$$\left|m - \delta N\frac{\eta_{(N)}^2(t_i)}{2}\right| \le \delta^2 N + (1+\delta)N^{2/3},$$

with probability approaching 1 exponentially fast as $2M^6 > N > M^6 \to \infty$. Here

$$\sum_{i=0}^{\frac{K}{\delta}-1} \frac{\eta_{j-1,N}^2(t_i)}{2} \cdot \delta$$

converges almost surely to $\int_0^K \frac{\eta_{j-1,N}^2(t)}{2}\,dt$ as $\delta \to 0$.

Hence, for $j \geq 2$, the total number $N_j(K)$ of emerging clusters of Horton–Strahler order $j$ within the time interval $[0, K]$ divided by $N$ is

$$N_j(K)/N = \int_0^K \frac{\eta_{j-1,N}^2(t)}{2}\,dt + \mathcal{O}(\delta^{1/3})$$

with probability approaching 1 as $M \to \infty$.

Fix $\varepsilon > 0$. We established that $\|\eta_{j,N} - \eta_j\|_{L^\infty[0,K]} \to 0$ in probability. Then

$$\left| \int_0^K \frac{\eta_{j-1}^2(t)}{2}\,dt - \int_0^K \frac{\eta_{j-1,N}^2(t)}{2}\,dt \right| \leq \frac{K}{2} \|\eta_{j-1} + \eta_{j-1,N}\|_{L^\infty[0,K]} \cdot \|\eta_{j-1} - \eta_{j-1,N}\|_{L^\infty[0,K]} \to 0.$$

Thus, $|N_j(K)/N - \int_0^K \frac{\eta_{j-1}^2(t)}{2}\,dt| < \varepsilon$ with probability $\mathcal{P}_{K,\varepsilon,N} \to 1$ as $N \to \infty$.

Now, for $K > 2(1-\varepsilon)/\varepsilon$,

$$\int_K^\infty \frac{\eta_{j-1}^2(t)}{2}\,dt \leq \int_K^\infty \frac{\eta^2(t)}{2}\,dt = \int_K^\infty \frac{2}{(t+2)^2}\,dt = \frac{2}{K+2} < \varepsilon$$

and

$$P\big(\eta_{(N)}(K) < \varepsilon\big) \geq 1 - \frac{2(1-\varepsilon)}{\varepsilon K}.$$

Therefore, the total number of emerging clusters of Horton–Strahler order $j$ within $[0, \infty)$ time interval divided by $N$ satisfies

$$\begin{aligned}
&P\left( \left| N_j/N - \int_0^\infty \frac{\eta_{j-1}^2(t)}{2}\,dt \right| < 3\varepsilon \right) \\
&\quad \geq P\left( \big(N_j - N_j(K)\big)/N < \varepsilon, \left| N_j(K)/N - \int_0^K \frac{\eta_{j-1}^2(t)}{2}\,dt \right| < \varepsilon \right) \\
&\quad \geq \min\left\{ P\big(\big(N_j - N_j(K)\big)/N < \varepsilon\big), P\left( \left| N_j(K)/N - \int_0^K \frac{\eta_{j-1}^2(t)}{2}\,dt \right| < \varepsilon \right) \right\} \\
&\quad \geq \min\left\{ 1 - \frac{2(1-\varepsilon)}{\varepsilon K}, \mathcal{P}_{K,\varepsilon,N} \right\} \\
&\quad \to 1 - \frac{2(1-\varepsilon)}{\varepsilon K}
\end{aligned}$$

as $N \to \infty$.

Thus, since we can take $K$ as large as we want,

$$P\left( \left| N_j/N - \int_0^\infty \frac{\eta_{j-1}^2(t)}{2}\,dt \right| < 3\varepsilon \right) \to 1.$$

□

## Acknowledgements

We are grateful to Ed Waymire for encouragement and continuing interest in this work. We thank the participants of the 2012 Oregon State University Workshop on Mathematical Problems in the Environmental Sciences for their constructive feedback. Thorough editing and criticism provided to us by an anonymous reviewer and the Associate Editor throughout the review process helped to significantly improve the original manuscript, enhancing its rigor and clarity.

## References

[1] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** (1999) 3–48. MR1673235
[2] N. Berestycki. Recent progress in coalescent theory. *Ensaios Mat.* **16** (2009) 1–193. MR2574323
[3] J. Bertoin. *Random Fragmentation and Coagulation Processes*. Cambridge University Press, Cambridge, 2006. MR2253162
[4] G. A. Burd, E. C. Waymire and R. D. Winn. A self-similar invariance of critical binary Galton–Watson trees. *Bernoulli* **6** (2000) 1–21. MR1781179
[5] R. Darling and J. Norris. Differential equation approximations for Markov chains. *Probab. Surv.* **5** (2008) 37–79. MR2395153
[6] L. Devroye and P. Kruszewski. A note on the Horton–Strahler number for random trees. *Inform. Process. Lett.* **56** (1995) 95–99. MR1359172
[7] P. S. Dodds and D. H. Rothman. Scaling, universality, and geomorphology. *Annu. Rev. Earth Planet. Sci.* **28** (2000) 571–610.
[8] M. Drmota. The height of increasing trees. *Ann. Comb.* **12** (2009) 373–402. MR2496124
[9] R. E. Horton. Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology. *Geol. Soc. Am. Bull.* **56** (1945) 275–370.
[10] W. I. Newman, D. L. Turcotte and A. M. Gabrielov. Fractal trees with side branching. *Fractals* **5** (1997) 603–614.
[11] J. R. Norris. Smoluchowski's coagulation equation: Uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. *Ann. Appl. Probab.* **9** (1) (1999) 78–109. MR1682596
[12] S. D. Peckham. New results for self-similar trees with applications to river networks. *Water Resour. Res.* **31** (1995) 1023–1029.
[13] J. Pitman. *Combinatorial Stochastic Processes*. *Lecture Notes in Mathematics* **1875**. Springer-Verlag, Berlin, 2006. MR2245368
[14] R. L. Shreve. Statistical law of stream numbers. *J. Geol.* **74** (1966) 17–37.
[15] R. L. Shreve. Infinite topologically random channel networks. *J. Geol.* **75** (1967) 178–186.
[16] A. N. Strahler. Quantitative analysis of watershed geomorphology. *Trans. – Am. Geophys. Union* **38** (1957) 913–920.
[17] X. G. Viennot. Trees everywhere. In *CAAP'90* 18–41. Springer, Berlin, 1990.
[18] I. Zaliapin and Y. Kovchegov. Tokunaga and Horton self-similarity for level set trees of Markov chains. *Chaos Solitons Fractals* **45** (3) (2012) 358–372. MR2881663
[19] S. Zanardo, I. Zaliapin and E. Foufoula-Georgiou. Are American rivers Tokunaga self-similar? New results on fluvial network topology and its climatic dependence. *J. Geophys. Res.* **118** (2013) 166–183.