Invasion Percolation on Galton-Watson Trees

Marcus Michelen, Robin Pemantle[1] and Josh Rosenberg
{marcusmi, pemantle, rjos}@math.upenn.edu

Abstract

We consider invasion percolation on Galton-Watson trees. On almost every Galton-Watson tree, the invasion cluster almost surely contains only one infinite path. This means that for almost every Galton-Watson tree, invasion percolation induces a probability measure on infinite paths from the root. We show that under certain conditions of the progeny distribution, this measure is absolutely continuous with respect to the limit uniform measure. This confirms that invasion percolation, an efficient self-tuning algorithm, may be used to sample approximately from the limit uniform distribution. Additionally, we analyze the forward maximal weights along the backbone of the invasion cluster and prove a limit law for the process.

Keywords: Backbone, incipient infinite cluster, limit uniform, Poisson point process, pivot, self-organized criticality.

[1]Partially supported by NSF grant DMS-1612674

1 Introduction

Given an infinite rooted tree, how might one sample, nearly uniformly, from the set of paths from the root to infinity? One motive for this question is that nearly uniform sampling leads to good estimates on the growth rate [JS89]. One might be trying to estimate the size of a search tree, or, in the case of [RS00], to determine the growth rate of the number of self-avoiding paths.

A number of methods have been studied. One is to do a random walk on the tree, with a "homesickness" parameter determining how much steps back toward the root are favored [LPP96]. The parameter needs to be tuned near criticality: too much homesickness and the walk gets stuck near the root; too little homesickness and the walk goes to infinity without taking the time to ensure that the path is well randomized. Randall and Sinclair [RS00] solve this by estimating the critical parameter as the walk progresses, re-tuning the homesickness to lie above this by an amount decreasing at an appropriate rate.

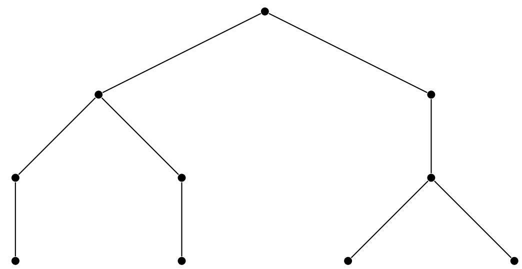
Another approach is to use percolation. One conditions the percolation cluster to survive to level N; as the percolation parameter decreases to criticality and N is taken to infinity, the law of this cluster approaches the law of the *incipient infinite cluster* (IIC). For many graphs—e.g. regular or Galton-Watson trees—the IIC almost surely contains a unique infinite path, thereby giving a mechanism for sampling such a path. In practice, the same considerations arise as with homesick random walks: tuning the percolation parameter too low yields too little likelihood of survival and too great a time cost to rejection sampling; too great a percolation parameter results in too many surviving paths and a selection problem which leads to poor randomization.

Invasion percolation was introduced as a model for how viscous fluid creeps through an environment in [WW83]. Each site is given an independent $U[0, 1]$ random variable, representing how great the percolation probability would have to be before the site would be open. The cluster then grows by adding, at each time step, the site with the least U value among sites neighboring the cluster but not in the cluster. On trees, it is not hard to see that the $\limsup$ of U-values of bonds chosen is equal to the critical percolation parameter. In other words, instead of running percolation at p_c and conditioning to survive, one allows slightly supercritical bonds but less and less as the cluster grows. As is the case for the IIC, the invasion cluster almost surely contains only one infinite path in the case of regular or Galton-Watson trees, and thus gives a different mechanism for sampling paths. Unlike the IIC and homesick random walk, invasion percolation requires no tuning to criticality and is an instance of self-organized criticality.

The invasion cluster has some properties in common with the IIC but not all. For example, results of Kesten [Kes86] and Zhang [Zha95] show that the growth exponents of the two are equal on the two-dimensional lattice; however the measures of the two clusters are mutually singular on the lattice [DSV09] as well as on a regular tree [AGdHS08]. Our focus is the comparison of the laws induced on paths by both the IIC and invasion percolation.

On a Galton-Watson tree T, there is a natural measure on paths, the limit-uniform measure μ_T, which although it does not restrict precisely to the uniform measure on each generation, approximates this as closely as possible. There is not, however, a fast algorithm for sampling from it. Rules such as "split equally at each node" lead to rapid sampling but the wrong entropy; in other words, the Radon-Nikodym derivative with respect μ_T on generation N will be exponential in N. It is not hard to show that on almost every Galton-Watson tree (assuming a $Z \log Z$ moment for the offspring distribution), the unique path in the IIC has law μ_T. Since sampling from the IIC is problematic, it is therefore natural to ask how close the law ν_T

of the path chosen by the invasion cluster is to μ_T. It is easy to see that the two laws are typically not equal. As an example, consider the set of trees with first three generations given by



When averaged over the remaining generations with the Galton-Watson measure—or equivalently, placing independent Galton-Watson trees at the terminal nodes—the limit-uniform measure splits equally at the root, while the invasion measure favors the left subtree regardless of the offspring distribution.

The best comparison one might hope for is that ν_T be absolutely continuous with respect to μ_T, perhaps even with Radon-Nikodym derivative in L^p. Our main result is as follows.

Theorem 1.1. *Suppose the offspring distribution Z has at least p moments and $\mathbf{P}[Z = 0] = 0$; set $p_1 := \mathbf{P}[Z = 1]$, let $\mu := \mathbf{E}[Z]$, and denote $q := \frac{\log \mu}{\log(1/p_1)}$. If*

$$2p^2q^2 + (3p^2 + 5p)q + (-p^2 + 11p - 4) < 0,$$

then $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$ almost surely.

The condition in Theorem 1.1 is a trade-off between p_1 and p. In the case of $p = \infty$, the condition becomes $p_1 < 1/\mu^{\frac{3+\sqrt{17}}{2}}$. In the case of $p_1 = 0$, the condition is $p > \frac{11+\sqrt{105}}{2}$.

A summary of the argument behind Theorem 1.1 is as follows. Let X_n be the KL-distance between the way that μ_T and ν_T split at the nth step γ_n of a path chosen from ν_T. A sufficient condition for absolute continuity is that $\sum_{n=1}^{\infty} \mathbf{E}X_n < \infty$. A precise statement is given in Lemma 2.7 below. A more detailed outline of the argument is given at the end of this section.

The reason we have a hope of estimating X_n is that there is a *backbone* decomposition for invasion percolation. Define the backbone to be the almost surely unique non-backtracking path $\gamma = (\mathbf{0}, \gamma_1, \gamma_2, \ldots)$ from the root to infinity. For any vertex v define the *pivot value* at v, denoted $\beta(v)$, to be the least p such that there is a path from v to infinity in the subtree at v with all U variables (not including the one at v) at most p. On a regular tree, invasion percolation was studied in [AGdHS08, AGdHS13]. For the purposes of studying ν_T, the regular tree is a degenerate case, because μ_T and ν_T are equal to each other and to the equally splitting measure. Further, on regular trees, the incipient infinite cluster stochastically dominates the invasion cluster; this fails to hold in the Galton-Watson case due to the fact that $\mu_T \neq \nu_T$. Despite these differences, the results on backbones and pivots in the regular case extend in a useful way to the Galton-Watson setting. In particular, [AGdHS08] prove that the process $\{\beta(\gamma_n) - p_c\}_{n\geq 0}$ converges to the *Poisson lower envelope* process when properly scaled; we prove similar results, and combine Theorem 6.2 and Corollaries 6.3 and 6.4 into the following:

Theorem 1.2. *Define $h_n := \beta(\gamma_n) - p_c$. Then*

(i) *Let* $\{U_j\}_{j=0}^\infty$ *be IID random variables each uniformly distributed on* $(0,1)$ *and define* $M_n = \min\{U_0, \ldots, U_n\}$*. Then for each* $\varepsilon > 0$*, the process* $\{h_n\}$ *may be coupled with* $\{M_n\}$ *so that with probability* 1*,* h_n *satisfies* $(1-\varepsilon)p_c M_n \le h_n \le (1+\varepsilon)p_c M_n$ *for all sufficiently large* n*.*

(ii) *For any* $\varepsilon > 0$ *as* $k \to \infty$*,*

$$(kh_{\lceil kt \rceil}/p_c)_{t\ge\varepsilon} \stackrel{*}{\Longrightarrow} (L(t))_{t\ge\varepsilon}$$

where $\stackrel{*}{\Longrightarrow}$ *denotes convergence in distribution of càdlàg paths in the Skorohod space* $D[0,\infty)$ *and* $L(t)$ *denotes the Poisson lower envelope process, defined in [AGdHS08] and Section 6.*

(iii) *The sequence* $n \cdot h_n$ *converges in distribution to* $p_c \cdot \exp(1)$*, where* $\exp(1)$ *is an exponential random variable with mean* 1*.*

Conditioning on T, the way the invasion measure splits at v depends on the whole tree. However, if one also conditions on the pivot at v, then the way the invasion measure splits at v becomes independent of everything outside of the subtree at v. A similar statement is true if one conditions on the pivot of v being less than or equal to a certain value; these are the Markov properties of Propositions 4.4 and 4.6. The limiting behavior of these values is given in Theorem 4.9 and Section 6. Further, Lemma 5.1 shows that this conditioned splitting measure is close to a ratio of survival probabilities under supercritical Bernoulli percolation. The problem is thus reduced to proving estimates of the survival probabilities of Galton-Watson trees under supercritical Bernoulli percolation as in Section 3.

The remainder of the paper is organized as follows. Section 2 sets up the notation and gives some preliminary results. Some care is required to set up the probability space so that we can easily speak of the random measures μ_T and ν_T, which are conditional on the Galton-Watson tree. Section 2 culminates in Lemma 2.7 and Corollary 2.8. Section 3 estimates near-critical survival probabilities for Galton-Watson trees. Section 4 proves two Markov properties for the subtree from γ_n together with $\beta(\gamma_n)$. The remainder of the section extends the work of [AGdHS08] by proving a limit law for $\beta(\gamma_n)$ which then implies an upper bound on the rate at which $\beta(\gamma_n) \downarrow p_c$. In particular, Corollary 6.3 shows convergence to the Poisson lower envelope process, as in [AGdHS08]. Section 5 completes the proof of Theorem 1.1 by comparing the conditional invasion measure to the ratio of survival probabilities and utilizing the estimates on survival probabilities from Section 3.

A glossary of notation by page of reference is included after the references.

Outline of Proof of Theorem 1.1

1. **Absolute continuity follows from summability of KL-divergence of splits**
   Let X_n be the KL-distance between the way that μ_T and ν_T split at the nth step γ_n of a path chosen from ν_T. A sufficient condition for absolute continuity is that $\sum_{n=1}^\infty X_n < \infty$. A precise statement is given in Lemma 2.7. In fact, we may replace the KL-distance with a process that differs from X_n for only finitely many n (Corollary 2.8).

2. **Shifting to the γ_n is the same as conditioning on the pivot being at most a certain value**
   We show that shifting to the γ_n is the same as examining a fresh Galton-Watson tree with the pivot of the root conditioned to be at most a certain random variable that we call the *dual pivot*, β_n^*. This is the content of the Markov property given in Proposition 4.4.

3. **Understanding how β_n^* behaves for large n**
   As $n \to \infty$, the variables β_n^* approach p_c. We in fact will have that the convergence is quite rapid, as shown by Theorem 4.9. The variables β_n^* are closely related to the pivots β_n whose rate of decay is given in Theorem 1.2; the process $\{\beta_n^*\}_n$ is difficult to study by itself, although the pair (β_n, β_n^*) is Markov with transition kernel given explicitly in Proposition 4.8.

4. **Conditioned on the pivot of the root being at most p, the split of the invasion measure is close to the ratio of survival probabilities**
   With Steps 2 and 3 in mind, we examine the split of the invasion measure conditioned on the root having pivot at most p. Lemma 5.1 shows that this splitting measure may be closely approximated by splitting according to the probability that the subtree survives p-percolation.

5. **The ratio of survival probabilities is close to the split of the limit-uniform measure**
   The last remaining step is to show that if p is close to p_c, the ratios of the probabilities of surviving p-percolation closely approximate the splits of the limit uniform measure (Proposition 5.2). In order to show this, much work needs to be done to approximate the near-critical survival probabilities of Galton-Watson trees. This is the content and focus of Section 3.

2 Construction and preliminary results

2.1 Galton-Watson trees

We begin with some notation we use for all trees, random or not. Let $\mathcal{U}$ be the canonical *Ulam-Harris* tree [ABF13]. The vertex set of $\mathcal{U}$ is the set $\mathbf{V} := \bigcup_{n=1}^{\infty} \mathbb{N}^n$, with the empty sequence $\mathbf{0} := \emptyset$ as the root. There is an edge from any sequence $\mathbf{a} = (a_1, \ldots, a_n)$ to any extension $\mathbf{a} \sqcup j := (a_1, \ldots, a_n, j)$. The depth of a vertex v is the graph distance between v and $\mathbf{0}$ and is denoted $|v|$. We work with trees T that are locally finite rooted subtrees of $\mathcal{U}$. The usual notations are in force: T_n denotes the set of vertices at depth n; $T(v)$ is the subtree of T at v, canonically identified with a rooted subtree of $\mathcal{U}$, in other words the vertex set of $T(v)$ is $\{w : v \sqcup w \in V(T)\}$; ∂T denotes the set of infinite non-backtracking paths from the root; if $\gamma \in \partial T$ then γ_n ($n \geq 0$) denotes the nth vertex in γ; the last common ancestor of v and w is denoted $v \wedge w$ and the last common vertex of γ and γ' is denoted $\gamma \wedge \gamma'$; $\overleftarrow{v}$ denotes the parent of v . Let μ_T^n denote the uniform measure on the nth generation of T. In some cases, for example for almost every Galton-Watson tree, the limit $\mu_T := \lim_{n\to\infty} \mu_T^n$ exists and is called the *limit-uniform measure* [LP17, Chapter 17.6].

Turning now to Galton-Watson trees, let $\phi(z) := \sum_{n=1}^{\infty} p_n z^n$ be the ordinary generating function for a supercritical branching process with no death, i.e., $\phi(0) = 0$. We recall,

$$\begin{aligned} \phi'(1) &= \mathbf{E}Z =: \mu \\ \phi''(1) &= \mathbf{E}[Z(Z-1)] \end{aligned}$$

where Z is a random variable with probability generating function ϕ. Throughout, we assume $\mathbf{E}[Z^2] < \infty$; in particular, this also means that $\phi''(1) < \infty$. Moreover, since our focus is on ∂T, the assumption of $\phi(0) = 0$ can be made without loss of generality by considering the reduced tree, as in [AN72, Chapter I.12].

We will work on the canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ where $\Omega = (\mathbb{N} \times [0,1])^{\mathbf{V}}$, $\mathcal{F}$ is the product Borel σ-field, and $\mathbf{P}$ is the probability measure making the coordinate functions $\omega_v = (\deg_v, U_v)$ IID with the law of (Z, U), where U is uniform on $[0,1]$ and independent of Z. The variables $\{\deg_v\}$—where $\deg_v$ is interpreted as the number of children of vertex v—will construct the Galton-Watson tree, while the variables $\{U_v\}$ will be used later for percolation. Let $\mathbf{T}$ be the random rooted subtree of $\mathcal{U}$ which is the connected component containing the root of the set of vertices that are either the root or are of the form $v \sqcup j$ such that $0 \le j < \deg_v$. This is a Galton-Watson tree with ordinary generating function ϕ.

As is usual for Galton-Watson branching processes, we denote $Z_n := |\mathbf{T}_n|$. Extend this by letting $Z_n(v)$ denote the number of offspring of v in generation $|v| + n$; similarly, extend the notation for the usual martingale $W_n := \mu^{-n} Z_n$ by letting $W_n(v) := \mu^{-n} Z_n(v)$. We know that $W_n(v) \to W(v)$ for all v, almost surely and in L^p if the offspring distribution has p moments. This is stated without proof for integer values of $p \ge 2$ in [Har63, p. 16] and [AN72, p. 33, Remark 3]; for a proof for all $p > 1$, see [BD74, Theorems 0 and 5]. Further extend this notation by letting $v^{(i)}$ denote the ith child of v, letting $Z_n^{(i)}(v)$ denote nth generation descendants of v whose ancestral line passes through $v^{(i)}$, and letting $W_n^{(i)}(v) := \mu^{-n} Z_n^{(i)}(v)$. Thus, for every v, $W(v) = \sum_i W^{(i)}(v)$. For convenience, define $p_c := 1/\mu$ and recall that p_c is almost surely the critical percolation parameter for $\mathbf{T}$ [Lyo90].

2.2 Bernoulli and Invasion Percolation

In this subsection we give the formal construction of percolation on random trees, and for invasion percolation. Our approach is to define a simultaneous coupling of invasion percolations on all subtrees T of $\mathcal{U}$ via the U variables, then specialize to the random tree $\mathbf{T}$. Let $\mathcal{T} := \sigma(\{\deg_v : v \in \mathbf{V}\})$ denote the σ-field generated by the tree variables. Because $\mathcal{T}$ is independent from the U variables, this means we have constructed a process whose law, conditional on $\mathcal{T}$, is invasion percolation on $\mathbf{T}$. We use the notation $\mathbf{E}_*$ to denote $\mathbf{E}[\cdot \mid \mathcal{T}]$; similarly $\mathbf{P}_*[\cdot] := \mathbf{P}[\cdot \mid \mathcal{T}]$. Moreover, we use $\mathtt{GW} := \mathbf{P}|_{\mathcal{T}}$ to denote the Galton-Watson measure on trees.

We begin with a similar construction for ordinary percolation. For $0 < p < 1$, simultaneously define Bernoulli(p) percolations on rooted subtrees T of $\mathcal{U}$ by taking the percolation clusters to be the connected component containing $\mathbf{0}$ of the induced subtrees of T on all vertices v such that $U_v \le p$; note that we always include the root $\mathbf{0}$, and thus the uniform variables U_v may equivalently be thought of as being edge-weights connecting the parent of v to v. Let $\mathcal{F}_n$ be the σ-field generated by the variables $\{U_v, \deg_v : |v| < n\}$. Let $p_c = 1/\mu = 1/\phi'(1)$ denote the critical probability for percolation. Write $v \leftrightarrow_{T,p} w$ if $U_u \le p$ for all u on the geodesic from v to w in T. Informally, $v \leftrightarrow_{T,p} w$ iff v and w are both in T and are connected in the p-percolation. The event of successful p percolation on T is $H_T(p) := \{\mathbf{0} \leftrightarrow_{T,p} \infty\}$ and the event of successful p percolation on the random tree $\mathbf{T}$, is denoted $H_{\mathbf{T}}(p)$ or simply $H(p)$. Let $g(T,p) := \mathbf{P}[H_T(p)]$ denote the probability of p percolation on T. The conditional probability $\mathbf{P}_*[H(p)]$ is measurable with respect to $\mathbf{T}$ and we may define $g(\mathbf{T}, p) := \mathbf{P}_*[H(p)]$. Furthermore, we may define $g(p) = \mathbf{P}[H(p)] = \mathbf{E} g(\mathbf{T}, p)$. Since $p_c = 1/\mu$ is the critical percolation parameter for a.e. $\mathbf{T}$, note that $g(\mathbf{T}, p) = 0$ for all $p \in [0, p_c]$.

Define invasion percolation on an arbitrary tree T as follows. Start with $I_0^T = \mathbf{0}$ where we recall that $\mathbf{0}$ is the root of T. Inductively define I_{n+1}^T to consist of I_n^T along with the vertex of minimal weight U_v adjacent to I_n^T. The invasion percolation cluster is defined as $I^T := \bigcup_n I_n^T$. Note that I^T is measurable with respect to the U variables. Let $I := I^{\mathbf{T}}$ denote the invasion cluster of the random tree $\mathbf{T}$. By independence of the U variables and $\mathcal{T}$, the conditional distribution of I given $\mathbf{T}$ agrees with that of invasion percolation.

Proposition 2.1. *For any $p > p_c$, I almost surely reaches some vertex v such that $v \leftrightarrow_p \infty$ in $\mathbf{T}(v)$.*

Proof: We consider the following coupling that generates I at the same time as $\mathbf{T}$: begin with the root, and generate children according to Z, giving each new edge a $(0,1)$ weight uniformly and independently. We denote this height 1 weighted tree as L_1. The sequence of weighted trees $\{L_n\}$ is now defined inductively as follows: for each $n \geq 1$, L_{n+1} is obtained by assigning Z children (using an independent copy of Z) to the boundary vertex of L_n with the smallest corresponding edge weight, and then giving each of the new edges a $(0,1)$ weight uniformly and independently.

For each $n \geq 1$, define $\mathcal{F}_n$ to be the Borel σ-field inside of $\mathcal{F}$ that is generated by L_n. Next, we define the increasing sequence of stopping times $N_1, N_2, \ldots$ in the following way: set N_1 equal to the minimum value of n such that all edges connected to boundary vertices of L_n have weight at least p, and if no such value exists, set $N_1 = \infty$. For $j \geq 1$, set N_{j+1} equal to infinity if either $N_j = \infty$, or there is no $n > N_j$ such that all edges connected to boundary vertices of L_n have weight at least p, and otherwise set N_{j+1} equal to the minimum $n > N_j$ for which this last condition is satisfied. Observing that $\{N_1 = \infty\}$ is simply the event that all edges of the invasion cluster I have weight less than p, we see that $\mathbf{P}(N_1 = \infty) = g(p)$. In addition, since no edges in L_{N_j} are considered until time N_{j+1}, we also find that for every j, k with $1 \leq j \leq k < \infty$, the random variable $N_{j+1} - N_j$ is independent of $\mathcal{F}_k$ with respect to the probability measure $\mathbf{P}(\cdot | N_j = k)$. Finally, noting that $(N_{j+1} - N_j | N_j = k)$ has the same distribution as N_1, we find that $\mathbf{P}(N_{j+1} = \infty | N_j < \infty) = g(p)$.

Now define $A_p \in \mathcal{F}$ to be the event that I eventually invades a vertex with corresponding edge weight less than p. Since having a j for which $N_j = \infty$ implies A_p, we can now conclude from the above observations that

$$\mathbf{P}(A_p) = \mathbf{E}[\mathbf{P}(A_p|\mathbf{T})] \geq \sum_{j=0}^{\infty} g(p)\big(1 - g(p)\big)^j = 1 \implies \mathbf{P}_*(A_p) = 1 \ \ \texttt{GW}\text{-a.s.},$$

thus completing the proof. □

Corollary 2.2. *For any $p > p_c$, the number of edges in I with weight greater than p is almost surely finite.*

This was proven for a large class of graphs by Häggström, Peres and Schonmann [HPS99], but this class doesn't cover the case of Galton-Watson trees conditioned on survival; they exploit quite a bit of symmetry that does not occur in the Galton-Watson case.

Proof: Let x be the first invaded vertex with an infinite subtree below with weights less than p. Then after x is invaded, no edges of weight larger than p will be invaded. □

Corollary 2.3. *There is almost surely only one infinite non-backtracking path from $\mathbf{0}$ in I. Equivalently, $\mathbf{T}$ is almost surely in the set of trees T such that I^T contains almost surely a unique infinite non-backtracking path from $\mathbf{0}$.*

Proof: Suppose that there are two distinct paths to infinity in I; by Corollary 2.2, there exist maximal weights M_1 and M_2 along these paths after they split, $\mathbf{P}$-almost surely. If $M_1 > M_2$, the second infinite path would be invaded before the edge containing M_1. Similarly, we cannot have $M_2 > M_1$. Finally, $M_1 = M_2$ has $\mathbf{P}$-probability 0, completing the proof. □

This proof is stated for invasion percolation on regular trees in [AGdHS08], but is identical for Galton-Watson trees once Corollary 2.2 is in place; the unique path guaranteed by Corollary 2.3 is typically called the backbone of I, and we continue this convention. Note that a regular tree is simply a Galton-Watson tree with Z almost-surely constant.

Definition 2.4 (the invasion path γ)**.** *Let $\gamma^T := (\mathbf{0}, \gamma_1^T, \gamma_2^T, \ldots)$ be the random sequence whose nth element is the unique v with $|v| = n$ such that $v \leftrightarrow \infty$ via a downward path in the invasion cluster I^T. Let ν_T denote the law of γ^T given T. Let $\nu_{\mathbf{T}}$ denote the random measure on the random space $(\mathbf{T}, \partial\mathbf{T})$ induced by the $\gamma^{\mathbf{T}}$. In other words, for measurable $A \subseteq \partial\mathcal{U}$, $\nu_{\mathbf{T}}(A, \omega) = \mathbf{P}[\gamma^T \in A]$ evaluated at $T = \mathbf{T}(\omega)$. By Corollary 2.3, this is a well defined probability measure for almost every ω.*

2.3 Preliminary comparison of limit-uniform and invasion measures

Our main goal is to see whether $\nu_{\mathbf{T}}$ is almost surely absolutely continuous with respect to $\mu_{\mathbf{T}}$. We give the summability criterion that establishes a sufficient condition for absolute continuity in terms of the KL-divergence of the two measures along a ray chosen from $\nu_{\mathbf{T}}$.

Definition 2.5 (the splits p and q at children of u, and their difference, X)**.** *Let v be a vertex of $\mathbf{T}$ and let u be the parent of v. Define*

$$\begin{aligned} p(v) &:= \mu_{\mathbf{T}}(v)/\mu_{\mathbf{T}}(u) \\ q(v) &:= \nu_{\mathbf{T}}(v)/\nu_{\mathbf{T}}(u) \\ X(u) &:= \sum_w q(w) \log[q(w)/p(w)] \end{aligned}$$

where the sum is over all children w of u and $\nu_{\mathbf{T}}(v) = \nu_{\mathbf{T}}(\{\gamma : v \in \gamma\})$ and $\mu_{\mathbf{T}}(v)$ is defined similarly. The quantity X is known as KL-divergence. The KL-divergence $\mathcal{K}(\rho, \rho')$ is defined between any two probability measures ρ and ρ' on a finite set $\{1, \ldots, k\}$ by the formula

$$\mathcal{K}(\rho, \rho') := \sum_{i=1}^k \rho'(i) \log \frac{\rho'(i)}{\rho(i)} .$$

It is a measure of the difference between the two distributions. It is always non-negative but not symmetric. The following inequality shows that $\mathcal{K}$ behaves like quadratic distance away from $\rho = 0$.

Proposition 2.6. *Let ρ and ρ' be probability measures on the set $\{1, \ldots, k\}$ and denote $\varepsilon_i := \rho'(i)/\rho(i) - 1$. Then*

$$\mathcal{K}(\rho, \rho') \leq \sum_{i=1}^k \rho(i) \varepsilon_i^2 . \tag{2.1}$$

Proof: Define the function R on $(-1, \infty)$ by

$$R(x) := \frac{(1 + x) \log(1 + x) - x}{x^2}$$

if $x \neq 0$ and $R(0) := 1/2$. This makes R continuous, positive, and decreasing from 1 to 0 on $(-1, \infty)$. When $\varepsilon = \rho'/\rho - 1$, we may compute

$$\frac{\rho' \log(\rho'/\rho) - (\rho' - \rho)}{\rho} = \frac{(1 + \varepsilon)\rho \log(1 + \varepsilon) - \varepsilon\rho}{\rho} = \varepsilon^2 R(\varepsilon) .$$

Because $\sum_{i=1}^k \rho(i) = \sum_{i=1}^k \rho'(i) = 1$, we see that

$$\mathcal{K}(\rho, \rho') = \sum_{i=1}^k (\rho'(i) - \rho(i)) + \rho(i)\varepsilon_i^2 R(\varepsilon_i) = \sum_{i=1}^k \rho(i)\varepsilon_i^2 R(\varepsilon_i)$$

and the result follows from $0 < R(\varepsilon_i) < 1$. $\square$

Applying Proposition 2.6 to $\rho' = q$ and $\rho = p$ gives

$$X(u) \leq \sum_w p(w)\varepsilon(w)^2 \tag{2.2}$$

where $\varepsilon(w) = \frac{q(w)}{p(w)} - 1$.

Lemma 2.7. *Let T be a fixed tree on which ν_T and μ_T are well defined on the Borel σ-field $\mathcal{B}$ on ∂T. If*

$$\sum_{n=1}^{\infty} \sum_{|v|=n} X(v)\nu_T(v) < \infty \tag{2.3}$$

then $\nu_T \ll \mu_T$.

Proof: On the measure space $(\partial T, \mathcal{B})$, define a filtration $\{\mathcal{G}_n\}$ by letting $\mathcal{G}_n$ denote the σ-field generated by the sets $\{\gamma : \gamma_n = v\}$. The Borel σ-field $\mathcal{B}$ is the increasing limit $\sigma(\bigcup_n \mathcal{G}_n)$. Let

$$M_n := \left.\frac{d\nu_T}{d\mu_T}\right|_{\mathcal{G}_n} .$$

In other words, $M_n(\gamma) = \nu_T(\gamma_n)/\mu_T(\gamma_n)$. Let $\overline{M} := \limsup_{n\to\infty} M_n$. The Radon-Nikodym martingale theorem [Dur10, Theorem 5.3.3] says that $\{M_n\}$ is a martingale with respect to $(\partial T, \mathcal{B}, \mu_T, \{\mathcal{G}_n\})$ and that $\nu_T \ll \mu_T$ is equivalent to $\nu_T(\{\overline{M} = \infty\}) = 0$. This is equivalent to $\nu_T(\{\underline{M} = 0\}) = 0$ where $\underline{M} = 1/\overline{M} = \liminf_n 1/M_n$. The sequence $\{1/M_n\}$ is a ν_T-martingale, therefore $\{\log(1/M_n)\}$ is a ν_T-supermartingale and to conclude that it ν_T-a.s. does not go to negative infinity, it suffices to show that its expectation is bounded from below.

We compute the conditional expected increment of $\log(1/M_n)$. Letting γ denote the ray $(\gamma_1, \gamma_2, \ldots)$,

$$\log \frac{1}{M_{n+1}(\gamma)} - \log \frac{1}{M_n(\gamma)} = \log \frac{\nu_T(\gamma_n)}{\mu_T(\gamma_n)} - \log \frac{\nu_T(\gamma_{n+1})}{\mu_T(\gamma_{n+1})} = -\log \frac{q(\gamma_{n+1})}{p(\gamma_{n+1})} .$$

Conditioning on $\mathcal{G}_n$, if $\gamma_n = u$, then the ν_T-probability of $\gamma_{n+1} = v$ is $q(v)$, whence

$$\mathbf{E}_{\nu_T}\left[\log \frac{1}{M_{n+1}} - \log \frac{1}{M_n} \,\middle|\, \mathcal{G}_n\right] = \sum_{v \text{ child of } u} -q(v) \log \frac{q(v)}{p(v)} = -X(u) .$$

Taking the unconditional expectation,

$$\mathbf{E}_{\nu_T}\left[\log \frac{1}{M_{n+1}} - \log \frac{1}{M_n}\right] = -\sum_{|v|=n} \nu_T(v) X(v)$$

and summing over n shows that (2.3) implies that $\log(1/M_n)$ has expectation bounded from below, establishing $\nu_T \ll \mu_T$. □

Corollary 2.8. *Recall that γ denotes the invasion path on $\mathbf{T}$ and let X_n denote $X(\gamma_n)$.*

(i) If $\sum_{n=1}^{\infty} \mathbf{E}X_n < \infty$ then $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$ `GW`-almost surely.

(ii) Define the filtration $\{\mathcal{G}'_n\}$ on $(\Omega, \mathcal{F})$ by letting $\mathcal{G}'_n$ be the σ-field generated by $\mathcal{T}$ together with $\gamma_1, \ldots, \gamma_n$. Let $Y(v)$ be non-negative random variables such that $Y(v) \in \mathcal{G}'_{|v|}$ and

$$\mathbf{P}[X(\gamma_n) \neq Y(\gamma_n) \text{ infinitely often}] = 0 .$$

Then $\sum_{n=1}^{\infty} \mathbf{E}Y_n < \infty$ implies that `GW`-almost surely, $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$.

Proof:

(i) Writing $\mathbf{E}X_n = \mathbf{E}\left[\mathbf{E}_* X_n\right]$ we see that the hypothesis of (i), namely $\sum_{n=1}^\infty \mathbf{E}X_n < \infty$, implies $\mathbf{E}\sum_{n=1}^\infty \mathbf{E}_* X_n < \infty$. This implies $\sum_{n=1}^\infty \mathbf{E}_* X_n < \infty$ almost surely. A version of $\mathbf{E}_* X_n$ is $\sum_{|v| \in \mathbf{T}_n} X(v) \nu_{\mathbf{T}}(v)$, whence (2.3) holds for GW-almost every $\mathbf{T}$, implying almost sure absolute continuity of $\mu_{\mathbf{T}}$ with respect to $\nu_{\mathbf{T}}$.

(ii) The argument used to prove Lemma 2.7 may be adapted as follows. Let $M_n := \left.\frac{d\nu_{\mathbf{T}}}{d\mu_{\mathbf{T}}}\right|_{\mathcal{G}_n'}$, a version of which is the function taking the value $\frac{\nu_{\mathbf{T}}(v)}{\mu_{\mathbf{T}}(v)}$ on $\{\gamma_n = v\}$. Again $\{M_n\}$ is a martingale and $\log(1/M_n)$ is a supermartingale which we need to show converges almost surely. The sequence

$$S_M := \sum_{n=1}^M \left(\log \frac{1}{M_{n+1}} - \log \frac{1}{M_n}\right) \mathbf{1}_{X(\gamma_n) = Y(\gamma_n)}$$

is a convergent supermartingale because its expected increments are either 0 or $-Y(\gamma_n)$; convergence of the unconditional expectations $\mathbf{E}Y(\gamma_n)$ implies almost sure convergence of the expected increments, implying almost sure convergence of the supermartingale $\{S_M\}$. The hypotheses of (ii) imply that the increments of S_M differ from the increments of $\log(1/M_n)$ finitely often almost surely, implying convergence of the supermartingale $\log(1/M_n)$ and hence $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$. $\square$

3 Survival function conditioned on the tree

This section is concerned with estimating the quenched survival function $g(\mathbf{T}, p)$. The ultimate goal will be to examine the behavior of $g(\mathbf{T}, p)$ for small $p - p_c$, as estimates on $g(\mathbf{T}, \cdot)$ will be central to step 5 of the outline. Before studying the random function $g(\mathbf{T}, p)$, we record some basic properties of the annealed function $g(p) = \mathbf{E}[g(\mathbf{T}, p)]$. For a more complete analysis of the function $g(\mathbf{T}, \cdot)$, see [MPR18].

3.1 Properties of the annealed function $g(p)$

Proposition 3.1. *The derivative from the right $K := g'(p_c)$ exists and is given by*

$$K := \frac{2}{p_c^3 \phi''(1)} . \tag{3.1}$$

Proof: Let $\phi_p(z) := \phi(1 - p + pz)$ be the offspring generating function for the Galton-Watson tree thinned by p-percolation for $p \in (p_c, 1)$. The fixed point of ϕ_p is $1 - g(p)$. In other words, $g(p)$ is the unique $s \in (0,1)$ for which $1 - \phi_p(1 - s) = s$. The first two derivatives of ϕ_p at 1 are given by

$$\phi_p'(1) = \frac{p}{p_c} ;$$
$$\phi_p''(1) = p^2 \phi''(1) .$$

By a Taylor expansion, this leads to

$$1 - \phi_p(1 - s) = \frac{p}{p_c} s - \frac{1 + o(1)}{2} \phi''(1) p^2 s^2$$

as $p \downarrow p_c$. Setting this equal to s and solving for $s \in (0,1)$ yields the conclusion of the proposition. $\square$

Corollary 3.2. *As* $p \downarrow p_c$*,* $g'(p) \to K$*.*

Proof: The existence of $g'(p)$ on $(p_c, 1)$ follows from the implicit function theorem. To obtain an expression for $g'(p)$, we differentiate both sides of the expression $\phi(1-p \cdot g(p)) = 1 - g(p)$ with respect to p, which gives

$$(-g(p) - p \cdot g'(p))\phi'(1 - p \cdot g(p)) = -g'(p).$$

Rearranging this expression to isolate $g'(p)$, while using Proposition 3.1, along with the fact that $\phi'(1-x) = \mu - \phi''(1)x + o(x)$ as $x \to 0$, we get

$$g'(p) \;=\; \frac{g(p)\phi'(1-p\cdot g(p))}{1 - p \cdot \phi'(1 - p \cdot g(p))} \;=\; \frac{2\mu^3}{\phi''(1)} + o(1)$$

as $p \downarrow p_c$. □

3.2 Preliminary estimates of $g(\mathbf{T}, p)$

We now move to estimating $g(\mathbf{T}, p)$, a random variable measurable with respect to $\mathcal{T}$. We first prove an upper bound on g which gives a uniform bound on the L^q norm of g. Additionally, we show that conditioning on only the first n levels gives a random variable exponentially close to g. Estimating this averaged random variable is a key element in the proof of Theorem 1.1, and is the content of Section 3.3.

The following result from [LP17] will be useful for obtaining an a.s. upper bound on $g(\mathbf{T}, p_c + \varepsilon)$.

Theorem 3.3 ([LP17, Theorem 5.24])**.** *For p-percolation, we have*

$$\frac{1}{\mathscr{R}(\mathbf{0} \leftrightarrow \infty) + 1} \le \mathbf{P}_*[\mathbf{0} \leftrightarrow \infty] \le \frac{2}{\mathscr{R}(\mathbf{0} \leftrightarrow \infty) + 1}$$

where $\mathscr{R}(\mathbf{0} \leftrightarrow \infty)$ *denotes the effective resistance from* $\mathbf{0}$ *to infinity when an edge connecting* $\overleftarrow{u}$ *to* u *is given resistance*

$$r(e) = \frac{1-p}{p^{|u|}}.$$

□

From this, we deduce:

Proposition 3.4. *For any* $\varepsilon \in (0, 1-p_c)$ *and* GW*-almost surely,*

$$g(\mathbf{T}, p_c + \varepsilon) < \frac{2\varepsilon \overline{W}}{(1 - p_c - \varepsilon)p_c} \tag{3.2}$$

where $\overline{W} := \sup_n W_n(\mathbf{T})$ *is almost surely finite because* $\lim_{n\to\infty} W_n$ *exists almost surely.*

Proof: To get an upper bound on g, we need a lower bound on the resistance. For each height n, short together all nodes at this height. For every $p = p_c + \varepsilon$ this gives a lower bound of

$$\mathscr{R}(\mathbf{0} \leftrightarrow \infty) \ge \sum_{n=1}^{\infty} \frac{1 - p_c - \varepsilon}{Z_n (p_c + \varepsilon)^n}$$

$$
\begin{aligned}
&= (1-p_c-\varepsilon)\sum_{n=1}^{\infty}\frac{p_c^n}{W_n(p_c+\varepsilon)^n} \\
&\geq \frac{(1-p_c-\varepsilon)}{\overline{W}}\sum_{n=1}^{\infty}\frac{p_c^n}{(p_c+\varepsilon)^n} \\
&= \frac{(1-p_c-\varepsilon)p_c}{\overline{W}\varepsilon}.
\end{aligned}
$$

Using Theorem 3.3, we get

$$
g(\mathbf{T}, p_c+\varepsilon) \leq \frac{2}{1+\frac{(1-p_c-\varepsilon)p_c}{\overline{W}\varepsilon}} \leq \frac{2\varepsilon\overline{W}}{(1-p_c-\varepsilon)p_c}. \tag{3.3}
$$

□

Proposition 3.5 (uniform L^q bound)**.** *Suppose the offspring distribution has a finite $q>1$ moment. Then for any $\delta>0$, there is a constant c_q such that for all $\varepsilon \in (0, 1-p_c-\delta)$,*

$$
\mathbf{E}g(\mathbf{T}, p_c+\varepsilon)^q \leq c_q\varepsilon^q
$$

where $c_q = c_q(\delta) > 0$.

Proof: First recall that if the offspring has a finite q-moment, then $M_q := \mathbf{E}W^q$ is finite as well. By the L^q maximal inequality (e.g., [Dur10, Theorem 5.4.3]), we have that

$$
\mathbf{E}\left[\left(\sup_{1\leq k\leq n} W_k\right)^q\right] \leq \left(\frac{q}{q-1}\right)^q \mathbf{E}W_n^q \leq \left(\frac{q}{q-1}\right)^q M_q
$$

because $\{W_n^q\}$ is a submartingale.

Note that $\left(\sup_{1\leq k\leq n} W_k\right)^q \uparrow \overline{W}^q$ as $n\to\infty$. By monotone convergence, this implies $\mathbf{E}[\overline{W}^q] \leq (q/(q-1))^q M_q$. In particular, for any $\varepsilon < 1-p_c-\delta$, this together with Proposition 3.4 implies

$$
\mathbf{E}[g(\mathbf{T}, p_c+\varepsilon)^q] \leq \left(\frac{2\varepsilon}{(1-p_c-\varepsilon)p_c}\right)^q \mathbf{E}[\overline{W}^q] \leq \left(\frac{2\varepsilon}{\delta p_c}\right)^q\left(\frac{q}{q-1}\right)^q M_q\,,
$$

proving the proposition with $c_q = \left(\frac{2q}{(q-1)\delta p_c}\right)^q M_q$. □

Let $\mathcal{T}_n$ denote $\sigma(\deg_v : |v| \leq n)$. Because $\mathcal{T}_n \uparrow \mathcal{T}$ and g is bounded, we know that $\mathbf{E}[g(\mathbf{T},p)\,|\,\mathcal{T}_n] \to g(\mathbf{T},p)$ almost surely and in L^1. It turns out that $g_n := \mathbf{E}[g(\mathbf{T},p)\,|\,\mathcal{T}_n]$ is much easier to estimate than g itself. Our strategy is to show this convergence is exponentially rapid, transferring the work from estimation of g to estimation of g_n.

Lemma 3.6. *For any $\delta>0$, there are constants $c_i>0$ such that for all $p\in(p_c, \sqrt{p_c}-\delta)$*

$$
\big|\, g(\mathbf{T},p) - g_n(\mathbf{T},p)\,\big| \leq c_1 e^{-c_2 n}
$$

with probability at least $1-e^{-c_3 n}$.

PROOF: Define a random set $S = S(n,p)$ to be the set of vertices $v \in \mathbf{T}_n$ such that $\mathbf{0} \leftrightarrow_p v$. Let $\pi_{\mathbf{T}}$ denote the law of the random variable S, an atomic probability measure on the subsets of the random set $\mathbf{T}_n$. Using

$$g(\mathbf{T},p) = \mathbf{P}[H(p) \,|\, \mathcal{T}] = \mathbf{E}\left[\mathbf{P}[H(p) \,|\, \mathcal{F}_n'] \,|\, \mathcal{T}\right]$$

where $\mathcal{F}_n'$ be the σ-field generated by $\mathcal{F}_n$ and $\mathcal{T}$, we obtain the explicit representation

$$g(\mathbf{T},p) = \sum_S \pi_{\mathbf{T}}(S) \left[1 - \left(\prod_{v \in S} (1 - g(\mathbf{T}(v),p))\right)\right] . \tag{3.4}$$

Order the vertices in $\mathbf{T}_n$ arbitrarily and define the revealed martingale $\{M_k\}$ by

$$M_k := \mathbf{E}\left[g(\mathbf{T},p) \,|\, \mathcal{T}_n, \{\mathbf{T}(v_j) : j \leq k\}\right] \tag{3.5}$$

as k ranges from 0 to $|\mathbf{T}_n|$. By definition, $M_0 = g_n$. Also, $M_{|\mathbf{T}_n|} = g(\mathbf{T},p)$ because from $\mathcal{T}_n$ together with $\{\mathbf{T}(v) : v \in \mathbf{T}_n\}$ one can reconstruct $\mathbf{T}$. Arguing as in (3.4), we obtain the explicit representation

$$M_k = \sum_S \pi_{\mathbf{T}}(S) \left[1 - \prod_{v \in S_{\leq k}} (1 - g(\mathbf{T}(v),p)(1-g(p))^{|S_{>k}|}\right] \tag{3.6}$$

where for a given set $S \subset \mathbf{T}_n$, $S_{\leq k}$ denotes the vertices in S indexed $\leq k$ and $S_{>k}$ denotes the set indexed $> k$.

We claim the increments of $\{M_k\}$ are bounded by p^n. Indeed, (3.6) implies

$$|M_k - M_{k-1}| \leq \sum_{S \ni v_k} \pi_{\mathbf{T}}(S) |g(\mathbf{T}(v_k),p) - g(p)| \leq \sum_{S \ni v_k} \pi_{\mathbf{T}}(S) = \mathbf{P}[\mathbf{0} \leftrightarrow_p v_k] = p^n .$$

Azuma's inequality [Azu67] implies that for any $c_1, c_2 > 0$, the bounded increments translate to an upper bound

$$\mathbf{P}\left[|g(\mathbf{T},p) - g_n(\mathbf{T},p)| > c_1 e^{-c_2 n} \,|\, \mathcal{T}_n\right] \leq \exp\left(-\frac{c_1^2 e^{-2c_2 n}}{2|\mathbf{T}_n| p^{2n}}\right) . \tag{3.7}$$

Recall that for any $\gamma > 0$,

$$\mathbf{P}[|\mathbf{T}_n| \geq (\mu(1+\gamma))^n] = \mathbf{P}[W_n \geq (1+\gamma^n)] \leq (1+\gamma)^{-n}$$

by Markov's inequality. Since $\mu p^2 < 1$ uniformly for $p \in [p_c, \sqrt{p_c} - \delta]$, we therefore may pick c_2 so that $e^{-2c_2 n}$ is exponentially larger in n than $|\mathbf{T}_n| p^{2n}$ with exponentially high probability. Conditioning on this event and applying (3.7) completes the proof. $\square$

3.3 Bounds on the difference between $g(T, p_c + \varepsilon)$ and $W g(p_c + \varepsilon)$

For the purposes of proving Theorem 1.1, we will show that $g(\mathbf{T}, p_c + \varepsilon)$ is close to $W g(p_c + \varepsilon)$ for small $\varepsilon > 0$. For a fixed vertex v in a tree $\mathbf{T}$ define $E(v, \varepsilon)$ by

$$g(\mathbf{T}(v), p_c + \varepsilon) = g(p_c + \varepsilon) \left(W(v) + E(v, \varepsilon)\right) .$$

Proposition 3.7. *Suppose the offspring distribution of Z has $p \geq 2$ moments. Then for any δ, ℓ for which both $0 < \delta < 1$ and $0 < \ell < \frac{1}{2}$, there exist constants $C_i > 0$ so that for all ε sufficiently small*

$$|E(\mathbf{0}, \varepsilon)| \leq C_1 W \varepsilon^{1-\delta} + C_2 \varepsilon^{1-2\ell} \sum_{j=1}^{\lceil \varepsilon^{-\delta} \rceil - 1} W_j \tag{3.8}$$

with probability at least $1 - C_3 \varepsilon^{p\ell - \delta}$.

Proof: Let c_1, c_2, c_3 be the constants from Lemma 3.6, and fix $\delta > 0$. Then for $m = \lceil \varepsilon^{-\delta} \rceil$, we have

$$|g_m(\mathbf{T}, p_c + \varepsilon) - g(\mathbf{T}, p_c + \varepsilon)| < c_1 e^{-c_2/\varepsilon^\delta} \tag{3.9}$$

with probability at least $1 - e^{-c_3/\varepsilon^\delta}$, which implies that (3.9) holds for the root and all children of the root with probability at least $1 - (\mu + 1)e^{-c_3/\varepsilon^\delta}$. Utilizing (3.9) and the fact that $g(p_c + \varepsilon) = \Theta(\varepsilon)$ as $\varepsilon \to 0^+$ (while also making sure to select $c_3 < c_2$) gives

$$\frac{1}{g(p_c + \varepsilon)} |g_m(\mathbf{T}, p_c + \varepsilon) - g(\mathbf{T}, p_c + \varepsilon)| < c_1 \frac{1}{g(p_c + \varepsilon)} e^{-c_2/\varepsilon^\delta} = O\left(e^{-c_3/\varepsilon^\delta}\right). \tag{3.10}$$

By [Dub71], there exist positive constants C_1' and c_2' so that

$$\mathbf{P}[W \leq a] \leq C_1' a^{c_2'}.$$

This implies that $C_1 e^{-c_3/\varepsilon^\delta} \leq W \varepsilon^{1-\delta}$ with probability at least $1 - Ce^{-c/\varepsilon^\delta}$ for some new constants. Thus, to show equation (3.8), it is sufficient to examine $g_m(\mathbf{T}, p_c + \varepsilon)$.

The Bonferroni inequalities imply that

$$\text{F.O.}_m(\mathbf{0}, \varepsilon) - \text{S.O.}_m(\mathbf{0}, \varepsilon) \leq g_m(\mathbf{T}, p_c + \varepsilon) \leq \text{F.O.}_m(\mathbf{0}, \varepsilon)$$

where

$$\text{F.O.}_m(\mathbf{0}, \varepsilon) := \left(1 + \frac{\varepsilon}{p_c}\right)^m W_m g(p_c + \varepsilon)$$
$$\text{and } \text{S.O.}_m(\mathbf{0}, \varepsilon) := g(p_c + \varepsilon)^2 \sum_{\substack{u,w \in \mathbf{T}_m \\ u \neq w}} (p_c + \varepsilon)^{2m - |u \wedge w|}.$$

To bound $g_m(\mathbf{T}, p_c + \varepsilon) - g(p_c + \varepsilon)W$, we first bound $\frac{\text{F.O.}_m(\mathbf{0}, \varepsilon)}{g(p_c + \varepsilon)} - W$. Write

$$\frac{\text{F.O.}_m(\mathbf{0}, \varepsilon)}{g(p_c + \varepsilon)} - W = W\left(\left(1 + \frac{\varepsilon}{p_c}\right)^m - 1\right) + [W_m - W](1 + \varepsilon/p_c)^m.$$

Note first that $|(1 + \varepsilon/p_c)^m - 1| \leq Cm\varepsilon/p_c$ for some $C > 0$. Recalling that $m = \lceil \varepsilon^{-\delta} \rceil$ gives a bound of $C\varepsilon^{1-\delta}$. Additionally, we have $(1 + \varepsilon/p_c)^m \leq 2$ for ε sufficiently small. We now look towards $|W_m - W|$.

By [AN72, Chapter I.13], we have that

$$\text{Var}[W_m - W \mid W_m] = \frac{W_m}{\mu^m} \left(\frac{\text{Var}[Z]}{\mu^2 - \mu}\right).$$

By the law of total variance, this implies that

$$\mathrm{Var}\left[W_m - W\right] = \frac{1}{\mu^m}\frac{\mathrm{Var}\left[Z\right]}{\mu^2-\mu} =: \frac{C_Z}{\mu^m}.$$

Chebyshev's inequality then gives

$$\mathbf{P}[|W_m - W| > \mu^{-m/3}] \le C_Z \mu^{-m/3}.$$

Since $\mu^{-m/3} \le \mu^{-\varepsilon^{-\delta}/3} \le C_2 e^{-c_1/\varepsilon^{c_2}}$ for some positive constants C_2 and c_1, c_2, we have that

$$\frac{|\mathrm{F.O.}_m(\mathbf{0},\varepsilon) - g(p_c+\varepsilon)W|}{g(p_c+\varepsilon)} \le C_1 W \varepsilon^{1-\delta} + C_2 e^{-c_1/\varepsilon^{c_2}} \tag{3.11}$$

with probability at least $1 - C_Z\mu^{-m/3} = 1 - C_3 e^{-c_3/\varepsilon^{c_4}}$.

By computing the lower probabilities of W again, recall that there exist constants C_1' and c_2' so that

$$\mathbf{P}[W \le a] \le C_1' a^{c_2'}.$$

This implies that $C_2 e^{-c_1/\varepsilon^{c_2}} < C_1 W \varepsilon^{1-\delta}$ with probability at least $1 - Ce^{-c_2' c_1/\varepsilon^{c_2}}$. Relabeling constants, this means that for sufficiently small ε, we can upgrade (3.11) to

$$\frac{|\mathrm{F.O.}_m(\mathbf{0},\varepsilon) - g(p_c+\varepsilon)W|}{g(p_c+\varepsilon)} \le C_1 W \varepsilon^{1-\delta} \tag{3.12}$$

with probability at least $1 - e^{-c_1/\varepsilon^{c_2}}$.

The last piece is to bound $\mathrm{S.O.}_m(\mathbf{0},\varepsilon)/g(p_c+\varepsilon)$. By Fubini's theorem,

$$\begin{aligned}
\frac{\mathrm{S.O.}_m(\mathbf{0},\varepsilon)}{g(p_c+\varepsilon)} &= g(p_c+\varepsilon) \sum_{\substack{u,w\in\mathbf{T}_m \\ u\ne w}} (p_c+\varepsilon)^{2m-|u\wedge w|} \\
&\le 2g(p_c+\varepsilon)\sum_{j=0}^{m-1} p_c^{2m-j} \sum_{u,w:|u\wedge w|=j} 1 \\
&\le 2g(p_c+\varepsilon)\sum_{j=0}^{m-1} p_c^{j} \sum_{v\in\mathbf{T}_j}\sum_{1\le i<k} W_{m-j-1}^{(i)}(v) W_{m-j-1}^{(k)}(v) \\
&\le g(p_c+\varepsilon)\sum_{j=0}^{m-1} p_c^{j} \sum_{v\in\mathbf{T}_j} W_{m-j}(v)^2
\end{aligned}$$

where the second inequality is from the bound $(1+\frac{\varepsilon}{p_c})^{2m} \le 2$ for sufficiently small ε.

Note that for each j the innermost sum is a sum of IID random variables. We utilize the Fuk-Nagaev inequality from [FN71] which states

$$\mathbf{P}\left[\sum_{u\in\mathbf{T}_j}[W_{m-j}(u)^2 - \mathbf{E}W_{m-j}^2] > t \,\middle|\, Z_j\right] \le C_p t^{-p/2} Z_j^{p/4} + \exp\left(-C\frac{t^2}{Z_j}\right).$$

Applying this bound for $t = \mathbf{E}W_{m-j}^2 Z_j \varepsilon^{-2\ell}$ gives

$$\mathbf{P}\left[\sum_{u \in \mathbf{T}_j} [W_{m-j}(u)^2 - \mathbf{E}W_{m-j}^2] > (\mathbf{E}W_{m-j}^2) Z_j \varepsilon^{-2\ell} \,\middle|\, Z_j \right] \le C_p' \varepsilon^{p\ell} (Z_j)^{-p/4} + \exp\left(-C' Z_j / \varepsilon^{4\ell}\right)$$
$$\le C_p'' \varepsilon^{p\ell}$$

for some choice of $C_p'' > C_p'$. By applying this bound and a union bound, we get

$$\frac{\text{S.O.}_{\cdot m}(\mathbf{0}, \varepsilon)}{g(p_c + \varepsilon)} \le g(p_c + \varepsilon)(1 + \varepsilon^{-2\ell}) \sum_{j=0}^{m-1} \left(\mathbf{E}W_{m-j}^2\right) Z_j p_c^j \le C g(p_c + \varepsilon) \varepsilon^{-2\ell} \sum_{j=0}^{m-1} W_j$$

with probability at least $1 - C_p'' m \varepsilon^{p\ell}$ for some new choice of C. This means that

$$\mathbf{P}\left[\frac{\text{S.O.}_{\cdot m}(\mathbf{0}, \varepsilon)}{g(p_c + \varepsilon)} > C g(p_c + \varepsilon) \varepsilon^{-2\ell} \sum_{j=0}^{m-1} W_j \right] \le m C_p'' \varepsilon^{p\ell}.$$

Recalling that $g(p_c + \varepsilon) = \Theta(\varepsilon)$ now gives

$$\frac{\text{S.O.}_{\cdot m}(\mathbf{0}, \varepsilon)}{g(p_c + \varepsilon)} \le C_2 \varepsilon^{1-2\ell} \sum_{j=0}^{m-1} W_j$$

with probability at least $1 - C\varepsilon^{p\ell - \delta}$ for some new C. Along with equations (3.9) and (3.12), this now implies the proposition. □

From here, we extract the estimate that will be used to prove Theorem 1.1:

Corollary 3.8. *Suppose the offspring distribution of Z has $p > 1$ moments and $p_1 := \mathbf{P}[Z = 1]$. Let δ, ℓ, d be positive constants such that*

$$\alpha = 1 - 3\ell - (1 + d)\delta \tag{3.13}$$

is greater than $\frac{1}{2}$. Then there exists a constant $C > 0$ such that for all $\varepsilon > 0$ sufficiently small

$$|E(v, \varepsilon)| \le C W(v) \varepsilon^{\alpha} \tag{3.14}$$

for the root and its children with probability at least $1 - C\varepsilon^{\delta'}$ for $\delta' = \min\left\{p\ell - \delta, \frac{\log(1/p_1)}{\log(\mu)} d\delta\right\}$.

Proof: The first term in equation (3.8) is always eventually smaller than $W(v)\varepsilon^{\alpha}$ since the exponent on ε is larger. The final term in equation (3.8) can now be dealt with separately.

By [BD74, Theorems 0 and 5], if Z is in L^p, then $W_k \xrightarrow{L^p} W$, implying $\mathbf{E}[|W_k - W|^p] \le C$ for some $C > 0$. Therefore,

$$\mathbf{P}[|W_k - W| > \varepsilon^{-\ell}] \le C\varepsilon^{p\ell}.$$

For $m = \lceil \varepsilon^{-\delta} \rceil$, condition on Z_1, apply a union bound, and take expectation to see that

$$\sum_{k=1}^{m} W_k \le m(\varepsilon^{-\ell} + W)$$

for the root and all of its children with probability at least $1 - C(1+\mu)\varepsilon^{p\ell-\delta}$. Applying this to the latter term in equation (3.8) gives

$$|E(v,\varepsilon)| \le C_1 W(v)\varepsilon^{1-2\ell-\delta} + C_2\varepsilon^{1-3\ell-\delta}$$

with probability at least $1 - C\varepsilon^{p\ell-\delta}$.

In the case where $p_1 = 0$, the lower tails on W provided by [Dub71] show that for any $r_1, r_2 > 0$ we have $\mathbf{P}[W(v) < \varepsilon^{r_1}] = o(\varepsilon^{r_2})$, thereby showing $W(v) < \varepsilon^{r_1}$ with probability less than ε^{r_2} for ε sufficiently small. Setting $r_1 = d\delta$ and $r_2 = p\ell - \delta$ completes the proof when $p_1 = 0$.

When $p_1 > 0$, there exists a constant C so that for all $a \in (0,1)$

$$\mathbf{P}[W < a] \le Ca^{\log(1/p_1)/\log(\mu)}.$$

This implies that for α as in (3.13),

$$\mathbf{P}[W(v) < \varepsilon^{1-3\ell-\delta-\alpha}] = O\left(\varepsilon^{\frac{\log(1/p_1)}{\log(\mu)}d\delta}\right). \tag{3.15}$$

Performing a union bound for the root and all of its children again completes the proof. □

4 Pivot Sequence on the Backbone

Define the shift function $\theta : \Omega \to \Omega$ by

$$(\theta(\omega))_v := \omega_{\gamma_1 \sqcup v}\,. \tag{4.1}$$

Informally, θ shifts the values of random variables at nodes $\gamma_1 \sqcup v$ in $T(\gamma_1)$ back to node v; these values populate the whole Ulam tree; values of variables not in $T(\gamma_1)$ are discarded; this is a tree-indexed version of the shift for an ordinary Markov chain. The n-fold shift θ^n shifts n steps down the backbone.

The main purpose of this section will be to understand the shift function θ, and thereby understand the behavior of the pivots. While this section contains many intermediate results—a fair number of which may be of independent interest—only a handful will be directly of use in the proof of Theorem 1.1: the pair of Propositions 4.4(i) and 4.5 demonstrating that shifting down the backbone is the same as conditioning on the pivot being at most a certain value (this is step 2 in the outline in the introduction); also of use will be Theorem 4.9, which accomplishes step 3 of the outline by showing that $\beta_n^* - p_c$ approaches 0 rapidly.

4.1 Markov properties

Before showing the necessary Markov properties, a fair bit of notation is necessary. We begin with the definition of the dual pivots β_n^*; these variables will be central to the proof of Theorem 1.1, primarily due to their appearance in Proposition 4.4.

Definition 4.1 (dual trees and pivots)**.** *Recall that $T(v)$ denotes the subtree from v, moved to the root. Let $T^*(v)$ denote the rooted subtree induced on all vertices $w \notin T(v)$, and let $\beta^*_{v,w}$ represent the pivot of the vertex w on $T^*(v)$, that is, the least x such that w is connected to infinity by a path with weights $\le x$ that avoids going through v. The dual pivot β_v^* is defined to be $\min_{w<v} \beta^*_{v,w}$. In keeping with the notation for pivots, we denote $\beta_n^* := \beta^*_{\gamma_n}$.*

Definition 4.2. *We define the following σ-fields.*

(i) For fixed $v \neq \mathbf{0}$, define $\mathcal{C}_v$ to be the σ-field generated by $\deg_w$ and U_w for all $w \neq v$ in $T(v)$ along with $\deg_v$. Define $\mathcal{B}_v^*$ to be the σ-field generated by all the other data: U_w and $\deg_w$ for all $w \in \mathcal{T}^*(v)$, along with U_v.

(ii) For $n \geq 1$, let $\mathcal{B}_n^*$ denote the σ-field containing γ_n and all sets of the form $\{\gamma_n = v\} \cap B$ where $B \in \mathcal{B}_v^*$. Informally, $\mathcal{B}_n^*$ is generated by γ_n and $\mathcal{B}_{\gamma_n}^*$.

(iii) Let $\mathcal{C}_n$ be the σ-field generated by $\theta^n \omega$; in other words it contains $\deg(\gamma_n)$ and all pairs $(\deg_{\gamma_n \sqcup x}, U_{\gamma_n \sqcup x})$. It is not important, but this definition does not allow $\mathcal{C}_n$ to know the identity of γ_n.

It is elementary that $\{\mathcal{B}_n^*\}$ is a filtration, that $\mathcal{B}_n^* \cap \mathcal{C}_n$ is trivial, and that $\mathcal{B}_n^* \vee \mathcal{C}_n = \mathcal{F}$.

Definition 4.3. *We define the following conditioned measures.*

(i) For $x \in (p_c, 1)$, let $\mathbf{Q}_x := (\mathbf{P} \,|\, \beta_0 \leq x)$ denote the conditional law given $\mathbf{0} \leftrightarrow_x \infty$, in other words, $\mathbf{Q}_x[A] = \frac{g_A(x)}{g(x)}$ where

$$g_A(x) := \mathbf{P}[A \cap \{\beta_0 \leq x\}] \,.$$

(ii) Let $\mathcal{L}$ denote the law of β_0, the pivot at the root. By [Dur10, Theorem 5.1.9], one may define regular conditional distributions $\mathbf{P}_x := (\mathbf{P} \,|\, \beta_0 = x)$. These satisfy $\mathbf{P}_x[\beta_0 = x] = 1$ and $\int \mathbf{P}_x \, d\mathcal{L}(x) = \mathbf{P}$. Also, $\mathbf{Q}_y = (1/g(y)) \int \mathbf{P}_x \, d\mathcal{L}|_{[0,y]}(x)$.

A common null set for all the conditioned measures is the set where either the invasion ray is not well defined or $\beta(v) = \beta_v^*$ for some v. Statements such as (4.3) below are always interpreted as holding modulo this null set.

Proposition 4.4 (Markov property for dual pivots)**.**

(i) For any $A \in \mathcal{F}$,

$$\mathbf{P}[\theta^n \omega \in A \,|\, \mathcal{B}_n^*] = \mathbf{Q}_{\beta_n^*}[A] \,.$$

(ii) More generally, if $0 < y \leq 1$ then for any $A \in \mathcal{F}$,

$$\mathbf{Q}_y[\theta^n \omega \in A \,|\, \mathcal{B}_n^*] = \mathbf{Q}_{\beta_n^* \wedge y}[A] \,.$$

(iii) Under $\mathbf{P}$, the sequence $\{\beta_n^*\}$ is a time homogeneous Markov chain adapted to $\mathcal{B}_n^*$ with transition kernel $p(x, S) = \mathbf{Q}_x[\beta_1^* \wedge x \in S]$ and initial distribution δ_1.

Proof: (i) By definition of conditional probability, the conclusion is equivalent to $\mathbf{P}[\theta^n \omega \in A; G] = \int_G \mathbf{Q}_{\beta_n^*}[A] \, d\mathbf{P}$ for all $G \in \mathcal{B}_n^*$. Writing G as the countable union of $\mathcal{B}_n^*$-measurable sets $\bigcup_v (G \cap \{\gamma_n = v\})$, it suffices to verify the previous identity for each piece $G \cap \{\gamma_n = v\}$. By the definition of $\mathcal{B}_n^*$, each of these may be written as $\{\gamma_n = v\} \cap B^*$ for $B^* \in \mathcal{B}_v^*$. Thus, it suffices to prove

$$\int_{\{\gamma_n = v\} \cap B^*} \mathbf{Q}_{\beta_v^*}[A] \, d\mathbf{P} = \int_{\{\gamma_n = v\} \cap B^*} \mathbf{1}_A(\theta^n \omega) \, d\mathbf{P} \tag{4.2}$$

for all $v \in T_n$, $B^* \in \mathcal{B}_v^*$ and $A \in \mathcal{F}$.

Fixing v, identify $(\Omega, \mathcal{F}, \mathbf{P})$ as a product space $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1) \times (\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ where $\Omega_1 = (\mathbb{N} \times [0,1])^{T(v)\setminus\{v\}} \times \mathbb{N}$ and $\Omega_2 = (\mathbb{N} \times [0,1])^{T^*(v)} \times [0,1]$. Let $\pi_i : \Omega \to \Omega_i$ denote the coordinate maps; then π_2 is a measure preserving map on $(\Omega, \mathcal{B}_v^*, \mathbf{P})$ and π_1 is a measure preserving map on $(\Omega, \mathcal{C}_v, \mathbf{P})$. In particular, $B^* = \pi_2^{-1} B$ for some $B \in \mathcal{F}_2$.

Working on the left-hand side of (4.2), observe that $\beta \circ \pi_2 = \beta_v^*$ (where β is defined on Ω_2 by ignoring $\times[0,1]$) and hence, if we let H represent the event that the invasion percolation ever gets to v, then we have

$$\{\gamma_n = v\} = \{\beta(v) \leq \beta_v^*\} \cap H = \{\beta_0(\pi_1\omega) < \beta_v^*(\pi_2\omega)\} \cap H\,. \tag{4.3}$$

Using this, we obtain

$$\begin{aligned}
\int_{\{\gamma_n = v\}\cap B^*} \mathbf{Q}_{\beta_v^*}[A]\, d\mathbf{P} &= \int_\Omega \mathbf{1}_B(\pi_2\omega) \mathbf{1}_{\{\beta(\pi_1\omega) < \beta_v^*(\pi_2\omega)\}\cap H} \mathbf{Q}_{\beta_v^*(\pi_2\omega)}[A]\, d\mathbf{P} \\
&= \int_{\Omega_2} \mathbf{1}_B(\omega_2) \mathbf{Q}_{\beta_v^*(\omega_2)}[A] \left[\int_{\Omega_1} \mathbf{1}_{\{\beta(\omega_1) < \beta_v^*(\omega_2)\}\cap H_2}\, d\mathbf{P}_1(\omega_1)\right] d\mathbf{P}_2(\omega_2) \\
&= \int_{\Omega_2} \mathbf{1}_B(\omega_2) \frac{g_A(\beta_v^*(\omega_2))}{g(\beta_v^*(\omega_2))} \left[\int_{\Omega_1} \mathbf{1}_{\{\beta(\omega_1) < \beta_v^*(\omega_2)\}\cap H_2}\, d\mathbf{P}_1(\omega_1)\right] d\mathbf{P}_2(\omega_2)
\end{aligned}$$

where H_2 above denotes $\pi_2(H)$. The integral over Ω_1 is equal to $g(\beta_v^*(\omega_2))\mathbf{1}_{H_2}(\omega_2)$ so we may simplify and continue. Writing $g_A(x)$ as $\int_{\Omega_1} \mathbf{1}_{A\cap\{\beta_0 < x\}}(\omega)\, d\mathbf{P}_1(\omega)$ in the second line, (4.2) is finished as follows.

$$\begin{aligned}
&= \int_{\Omega_2} \mathbf{1}_B(\omega_2) g_A(\beta_v^*(\omega_2)) \mathbf{1}_{H_2}(\omega_2)\, d\mathbf{P}_2 \\
&= \int_{\Omega_1\times\Omega_2} \mathbf{1}_B(\omega_2) \mathbf{1}_A(\omega_1) \mathbf{1}_{\beta(\omega_1) < \beta_v^*(\omega_2)} \mathbf{1}_{H_2}(\omega_2)\, d\mathbf{P}_1(\omega_1)\, d\mathbf{P}_2(\omega_2) \\
&= \int_{\{\gamma_n = v\}\cap B^*} \mathbf{1}_A(\pi_1\omega)\, d\mathbf{P} \\
&= \int_{\{\gamma_n = v\}\cap B^*} \mathbf{1}_A(\theta^n\omega)\, d\mathbf{P}
\end{aligned}$$

because $\pi_1\omega = \theta^n\omega$ on $\{\gamma_n = v\}$.

(ii) Begin with the observation that

$$\{\beta_0 < y\} \cap \{\gamma_n = v\} = \{\beta(\pi_1\omega) < y\} \cap \{U_w < y \text{ for all } w \leq v\} \cap \{\beta(\pi_1\omega) < \beta_v^*(\pi_2\omega)\} \cap H\,. \tag{4.4}$$

As before, letting $G = \{\gamma_n = v\} \cap \pi_2^{-1}(B)$, we aim to prove the second identity in

$$\int_G \mathbf{Q}_{\beta_v^* \wedge y}[A]\, d\mathbf{Q}_y = \int_G \frac{g_A(\beta_v^* \wedge y)}{g(\beta_v^* \wedge y)}\, d\mathbf{Q}_y = \int_G \mathbf{1}_A(\theta^n\omega)\, d\mathbf{Q}_y\,, \tag{4.5}$$

the first being definitional. Also by definition, $\mathbf{Q}_y[\cdot] = (1/g(y))\mathbf{P}[\cdot \cap \{\beta_0 < y\}]$, whence, using (4.4), the left-hand side of (4.5) becomes

$$\frac{1}{g(y)} \int_{\Omega_1\times\Omega_2} \mathbf{1}_{B\cap H_2}(\omega_2)\, \mathbf{1}_{U_w < y \forall w \leq v}(\omega_2)\, \frac{g_A(\beta_v^*(\omega_2) \wedge y)}{g(\beta_v^*(\omega_2) \wedge y)}\, \mathbf{1}_{\beta(\omega_1) < y}\, \mathbf{1}_{\beta(\omega_1) < \beta_v^*(\omega_2)}\, d(\mathbf{P}_1 \times \mathbf{P}_2)\,.$$

Integrating over Ω_1 turns the last two indicator functions into $g(\beta_v^*(\omega_2) \wedge y)$, again canceling the denominator and yielding

$$\frac{1}{g(y)} \int_{\Omega_2} \mathbf{1}_B(\omega_2)\, \mathbf{1}_{U_w < y \forall w \leq v}(\omega_2)\, g_A(\beta_v^*(\omega_2) \wedge y) \mathbf{1}_{H_2}(\omega_2)\, d\mathbf{P}_2(\omega_2)\,.$$

Rewriting $g_A(x)$ as $\int_{\Omega_1} \mathbf{1}_{A\cap\{\beta_0<x\}}\, d\mathbf{P}_1$, this becomes

$$\frac{1}{g(y)} \int_\Omega \mathbf{1}_B(\pi_2\omega)\; \mathbf{1}_{U_w<y\forall w\le v}(\pi_2\omega)\; \mathbf{1}_{\beta(\pi_1\omega)<\beta_v^*(\pi_2\omega)}\; \mathbf{1}_{\beta(\pi_1\omega)<y}\; \mathbf{1}_A(\omega_1)\mathbf{1}_{H_2}(\omega_2)\; d\mathbf{P}\,.$$

Observing that the first, third and last indicator functions define G, this simplifies to

$$\begin{aligned}\frac{1}{g(y)} \int_{G\cap\{\beta_0(\omega)<y\}} \mathbf{1}_A(\pi_1\omega)\, d\mathbf{P} &= \frac{1}{g(y)} \int_G \mathbf{1}_A(\pi_1\omega)\mathbf{1}_{\beta_0<y}(\omega)\, d\mathbf{P} \\ &= \int_G \mathbf{1}_A(\pi_1\omega)\, d\mathbf{Q}_y \\ &= \int_G \mathbf{1}_A(\theta^n\omega)\, d\mathbf{Q}_y\end{aligned}$$

where the last inequality follows from the fact that $\gamma_n = v$ on G, which implies $\pi_1\omega = \theta^n\omega$. Hence, that completes the proof of (ii).

(iii) Begin by observing that $\gamma_{n+1} = v \sqcup j$ if and only if $\gamma_n = v$ and $\beta(v\sqcup j) < \beta_1^* \circ \pi_1$. In other words, given that the backbone contains v, the next backbone vertex depends only on $\theta^n\omega$ and is chosen in the same way the first backbone vertex of ω was chosen. From this it follows that

$$\beta_{n+1}^* = \beta_n^* \wedge \beta_1^* \circ \theta^n\,.$$

Therefore,

$$\begin{aligned}\mathbf{P}[\beta_{n+1}^* \in S \,|\, \mathcal{B}_n^*] &= \mathbf{P}[\beta_n^* \wedge (\beta_1^* \circ \theta^n) \in S \,|\, \mathcal{B}_n^*] \\ &= \mathbf{P}[\theta^n\omega \in \{\omega' : \beta_1^*(\omega') \wedge \beta_n^*(\omega) \in S\} \,|\, \mathcal{B}_n^*] \\ &= \mathbf{Q}_y[\beta_1^* \wedge y \in S]\,\big|_{\,y=\beta_n^*}\end{aligned}$$

as desired. □

Remark. Note that the final equality in the proof of (iii) does not immediately follow from the proof of (i) since $\{\omega' : \beta_1^*\omega'\wedge\beta_n^*\omega \in S\}$ is not a fixed set, but rather depends on β_n^*. Nevertheless, the proof of this equality follows from a slight modification of the proof of (i), where we simply replace the expressions $\mathbf{Q}_{\beta_v^*(\pi_2\omega)}[A]$, $g_A(\beta_v^*(\omega_2))$, and $\mathbf{1}_A(\pi_1\omega)$ by the expressions $\mathbf{Q}_{\beta_v^*(\pi_2\omega)}[\beta_1^* \wedge \beta_v^* \in S]$, $\mathbf{P}[\{\beta_0 < \beta_v^*(\omega_2)\} \cap \{\beta_1^* \wedge \beta_v^* \in S\}]$, and $\mathbf{1}_{\{\beta_1^*(\pi_1\omega)\wedge\beta_v^*(\omega)\in S\}}$ respectively.

It is immediate that $\mathbf{Q}_x \ll \mathbf{P}$ for all x. The following more quantitative statement will be useful, especially when used in conjunction with Proposition 4.4(i).

Proposition 4.5. *Let $q > 1$ and suppose that the offspring distribution has a finite q-moment. Then there exists a constant C_q such that for all $A \in \mathcal{T}$ and for all $\delta > 0$ and all $x \in (p_c, 1)$,*

$$\mathbf{P}[A] \le \delta \;\; \textit{implies} \;\; \mathbf{Q}_x[A] \le C_q \delta^{1-1/q}\,.$$

Proof: On $\mathcal{T}$, the density of $\mathbf{Q}_x$ with respect to $\mathbf{P}$ is given by

$$\frac{d\mathbf{Q}_x}{d\mathbf{P}}(T) = \frac{g(T,x)}{g(x)}\,.$$

Combining Corollary 3.2, which implies $g(x) \sim K(x-p_c)$, with Proposition 3.5, which shows $\int g(T, p_c + \varepsilon)^q \, d\mathtt{GW}(T) \le c_q \varepsilon^q$ provided $p_c + \epsilon$ is bounded away from 1, we see that

$$\int \left| \frac{d\mathbf{Q}_x}{d\mathbf{P}}(T) \right|^q \, d\mathtt{GW}(T) \le c'_q$$

for some constant c'_q and all $x \in (p_c, 1)$. Applying Hölder's inequality with $1/p + 1/q = 1$ then gives

$$\mathbf{Q}_x[A] = \int \mathbf{1}_A \frac{d\mathbf{Q}_x}{d\mathbf{P}} \, d\mathbf{P} \le \left[\int \mathbf{1}_A \, d\mathbf{P} \right]^{1/p} \left[\int \left(\frac{d\mathbf{Q}_x}{d\mathbf{P}} \right)^q \, d\mathbf{P} \right]^{1/q} \le C_q \delta^{1-1/q}$$

when $C_q = (c'_q)^{1/q}$. □

The measures $\mathbf{P}_x$ are in some sense more difficult to compute with than $\mathbf{Q}_x$ because of the conditioning on measure zero sets. Relations such as the Markov property, however, are conceptually somewhat simpler. The following statement of the Markov property generalizes what was proved in [AGdHS08, Theorem 1.2 and Proposition 3.1], with $\mathcal{B}_n^+$ representing the σ-field generated by $\mathcal{B}_n^*$ together with β_n. Note, however, that the only role Propositions 4.6, 4.7 and 4.8 play in the proof of Theorem 1.1 is that they are utilized to prove Theorem 4.9. The proposition below is also of independent interest, and will be crucial for studying the forward maximal weight process in Section 6.

Proposition 4.6 (Markov property for pivots). *For any $A \in \mathcal{F}$,*

$$\mathbf{P}\left[\theta^n \omega \in A \,|\, \mathcal{B}_n^+\right] = \mathbf{P}_{\beta_n}[A]$$

on $(\Omega, \mathcal{F}, \mathbf{P}_x)$.

Proof: By definition of conditional probability, the conclusion is equivalent to

$$\mathbf{P}[B \cap \{\theta^n \omega \in A\}] = \int_B \mathbf{P}_{\beta_n}[A] \, d\mathbf{P} \tag{4.6}$$

holding for all $B \in \mathcal{B}_n^+$ and $A \in \mathcal{F}$. It is enough to prove (4.6) for sets that are subsets of $\{\gamma_n = v\}$ for some v: if it holds for sets of this form, then

$$\begin{aligned}
\mathbf{P}[B \cap \{\theta^n \omega \in A\}] &= \sum_v \mathbf{P}[B \cap \{\gamma_n = v\} \cap \{\theta^n \omega \in A\}] \\
&= \sum_v \int_{B \cap \{\gamma_n = v\}} \mathbf{P}_{\beta_n}[A] \, d\mathbf{P} \\
&= \int_B \mathbf{P}_{\beta_n}[A] \, d\mathbf{P} \, .
\end{aligned}$$

We now fix v and assume without loss of generality that $B \subseteq \{\gamma_n = v\}$. The identity (4.6) we need to prove now reduces to

$$\mathbf{P}[B \cap \{\sigma^v \omega \in A\}] = \int_B \mathbf{P}_{\beta(v)}[A] \, d\mathbf{P} \tag{4.7}$$

and we need to show it holds for all $B \in \mathcal{B}_v^+$ where $\mathcal{B}_v^+$ denotes the σ-field generated by $\mathcal{B}_v^*$ together with β_v and σ^v denotes shifting to v.

We claim it is enough to prove (4.7) for sets B of the form $B_1 \times B_2$ where $B_2 = \{\beta(v) \le b\}$ and B_1 is an element of $\mathcal{B}_v^*$ contained in the event $\{\beta_v^* \ge a\}$ for real numbers $0 < b < a < 1$. To see why this is enough, observe that the set of all B for which (4.7) holds is a λ-system, meaning it is closed under increasing union and set theoretic difference of nested sets. The class of sets of the form $B_1 \times B_2$ above are closed under intersection, whence by Dynkin's Theorem [Dur10, Theorem 2.1.2], if (4.6) holds when B is in this class, then it holds for all B in the σ-field generated by this class, which is $\mathcal{B}_v^+$.

Working on the left-hand side of (4.7),

$$\begin{aligned} \mathbf{P}[B \cap \{\sigma^v \omega \in A\}] &= \mathbf{P}[B_1 \cap \{\sigma^v \omega \in A, \beta(v) \le b\}] \\ &= \mathbf{P}[B_1]\mathbf{P}[\sigma^v \omega \in A, \beta(v)(\omega) \le b] \\ &= \mathbf{P}[B_1]\mathbf{P}[\sigma^v \omega \in A \cap \{\beta_0 \le b\}] \\ &= \mathbf{P}[B_1]\mathbf{P}[A \cap \{\beta_0 \le b\}] \,. \end{aligned}$$

Here, the first equality is definitional, the second uses independence of $\mathcal{B}_v^*$ and $\beta(v)$, the third uses $\beta(v)(\omega) \le b$ if and only if $\beta(\mathbf{0})(\sigma^v \omega) \le b$, and the last holds because σ^v preserves the measure $\mathbf{P}$.

Working on the right-hand side of (4.7), identify $(\Omega, \mathcal{F}, \mathbf{P})$ as a product $(\Omega^{(1)}, \mathcal{B}_v^*, \mathbf{P}^{(1)}) \times (\Omega^{(2)}, \mathcal{C}_v, \mathbf{P}^{(2)})$ in the obvious way and compute

$$\begin{aligned} \int_B \mathbf{P}_{\beta(v)}[A] \, d\mathbf{P} &= \left(\int_{B_1} 1 \, d\mathbf{P}_1 \right) \cdot \left(\int_{\beta(v) \le b} \mathbf{P}_{\beta(v)}[A] \, d\mathbf{P}_2 \right) \\ &= \mathbf{P}[B_1] \int \mathbf{P}_x[A \cap \{\beta_0 \le b\}] \, d\mathcal{L}(x) \\ &= \mathbf{P}[B_1]\mathbf{P}[A \cap \{\beta_0 \le b\}] \end{aligned}$$

because the $\mathbf{P}^{(2)}$-law of $\beta(v)$ is $\mathcal{L}$. This finishes the proof. $\square$

Proposition 4.7. *The sequence $\{\beta_n, \beta_n^*\}$ is a time-homogeneous Markov chain adapted to $\{\mathcal{B}_n^+\}$ with initial distribution $\mathcal{L} \times \delta_1$.*

Proof: For any Borel $A \subseteq [0,1] \times [0,1]$ and any $x, y \in [0,1]$, define

$$\begin{aligned} \tilde{A} &:= \{\omega \in \mathcal{F} : (\beta_1(\omega), \beta_1^*(\omega)) \in A\} \\ \tilde{A}_y &:= \{\omega \in \mathcal{F} : (\beta_1(\omega), \beta_1^*(\omega) \wedge y) \in A\} \\ \text{and } \mathbf{P}_{x,y}[A] &:= \mathbf{P}_x[\tilde{A}_y] \,. \end{aligned}$$

To show that $\{\beta_n, \beta_n^*\}$ is a time homogeneous Markov chain adapted to $\mathcal{B}_n^+$, it will suffice to show that for any Borel $A \subseteq [0,1] \times [0,1]$ and any $B \in \mathcal{B}_n^+$,

$$\mathbf{P}\left[\left\{(\beta_{n+1}, \beta_{n+1}^*) \in A\right\} \cap B\right] = \int_B \mathbf{P}_{\beta_n, \beta_n^*}[A] \, d\mathbf{P} \,. \tag{4.8}$$

Starting with the case where $A = [0,x] \times [0,y]$ for $x, y \in (0,1]$, we have

$$\begin{aligned} \left\{(\beta_{n+1}, \beta_{n+1}^*) \in A\right\} \cap B &= \{\beta_{n+1} \le x\} \cap \left\{\beta_{n+1}^* \le y\right\} \cap B \\ &= \{\beta_{n+1} \le x\} \cap (\{\beta_n^* \le y\} \cap B) \cup \left(\{\beta_{n+1} \le x\} \cap \left\{\beta_{n+1}^* \le y\right\}\right) \cap (\{\beta_n^* > y\} \cap B) \\ &= \left\{\omega : \theta^n \omega \in \tilde{A}'\right\} \cap (\{\beta_n^* \le y\} \cap B) \cup \left\{\omega : \theta^n \omega \in \tilde{A}\right\} \cap (\{\beta_n^* > y\} \cap B) \end{aligned}$$

where $A' := [0,x] \times [0,1]$. It now follows from Proposition 4.6 that

$$\begin{aligned}
\mathbf{P}\left[\left\{(\beta_{n+1}, \beta^*_{n+1}) \in A\right\} \cap B\right] &= \int_{\{\beta^*_n \le y\} \cap B} \mathbf{P}_{\beta_n}[\tilde{A}']\, d\mathbf{P} + \int_{\{\beta^*_n > y\} \cap B} \mathbf{P}_{\beta_n}[\tilde{A}]\, d\mathbf{P} \\
&= \int_B \mathbf{P}_{\beta_n}[\tilde{A}'] \mathbf{1}_{\beta^*_n < y} + \mathbf{P}_{\beta_n}[\tilde{A}] \mathbf{1}_{\beta^*_n > y}\, d\mathbf{P} \\
&= \int_B \mathbf{P}_{\beta_n, \beta^*_n}[A]\, d\mathbf{P}\,.
\end{aligned}$$

Hence, (4.8) has been proven for all A of the form $[0,x] \times [0,y]$. Since the set of all A for which (4.8) holds is a λ-system, and the collection of all sets of the form $[0,x] \times [0,y]$ is a π-system that generates the Borel sets in $[0,1] \times [0,1]$, it now follows from Dynkin's π-λ Theorem that (4.8) holds for all Borel sets A. $\square$

4.2 Decay of $\beta^*_n - p_c$

To study the decay rate of β^*_n, we work with the pair (β_n, β^*_n) rather than β^*_n individually. In this vein, we show that the pair $(\beta_n - p_c, \beta^*_n - p_c)$ is Markov, and compute the transition probabilities.

Proposition 4.8. *Define $h^*_n := \beta^*_n - p_c$ and $f(x) := \phi'(1-(p_c+x)g(p_c+x))$. Then $\{h_n, h^*_n\}$ has transition probabilities given by $p(\{a,b\},\cdot) = \nu_a \times \tilde{\nu}_{a,b}$ where*

$$\begin{aligned}
\frac{d\nu_a}{dx} &= \frac{f(a)g'(p_c+x)}{g'(p_c+a)} \mathbf{1}_{x<a} + C_a \delta_a \\
\text{and} \quad \frac{d\tilde{\nu}_{a,b}}{dx} &= -\frac{f'(x)}{f(a)} \mathbf{1}_{a<x<b} + \tilde{C}_{a,b}\delta_b
\end{aligned}$$

with $C_a = f(a)(p_c+a)$ and $\tilde{C}_{a,b} = \frac{f(b)}{f(a)}$.

Proof: Note first that $g(\cdot)$ is differentiable on $(p_c, 1)$ as described in Corollary 3.2.

(i) Since $h^*_{n+1} = \min\{\beta^*_n - p_c, \beta^*_{\gamma_{n+1},\gamma_n} - p_c\}$, it follows that

$$\begin{aligned}
&\mathbf{P}[h_{n+1} \in (a, a+da), h^*_{n+1} \in (b, b+db) \,|\, h_n \in (a, a+da), h^*_n \in (b, b+db)] \\
&= \mathbf{P}[\beta_{n+1} \in (p_c+a, p_c+a+da), \beta^*_{\gamma_{n+1},\gamma_n} > p_c+b \,|\, \beta_n \in (p_c+a, p_c+a+da)] \\
&= \mathbf{P}[\beta_1 \in (p_c+a, p_c+a+da), \beta^*_{\gamma_1,\mathbf{0}} > p_c+b \,|\, \beta_0 \in (p_c+a, p_c+a+da)] \\
&= \frac{\left(\sum_{k=1}^{\infty} \mathbf{P}[Z=k] k(p_c+a) g'(p_c+a)(1-(p_c+b)g(p_c+b))^{k-1}\, da\right) + o(da)}{g'(p_c+a)\, da + o(da)} \\
&= (p_c+a)\phi'(1-(p_c+b)g(p_c+b)) + o(1) \\
&= (p_c+a)f(b) + o(1).
\end{aligned}$$

(ii) For $a < x < b$ we have

$$\begin{aligned}
&\mathbf{P}[h_{n+1} \in (a, a+da), h^*_{n+1} \in (x, x+dx) \,|\, h_n \in (a, a+da), h^*_n (b, b+db)] \\
&= \mathbf{P}[\beta_1 \in (p_c+a, p_c+a+da), \beta^*_{\gamma_1,\mathbf{0}} \in (p_c+x, p_c+x+dx) \,|\, \beta_0 \in (p_c+a, p_c+a+da)] \\
&= \frac{\left(\sum_{k=2}^{\infty} \mathbf{P}[Z=k] k(p_c+a) g'(p_c+a) \left(-\frac{d}{dx}(1-(p_c+x)g(p_c+x))^{k-1}\right) dx\, da\right) + o(dx\, da)}{g'(p_c+a) da + o(da)}
\end{aligned}$$

$$
\begin{aligned}
&= (p_c + a)(g(p_c + x) + (p_c + x)g'(p_c + x))\phi''(1 - (p_c + x)g(p_c + x))\, dx + o(dx) \\
&= -(p_c + a) f'(x)\, dx + o(dx)
\end{aligned}
$$

where we used $f'(x) = -\left[g(p_c + x) + (p_c + c)g'(p_c + x)\right]\phi''(1 - (p_c + x)g(p_c + x))$.

(iii) For $z < a < b$ we have

$$
\begin{aligned}
&\mathbf{P}[h_{n+1} \in (z, z + dz), h^*_{n+1} \in (b, b + db) \,|\, h_n \in (a, a + da), h^*_n \in (b, b + db)] \\
&\quad = \mathbf{P}[\beta_1 \in (p_c + z, p_c + z + dz), \beta^*_{\gamma_1,\mathbf{0}} > b \,|\, \beta_0 \in (p_c + a, p_c + a + da)] \\
&\quad = \frac{\left(\sum_{k=1}^{\infty} \mathbf{P}[Z = k] k g'(p_c + z)(1 - (p_c + b)g(p_c + b))^{k-1}\, dz\, da\right) + o(dz\, da)}{g'(p_c + a)\, da + o(da)} \\
&\quad = \frac{g'(p_c + z)}{g'(p_c + a)} \phi'(1 - (p_c + b)g(p_c + b))\, dz + o(dz) \\
&\quad = \frac{g'(p_c + z)}{g'(p_c + a)} f(b)\, dz + o(dz).
\end{aligned}
$$

(iv) For $z < a < x < b$, we have

$$
\begin{aligned}
&\mathbf{P}[h_{n+1} \in (z, z + dz), h^*_{n+1} \in (x, x + dx) \,|\, h_n \in (a, a + da), h^*_n \in (b, b + da)] \\
&\quad = \mathbf{P}[\beta_1 \in (p_c + z, p_c + z + dz), \beta^*_{\gamma_1,\mathbf{0}} \in (p_c + x, p_c + x + dx) \,|\, \beta_0 \in (p_c + a, p_c + a + da)] \\
&\quad = \frac{\left(\sum_{k=2}^{\infty} \mathbf{P}[Z = k] k g'(p_c + z) \left(-\frac{d}{dz}\left(1 - (p_c + x)g(p_c + x)\right)^{k-1}\right) da\, dx\, dz\right) + o(da\, dx\, dz)}{g'(p_c + a)\, da + o(da)} \\
&\quad = \frac{g'(p_c + z)}{g'(p_c + a)} \left(-\frac{d}{dx}(\phi'(1 - (p_c + x)g(p_x + x)))\right) dx\, dz + o(dx\, dz) \\
&\quad = -\frac{g'(p_c + z)}{g'(p_c + a)} f'(x)\, dx\, dz + o(dx\, dz)\,.
\end{aligned}
$$

Noting that $h_n < h^*_n$, it follows from $(i), (ii), (iii)$, and (iv) that

$$
\begin{aligned}
p\left(\{a, b\}, \{A, B\}\right) = &(p_c + a)\mathbf{1}_{a \in A} \cdot f(b)\mathbf{1}_{b \in B} \\
&- \left(\int_B \mathbf{1}_{a<x<b} f'(x)\, dx\right)(p_c + a)\mathbf{1}_{a \in A} + \left(\int_A \mathbf{1}_{z<a} \frac{g'(p_c + z)}{g'(p_c + a)}\, dz\right) f(b)\mathbf{1}_{b \in B} \\
&- \int_{A \times B} \mathbf{1}_{a<x<b} f'(x) \mathbf{1}_{z<a} \frac{g'(p_c + z)}{g'(p_c + a)}\, dx\, dz\,.
\end{aligned}
$$

Hence, we see that $p(\{a, b\}, \cdot) = \mu_a \times \tilde{\mu}_{a,b}$ where

$$
\begin{aligned}
\frac{d\mu_a}{dx} &= \mathbf{1}_{x<a} \frac{g'(p_c + x)}{g'(p_c + a)} + (p_c + a)\delta_a \quad \text{and} \\
\frac{d\tilde{\mu}_{a,b}}{dx} &= -\mathbf{1}_{a<x<b} \cdot f'(x) + f(b)\delta_b\,.
\end{aligned}
$$

Noting that $\mu_a((0, a]) = \frac{g(p_c+a)}{g'(p_c+a)} + (p_c + a) = \frac{1}{f(a)}$ and $\tilde{\mu}_{a,b}((a, b]) = f(a)$, we define probability measure $\nu_a = f(a)\mu_a$ and $\tilde{\nu}_{a,b} = \frac{1}{f(a)}\tilde{\mu}_{a,b}$, and we see that ν_a and $\tilde{\nu}_{a,b}$ satisfy the statement in the proposition. $\square$

The decay rate of $\beta^*_n - p_c = h^*_n$ follows from analyzing this Markov chain; the following Theorem accomplishes Step 3 of the outline.

Theorem 4.9. *There exists $C > 0$ such that for any $t \in (1/2, 1)$, $\mathbf{P}[h_n^* > n^{-t}]$ is $O(e^{-Cn^{1-t}})$.*

Proof: We start by looking at $\frac{d\nu_a}{dx}$. First we want to show that $\frac{f(a)}{g'(p_c+a)}$ is bounded below by something positive. Noting that $1 - g(p)$ is the unique non trivial fixed point of $\phi_p(x) := \phi(px + 1 - p)$, it follows from implicit differentiation that

$$g'(p_c + x) = \frac{g(p_c + x)\phi'(1 - (p_c + x)g(p_c + x))}{1 - (p_c + x)\phi'(1 - (p_c + x)g(p_c + x))}$$

which then implies that

$$\frac{f(a)}{g'(p_c + a)} = \frac{1 - (p_c + a)\phi'(1 - (p_c + a)g(p_c + a))}{g(p_c + a)}. \tag{4.9}$$

As $a \to 0$, this expression equals

$$\begin{aligned}\frac{1 - (p_c + a)(\mu - \phi''(1)(p_c + a)g(p_c + a) + o(g(p_c + a)))}{g(p_c + a)} &= \frac{p_c\phi''(1)(p_c + a)g(p_c + a) - a\mu + o(a)}{g(p_c + a)} \\ &= p_c^2\phi''(1) - \frac{\mu}{g'(p_c)} + o(1) \\ &= p_c^2\phi''(1) - \frac{p_c^2\phi''(1)}{2} + o(1) \\ &= \frac{p_c^2\phi''(1)}{2} + o(1).\end{aligned}$$

Hence, there must exist some $r' > 0$ such that if $a < r'$ then (4.9) is greater than $\frac{p_c^2\phi''(1)}{3}$. Next observe that the numerator of (4.9) is equal to one minus the derivative of $\phi_p(x)$ evaluated at the fixed point $1 - g(p)$ (where $p = p_c + a$) from which it follows that this numerator, and therefore (4.9) itself, is positive whenever $a > 0$. Since we can also see that (4.9) is continuous on the compact set $[r', 1-p_c]$, it follows that (4.9) must be bounded below by some value $C' > 0$ on $[r', 1-p_c]$. Now setting $C'' = \min\{C', \frac{p_c^2\phi''(1)}{3}\}$, we find that (4.9) is greater than or equal to C'' on $[0, 1-p_c]$. Finally, if we couple this with the fact that $g'(p_c + x) \to \frac{2}{p_c^3\phi''(1)}$ as $x \to 0$, which in turn implies that $\exists\ r > 0$ such that $g'(p_c + x) > \frac{1}{p_c^3\phi''(1)}$ on $[0, r]$, we find that if we set $\tilde{C} = \frac{C''}{p_c^3\phi''(1)}$, then for any x, a where $x < r$ and $x < a$ we have $\frac{d\nu_a}{dx} \geq \tilde{C}$.

Now turning our focus towards $\frac{d\tilde{\nu}_{a,b}}{dx}$, observe that in the expression $\frac{-f'(x)}{f(a)}$, the numerator goes to $2\mu^2$ as $x \to 0$, which can be seen by differentiating and noting that $g'(p_c) = \frac{2}{p_c^3\phi''(1)}$; additionally, the denominator goes to μ as $a \to 0$. Hence, the ratio goes to 2μ as $x \to 0, a \to 0$. From this it follows that $\exists\ \ell > 0$ such that if $a < x < \ell$ and $x < b$ then $\frac{d\tilde{\nu}_{a,b}}{dx} > \mu$. Next we note the following string of inequalities.

$$\begin{aligned}\mathbf{P}[h_n^* \geq n^{-t}] &= \mathbf{P}\left[h_{\lfloor \frac{n}{2} \rfloor} \geq \frac{n^{-t}}{2}\right]\mathbf{P}\left[h_n^* \geq n^{-t} \,\middle|\, h_{\lfloor \frac{n}{2} \rfloor} \geq \frac{n^{-t}}{2}\right] + \mathbf{P}\left[h_{\lfloor \frac{n}{2} \rfloor} < \frac{n^{-t}}{2}\right]\mathbf{P}\left[h_n^* \geq n^{-t} \,\middle|\, h_{\lfloor \frac{n}{2} \rfloor} < \frac{n^{-t}}{2}\right] \\ &\leq \mathbf{P}\left[h_{\lfloor \frac{n}{2} \rfloor} \geq \frac{n^{-t}}{2}\right] + \mathbf{P}\left[h_n^* \geq n^{-t} \,\middle|\, h_{\lfloor \frac{n}{2} \rfloor} < \frac{n^{-t}}{2}\right] \\ &\leq \mathbf{P}\left[h_{\lfloor \frac{n}{2} \rfloor} \geq \frac{n^{-t}}{2}\right] + \prod_{j=\lceil \frac{n}{2} \rceil}^{n} \mathbf{P}\left[h_j^* \geq n^{-t} \,\middle|\, h_{\lfloor \frac{n}{2} \rfloor} < \frac{n^{-t}}{2}, h_{j-1}^* \geq n^{-t}\right].\end{aligned} \tag{4.10}$$

Now using (4.10), along with Proposition 4.8 and the results from the previous paragraph, we find that if $\frac{n^{-t}}{2} < r$ and $n^{-t} < \ell$, then $\mathbf{P}[h_n^* \geq n^{-t}] \leq \left(1 - \frac{\tilde{C}}{2}n^{-t}\right)^{\lfloor \frac{n}{2} \rfloor} + \left(1 - \frac{\mu}{2}n^{-t}\right)^{\lceil \frac{n}{2} \rceil}$. Defining $C = \min\left\{\frac{\tilde{C}}{4}, \frac{\mu}{4}\right\}$, we finally get that $\mathbf{P}[h_n^* \geq n^{-t}]$ is $O\left(e^{-Cn^{1-t}}\right)$, thus completing the proof. □

5 Proof of Theorem 1.1

This section is devoted to the proof of Theorem 1.1. For a non-root vertex $v \in \mathbf{T}_{n+1}$ with $|v| = n+1$ and $p > p_c$, define

$$\tilde{q}(v,p) := \mathbf{Q}_p[\sigma_{\overleftarrow{v}}^{-1}\{v = \gamma_1\}]. \tag{5.1}$$

In words, $\tilde{q}(v,p)$ considers the tree rooted at $\overleftarrow{v}$ and finds the probability that v is in the backbone conditioned on the root having pivot at most p. We then have $q(v) = \mathbf{E}_*^{(n)}[\tilde{q}(v, \beta_n^*)]$, where β_n^* is as defined in Definition 4.1 and $\mathbf{E}_*^{(n)} := \mathbf{E}[\cdot|\mathcal{T}, \gamma_n]$.

5.1 Comparing $\tilde{q}$ and the ratio of survival functions

The goal of this section is to accomplish step 4 of the outline. This takes the form of

Lemma 5.1. *Let $\{w_k\}_{k=1}^d$ be an enumeration of the children of v. Then for any $p > p_c$ and j,*

$$\left|\tilde{q}(w_j,p) - \frac{g(T(w_j),p)}{\sum_{k=1}^d g(T(w_k),p)}\right| \leq \frac{g(T(v),p)}{1-g(T(v),p)} \cdot \frac{g(T(w_j),p)}{\sum_{k=1}^d g(T(w_k),p)}. \tag{5.2}$$

Proof: Define

$$A_j = \tilde{q}(w_j,p) - \frac{g(T(w_j),p)}{\sum_{k=1}^d g(T(w_k),p)}$$

and write

$$\begin{aligned}
\tilde{q}(w_j,p) &= \frac{\mathbf{P}_*[U_{w_j} \vee \beta(w_j) \text{ is smallest } \mid \beta(v) \leq p]}{\sum_{i=1}^d \mathbf{P}_*[U_{w_i} \vee \beta(w_i) \text{ is smallest } \mid \beta(v) \leq p]} \\
&= \frac{\mathbf{P}_*[U_{w_j} \vee \beta(w_j) \text{ smallest and } U_{w_j} \vee \beta(w_j) \leq p]}{\sum_{i=1}^d \mathbf{P}_*[U_{w_i} \vee \beta(w_i) \text{ smallest and } U_{w_i} \vee \beta(w_i) \leq p]}. \qquad (5.3)
\end{aligned}$$

For each j, we observe that

$$\begin{aligned}
p \cdot g(T(w_j),p)(1-B_j) &\leq \mathbf{P}_*[U_{w_j} \vee \beta(w_j) \text{ smallest and } U_{w_j} \vee \beta(w_j) \leq p] \\
&\leq p \cdot g(T(w_j),p)
\end{aligned}$$

where $1 - B_j = \prod_{1 \leq i \neq j \leq d}(1 - pg(T(w_i),p))$. The upper bound is the probability that $U_{w_j} \vee \beta(w_j) \leq p$, while the lower bound is the probility that $U_{w_j} \vee \beta(w_j) \leq p$, and that this does not hold for any of the siblings of w_j.

This gives the bounds

$$\frac{g(T(w_j),p)(1-B_j)}{\sum_{k=1}^d g(T(w_k),p)} \leq \tilde{q}(w_j,p) \leq \frac{g(T(w_j),p)}{\sum_{k=1}^d g(T(w_k),p)(1-B_k)}. \tag{5.4}$$

Sandwich bounds on the difference with survival ratios follow:

$$\frac{-B_j g(T(w_j),p)}{\sum_{k=1}^d g(T(w_k),p)} \leq A_j \leq \frac{\sum_{k=1}^d [g(T(w_k),p)g(T(w_j),p)B_k]}{(\sum_{k=1}^d g(T(w_k),p))(\sum_{k=1}^d [g(T(w_k),p)(1-B_k)])}. \tag{5.5}$$

Finally, the simple bound of

$$B_k \leq 1 - \prod_{i=1}^{d}(1 - pg(T(w_i), p)) = g(T(v), p)$$

allows us to rewrite equation (5.5) as

$$-\frac{g(T(v),p)g(T(w_j),p)}{\sum_{k=1}^{d} g(T(w_k),p)} \leq A_j \leq \frac{g(T(v),p)}{1-g(T(v),p)} \frac{g(T(w_j),p)}{\sum_{k=1}^{d} g(T(w_k),p)} . \tag{5.6}$$

□

5.2 Completing the Argument

The main ingredients for showing that p and q are close are in place: Corollary 3.8 bounds the fluctuations of $g(\mathbf{T}, \cdot)$, which will allow us to complete step 5 of the outline; Lemma 5.1 shows that $\tilde{q}$ is close to the ratio of survival probabilities for a fixed p (step 4); and Propositions 4.4(i), 4.5 and Theorem 4.9 will allow us to translate bounds for a fixed p into a bound for the random variable β_n^* (steps 3 and 2 respectively). We now put these pieces together for one final bound:

Proposition 5.2. *Letting* $q := \frac{\log(\mu)}{\log(1/p_1)}$, *if*

$$2p^2q^2 + (3p^2 + 5p)q + (-p^2 + 11p - 4) < 0, \tag{5.7}$$

then there exists $M > 0$ *and* $t \in (1/2, 1)$ *such that, with probability* 1, *the set*

$$\bigcup_{n=1}^{\infty} \left\{ v \in \mathbf{T}_{n+1} : \left| \frac{q(v)}{p(v)} - 1 \right| > 3Mn^{-t}, \overleftarrow{v} = \gamma_n \right\}$$

is finite.

Proof: Define $\mathbf{P}_*^{(n)} := \mathbf{P}[\cdot \mid \mathcal{T}, \gamma_n]$; we start by noting that for a tree T and vertex v, $\left|\frac{q(v)}{p(v)} - 1\right| = \left|\mathbf{E}_*^{(n)} \left[\frac{\tilde{q}(v,\beta_n^*)}{p(v)} - 1\right]\right|$. Now define

$$A_n := \left\{ h_n^* \leq n^{-\frac{t}{\alpha}} \right\} \bigcap_{v:\overleftarrow{v}=\gamma_n} \left\{ \frac{1}{p(v)} \left| \tilde{q}(v, \beta_n^*) - \frac{g(T(v), \beta_n^*)}{\sum\limits_{w:\overleftarrow{w}=\gamma_n} g(T(w), \beta_n^*)} \right| \leq 3Cn^{-t} \right\} \bigcap_{v:\overleftarrow{v}=\gamma_n} \left\{ \left| \frac{\frac{g(T(v),\beta_n^*)}{\sum g(T(w),\beta_n^*)}}{p(v)} - 1 \right| \leq 3n^{-t} \right\}$$

where C is as in Corollary 3.8 and $\frac{1}{2} < t < \alpha < 1$ with $\alpha = 1 - 3\ell - (1+d)\delta$. Observing that

$$\left| \frac{q(v)}{p(v)} - 1 \right| \leq \mathbf{E}_*^{(n)} \left[\left| \frac{\tilde{q}(v, \beta_n^*)}{p(v)} - 1 \right| \cdot \mathbf{1}_{A_n} \right] + \mathbf{P}_*^{(n)}[A_n^c] \max\left\{ \frac{1}{p(v)}, 1 \right\} \leq (3C+3)n^{-t} + \mathbf{P}_*^{(n)}[A_n^c] \max\left\{ \frac{1}{p(v)}, 1 \right\},$$

we see that to complete the proof we simply need to establish the following claim.

Claim: With probability 1, $\mathbf{P}_*^{(n)}[A_n^c] \max\left\{\frac{1}{p(v)}, 1\right\} > n^{-t}$ for only finitely many children of the backbone.

Proof of Claim: Recall first that $p(v)=\frac{W(v)}{\sum W(w)}$. Then compute

$$\left|\frac{\frac{g(\mathbf{T}(v),p_c+\epsilon)}{\sum g(\mathbf{T}(w),p_c+\epsilon)}}{p(v)}-1\right|=\left|\frac{\frac{W(v)+E(v,\epsilon)}{\sum W(w)+E(w,\epsilon)}}{\frac{W(v)}{\sum W(w)}}-1\right|=\left|\frac{E(v,\epsilon)\sum W(w)-W(v)\sum E(w,\epsilon)}{W(v)\sum[W(w)+E(w,\epsilon)]}\right|. \tag{5.8}$$

Now using Corollary 3.8, we see that if we start with a fresh Galton-Watson tree then

$$\begin{aligned}\left|\frac{\frac{g(\mathbf{T}(v),p_c+n^{-\frac{t}{\alpha}})}{\sum g(\mathbf{T}(w),p_c+n^{-\frac{t}{\alpha}})}}{p(v)}-1\right| &\le C\frac{W(v)n^{-t}\sum W(w)+\sum W(w)n^{-t}W(v)}{W(v)\sum W(w)(1-Cn^{-t})}\\ &=2C\frac{n^{-t}}{1-Cn^{-t}}\end{aligned} \tag{5.9}$$

for every child v of the root, with probability at least $1-Cn^{-\frac{t}{\alpha}\delta'}$. If we now condition on $h_n^*\le n^{-\frac{t}{\alpha}}$ and combine (5.9) with Propositions 4.4(i) and 4.5 we find that

$$\left|\frac{\frac{g(\mathbf{T}(v),\beta_n^*)}{\sum g(\mathbf{T}(w),\beta_n^*)}}{p(v)}-1\right|\le 2C\frac{n^{-t}}{1-Cn^{-t}} \tag{5.10}$$

for every v such that $\overleftarrow{v}=\gamma_n$ with probability at least $1-Cn^{-\frac{t}{\alpha}(1-\frac{1}{p})\delta'}$.

For the next step, recall that Lemma 5.1 shows

$$\frac{1}{p(v)}\left|\tilde{q}(v,\beta_n^*)-\frac{g(\mathbf{T}(v),\beta_n^*)}{\sum\limits_{w:\overleftarrow{w}=\gamma_n}g(\mathbf{T}(w),\beta_n^*)}\right|\le\frac{g(\mathbf{T}(\gamma_n),\beta_n^*)}{1-g(\mathbf{T}(\gamma_n),\beta_n^*)}\frac{\frac{g(\mathbf{T}(v),\beta_n^*)}{\sum g(\mathbf{T}(w),\beta_n^*)}}{p(v)}. \tag{5.11}$$

Using (5.10), we see that when we condition on $h_n^*\le n^{-\frac{t}{\alpha}}$, the latter fraction in (5.11) is bounded by, say 2, for every child of γ_n, with probability at least $1-Cn^{-\frac{t}{\alpha}(1-\frac{1}{p})\delta'}$. For the former fraction, we note that because $g(\mathbf{T}(v),p_c+\varepsilon)\le C'\varepsilon\overline{W}(v)$ for all ε bounded uniformly away from $1-p_c$ (see Proposition 3.4), it follows that for a fresh Galton-Watson tree and for s bounded uniformly away from 0, we have

$$\mathbf{P}[g(\mathbf{T}(v),p_c+n^{-s})>n^{-t}]=O(n^{-p(s-t)}).$$

Now setting $s=\frac{t}{\alpha}$ and combining the above string of inequalities with Propositions 4.4(i) and 4.5 we find that

$$\mathbf{P}[g(\mathbf{T}(\gamma_n),\beta_n^*)>n^{-t}\,|\,h_n^*\le n^{-\frac{t}{\alpha}}]=O(n^{-(p-1)(\frac{1}{\alpha}-1)t}).$$

Combining this with what we determined about the second fraction in (5.11), it now follows that if we condition on $h_n^*\le n^{-\frac{t}{\alpha}}$, then

$$\frac{1}{p(v)}\left|\tilde{q}(v,\beta_n^*)-\frac{g(\mathbf{T}(v),\beta_n^*)}{\sum\limits_{w:\overleftarrow{w}=\gamma_n}g(\mathbf{T}(w),\beta_n^*)}\right|\le 3n^{-t} \tag{5.12}$$

for every child of γ_n, with probability at least $1-C''n^{-\frac{t}{\alpha}(1-\frac{1}{p})\delta'}$ (where we're using the fact that $\frac{1}{\alpha}(1-\frac{1}{p})\delta'\le(p-1)(\frac{1}{\alpha}-1)$). Finally, putting (5.12) together with (5.10) and Theorem 4.9, and defining $t':=\frac{t}{\alpha}(1-\frac{1}{p})\delta'$, we get that $\mathbf{E}\left[\mathbf{P}_*^{(n)}[A_n^c]\right]$ is $O(n^{-t'})$.

From this last result involving $\mathbf{E}\left[\mathbf{P}_*^{(n)}[A_n^c]\right]$, we know that for any constant C_1 such that $0 < C_1 < 1$ we have

$$\mathbf{P}\left[\mathbf{P}_*^{(n)}[A_n^c] > n^{-C_1 t'}\right] = O(n^{-(1-C_1)t'}). \tag{5.13}$$

For the next step, we utilize Propositions 4.4(i) and 4.5 to note that for any $t'' > 0$ and any constant C_2 with $0 < C_2 < 1$, the probability $\frac{1}{p(v)} > n^{t''}$ for any child of γ_n, is bounded by

$$\mathbf{P}[W(v) < \mu n^{-C_2 t''} \text{ for at least 1 child of } \gamma_n] + \mathbf{P}[W(\gamma_n) \geq n^{(1-C_2)t''}], \tag{5.14}$$

which is $O\left(\max\left\{n^{-(p-1)(1-C_2)t''}, n^{-(1-\frac{1}{p})\frac{1}{q}C_2 t''}\right\}\right)$, as discussed at the end of Section 3.3.

To finish establishing the claim, we now need to show that $t, \delta, d, \ell, t'',$ and C_1 can be chosen so that

(i) $\frac{1}{2} < t < \alpha < 1$

(ii) $\frac{1}{p(v)} > n^{t''}$ only finitely often with probability 1

(iii) $\mathbf{P}_*^{(n)}[A_n^c] > n^{-C_1 t'}$ only finitely often with probability 1

(iv) $C_1 t' - t'' > t$, i.e. $n^{-C_1 t'} \cdot n^{t''} \leq n^{-t}$.

To accomplish this, we first note that it follows from (5.14) and the Borel-Cantelli lemma that the second condition will hold if $t'' > \frac{1+pq}{p-1}$. Hence, the fourth condition then reduces to $C_1 t' - \frac{1+pq}{p-1} > t$. Combining this with the third condition, which by (5.13) and Borel-Cantelli will be satisfied if $(1 - C_1)t' > 1$, we find that proving our claim is reduced to finding $t, \delta, \ell,$ and d with $\frac{1}{2} < t < \alpha < 1$ such that

$$t' > 1 + \frac{1+pq}{p-1} + t = \frac{p}{p-1}(1+q) + t.$$

Using our formulas for t' and δ', this can be written as

$$\left(\frac{1}{\alpha}\left(1 - \frac{1}{p}\right)\min\left\{p\ell - \delta, \frac{d\delta}{q}\right\} - 1\right) t > \frac{p}{p-1}(1+q) \tag{5.15}$$

It now suffices to show that (5.15) can be made to hold for $t = \alpha = \frac{1}{2}$. Substituting $\frac{1}{2}$ for t and α in (5.15) and noting that $\alpha = \frac{1}{2} \implies \delta = \frac{1}{1+d}\left(\frac{1}{2} - 3\ell\right)$, (5.15) becomes

$$\left(\left(1 - \frac{1}{p}\right)\min\left\{p\ell - \frac{1}{1+d}\left(\frac{1}{2} - 3\ell\right), \frac{d}{1+d}\cdot\frac{1}{q}\left(\frac{1}{2} - 3\ell\right)\right\} - \frac{1}{2}\right) > \frac{p}{p-1}(1+q).$$

Observing that the expression on the left is increasing with respect to d, we take $d \to \infty$, which gives

$$\left(\left(1 - \frac{1}{p}\right)\min\left\{p\ell, \frac{1}{q}\left(\frac{1}{2} - 3\ell\right)\right\} - \frac{1}{2}\right) > \frac{p}{p-1}(1+q).$$

Expressing this as a pair of inequalities and then simplifying we get

$$\frac{p}{(p-1)^2}(1+q) + \frac{1}{2(p-1)} < \ell < \frac{1}{3}\left(\frac{1}{2} - \frac{p^2 q(1+q)}{(p-1)^2} - \frac{pq}{2(p-1)}\right).$$

For such an ℓ to exist it suffices to have

$$\frac{p}{(p-1)^2}(1+q) + \frac{1}{2(p-1)} < \frac{1}{3}\left(\frac{1}{2} - \frac{p^2 q(1+q)}{(p-1)^2} - \frac{pq}{2(p-1)}\right).$$

Now simplifying the above inequality, we get (5.7), thus completing the proof of the proposition. □

Proof of Theorem 1.1: As guaranteed by Proposition 5.7, let $M > 0$ and $t \in (1/2, 1)$ so that with probability 1, the set

$$\bigcup_{n=1}^{\infty} \left\{ v \in \mathbf{T}_{n+1} : \left| \frac{q(v)}{p(v)} - 1 \right| > 3Mn^{-t}, \overleftarrow{v} = \gamma_n \right\}$$

is finite. Define the event

$$A_n := \left\{ \left| \frac{q(v)}{p(v)} - 1 \right| \leq 3Mn^{-t} \text{ for all } v \in \mathbf{T}_{n+1} \text{ with } \overleftarrow{v} = \gamma_n \right\} .$$

Define $Y_n := X_n \mathbf{1}_{A_n}$; by the choice of M, t, $Y_n = X_n$ all but finitely often almost surely. Therefore, by Corollary 2.8, it is sufficient to show that $\sum \mathbf{E} Y_n < \infty$. By definition of A_n, Proposition 2.6 gives an upper bound of $Y_n \leq 9M^2 n^{-2t}$. Taking expectation and recalling $t \in (1/2, 1)$ completes the proof. □

6 The Forward Maximal Weight Process

This section will be devoted to describing the limiting behaviour of the process $\{\beta_n - p_c\}$. We begin by showing that $\{\beta_n\}$ is a time-homogeneous Markov chain and computing the transition probabilities.

Lemma 6.1.

(i) *The sequence $\{\beta_n := \beta(\gamma_n)\}$ is a time-homogeneous Markov chain adapted to $\{\mathcal{B}_n^+\}$ with initial distribution $\mathcal{L}$.*

(ii) *Reparametrizing by letting $h_n := \beta_n - p_c$, a formula for the transition kernel of the chain $\{h_n\}$ is given in terms of the OGF ϕ by $p(a, \cdot) = \mu_a$ where*

$$\frac{d\mu_a}{dx} = C_a \delta_a + \frac{\phi' \left(1 - (p_c + a) g(p_c + a)\right) g'(p_c + x)}{g'(p_c + a)} \mathbf{1}_{(0,a)}(x)$$

and

$$C_a = 1 - \frac{\phi' \left(1 - (p_c + a) g(p_c + a)\right) g(p_c + a)}{g'(p_c + a)} .$$

Proof: Conclusion (i) follows from the recursion $\beta_{n+1} = \beta_n \circ \theta$ by applying the Markov property of Proposition 4.7 with $n = 1$ and A of the form $\{\beta_1 \in S\}$. Conclusion (ii) follows from the recursive description of β_n as the minimum of $\max\{U(v), \beta(v)\}$ over children of γ_{n-1} and γ_n is the argmin. More specifically, fix $0 < x < a$; then

$$\mathbf{P}[\beta_1 \in [p_c{+}x, p_c{+}x{+}dx] \,|\, \beta_0 \in [p_c{+}a, p_c{+}a{+}da]] = \frac{\mathbf{P}[\beta_1 \in [p_c + x, p_c + x + dx] \cap \beta_0 \in [p_c + a, p_c + a + da]]}{\mathbf{P}[\beta_0 \in [p_c + a, p_c + a + da]]}.$$

We note $\mathbf{P}[\beta_0 \in [p_c{+}a, p_c{+}a{+}da]] = g'(p_c{+}a)da + o(da)$. To calculate the numerator, note that in order for this event to occur, up to a term of $O(dx^2)$, only one child of the root may have pivot in $[p_c + x, p_c + x + dx]$. This child v must have $U_v \in [p_c + a, p_c + a + da]$ and all other children must have pivot above $p_c + a + da$. This gives

$$\mathbf{P}[\beta_1 \in [p_c + x, p_c + x + dx] \cap \beta_0 \in [p_c + a, p_c + a + da]]$$

$$= \sum_{k=1}^{\infty} \mathbf{P}[Z = k] k g'(p_c + x) dx da (1 - (p_c + a) g(p_c + a))^{k-1} + o(dxda)$$
$$= \phi'(1 - (p_c + a) g(p_c + a)) g'(p_c + x)\, dx da + o(dxda).$$

Combining the two and taking $da \to 0^+$ completes the proof. □

Theorem 6.2. *Let $U_0, U_1, \dots$ be a sequence of IID random variables each uniformly distributed on $(0,1)$, and let $M_n = \min\{U_0, U_1, \dots, U_n\}$. For each C_1, C_2 such that $0 < C_1 < p_c < C_2$, the process $\{h_n\}$ can be coupled with the process $\{M_n\}$ so that, with probability 1, h_n eventually (meaning for all sufficiently large n) satisfies $C_1 \cdot M_n \le h_n \le C_2 \cdot M_n$.*

Proof: We start by looking at the function

$$f_a(u) = \begin{cases} 0 & \text{if } u \ge a \\ \frac{\phi'(1-(p_c+a)g(p_c+a))g'(p_c+u)}{g'(p_c+a)} & \text{otherwise} \end{cases}.$$

Writing u as $u = s \cdot a$ (for $s \in [0,1)$) and using Corollary 3.2, we find that

$$\lim_{a\to 0} \frac{\phi'\left(1 - (p_c + a) g(p_c + a)\right) g'(p_c + sa)}{g'(p_c + a)} = \lim_{a \to 0} \phi'\left(1 - (p_c + a) g(p_c + a)\right) \cdot \lim_{a\to 0} \frac{g'(p_c + sa)}{g'(p_c + a)} = \mu \tag{6.1}$$

with the convergence clearly being uniform with respect to s. Turning now to the process $\{M_n\}$, if we define

$$\tilde{f}_a(u) = \begin{cases} 0 & \text{if } u \ge a \\ 1 & \text{otherwise} \end{cases}$$

and the measures ν_a, where $\nu_a(A) = (1-a)\mathbf{1}_{(a\in A)} + \int_0^1 \tilde{f}_a(u)\mathbf{1}_{(u\in A)}\, du$, then we see that $\{M_n\}$ is a Markov chain with transition kernel $\tilde{p}(x, \cdot) = \nu_x(\cdot)$.

Note that (6.1) implies there must exist $\delta > 0$ such that for $a < \delta$ we have $\frac{1}{C_2} < f_a(u) < \frac{1}{C_1}$ on $(0, a)$. Define $N_\delta := \min\{n : h_n < \delta\}$ and note that since $h_n \to 0$ a.s., it follows that $N_\delta < \infty$ a.s. Define the family of functions Q_r for $r \in [0, \delta)$ where $Q_r : [0, r) \to \mathbb{R}$ is defined as

$$Q_r(x) = \sum_{j=0}^{\infty} q^{j+1} \left(C_2 \cdot f_r\left(q^j x\right) - 1 \right) \tag{6.2}$$

where $q := \frac{C_1}{C_2}$. Observe that because $\frac{1}{C_2} < f_a(u) < \frac{1}{C_1}$ on $(0, a)$ for $a < \delta$, it follows that

$$Q_r(x) > \sum_{j=0}^{\infty} q^{j+1} \left(C_2 \cdot C_2^{-1} - 1 \right) = 0 \tag{6.3}$$

and that

$$Q_r(x) < \sum_{j=0}^{\infty} q^{j+1} \left(C_2 \cdot C_1^{-1} - 1 \right) = \frac{q}{q-1} \cdot \frac{q-1}{q} = 1\,. \tag{6.4}$$

In addition, note that it follows from (6.2) that we have

$$\frac{1}{C_1} Q_r(x) + \frac{1}{C_2}\left(1 - Q_r(qx)\right) \quad = \quad C_2^{-1} + \sum_{j=0}^{\infty} q^{j+1}\left(q^{-1} \cdot f_r\left(q^j x\right) - C_1^{-1}\right) - \sum_{j=0}^{\infty} q^{j+1}\left(f_r\left(q^{j+1}x\right) - C_2^{-1}\right)$$

$$
\begin{aligned}
&= C_2^{-1} - \frac{C_1}{C_2 - C_1} \cdot \left(C_1^{-1} - C_2^{-1}\right) + \sum_{j=0}^{\infty} q^j f_r\left(q^j x\right) - \sum_{j=1}^{\infty} q^j f_r\left(q^j x\right) \\
&= f_r(x)\,. \qquad (6.5)
\end{aligned}
$$

We'll now use the family of functions Q_r, along with the process $\{h_n\}$ and the sequence $\{U_k\}$ defined in the statement of the theorem, to define a new sequence $\{V_k\}$. Letting $V_0 = h_{N_\delta}$, we define V_j for $j \geq 1$ as follows. First let $L_n = \min\{V_0, V_1, \ldots, V_n\}$. Now if $C_1 \cdot U_{N_\delta+j} \geq L_{j-1}$, set $V_j = C_1 \cdot U_{N_\delta+j}$. If instead $C_1 \cdot U_{N_\delta+j} < L_{j-1}$, then with probability $Q_{L_{j-1}}(C_1 \cdot U_{N_\delta+j})$ set V_j equal to $C_1 \cdot U_{N_\delta+j}$, and with probability $1 - Q_{L_{j-1}}(C_1 \cdot U_{N_\delta+j})$ set V_j equal to $C_2 \cdot U_{N_\delta+j}$. Next we define the process $\left\{\tilde{h}_n\right\}$ as

$$
\tilde{h}_n = \begin{cases} h_n & \text{if } n < N_\delta \\ L_{(n-N_\delta)} & \text{otherwise} \end{cases} .
$$

Observe that in order to show that $\{\tilde{h}_n\}$ has the same joint distribution as $\{h_n\}$, it will suffice to establish that for any $n > 0$ and $0 < x < y < r$

$$
\mathbf{P}[\tilde{h}_{n+1} \in [x,y) \,|\, \tilde{h}_n = r] = \mathbf{P}[h_{n+1} \in [x,y) \,|\, h_n = r] = \int_x^y f_r(t)\,dt. \qquad (6.6)
$$

In the case where $r \geq \delta$ we see that (6.6) follows immediately from the definition of $\{\tilde{h}_n\}$. Alternatively, if $r < \delta$ then it will follow from the definition of $\{\tilde{h}_n\}$ that

$$
\mathbf{P}[\tilde{h}_{n+1} \in [x,y) \,|\, \tilde{h}_n = r] = \int_x^y C_1^{-1} Q_r(x) + C_2^{-1}(1 - Q_r(qx))\,dx = \int_x^y f_r(x)\,dx. \qquad (6.7)
$$

Defining the times $\tau_1 = \min\{n : V_n < V_0\}$, $\tau_2 = \min\{n : U_{N_\delta+n} < M_{N_\delta}\}$, and $\tau = \max\{\tau_1, \tau_2\}$, we see that $\tau_1 < \infty$ a.s. due to the fact that $\tilde{h}_n \to 0$ a.s. since $\left\{\tilde{h}_n\right\}$ has the same joint distribution as $\{h_n\}$, and $\tau_2 < \infty$ a.s. due to the U_j's being IID uniform on $(0,1)$. Since $\tilde{h}_n = \min\{V_1, V_2, \ldots, V_{n-N_\delta}\}$ for $n \geq N_\delta + \tau$, $M_n = \min\{U_{N_\delta+1}, U_{N_\delta+2}, \ldots, U_n\}$ for $n \geq N_\delta + \tau$, and $C_1 U_{N_\delta+j} \leq V_j \leq C_2 U_{N_\delta+j}$ for all $j \geq 1$, it can be concluded that $C_1 M_n \leq \tilde{h}_n \leq C_2 M_n$ for all $n \geq N_\delta + \tau$, thus establishing that $(\tilde{h}_n, M_n)$ gives us our desired coupling. □

This coupling is enough to prove convergence on the level of paths. Let $\mathcal{P}$ be an intensity 1 Poisson point process on the upper-half-plane; define the *Poisson lower envelope process* by

$$
L(t) := \min\{y > 0 : (x,y) \in \mathcal{P} \text{ for some } x \in [0,t]\}.
$$

Then we have

Corollary 6.3. *For any $\varepsilon > 0$ as $k \to \infty$,*

$$
(kh_{\lceil kt \rceil}/p_c)_{t \geq \varepsilon} \overset{*}{\Longrightarrow} (L(t))_{t \geq \varepsilon} \qquad (6.8)
$$

where $\overset{*}{\Longrightarrow}$ denotes convergence in distribution of càdlàg paths in the Skorohod space $D[\varepsilon, \infty)$.

Proof: Note that by taking $C_1, C_2 \to p_c$ in Theorem 6.2, it is sufficient to show convergence of $(kM_{\lceil kt \rceil})_{t \geq \varepsilon}$ to $(L(t))_{t \geq \varepsilon}$.

Showing convergence in distribution in the Skorohod topology is equivalent to establishing tightness and convergence in distribution of all finite dimensional projections to the corresponding projections of $(L(t))_{t\geq\varepsilon}$ [Bil99, Theorem 13.1]. To accomplish this, we will show that for any $\gamma > 0$, we will be able to couple $kM_{\lceil kt\rceil}$ and $L(t)$ so that we have

$$L(t(1+\gamma)) \leq kM_{\lceil kt\rceil} \leq L(t(1-\gamma)), \qquad t \geq \varepsilon \tag{6.9}$$

for sufficiently large k. From here, convergence of the finite-dimensional projections will follow.

Step 1: Finite dimensional projections: We proceed in a similar fashion to [AGdHS08]: we Poissonize, sandwich the Poissonized version of $kh_{\lceil kt\rceil}$ between two scaled copies of $L(t)$, and then use the strong law of large numbers to depoissonize.

Consider an intensity 1 Poisson process on $[0,\infty)$ and define $N(t)$ to be the number of points in $[0,t]$. Define

$$\tilde{M}(t) = M_{N(t)}, \quad t \geq 0$$

to be the Poissonized version of the min-uniform process defined by $M_n = \min\{U_1,\ldots,U_n\}$ for $n \geq 1$ and $M_0 = 1$; note that this differs slightly from the M_n in Theorem 6.2 and is entirely distinct from the M_n appearing in the glossary. Then note that both $\tilde{M}(t)$ and $L(t)$ are continuous-time Markov processes that jump from height z to height $zU[0,1]$ at exponential rate z. Moreover, the processes $L_1(t) := 1 \wedge L(t)$ and $\tilde{M}(t)$ have the same starting value and jump from z to $zU[0,1]$ at exponential rate z. Using the same exponential clock and uniforms for both processes gives

$$L_1(t) = \tilde{M}(t)$$

for all $t \geq 0$. Since $L(t)$ is eventually less than 1, we have that there exists an almost-surely finite time τ so that

$$L(t) = \tilde{M}(t), \qquad t \geq \tau.$$

Thus, for all k and $t \geq \tau$, we have

$$k\tilde{M}(kt) = kL(kt) =: L'_k(t) \overset{d}{=} L(t) \tag{6.10}$$

since for all k the process $(kL(kt))_{t\geq 0}$ has the same law as $(L(t))_{t\geq 0}$.

By the strong law of large numbers, for any fixed $\varepsilon > 0$ and $\gamma > 0$, there exists an almost-surely finite random variable K so that

$$N(kt) \in \left[\frac{kt}{1+\gamma}, \frac{kt}{1-\gamma}\right], \qquad k \geq K$$

uniformly in $t \geq \varepsilon$. Since $\tilde{M}(t)$ is decreasing, this implies for all $k \geq K/(1-\gamma)$ and uniformly in $t \geq \varepsilon$ we have

$$kM_{\lceil kt\rceil} \in [k\tilde{M}((1+\gamma)kt), k\tilde{M}((1-\gamma)kt)].$$

Combining this with equation (6.10) gives

$$L'_k(t(1+\gamma)) \leq kM_{\lceil kt\rceil} \leq L'_k(t(1-\gamma)), \qquad k \geq K/(1-\gamma),\ t \geq \varepsilon\,. \tag{6.11}$$

Taking $\gamma \to 0$ and utilizing (6.11) proves that for any sequence $\varepsilon \leq t_1 < t_2 < \ldots < t_n \leq R$ we have $(kM_{\lceil kt_1\rceil},\ldots,kM_{\lceil kt_n\rceil}) \implies (L'_k(t_1),\ldots,L'_k(t_n))$ as $k \to \infty$.

Step 2: Tightness: By definition of the Skorohod space $D[\varepsilon,\infty)$, it is sufficient to show convergence in distribution in the space $D[\varepsilon, R]$ for each $R > \varepsilon$. Let $\mathbf{P}_k$ denote the law of the path $(kM_{\lceil kt \rceil})_{t\geq\varepsilon}$. Then tightness of $\{\mathbf{P}_k\}_{k\geq 1}$ in $D[\varepsilon, R]$ is equivalent to showing that for each $\eta > 0$ the following two conditions hold:

$$\text{there exists } a > 0 \text{ so that } \limsup_{k\to\infty} \mathbf{P}_k[x \in D[\varepsilon, R] : \sup_t |x(t)| > a] < \eta \tag{6.12}$$

$$\text{for each } r > 0 \text{ there exists } \delta \in (0,1) \text{ so that } \limsup_{k\to\infty} \mathbf{P}_k\left[x \in D[\varepsilon, R] : \overline{w}_x(\delta) > r\right] < \eta \tag{6.13}$$

where $\overline{w}_x(\delta)$ is the *càdlàg modulus* defined by

$$\overline{w}_x(\delta) := \inf_{\Pi\in P_\delta} \max_{1\leq i\leq j} \sup_{s,t\in[t_i,t_{i+1})} |x(t) - x(s)| \tag{6.14}$$

and where the infimum is over partitions $\varepsilon = t_0 < t_1 < \ldots < t_j = R$ with mesh greater than δ [Bil99, Theorem 13.2].

Since $(kM_{\lceil kt\rceil})_{t\in[\varepsilon,R]}$ is monotone decreasing in t, the existence of an $a > 0$ to satisfy (6.12) follows from (6.11). Note also that the path $(kM_{\lceil kt\rceil})_{t\in[\varepsilon,R]}$—like $(L(t))_{t\in[\varepsilon,R]}$—is piecewise constant. For a piecewise constant path $x \in D[\varepsilon, R]$, we have that $\overline{w}_x(\delta) = 0$ if all jumps are spaced at least δ apart and no jumps occur in the intervals $[\varepsilon, \varepsilon+\delta]$ or $[R-\delta, R]$. We will show that as $\delta \to 0$, the $\liminf_k$ of the probability that all jumps of $(kM_{\lceil kt\rceil})_{t\in[\varepsilon,R]}$ are spaced more than δ apart and no jumps occur in $[\epsilon, \epsilon+\delta]$ or $[R-\delta, R]$ tends to 1.

For simplicity of notation, we omit ceiling and floor functions. Note that for any constant $C > 0$,

$$\mathbf{P}[M_{k\varepsilon} \leq C/(k\varepsilon)] = \left(1 - \frac{C}{k\varepsilon}\right)^{k\varepsilon} = O(e^{-C}),$$

implying $\mathbf{P}[M_{k\varepsilon} \leq C/(k\varepsilon)] = 1 - O(e^{-C})$. Partition the interval $[\varepsilon, R]$ into n intervals of size on the order of 2δ, labeled $A_1, A_2, \ldots, A_n$; similarly, set $B_1 = [\varepsilon, \varepsilon+\delta], B_2 = [\varepsilon+\delta, \varepsilon+3\delta]$, and subsequently set $B_i = 2\delta + B_{i-1}$. If (kM_{kt}) has two jumps within a distance of δ from each other, then some interval A_i or B_i has two jumps within it. Since each interval A_i and B_i contains $O(k\delta)$ integers, the probability that an interval contains two jumps is equal to the probability that there are 2 independent uniforms out of $O(k\delta)$ many that are both less than $C/(k\varepsilon)$. This implies

$$\mathbf{P}[\text{a fixed interval of length } O(\delta) \text{ contains two jumps}] = O\left(k^2\delta^2\right) \cdot O\left((C/(k\varepsilon))^2\right) = O\left(C^2\delta^2/\varepsilon^2\right).$$

Applying a union bound over all $2n = O(1/\delta)$ intervals shows that the probability there are two jumps in some A_i or B_i is $O(C^2\delta/\varepsilon^2)$. Likewise, the probability that the first interval $[\varepsilon, \varepsilon+\delta]$ or $[R-\delta, R]$ contains a jump is $O(C\delta/\varepsilon)$ as well. Taking $\delta \to 0$ followed by $C \to \infty$ completes the proof.

□

Corollary 6.4. *The sequence $n \cdot h_n$ converges in distribution to $p_c \cdot \exp(1)$, where $\exp(1)$ is an exponential random variable with mean 1.*

Proof: It suffices to show that for every $x \in (0,\infty)$, we have $\lim_{n\to\infty} \mathbf{P}[n \cdot h_n > x] = e^{-\mu x}$. Let N_δ and τ be the stopping times defined in the proof of Theorem 6.2 and recall that $N_\delta + \tau < \infty$ a.s. and for $n > N_\delta + \tau$ we have $C_1 M_n < h_n < C_2 M_n$. It follows that

$$\mathbf{P}[n \cdot h_n > x] \quad = \quad \mathbf{P}\left[h_n > \frac{x}{n}\right]$$

$$
\begin{aligned}
&\geq \mathbf{P}\left[C_1 \cdot M_n > \frac{x}{n}\right] - \mathbf{P}[N_\delta + \tau > n] \\
&= \left(1 - \frac{x/C_1}{n}\right)^{n+1} - \mathbf{P}[N_\delta + \tau > n]\,. \qquad (6.15)
\end{aligned}
$$

Taking the lim inf of the expressions on the right and left sides in (6.15), while recalling that $N_\delta + \tau < \infty$ a.s., we find that $\liminf_{n\to\infty} \mathbf{P}[n \cdot h_n > x] \geq e^{-\frac{x}{C_1}}$. Since $C_1 < p_c$ is arbitrary, it then follows that $\liminf_{n\to\infty} \mathbf{P}[n \cdot h_n > x] \geq e^{-\mu x}$. Conversely, Theorem 6.2 also implies that

$$
\begin{aligned}
\mathbf{P}[n \cdot h_n > x] &= \mathbf{P}\left[h_n > \frac{x}{n}\right] \\
&\leq \mathbf{P}\left[C_2 \cdot M_n > \frac{x}{n}\right] + \mathbf{P}[N_\delta + \tau > n] \\
&= \left(1 - \frac{x/C_2}{n}\right)^{n} + \mathbf{P}[N_\delta + \tau > n]. \qquad (6.16)
\end{aligned}
$$

Taking the lim sup of the expressions on the right and left sides in (6.16) then gives $\limsup_{n\to\infty} \mathbf{P}[n \cdot h_n > x] \leq e^{-\frac{x}{C_2}}$ which, since $C_2 > p_c$ is arbitrary, implies $\limsup_{n\to\infty} \mathbf{P}[n \cdot h_n > x] \leq e^{-\mu x}$. Combining this with the lower bound on the lim inf gives $\lim_{n\to\infty} \mathbf{P}[n \cdot h_n > x] = e^{-\mu x}$. □

References

[ABF13] L. Addario-Berry and K. Ford. Poisson-Dirichlet branching random walks. *Ann. Appl. Prob.*, 23:283–307, 2013.

[AGdHS08] O. Angel, J. Goodman, F. den Hollander, and G. Slade. Invasion percolation on regular trees. *Ann. Probab.*, 36:420–466, 2008.

[AGdHS13] O. Angel, J. Goodman, F. den Hollander, and G. Slade. Scaling limit of the invasion percolation cluster on a regular tree. *Ann. Probab.*, 41:229–261, 2013.

[AN72] K. Athreya and P. Ney. *Branching Processes*. Springer-Verlag, New York, 1972.

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19:357–367, 1967.

[BD74] N. Bingham and R. Doney. Asymptotic properties of supercritical branching processes I: the Galton-Watson process. *Adv. Appl. Prob.*, 6:711–731, 1974.

[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[DSV09] M. Damron, A. Sapozhnikov, and B. Vágvölgyi. Relations between invasion percolation and critical percolation in two dimensions. *Ann. Probab.*, 37:2297–2331, 2009.

[Dub71] Serge Dubuc. Problèmes relatifs à l'itération de fonctions suggérés par les processus en cascade. *Ann. Inst. Fourier (Grenoble)*, 21(1):171–251, 1971.

[Dur10] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, New York, NY, fourth edition, 2010.

[FN71] D. H. Fuk and S. V. Nagaev. Probabilistic inequalities for sums of independent random variables. *Teor. Verojatnost. i Primenen.*, 16:660–675, 1971.

[Har63] T. Harris. *The theory of branching processes.* Springer, Berlin, 1963.

[HPS99] O. Häggström, Y. Peres, and R. Schonmann. Percolation on transitive graphs as a coalescent process: relentless merging followed by simultaneous uniqueness. In *Perplexing problems in probability*, volume 44, pages 69–90. Birkhäuser, Boston, 1999.

[JS89] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing markov chains. *Inform. and Comput.*, 82:93–133, 1989.

[Kes86] H. Kesten. The incipient infinite cluster in two-dimensional percolation. *Prob Theory Rel. Fields*, 73:369–394, 1986.

[LP17] R. Lyons and Y. Peres. *Probability on trees and networks.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017.

[LPP96] R. Lyons, R. Pemantle, and Y. Peres. Biased random walks on galton-watson trees. *Prob. Th. Rel. Fields*, 106:249–264, 1996.

[Lyo90] R. Lyons. Random walks and percolation on trees. *Ann. Probab.*, 18:931–958, 1990.

[MPR18] Marcus Michelen, Robin Pemantle, and Josh Rosenberg. Quenched survival of Bernoulli percolation on Galton-Watson trees. *arXiv preprint arXiv:1805.03693*, 2018.

[RS00] D. Randall and A. Sinclair. Self-testing algorithms for self-avoiding walks. *Journal of Mathematical Physics*, 41:1570–1584, 2000.

[WW83] D. Wilkinson and J. Willemsen. Invasion percolation: a new form of percolation theory. *J. Phys., ser. A: Math. Gen.*, 16:3365–3376, 1983.

[Zha95] Y. Zhang. The fractal volume of the two-dimensional invasion percolation cluster. *Comm. Math.Phys.*, 167:237–254, 1995.

Notation

Trees

$\mathbf{0}$	root of tree	5
$\mathcal{U}$	canonical tree	5
$\mathbf{V}$	canonical vertex set	5
$\|v\|$	depth of a vertex	5
$\overleftarrow{v}$	parent of a vertex	5
$v \sqcup i$	ith child of v	5
$v \wedge w$	least common ancestor of two nodes	5
$\gamma \wedge \gamma'$	least common vertex of two paths	5
γ_n	nth vertex in path γ	5
σ^v	shift to vertex v	21

$T(v)$	subtree of v rooted at v	5
$T^*(v)$	$T \setminus T(v)$	17
T_n	set of nodes of T at depth n	5
∂T	set of infinity non-backtracking paths from root	5
Z_n	number of descendants at height n	6
$Z_n(v)$	number of offspring of v in generation $n + \lvert v \rvert$	6
$Z_n^{(i)}(v)$	number of nth generation descendents of v that pass through $v \sqcup i$	6

Branching processes

$(\Omega, \mathcal{F}, \mathbf{P})$	the probability space	5
ϕ	probability generating function for progeny distribution	5
Z	generic random variable with p.g.f. ϕ	5
μ	$= \mathbf{E}Z = \phi'(1)$	5
p_c	$= 1/\mu$	6
$\{\deg_v\}$	IID $\sim \phi$ variables that construct the Galton-Watson tree	6
$\mathcal{T}$	σ-algebra generated by $\{\deg_v\}$	6
$\mathtt{GW}$	$= \mathbf{P}\vert_{\mathcal{T}}$, the Galton-Watson measure	6
$\mathbf{T}$	random rooted subtree chosen with Galton-Watson measure	6
$\mathbf{E}_*$	$= \mathbf{E}[\cdot \,\vert\, \mathcal{T}]$	6
$\mathbf{P}_*$	$= \mathbf{P}[\cdot \,\vert\, \mathcal{T}]$	6
W_n	the martingale $\mu^{-n} Z_n$	6
$W_n(v)$	$= \mu^{-n} Z_n(v)$	6
W	$= \lim W_n$	6
$\overline{W}$	$= \max_n W_n$	11
$W_n^{(i)}(v)$	$= \mu^{-n} Z_n^{(i)}(v)$	6

Percolation

U_v	uniform random variables defining percolation	6
$\mathcal{F}_n$	σ-algebra up to level n	6
$\mathcal{F}'_n$	σ-algebra of pivots up to n and entire tree	13
I	invasion cluster	6
γ	the backbone of invasion percolation	8
$v \leftrightarrow_p w$	event that v and w connected in p percolation	6
$H(p)$	event that root is connected to infinity in p percolation	6
$g(T, p)$	probability that root of T is connected to infinity in p percolation	6
$g(p)$	$\mathbf{E}g(\mathbf{T}, p)$	6
K	$\lim_{\varepsilon \to 0^+} g(p_c + \varepsilon)/\varepsilon$	10
$\mathcal{C}_v$	$\sigma(T(v))$	18
$\mathcal{C}_n$	$\sigma(T(\gamma_n))$	18
$\mathcal{B}_v^*$	$\sigma(T \setminus T(v))$	18
$\mathcal{B}_n^*$	$\sigma(T \setminus T(\gamma_n))$	18
$\mathcal{B}_n^+$	$\sigma(\mathcal{B}_n^* \cup \{\beta_n\})$	21

$\mathcal{G}_n$	$= \sigma(T(v); \lvert v \rvert = n)$	9
$\beta(v)$	pivot of v	3
β_n	pivot of nth vertex of backbone	**??**
β^*_v	dual pivot of v	17
β^*_n	dual pivot of nth vertex of backbone	17
$\mathbf{E}^{(n)}_*$	$= \mathbf{E}[\cdot \vert \mathcal{T}, \gamma_n]$	26
$\mathbf{P}^{(n)}_*$	$= \mathbf{P}[\cdot \vert \mathcal{T}, \gamma_n]$	27

Random measures

μ^n_T	uniform measure on T_n	5
μ_T	limit uniform measure on ∂T	5
ν_T	invasion measure ∂T	8
M_n	$= \frac{d\nu_T}{d\mu_T}\Big\rvert_{\mathcal{G}_n}$	9
$p(v)$	probability that limit uniform measure splits from $\overleftarrow{v}$ to v	8
$q(v)$	probability that invasion measure splits from $\overleftarrow{v}$ to v	8
$\tilde{q}(v, p)$	probability that invasion measure splits from $\overleftarrow{v}$ to v, given $\beta(\overleftarrow{v}) \leq p$ ...	26
$X(v)$	$= \sum_w q(w) \log[q(w)/p(w]$	8