hexagon:2611.00820v1

The speed of biased random walk on leafless Bienaymé trees is monotone

ChatGPT, directed by Igor Kortchemski *

Abstract

Let ν be the offspring variable of a Bienaymé tree without leaves, with finite mean $m > 1$. We prove that the speed $v(\lambda)$ of the λ-biased random walk is continuously differentiable on $[0, m)$, with $v'(\lambda) < 0$ for every $0 < \lambda < m$. In particular v is strictly decreasing, which answers a question posed by Lyons, Pemantle and Peres in 1996. Heng Ma recently gave an independent proof of strict monotonicity. Our proof is entirely analytic when $\mathbb{E}[1/\nu] \leq 31/49$. In general it also uses two interval-arithmetic computations, for biases between 1 and $9/5$. The main theorem, including both computations, has been formalized in Lean.

1 Introduction

Let $\mathcal{T}$ be a Bienaymé tree[1]. Starting with a single root, each vertex independently produces a number of children distributed as ν. We assume that $\mathbb{P}(\nu = 0) = 0$ and $1 < m = \mathbb{E}[\nu] < \infty$. At a vertex with k children, the λ-biased walk assigns weight λ to its parent and weight one to each child. Lyons [13] proved transience for $\lambda < m$ and recurrence for $\lambda > m$. For $0 \leq \lambda < m$ there is a deterministic speed $v(\lambda) > 0$ such that almost surely, for almost every tree, $|X_n|/n \to v(\lambda)$, where $|X_n|$ is the distance from the root [14, 15]. On the regular tree with k children per vertex, the speed for $0 \leq \lambda < k$ equals the local drift $(k - \lambda)/(k + \lambda)$ [15].

On a random tree this local drift still decreases with λ, but the walk also changes the degrees it visits. Lyons, Pemantle and Peres [15, Section 6, Question 1] asked whether the speed decreases with λ. We answer it.

Theorem 1.1. *For every offspring law ν with $\mathbb{P}(\nu = 0) = 0$ and $1 < m = \mathbb{E}[\nu] < \infty$, the speed $v(\lambda)$ of the λ-biased random walk on the Bienaymé tree is of class C^1 on $[0, m)$ and strictly decreasing there. Its derivative satisfies $v'(\lambda) < 0$ for $0 < \lambda < m$ and $v'_+(0) = -2\mathbb{E}[1/\nu]$.*

*CNRS & DMA, École normale supérieure, PSL University, 75005 Paris, France, igor.kortchemski@math.cnrs.fr

[1] Also known as Galton–Watson trees, but we prefer the terminology of Bienaymé trees for historical reasons [9].

Independently, Heng Ma [11] proved strict decrease throughout the transient regime under the same assumptions. His proof is analytic when every vertex has at least two children. For laws with $\mathbb{P}(\nu = 1) > 0$, he uses a computer-assisted argument for $1 < \lambda < \min\{m, (3+\sqrt{5})/2\}$. Our proof is analytic for $9/5 \le \lambda < m$ (Theorem 3.7). We also prove that v is C^1 at $\lambda = 1$ and that $v'(1) < 0$.

1.1 Relation to earlier work

For trees with leaves, Lyons, Pemantle and Peres [15, Section 4] showed that trapping can give zero speed at small positive λ. Ben Arous, Fribergh, Gantert and Hammond [4] studied the depth of these traps and the resulting sublinear displacement. Even without leaves, monotonicity can fail on multitype and deterministic trees, as shown by Lyons, Pemantle and Peres [16, Section 2].

For leafless Bienaymé trees, Ben Arous, Fribergh and Sidoravicius [5] proved monotonicity for $\lambda \le 1/1160$, and for $\lambda \le d/1160$ if every vertex has at least d children. The finite-difference asymptotic in Theorem 1.3 of their preprint gives the right derivative $v'_+(0) = -2\mathbb{E}[1/\nu]$ by summing over unit increments of the reciprocal bias. Aïdékon [1, Section 2] proved monotonicity up to $\lambda = 1/2$. Song, Wang and Xiang [21] differentiated the escape recursion on finite trees and proved strict decrease up to $m_1/(1+\sqrt{1-1/m_1})$ when the minimum offspring number is $m_1 \ge 2$. Mandarapu and Kunkunuru [17] compared conductance laws at two biases and obtained strict decrease on $[0, 1.755]$ for the uniform law on $\{2, 3\}$ by a computer-assisted proof. Schwarzfischer [20, Theorem 1.1] proved $v'(\lambda) < 0$ on $(0, m)$ when every vertex has at least four children by differentiating the law of the escape probability. He also observed that his method should extend to minimum offspring two. Corollary 4.7 below gives a fully analytic proof for $\mathbb{E}[1/\nu] \le 31/49$, including every law with $\mathbb{P}(\nu = 1) \le 13/49$. For laws with unary vertices, the escape probability can be arbitrarily small near bias one.

For the continuous-time walk, Ben Arous, Hu, Olla and Zeitouni [6] proved an Einstein relation at critical bias under an exponential offspring moment assumption. Ben Arous and Fribergh [3] discussed monotonicity near critical bias and recorded Aïdékon's observation that v is C^1 on $(0, 1)$. Mehrdad, Sen and Zhu [18] studied a different comparison, varying the offspring distribution at a fixed strong outward bias.

Our starting point is the speed formula of Aïdékon [2, Theorem 1.1], in the two-copy form of Rousselin [19, Eq. (4.8)]. Aïdékon's theorem assumes $\mathbb{P}(\nu = 1) < \lambda < m$ in the leafless case. In Section 2 we cover the remaining biases by averaging the drift in the stationary environment seen by the walk. Bowditch and Tokushige [8, Theorem 1.1] proved differentiability under an exponential offspring moment assumption. Here we need a formula for the derivative under the sole moment condition $m < \infty$.

1.2 Idea of the proof

Let β be the probability that a walk started at the root never visits its parent, and let β' be the escape probability of an independent tree. Aïdékon's speed formula has the form $v = N/Z$, where N and Z are expectations over (β, β'). There are two contributions to its derivative.

Changing λ changes the integrands even at fixed escape probabilities. It also changes β on each tree. With $u = -\lambda\partial_\lambda\beta$, the derivative identity of Proposition 2.6 reads

$$-\lambda Z v'(\lambda) = \lambda H + 2\mathbb{E}[u\,\Gamma(\beta)].$$

Here H comes from differentiating at fixed escapes and Γ is the derivative in one escape variable averaged over the other tree. For $\lambda \le 1$, comparison with the drift at the root and a first lower bound on u give the sign directly (Theorem 2.7).

Above bias one, the recursion takes a simpler form in the variable $Y = \beta/(\lambda - 1 + \beta)$. We use it in two ways. First, exchanging one child subtree with the independent tree gives an exact weighted average of u. The remaining coefficient of u is nonnegative, so a lower bound on u suffices. To control this coefficient and the speed, we compare the moments of Y with those of λ^{-G_*}, where the geometric variable G_* is chosen to match the mean of Y. The resulting bounds give a proof for $\lambda \ge 9/5$ (Theorem 3.7).

Second, at a vertex with one child the transformed escape is simply the child's transformed escape divided by λ. Let $p = \mathbb{P}(\nu \ge 2)$. Following the path to its first branching vertex gives $Y = \lambda^{-G}W$, where G is geometric of parameter p and independent of the contribution W from the split. The scale $s = (\lambda - 1)/p$ appears because $\mathbb{E}[\lambda^{-G}] = 1/(1+s)$. The proof in the remaining range $1 < \lambda < 9/5$ uses the following consequences of this decomposition.

- For $s \le 1/8$, we argue at a first possible zero of v'. The speed there is at most its explicit value at bias one. Fluctuations in the lengths of two independent unary paths then force $v' < 0$.
- For $s \ge 4$, differentiation at the first split gives a lower bound in terms of the mean and variance of W. The same argument works analytically for every s when $\mathbb{E}W$ is small, in particular near the critical bias m. Together with the geometric comparison this also gives a proof at every bias for $\mathbb{E}[1/\nu] \le 31/49$ (Corollary 4.7).
- For $1/8 \le s \le 4$, we keep track of several moments of W together. At fixed first and second moments, the quotient for an explicit two-point law bounds the speed. We can then estimate each factor in the derivative separately.

Sections 4 and 5 give these arguments. Appendix A describes the two interval verifications.

1.3 AI use

The author started using AI tools to tackle this problem in the beginning of November 2026. He directed the investigation and set the standards for checking the proof. AI agents developed the arguments, wrote the verification programs and drafted the manuscript.

The first step was an exploratory investigation conducted mainly with OpenAI's GPT models through Codex and consisted in eighteen rounds. It produced the derivative identity of Proposition 2.6, the proof of monotonicity on $[0,1]$ (Theorem 2.7), lower bounds on the response, and an early form of the subtree exchange used in Section 3. It also gave a five-copy inequality for large biases and proved monotonicity for several families of offspring laws. These families later served as tests for the general argument.

In a second step, using Claude Code the author directed a campaign to prove the conjecture for every leafless offspring law of finite mean $m > 1$. Anthropic's Claude Opus 5.5, running in Claude Code, coordinated the work. The campaign recorded 134 tasks, including proof searches, independent reviews, computational checks and the preparation of the first draft. The agents used OpenAI's GPT-6.2 Astra and GPT-6.2 Sol through Codex, and Anthropic's Claude Opus 5.5, Claude Fable 5.2 and Claude Sonnet 5.5. The author also submitted six questions to ChatGPT Pro and incorporated the replies into the investigation. By the end of this second step, the first draft contained a complete proof of the conjecture, including the computational verifications needed to cover the full range of biases.

In a third step, the author revised the 75-page first draft over several rounds with OpenAI's GPT-6.2 Astra through Codex. ChatGPT Pro reviewed intermediate drafts and suggested ways to simplify the computations or replace them with analytic arguments. For example, the first proof for $2 \le \lambda < m$ required checking more than 28 million polynomial coefficients. A comparison with the moments of a geometric law now gives an analytic proof for $9/5 \le \lambda < m$ (Theorem 3.7). Eight computations accompanied the first draft. The present manuscript has 25 pages and uses the two interval verifications described in Appendix A.

In the last step, Claude Opus 5.5, running in Claude Code, coordinated the formalization of Theorem 1.1 in Lean 4 with Mathlib. The work took 18 hours and 26 minutes. The agents that wrote and reviewed the proofs used Kimi K3.1 Pro through Open Math OS, GPT-6.2 Sol, GPT-6 Luna and GPT-6.2 Astra through Codex, and Claude Opus 5.5 and Claude Sonnet 5.5 through Claude Code.

The Lean statement identifies $v(\lambda)$ as the almost-sure limit of $|X_n|/n$ for the biased random walk on almost every tree. Two models compared the Lean statement with the theorem in the paper. A separate check translated the Lean statement back into ordinary mathematical language and compared the result with the paper. After eight and a half hours, the agents had proved Corollary 4.7, whose proof is entirely analytic. After ten and a half hours, they had reduced Theorem 1.1 to the two lemmas of Appendix A.

For these two lemmas, the formalization starts from the subdivisions used by the C++ programs. The Lean kernel checks that the boxes cover the required domains and verifies the interval bounds using exact integer arithmetic with outward rounding on dyadic grids. For Lemma A.1, the dyadic bounds required a finer subdivision in some places, giving 41 919 boxes in all. For Lemma A.2, the verification uses the same 8 185 rectangles and 2 102 159 positive leaves as in the paper. Compiling the two certificates took about 620 CPU-hours, or four and a half hours of elapsed time on four servers.

The Lean proof of Theorem 1.1 uses only the axioms `propext`, `Classical.choice` and `Quot.sound`. The kernel checked each module at compilation. At the close of the campaign, a check of the full development with an independent checker was still pending. The campaign used about 2.7 billion tokens, including at least 2.2 billion cached input tokens and 9.4 million output tokens (approximate token counts by model were 1.7 billion for Kimi K3.1 Pro, 735 million for Claude Opus 5.5, 125 million for GPT-6 Luna, 112 million for GPT-6.2 Sol, 13 million for Claude Sonnet 5.5 and 1 million for GPT-6.2 Astra). The formalization and instructions for reproducing the checks are available at `https:`

//github.com/igor-kortchemski/monotonicity-speed-biased-random-walk.

The author takes full responsibility for the results stated in this paper and for their presentation. He hopes to contribute to a revised digested version for arXiv.

2 Escape probabilities and the derivative of the speed

We derive the speed from two independent copies of the tree and differentiate it with respect to the bias. The first application is Theorem 2.7, which treats $0 < \lambda \leq 1$.

2.1 The model and the escape probabilities

Write $p_1 = \mathbb{P}(\nu = 1)$ and $p = 1 - p_1 > 0$. Let J have the law of ν given $\nu \geq 2$. Write ϱ for the root of the Bienaymé tree $\mathcal{T}$, $|x|$ for the generation of a vertex x, $\nu(x)$ for its number of children, and $\mathcal{T}_x$ for its descendant tree. We add a parent ϱ_* to the root, with $|\varrho_*| = -1$. On the resulting tree the walk jumps from a vertex with j children to its parent with probability $\lambda/(\lambda + j)$ and to each child with probability $1/(\lambda + j)$. From ϱ_* it returns to ϱ. At bias zero its speed is one.

Let $P_x^{\mathcal{T}}$ denote the law of the walk started at x given the tree. We write $\mathbb{P}$ for the joint law of the tree and the walk started at ϱ. The escape probability of the tree is

$$\beta = \beta_\lambda(\mathcal{T}) = P_\varrho^{\mathcal{T}}(X_n \neq \varrho_* \text{ for all } n \geq 0).$$

We use β, β' for the escapes of two independent trees and $\mathcal{T}_1, \dots, \mathcal{T}_\nu$ for the child subtrees of the root. For $|x| \leq n$, let $\beta^{(n)}(x)$ be the probability of reaching generation n from x before visiting its parent. Its value at the root is denoted by $\beta^{(n)}$.

Lemma 2.1. *Let $\lambda > 0$ and let $\mathcal{T}$ be a tree without leaves.*

(i) With $S = \sum_{i=1}^{\nu} \beta(\mathcal{T}_i)$, one has $\beta(\mathcal{T}) = S/(\lambda + S)$. Likewise $\beta^{(n)}(x) = S^{(n)}(x)/(\lambda + S^{(n)}(x))$ for $|x| < n$, where $S^{(n)}(x)$ is the sum of $\beta^{(n)}$ over the children of x.

(ii) As $n \to \infty$, $\beta^{(n)}(x)$ decreases to $\beta(\mathcal{T}_x)$. Both $\beta^{(n)}(x)$ and $\beta(\mathcal{T}_x)$ are nonincreasing functions of λ.

(iii) If $\lambda < 1$, then $\beta^{(n)}(x) \geq 1 - \lambda$ and $\beta(\mathcal{T}_x) \geq 1 - \lambda$.

(iv) If $\mathcal{T}$ is a Bienaymé tree and $0 < \lambda < m$, then $0 < \beta < 1$ almost surely.

Proof. (i) Conditioning on the first step and on the first return from a child gives $(\lambda + \nu)\beta = S + (\nu - S)\beta$, and likewise for the walk stopped at generation n.

(ii)–(iii) The events defining $\beta^{(n)}(x)$ decrease to escape up to a null event, since a walk confined to finitely many generations has a uniformly positive chance to hit the parent within a bounded number of steps. The map $S/(\lambda + S)$ increases in S and decreases in λ. Starting with boundary value 1, induction gives monotonicity in λ and, for $\lambda < 1$, the lower bound $1 - \lambda$, since every vertex has a child and $(1 - \lambda)/(\lambda + 1 - \lambda) = 1 - \lambda$. Both properties pass to the limit.

(iv) Finiteness of S gives $\beta < 1$. By Lyons [13, Proposition 6.4 and Theorem 4.3], the branching number equals m almost surely and the walk is transient for $\lambda < m$. Its probability β of never returning to ϱ_* after leaving it is therefore positive. □

2.2 The speed formula

For $0 < \lambda < m$ put $a = \lambda - 1$. For two independent escape probabilities β, β' write

$$\tau = \beta + \beta' - \beta\beta', \qquad D = a + \tau, \qquad N = \mathbb{E}\Big[\frac{\beta\beta'}{D}\Big], \qquad Z = \mathbb{E}\Big[\frac{\tau}{D}\Big],$$

with $D(b, b') = a + b + b' - bb'$ when its arguments need to be specified.

Let $(\beta_i)_{i\geq 0}$ be independent copies of β, independent of ν, and put $S = \sum_{i=1}^{\nu} \beta_i$.

Theorem 2.2 (Speed formula). *For every $0 < \lambda < m$, $|X_n|/n$ converges almost surely to a deterministic limit $v(\lambda) > 0$. With the notation above,*

$$v(\lambda) = \frac{\mathbb{E}[(\nu - \lambda)\beta_0/(a + \beta_0 + S)]}{\mathbb{E}[(\nu + \lambda)\beta_0/(a + \beta_0 + S)]} = \frac{N}{Z}. \tag{2.1}$$

Proof. First note the bounds

$$\min(1, \lambda)\max(\beta, \beta') \leq D = \lambda - (1 - \beta)(1 - \beta') \leq \lambda. \tag{2.2}$$

For $\lambda \geq 1$ we have $D \geq \tau \geq \beta$. Below one, Lemma 2.1(iii) gives $1 - \beta' \leq \lambda$ and hence $D \geq \lambda\beta$. By symmetry the bounds follow. Since $0 < \beta\beta' < \tau \leq 2\max(\beta, \beta')$, we have $0 < N < Z < \infty$.

For $p_1 < \lambda < m$ the degree quotient is the speed by Aïdékon [2, Theorem 1.1]. For $0 < \lambda \leq p_1$, use the occupation argument of Lin [12, Definition 1, Proposition 4 and Section 6]. In the stationary environment seen by the walk, the number of children at the current vertex has the law of ν weighted by $(\nu + \lambda)\beta_0/(a + \beta_0 + S)$. Here S is the conductance of the descendant tree. Lin obtains the time average of the number of children along the original walk. The same ergodic argument applies to any bounded function of that number. Take the conditional drift $j \mapsto (j - \lambda)/(j + \lambda)$. Its time average is the degree quotient. The difference from the average height increment tends to zero by the strong law for bounded martingale differences, since the artificial parent is visited only finitely often by transience. This proves the speed formula also in this range.

To obtain the two-copy form of Rousselin [19, Eq. (4.8)], set $b = S/(\lambda + S)$. By the escape recursion, b has the escape law and is independent of β_0. Since the $\nu + 1$ escapes are exchangeable given ν, and $D(b, \beta_0) = (1 - b)(a + \beta_0 + S)$, we have

$$\mathbb{E}\left[\frac{(\nu \pm \lambda)\beta_0}{a + \beta_0 + S}\right] = \mathbb{E}\left[\frac{S \pm \lambda\beta_0}{a + \beta_0 + S}\right] = \lambda\mathbb{E}\left[\frac{b \pm \beta_0(1 - b)}{D(b, \beta_0)}\right].$$

By symmetry these last expectations are λZ and λN, respectively. □

Remark 2.3. At the critical bias the speed extends continuously to $v(m) = 0$. More precisely,

$$v(\lambda) \leq \frac{m - \lambda}{m + \lambda}, \qquad \limsup_{\lambda \uparrow m} \frac{v(\lambda)}{m - \lambda} \leq \frac{m}{2\mathbb{E}[\nu^2]}.$$

In particular the left derivative at m is zero if $\mathbb{E}[\nu^2] = \infty$.

Proof. Write $b = \mathbb{E}\beta$. We first recover the bound $v \le b/(2-b)$ of Schwarzfischer [20, Lemma 3.1]. Since $t \mapsto \mathbb{E}[\beta'/D(t,\beta')]$ decreases, its covariance with β when evaluated at $t = \beta$ is nonpositive. Thus $N \le b\mathbb{E}[\beta'/D] = b(N+Z)/2$ by symmetry.

Conditional Jensen in the escape recursion gives $b \le \mathbb{E}[\nu b/(\lambda+\nu b)] \le mb/(\lambda+mb)$. The last bound gives $b \le (m-\lambda)/m$, hence the first speed bound. The first inequality, after division by b and rearrangement, gives $(m-\lambda)/b \ge \mathbb{E}[\nu^2/(\lambda+\nu b)]$. As $\lambda \uparrow m$, we have $b \to 0$. Fatou's lemma and $v \le b/(2-b)$ give the asserted limit bound. □

2.3 Differentiating the escape probability

Lemma 2.4. *Write $U = \mathbb{E}[1/\nu]$. If $\lambda > 0$ and $\lambda U < 1$, then $\mathbb{E}[1/\beta_\lambda] \le (1-\lambda U)^{-1}$. This includes a neighbourhood of bias one since $1/m \le U < 1$.*

Proof. Starting at $R_0 = 1$, the finite-depth reciprocals $R_n = 1/\beta^{(n)}$ satisfy

$$R_n = 1 + \frac{\lambda}{\sum_{i=1}^{\nu} 1/R_{n-1,i}} \le 1 + \frac{\lambda}{\nu^2} \sum_{i=1}^{\nu} R_{n-1,i}.$$

Taking expectations and iterating gives $\mathbb{E}R_n \le \sum_{j=0}^{n} (\lambda U)^j$. Let $n \to \infty$ by monotone convergence. Jensen and $\mathbb{P}(\nu \ge 2) > 0$ give the bounds on U. □

We call $u = -\lambda \partial_\lambda \beta$ the *response* of the tree. Aïdékon [1, Section 1] differentiated the escape recursion below bias one. For the finite-depth argument see Song, Wang and Xiang [21, Section 2], and for its expansion along a path see Schwarzfischer [20, Remark 2.7].

Proposition 2.5 (Escape response)**.** *Almost surely the escape of every subtree is C^1 on $(0,m)$. On each compact interval $I \subset (0,m)$ there is an integrable $K_I \ge 0$ such that $0 \le -\partial_\lambda \beta \le K_I \beta$ throughout I. Writing $S = \sum_i \beta_i$ and u_i for the child responses, we have*

$$u = \beta(1-\beta) + \frac{\lambda}{(\lambda+S)^2} \sum_i u_i.$$

In particular $u \ge \beta(1-\beta)$, and $2u \ge 1-\beta$ for $\lambda \le 1$.

Proof. At finite depth write $b_x = \beta^{(n)}(x)$ and $w_x = -\lambda \partial_\lambda \log b_x$. Differentiating the recursion gives

$$w_x = (1-b_x)\Big(1 + \sum_y \frac{b_y}{\sum_z b_z} w_y\Big),$$

where y, z run over the children of x and $w_x = 0$ at the boundary. Follow a downward path by choosing a child with probability proportional to b_y. Iteration expresses w_ϱ as the sum, over $1 \le k \le n$, of the path averages of the product of $1 - b_x$ along its first k edges, one factor at the parent end of each edge.

Fix $I = [\lambda_-, \lambda_+] \subset (0,m)$. On the edge to child y, bound its factor by $\sigma(y) = \lambda_+/(\lambda_+ + \sum_z \beta_{\lambda_+}(\mathcal{T}_z))$, where the sum is over the siblings of y. This bound is independent of the subtree at y and holds for every n and $\lambda \in I$. Let M_k be the largest product of σ along a

path of length k from the root. Let $\gamma \geq 1$ be an integer. Bounding M_k^γ by the sum of the γth powers and conditioning successively on the siblings of a marked child gives

$$\mathbb{E}M_k^\gamma \leq r_\gamma^k, \qquad r_\gamma = \mathbb{E}\left[\nu \left(\frac{\lambda_+}{\lambda_+ + \sum_{i=2}^{\nu} \beta_{\lambda_+}(\mathcal{T}_i)}\right)^\gamma\right].$$

Since the escapes are positive and $\mathbb{E}[\nu] < \infty$, dominated convergence gives $r_\gamma \to \mathbb{P}(\nu = 1) < 1$. Choose $\gamma \geq 1$ with $r_\gamma < 1$. Then $\mathbb{E}M_k \leq r_\gamma^{k/\gamma}$ and $K_I = \lambda_-^{-1} \sum_{k\geq 1} M_k$ is integrable.

It follows that $0 \leq -\partial_\lambda \log b_\varrho \leq K_I$ uniformly in depth. On every subtree the same bound gives equicontinuity, hence uniform convergence of the escapes on I to their positive limits. Each path average of fixed length now converges uniformly, since it involves finitely many positive child sums. Their summable bounds M_k give uniform convergence of w_ϱ. Passing to the limit in the logarithmic derivatives proves the first assertions on a countable compact exhaustion of $(0, m)$.

The recursion for u follows by differentiation. At finite depth take $u_x = -\lambda \partial_\lambda b_x$, $S = \sum_y b_y$ and $d_x = 2u_x - (1 - b_x)$. Then $(\lambda + S)^2 d_x/\lambda = \nu(x) - \lambda + \sum_y d_y$. For $\lambda \leq 1$ each source is nonnegative and the boundary values are zero. Induction and passage to the limit give $2u \geq 1 - \beta$. □

2.4 Differentiating the speed

Write $D(b, b') = a + b + b' - bb'$ as before. The speed is the unique c for which $F_c(b, b') = (bb' - c(b + b' - bb'))/D(b, b')$ has mean zero at $(b, b') = (\beta, \beta')$. A change of bias acts both on D and on the two escapes. For b in the range $[\max(0, 1 - \lambda), 1]$, set

$$H = \mathbb{E}\frac{\beta\beta' - v\tau}{D^2}, \qquad \Gamma(b) = \mathbb{E}\frac{P(\beta')}{D(b, \beta')^2}, \qquad P(t) = t^2 + a(1 + v)t - av.$$

Proposition 2.6 (Derivative of the speed)**.** *The functions N, Z and v are C^1 on $(0, m)$. The speed is C^1 up to zero with $v'_+(0) = -2\mathbb{E}[1/\nu]$. For $0 < \lambda < m$,*

$$\mathcal{C}(\lambda) := -\lambda Z v'(\lambda) = \lambda H + 2\mathbb{E}[u\,\Gamma(\beta)]. \tag{2.3}$$

For $K_0 = \mathbb{E}[1/D]$ and $K_1 = \mathbb{E}[\beta/D]$,

$$\lambda H + \mathbb{E}[(1 - \beta)\Gamma(\beta)] = K_1 - vK_0, \tag{2.4}$$

$$2(K_1 - vK_0) = 1 + v - \big(a + (\lambda + 1)v\big)K_0. \tag{2.5}$$

Proof. For fixed $c \in [0, 1]$, the partial derivatives of F_c in λ and b are $-(bb' - c\tau)/D^2$ and $P_c(b')/D^2$, where $P_c(t) = t^2 + a(1 + c)t - ac$. To differentiate the expectation, take a compact interval $I = [\lambda_-, \lambda_+] \subset (0, m)$ and put $B = \max(\beta, \beta')$. By (2.2), $D \geq \min(1, \lambda_-)B$. Also $|a| \leq D$ above bias one and $|a| \leq B$ below it. Proposition 2.5 then bounds $|\partial_\lambda \beta\, P_c(\beta')/D^2|$ by $C_I K_I$. The partial derivative at fixed escapes is bounded by τ/D^2. Away from bias one, D is bounded below on compact intervals. Near one, choose I with $\lambda_+\mathbb{E}[1/\nu] < 1$. Then $\tau/D^2 \leq 2/[\min(1, \lambda_-)^2\beta_{\lambda_+}]$ is integrable by Lemma 2.4. The same bounds give $\mathbb{E}[1/D] < \infty$.

Differentiation under the expectation for $c = 0, 1$ gives the regularity of N, Z and their quotient. At $c = v(\lambda)$, differentiating $N - vZ = 0$ gives (2.3) by symmetry of the two trees. For (2.4), use

$$\lambda(bb' - v\tau) + \tfrac{1}{2}[(1-b)P(b') + (1-b')P(b)] = D(b,b')\big((b+b')/2 - v\big).$$

Divide by D^2 and average. Finally $2K_1 = N + Z = (1+v)Z$ and $Z = 1 - aK_0$, proving (2.5).

At zero put $f = \beta_0/(a + \beta_0 + S)$. By (2.1), $1 - v = 2\lambda\mathbb{E}f/\mathbb{E}[(v+\lambda)f]$. For $\lambda \le 1/2$ the escapes lie in $[1-\lambda, 1]$, so $f \to 1/v$ as $\lambda \downarrow 0$. Each path factor in the proof of Proposition 2.5 is at most λ, giving $|\partial_\lambda \beta| \le \lambda^{-1}\sum_{k\ge 1}\lambda^k \le 2$. Since $a + \beta_0 + S \ge v/2$, the quotient rule gives $|f| + |\partial_\lambda f| \le C/v$. Thus $\mathbb{E}f/\mathbb{E}[(v+\lambda)f]$ tends to $\mathbb{E}[1/v]$ with bounded derivative. The assertion at zero follows. □

2.5 Bias at most one

Theorem 2.7. *For every $0 < \lambda \le 1$, we have $v'(\lambda) < 0$.*

Proof. We first compare the speed with the expected drift at the root. Put $d_j = (j-\lambda)/(j+\lambda)$ and let T_n be the sum of n independent escapes. Given $v = n - 1$, the weight of d_{n-1} in (2.1) is proportional to

$$(n + \lambda - 1)\mathbb{E}\frac{\beta_1}{T_n + \lambda - 1} = 1 + (1-\lambda)\mathbb{E}\frac{1-\beta_1}{T_n + \lambda - 1}$$

by exchangeability. Couple $T_{n+1} = T_n + \beta_{n+1}$ as in Lin [12, Section 6]. The weight decreases with n, whereas d_{n-1} increases. The covariance inequality gives

$$v \le \mathbb{E}d_v = 1 - 2\lambda\mathbb{E}\frac{1}{v+\lambda} < 1 - \lambda\mathbb{E}[1/v] \le \frac{1}{\mathbb{E}[1/\beta]}.$$

The strict inequality uses $v \ge \lambda$ and $\mathbb{P}(v \ge 2) > 0$, and the last inequality is Lemma 2.4.

It remains to relate this drift comparison to v'. Put $t = 1 - \lambda$. Since $b' \ge t$ on the escape law, $P(b') = b'(b'-t) + vt(1-b') \ge 0$, so $\Gamma \ge 0$ there. The bound $2u \ge 1 - \beta$ of Proposition 2.5 and (2.4) give $-\lambda Z v' \ge K_1 - vK_0$. For each $b' \ge t$, the function $b/D(b,b')$ has derivative $(b'-t)/D(b,b')^2 \ge 0$. Its covariance with $1/b$ is therefore nonpositive. Conditioning on β' gives $K_0 \le K_1\mathbb{E}[1/\beta]$. The strict drift bound above now gives $K_1 - vK_0 > 0$. □

3 Comparison of escapes above bias one

Throughout this section $1 < \lambda < m$, $a = \lambda - 1$ and $r = v/a$. We first control the speed using the mean of Y. An exchange of two subtrees then gives a weighted average of the response exactly.

At a vertex with one child, the following change of variable turns the escape recursion into multiplication by $1/\lambda$. Define

$$Y = \frac{\beta}{a + \beta'}, \qquad \frac{Y}{1-Y} = \frac{\beta}{a}.$$

By Lemma 2.1, $0 < Y < 1/\lambda$ almost surely.

Proposition 3.1 (Transformed escape and speed kernel). *If Y_i are the transformed escapes at the children, then*

$$\frac{\lambda Y}{1-\lambda Y} = \sum_{i=1}^{\nu} \frac{Y_i}{1-Y_i}. \tag{3.1}$$

For two independent copies Y, Y' we have

$$1-\lambda YY' = \frac{aD}{(a+\beta)(a+\beta')}, \qquad \frac{\beta\beta'}{D} = \frac{aYY'}{1-\lambda YY'},$$

$$\frac{\tau}{D} = \frac{Y(1-Y')+Y'(1-\lambda Y)}{1-\lambda YY'} = 1 - \frac{(1-Y)(1-Y')}{1-\lambda YY'},$$

and $\max(Y, Y') \le \tau/D \le Y + Y'$. Consequently $\mathbb{E}Y \le Z \le 2\mathbb{E}Y$.

Proof. For $S = \sum_i \beta_i$ the root recursion gives $\lambda Y = S/(a+S)$, which proves (3.1). Substitution of $\beta = aY/(1-Y)$ gives the identities. Finally,

$$\frac{\tau}{D} - Y = \frac{Y'(1-Y)(1-\lambda Y)}{1-\lambda YY'} \in [0, Y']$$

since $1-\lambda Y \le 1-\lambda YY'$. Symmetry gives the remaining bounds. □

3.1 A geometric comparison at fixed mean

Following the first child gives a geometric comparison for the moments of Y. The parameter is chosen to match its mean. Write $Q_n = \sum_{i=0}^{n-1} \lambda^i$.

Lemma 3.2 (Geometric moment comparison). *Let $\mu = \mathbb{E}Y$, $b = p_{\mathrm{eff}} = a\mu/(1-\mu)$ and let G_* have the geometric law on $\{1, 2, \ldots\}$ with parameter b. Then, for every integer $n \ge 1$,*

$$\mathbb{E}Y^n \le \mathbb{E}\lambda^{-nG_*} = \frac{b}{\lambda^n - 1 + b}, \qquad \mathbb{E}\lambda^{-G_*} = \mu.$$

In particular $\mathbb{E}Y^2/\mu \le [1+\lambda(1-\mu)]^{-1}$.

Proof. We have $0 < \mu < 1/\lambda$, so $0 < b < 1$. Write Y_1 for the first child's transformed escape and $T = a^{-1}\sum_{i=2}^{\nu} \beta_i$. By the escape recursion,

$$1-\lambda Y = (1-Y_1)h_T(Y_1), \qquad h_t(y) = [1+t(1-y)]^{-1}.$$

In particular $\lambda Y \ge Y_1$. Multiplying by $\sum_{j=0}^{n-1}(\lambda Y)^j$ gives $1-(\lambda Y)^n \ge (1-Y_1)h_T(Y_1)\sum_{j=0}^{n-1} Y_1^j$. Since T is independent of Y_1, the conditional mean of $h_T(Y_1)$ is $h(Y_1)$, where $h(y) = \mathbb{E}h_T(y)$ increases in y. Under the law with density $(1-Y_1)/(1-\mu)$, the mean of $h(Y_1)$ is $(1-\lambda\mu)/(1-\mu) = 1-b$. Its covariance with the increasing sum $\sum_{j=0}^{n-1} Y_1^j$ is nonnegative. Consequently

$$1-\lambda^n\mathbb{E}Y^n \ge (1-b)(1-\mathbb{E}Y^n).$$

Rearranging and summing the geometric law gives the moment comparison. For $n = 1$ there is equality, and $n = 2$ gives the last assertion. □

Lemma 3.3 (Speed ceiling and averaged gradient). *Let $1 < \lambda < m$, $\mu = \mathbb{E}Y$ and $r = v/a$. Then:*

(i) writing $b = p_{\mathrm{eff}}$ and $S(q) = \sum_{j\ge1} jq^{j-1}/Q_j$,

$$v \le \frac{abS(1-b)}{2\lambda - (1+\lambda)bS(1-b)} < b;$$

(ii) with $f_0(y) = y - r(1-y)(1-\lambda y)$, we have $\Gamma(0) = \mathbb{E}f_0(Y)$. For every $b \in [0,1]$, $0 < \Gamma(b) \le \mu$, and $\mathbb{E}[(\sqrt{\lambda}Y)^j f_0(Y)] > 0$ for every integer $j \ge 0$.

Proof. (i) For $0 \le w \le b$ let

$$\mathcal{H}_w(y,t) = \frac{ayt - w(y+t-(1+\lambda)yt)}{1-\lambda yt}, \qquad B_w(t) = at - w(1-t)(1-\lambda t).$$

By Proposition 3.1, $\mathbb{E}\mathcal{H}_w(Y,Y') = N - wZ$. For an independent variable $\xi \in [0, 1/\lambda]$ the expansion

$$\mathbb{E}\mathcal{H}_w(y,\xi) = -w\mathbb{E}\xi + \sum_{n\ge1} \lambda^{n-1}\mathbb{E}[\xi^{n-1}B_w(\xi)]\, y^n \tag{3.2}$$

converges uniformly for $0 \le y \le 1/\lambda$, since $\lambda y\xi \le 1/\lambda$. All its nonconstant coefficients are nonnegative as soon as $0 \le w \le b_0 := am_0/(1-m_0)$, where $m_0 = \mathbb{E}\xi > 0$. Indeed $\mathbb{E}B_w(\xi) \ge \mathbb{E}B_{b_0}(\xi) = \lambda b_0(m_0 - \mathbb{E}\xi^2) > 0$, and ξ^{n-1} and $B_w(\xi)$ are increasing functions of ξ. Thus increasing the moments in either escape variable increases the expectation of the kernel whenever the other variable satisfies the mean condition. By Lemma 3.2 we may replace Y, Y' in turn by independent copies of $Y_* = \lambda^{-G_*}$, since their means agree.

Write $q = 1-b$ and $S_* = S(q)$. To compute the quotient for the geometric law, group the pairs of lengths (G_*, G_*') by $k = G_* + G_*' - 1$. The k pairs each have probability $b^2(1-b)^{k-1}$, and $\sum_{i=1}^{k} \lambda^{-i} = (1-\lambda^{-k})/a$. Consequently

$$\mathbb{E}\frac{Y_*}{1-\lambda Y_*Y_*'} = \frac{b}{a}, \qquad \mathbb{E}\frac{Y_*Y_*'}{1-\lambda Y_*Y_*'} = \frac{b^2S_*}{a\lambda}.$$

The quotient for the geometric comparison law is $v_* = abS(q)/[2\lambda - (1+\lambda)bS(q)]$. Let U be uniform on $[1,\lambda]$. Since $Q_j/j = \mathbb{E}U^{j-1}$, Jensen's inequality gives $Q_j/j \ge ((1+\lambda)/2)^{j-1}$ and hence

$$S(q) \le \frac{1+\lambda}{1+\lambda-2q}. \tag{3.3}$$

It follows that $v_* \le b(1+\lambda)/(2\lambda - ab) < b$. The comparison at $w = v_*$ thus gives $N - v_*Z \le 0$.

(ii) Substitution gives $P(\beta')/(a+\beta')^2 = f_0(Y') = B_v(Y')/a$. The preceding bound on $\mathbb{E}B_b$ gives strict positivity. In $\Gamma(t)$ the additional factor $[(a+\beta')/D(t,\beta')]^2 = [1 + t(1-\beta')/(a+\beta')]^{-2}$ increases with β'. Since f_0 is nondecreasing, Chebyshev's inequality gives the two strict lower bounds, first with this factor and then with $(\sqrt{\lambda}Y')^j$. The factor is at most one and $f_0(Y') \le Y'$, so $\Gamma(t) \le \mu$. □

3.2 Exchanging two subtrees

In (2.3) we need to estimate $\mathbb{E}[u\Gamma(\beta)]$. Exchanging a child subtree with an independent tree determines a second average of u exactly. This exchange uses the kernel symmetry of Schwarzfischer [20, Lemma 2.6 and proof of Proposition 2.9]. Set

$$\Gamma_2(b) = \mathbb{E}\left[\frac{\beta'^2}{D(b,\beta')^2}\right], \qquad R_W = \mathbb{E}\left[\frac{\beta\beta'(1-\beta')}{D^2}\right].$$

If we subtract a multiple $\theta\Gamma_2$ from Γ, we can thus estimate the response only in the remaining term.

Write $y_b = b/(a+b)$ and

$$f_\theta(y) = y - \theta y^2 - r(1-y)(1-\lambda y).$$

Call θ *admissible* if $\theta \le 1$ and $\mathbb{E}f_\theta(Y) \ge 0$. Lemma 3.3(ii) gives admissibility of $\theta = 0$.

Lemma 3.4 (Single crossing). *Suppose a bounded function g is nonpositive below c and nonnegative above c. For every bounded nonnegative nondecreasing h,*

$$\mathbb{E}[g(Y)h(Y)] \ge h(c)\mathbb{E}[g(Y)].$$

For $b \in [0,1]$ we have

$$\Gamma(b) - \theta\Gamma_2(b) = (1-y_b)^2\mathbb{E}\left[\frac{f_\theta(Y)}{(1-\lambda y_b Y)^2}\right].$$

In particular this difference is nonnegative for every admissible θ.

Proof. The first inequality follows from $g(y)(h(y)-h(c)) \ge 0$. For the identity, divide $P(t) - \theta t^2$ by $(a+t)^2$ to obtain $f_\theta(y_t)$, and use the factorization of D in Proposition 3.1. At the two endpoints $0, 1/\lambda$ the quadratic f_θ has values $-r$ and $(\lambda-\theta)/\lambda^2$. For $\theta \le 1$ these have opposite signs, so f_θ crosses zero once. The weight $(1-\lambda y_b Y)^{-2}$ increases with Y. □

Theorem 3.5 (Subtree exchange). *We have $\mathbb{E}[u\Gamma_2(\beta)] = R_W$. For every admissible θ and every random lower bound $0 \le \underline{u} \le u$ on the same tree,*

$$\mathcal{C} \ge \lambda H + 2\theta R_W + 2\mathbb{E}[\underline{u}(\Gamma - \theta\Gamma_2)(\beta)].$$

Proof. Put $s_b = b(1-b)$, $S = \sum_{i=1}^{\nu}\beta_i$ and $T = a + S + \beta'$. The response recursion and $D = \lambda T/(\lambda+S)$ give

$$\mathbb{E}\left[\frac{(u-s_\beta)\beta'}{D^2}\right] = \frac{1}{\lambda}\mathbb{E}\left[\frac{(\sum_i u_i)\beta'}{T^2}\right] = \frac{1}{\lambda}\mathbb{E}\left[\frac{u'S}{T^2}\right] = \mathbb{E}\left[\frac{u's_\beta}{D^2}\right].$$

Conditionally on ν, the middle equality exchanges each child subtree with the independent tree. Their total escape $S+\beta'$ is preserved. All terms are integrable since $D \ge a$ and $\mathbb{E}u < \infty$. Use symmetry and $\beta' - s_{\beta'} = \beta'^2$ to obtain the identity. In (2.3) subtract $2\theta\mathbb{E}[u\Gamma_2(\beta)]$ and add $2\theta R_W$. By Lemma 3.4 the coefficient of the remaining response is nonnegative, which gives the lower bound. □

3.3 A test using the first two moments

The response produced at the root alone is $\beta(1-\beta)$. For this choice of $\underline{u}$, the subtree exchange gives the following test. Write $\mu = \mathbb{E}Y$, $m_2 = \mathbb{E}Y^2$ and $R_0 = \mu - \lambda m_2$.

Corollary 3.6. *Let $c = m_2/\mu$ and set*

$$Q = 2 - c - \lambda c^2 - r\{2/\mu - 2(1+\lambda) + (1+\lambda)c + a\lambda c^2\}. \tag{3.4}$$

If $Q > 0$, then $v'(\lambda) < 0$. The same conclusion holds whenever $\theta = 1$ is admissible.

Proof. Set $\theta = 1 - (1+v)R_0/(2\mu) \in (0,1)$. Substitution gives $2\mathbb{E}f_\theta(Y) = \mu Q$. In particular $Q > 0$ makes θ admissible. The alternative assumption $\mathbb{E}f_1(Y) \geq 0$ also gives $\mathbb{E}f_\theta(Y) > 0$, since $\theta < 1$. Take $\underline{u} = s_\beta = \beta(1-\beta)$ in Theorem 3.5. Then $R_W - \mathbb{E}[s_\beta \Gamma_2(\beta)] = \mathbb{E}[s_\beta s_{\beta'}/D^2]$. Since $\mathbb{E}[(\beta\beta' - v\tau)/D] = 0$, replace λ in λH by $\lambda - D = (1-\beta)(1-\beta')$. Substituting $P(t) = \lambda[t - v(1-t)] + (1-t)(v-t)$ and using symmetry gives

$$\mathcal{C} \geq \mathbb{E}\left[\frac{s_\beta}{D^2}\{2\lambda[\beta' - v(1-\beta')] + (2\theta - 1 + v)s_{\beta'}\}\right].$$

Indeed the remaining terms simplify by $\beta\beta' - v\tau + \beta(v-\beta') + \beta'(v-\beta) = (v-1)\beta\beta'$. In the transformed variables the lower bound is

$$\mathbb{E}\left[Y(1-\lambda Y)\frac{g(Y')}{(1-\lambda YY')^2}\right], \qquad g(y) = 2\lambda f_\theta(y) + \lambda(1+v)y(c-y).$$

Our choice of θ makes the last summand in $g(Y')$ centered. The quadratic g is negative at zero and positive at $1/\lambda$. Conditionally on Y the weight $(1-\lambda YY')^{-2}$ increases with Y' and is at least one. Lemma 3.4 thus bounds the preceding expectation below by $R_0\mathbb{E}g(Y')$, which is positive. □

3.4 An analytic test for high bias

The next test will be used twice. A large bias suffices on its own. For smaller biases, the first branching vertex will give a lower bound on the geometric parameter p_{eff} in Corollary 4.7.

Theorem 3.7. *Let $1 < \lambda < m$, $a = \lambda - 1$ and $b = p_{\text{eff}}$. If $24b + 10a \geq 8$, then $v'(\lambda) < 0$. In particular this holds for $9/5 \leq \lambda < m$.*

Proof. Write $L = \lambda$, $\mu = \mathbb{E}Y$, $c = \mathbb{E}Y^2/\mu$ and $r = v/a$. By Corollary 3.6, it suffices to prove $Q > 0$ in (3.4).

The expression Q decreases in c and r. Indeed the bracket increases in c and at $c = \mu$ equals $(1-L\mu)(2-2\mu-a\mu^2)/\mu > 0$. Write $q = 1 - b$. Lemmas 3.2 and 3.3 give $c \leq (L-q)/(L^2-q)$ and $r \leq bS(q)/[2L - (1+L)bS(q)]$. Substitution, using $\mu = b/(L-q)$, reduces the assertion to $F(S(q)) > 0$, where F is the decreasing affine function

$$F(t) = L(2L^3 - Lq - q^2) - (L-q)\{L^2(L+1) - Lq - q^2\}t.$$

Since $(1-L^{-j})j/Q_j = ajL^{-j}$, summing over j gives

$$S(q) = \frac{aL}{(L-q)^2} + \frac{1}{L}S(q/L).$$

Iterating once more and applying (3.3) to the remainder gives $S(q) \le S_2$, where

$$S_2 = \frac{aL}{(L-q)^2} + \frac{aL^2}{(L^2-q)^2} + \frac{L+1}{L^2(L+1)-2q}.$$

It remains to prove $F(S_2) > 0$. Set $z = aL/(a+b)$ and $C = 1 + 3z + z^2$. Then $a \le z \le L$ and $b = (L-1)(L-z)/z$. Substitution gives $F(S_2) = aL^2(C-L)\Phi/z^2$, where

$$\Phi = L - z - \frac{Lz}{(1+z)^2} - \frac{1+L}{2+(L+2)z} + \frac{Lz^2}{C-L}.$$

Here $C - L \ge 2z + z^2 > 0$. At fixed z, the last term increases with L. The derivative of the preceding terms is

$$1 - \frac{z}{(1+z)^2} - \frac{2+z}{[2+(L+2)z]^2} \ge 1 - \frac{1}{4} - \frac{1}{2} > 0.$$

As L decreases towards $\max(1, z)$, the product $b = (L-1)(L-z)/z$ decreases to zero. We therefore reach either the line $24b + 10a = 8$ or the boundary $b = 0$ while $24b + 10a \ge 8$. At $b = 0$ we have $L = z \ge 9/5$, and $\Phi > 0$ amounts to

$$\frac{z^2(z-1)}{(z+1)^2} > \frac{z+1}{(z+1)^2+1}.$$

The left side increases and the right side decreases for $z \ge 1$. At $9/5$ their values are $81/245 > 70/221$.

On $24b + 10a = 8$ put $u = 8b/(3a)$, so $(a, b) = (8, 3u)/(10 + 9u)$. After clearing positive denominators, put

$$H(u) = \frac{(10+9u)^{12}}{2^{20} \cdot 10^5}(L-q)(L^2-q)^2\{L^2(L+1)-2q\}F(S_2).$$

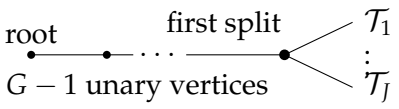
This polynomial has degree eleven. Expanding and rounding each coefficient down to the nearest integer gives, in increasing degree,

$$(23, 47, 0, -70, -59, 3, 31, 21, 7, 1, 0, 0).$$

Let h be the polynomial with these coefficients. In $h(u)/u^3$ the two negative terms are affine, so this function is convex for $u > 0$. Its tangent at 1 is $2 + 2u$. Thus $H(u) \ge h(u) > 0$ for $u \ge 0$, with $h(0) = 23$. Since F decreases, $F(S(q)) \ge F(S_2) > 0$ as required. □

4 The first branching vertex

Fix $1 < \lambda < m$. Write $q = 1 - p$, $a = \lambda - 1$ and $s = a/p$. We follow the only child until a vertex has two or more children. Let G count the vertices visited up to and including the first branching vertex. If the root branches, then $G = 1$. The split contributes the last factor of λ^{-1} in the formula below.



Lemma 4.1. *Let J be the degree of the first branching vertex and $\mathcal{T}_i$ its child subtrees. Put $x = 1/s$, $X_i = \beta(\mathcal{T}_i)/p$, $S = \sum_{i=1}^{J} X_i$, $W = S/(s+S)$ and $w = \mathbb{E}W$. Almost surely, simultaneously for $1 < \lambda < m$ on the same tree,*

$$Y = LW, \qquad L = \lambda^{-G}.$$

Here $\mathbb{P}(G = n) = p(1-p)^{n-1}$ for $n \geq 1$, independently of the split degree and its child trees. The split degree has the law of ν given $\nu \geq 2$. Given J, the X_i are independent with the law of β/p. In particular L and W are independent. With $Q_0 = 0$ and $\iota_n = (x + Q_n)^{-1}$ for $n \geq 0$, we have

$$\mathbb{E}L^n = x\iota_n, \qquad \mathbb{E}Y = \frac{w}{1+s}, \qquad 1 - w \leq \frac{se}{1-e} \leq s, \qquad e = \mathbb{E}[1/J].$$

Proof. At a unary vertex, (3.1) gives $Y = Y_1/\lambda$. At the split it gives $Y = W/\lambda$. Independence of the offspring and of the child subtrees gives the representation, and summing the geometric series gives the moments of L.

For the deficit $d = 1 - w$, write $\mu = \mathbb{E}Y = (1-d)/(1+s)$. At a split, $1 - W = (1 + \sum_{i=1}^{J} Y_i/(1-Y_i))^{-1}$ is concave in each child coordinate. Indeed, with the other odds summing to t, it is $(1-y)/(1+t(1-y))$. Jensen's inequality, first in the children given J, then in $1/J$, yields

$$d \leq \mathbb{E}\frac{1-\mu}{1+(J-1)\mu} \leq \frac{(1-\mu)e}{\mu + (1-\mu)e}.$$

The last step uses concavity of $z/(\mu + (1-\mu)z)$ on $[0,1]$. Substituting the value of μ gives $(1-e)d \leq se$, as required. □

4.1 A logarithmic lower bound for the response

Write $Y = \lambda^{-\ell}$. A unary vertex adds one to ℓ. We bound the response of the tree from below by the response obtained by holding this length fixed.

Lemma 4.2. *Let $1 < \lambda < m$ and $d_0 = a/\log\lambda$. For $0 < b \leq 1$ define*

$$\varphi_\lambda(b) = (1-b)d_0 \int_0^1 \frac{1-t}{(1+at)(1+at/b)}\,\mathrm{d}t. \tag{4.1}$$

Then $u \geq \varphi_\lambda(\beta) \geq 0$ almost surely, and

$$(1-Y)^2\varphi_\lambda(\beta) = Y[-d_0 \log Y - \lambda(1-Y)].$$

Proof. Put $Q = -\lambda\log\lambda\,\partial_\lambda$ and $t = \lambda Y$. Since $QY \geq -Y\log Y$ is equivalent to $Qt \geq -t\log t$, it suffices to propagate this inequality from the children to the parent. By (3.1), $t/(1-t) = \sum_i Y_i/(1-Y_i)$. If $QY_i \geq -Y_i\log Y_i$ for every child, differentiating gives

$$Qt \geq (1-t)^2 \sum_i \frac{Y_i}{1-Y_i}\frac{-\log Y_i}{1-Y_i} \geq -t\log t.$$

The last inequality uses $Y_i \leq t$ and the fact that $-\log y/(1-y)$ decreases in y. At the finite-depth boundary $Y = 1/\lambda$ gives equality. Induction and the convergence of derivatives in Proposition 2.5 give $QY \geq -Y\log Y$ at the root.

Differentiating $Y = \beta/(a+\beta)$ turns this into $(1-Y)^2 u \geq Y[-d_0 \log Y - \lambda(1-Y)]$. To identify the right-hand side, write the factor $(1-b)/[(1+at)(1+at/b)]$ in (4.1) as $(1+at/b)^{-1} - b(1+at)^{-1}$. Integration gives

$$\varphi_\lambda(b) = \frac{b}{a}\left[(a+b)\frac{\log(1+a/b)}{\log\lambda} - \lambda\right].$$

Substitute $b = \beta = aY/(1-Y)$ to conclude. □

For the next two arguments put $V = v/p$, $x = 1/s$ and $k = \sqrt{\lambda}$.

4.2 Bias just above one

Theorem 4.3. *For $0 < s \leq 1/8$ we have $v'(1+ps) < 0$.*

Proof. It suffices to consider biases for which $v(1+ps) \leq v(1)$. Indeed $v'(1) < 0$ by Theorem 2.7. At a first zero of v' in this interval the speed would still be at most $v(1)$. At bias one, exchangeability in (2.1) recovers the formula $v(1) = \mathbb{E}[(\nu-1)/(\nu+1)]$ of Lyons, Pemantle and Peres [14, Theorem 3.2]. Thus, writing $e = \mathbb{E}[1/J]$,

$$V = \frac{v(1+ps)}{p} \leq \mathbb{E}\frac{J-1}{J+1} \leq c := 1 - 4e/3,$$

since $1/(J+1) \geq 2/(3J)$ for $J \geq 2$.

Put $U = q + pe$, $A = 1 - e - sU$ and $R = p/\beta$. Then $A \geq 1 - e - s \geq 3/8$ and $M = \mathbb{E}R \leq 1/A$ by Lemma 2.4, since $1 - \lambda U = pA$. We first bound the harmonic mean of two independent copies R, R'. By Lemma 4.1, $R + 1/s = \lambda^G(1/s + 1/S)$ with G independent of S. In particular $R \geq pG$. Jensen's inequality conditional on G, G' and the geometric laws of G and $\min(G, G')$ give

$$\mathbb{E}|R - R'| \geq 2(M + 1/s)\left(1 - \frac{\mathbb{E}\lambda^{\min(G,G')}}{\mathbb{E}\lambda^G}\right) = \frac{2q(1+sM)}{1+q-sq^2} \geq \frac{2q(1+sM)}{2-s}.$$

Write $\Delta = \frac{1}{2}\mathbb{E}[(R-R')^2/(R+R')]$. Cauchy–Schwarz gives $\Delta \geq q^2 A(1+s/A)^2/(2-s)^2$. Here $(1+su)^2/u$ decreases up to $u = 1/s$ and $sM \leq s/A \leq 1/3$. For $j \geq 2$, the mean of $1/R_1, \ldots, 1/R_j$ is the average of their two-point means. By convexity of the reciprocal,

$$\mathbb{E}\frac{j}{\sum_{i=1}^j 1/R_i} \leq \mathbb{E}\frac{2RR'}{R+R'} = M - \Delta.$$

Averaging the recursion $R = p + \lambda/(\sum_{i=1}^\nu 1/R_i)$ gives $AM \leq 1 - \lambda e\Delta$.

Now $T = p/\tau = RR'/(R+R'-p)$. The identity

$$4T = R + R' + p + \frac{p^2 - (R-R')^2}{R+R'-p}$$

and $\mathbb{E}[p/(R+R'-p)] \leq \mathbb{E}[1/(G+G'-1)] = p$ give $2\mathbb{E}T - p \leq M - \Delta - pq/2$. Here $\mathbb{P}(G+G'-1=j) = jp^2q^{j-1}$. Since $A + \lambda e = 1 - sq$,

$$(1+\lambda e/A)\Delta \geq q^2\frac{(1-s)(1+s/A)^2}{(2-s)^2} \geq q^2\left(\frac{1}{4} + \frac{s}{2A}\right).$$

For the last inequality put $t = s/A \in [s, 1/3]$ and clear denominators. It becomes $4(1-s)t^2 \geq s^2(1+2t)$, which follows from $s \leq 1/8$. Consequently

$$2\mathbb{E}T - p \leq d := \frac{1 - sq^2/2}{A} - \frac{q(1+p)}{4}.$$

To bound the loss from using the response $(1-\beta)/2$ in (2.4), apply Jensen in (4.1) under the density $2(1-t)$. The integrand after removing this weight is a product of positive decreasing convex functions, and the mean of t is $1/3$. Using $d_0 \geq \sqrt{1+a} \geq 1 + a/3$ gives $1 - \beta - 2u \leq s(R-p)/(3+sR)$. By Lemma 3.3(ii), $0 < \Gamma \leq \mathbb{E}Y \leq 1/(1+s)$. Concavity in R therefore bounds the loss by $\mathcal{L} = s(1/A - p)/[(1+s)(3+s/A)]$. Also $\mathbb{E}[p/D] = \mathbb{E}[T/(1+sT)] \leq (p+d)/h$, where $h = 2 + s(p+d)$. Equations (2.4) and (2.5) now give $h\mathcal{C} \geq 1 - Vd - h\mathcal{L}$.

It remains to check the sign. We have $3p + 1/A \leq 6$, hence $h/[(1+s)(3+s/A)] \leq 2/3$. Using $d > 0$, $V \leq c$ and $1 - pA = (1+ps)U$ gives

$$Ah\mathcal{C} \geq e/3 - s(5+2ps)U/3 + c\,[Aq(1+p)/4 + sq^2/2].$$

Split $U = q + pe$ and use $e/3 = pe/3 + qe/3$. The coefficient of pe is at least $1/3 - 7/32 > 0$. The terms containing pq have coefficient $c(A-2s)/4 - 2s^2/3 \geq 0$, since $c \geq 1/3$ and $A - 2s \geq 1/8$. The remaining coefficient of q is at least

$$\frac{e}{3} - \frac{5s}{3} + \frac{c(1-e+s)}{4} \geq \frac{32e^2 - 28e + 7}{96} > 0.$$

The minimum in s is at $1/8$, and the quadratic has negative discriminant. We have proved $\mathcal{C} = -\lambda Z v' > 0$, which rules out a first zero of v'. □

4.3 Differentiation at the first split

We now keep the offspring law fixed and vary $x = p/(\lambda - 1)$. Increasing x decreases the bias, so the desired sign is $\partial_x V > 0$. Let $\eta_n = \mathbb{E}W^n$ and put

$$A_n = k^{n-1}\iota_n, \qquad m_n = A_n\eta_n, \qquad S = \sum_{n\geq 1} m_n^2, \qquad T = \sum_{n\geq 1} m_n m_{n+1}.$$

Here $\mathbb{E}Y^n = xk^{1-n}m_n$ by the independence of the prefix and the first split. Thus m_n is the terminal moment with its prefix factor included.

Lemma 4.4 (Differentiating the speed in its moments). *Let $h = m_1 - x[(1+\lambda)S/2 - kT]$ and $\gamma = (1+\lambda)/(2k)$. Then $V = S/(2h)$, $m_1/2 \leq h \leq m_1$, and*

$$h\partial_x V = V(S - \gamma T) + \sum_{n\geq 1} D_n \partial_x m_n, \qquad D_n = m_n - VC_n, \quad 0 \leq D_n \leq m_n, \tag{4.2}$$

where $m_0 = 0$ and

$$C_n = \mathbf{1}_{\{n=1\}} + x\big[k(m_{n-1} + m_{n+1}) - (1+\lambda)m_n\big].$$

Moreover $S - \gamma T \geq aS/(2\lambda) > 0$.

Proof. Expand $(1-\lambda YY')^{-1}$ in Proposition 3.1. Substitution of $\mathbb{E}Y^n = xk^{1-n}m_n$ gives $N/p = xS$ and $Z = 2xh$. Since $\mathbb{E}Y = xm_1$, the bounds on h follow from $\mathbb{E}Y \le Z \le 2\mathbb{E}Y$.

We differentiate the quotient $S/(2h)$ first with respect to x at fixed moments, then with respect to its individual coordinates. Since $x\partial_x\lambda = -a$, the first derivative is $V(S-\gamma T)/h$. For the coordinate derivatives, $\partial_{m_n}h = C_n$, so $h\partial_{m_n}V = m_n - VC_n$. To see the signs, expand the two deficit factors to obtain

$$C_n = k^{n-1}\mathbb{E}[Y^{n-1}(1-Y)(1-\lambda Y)] \ge 0.$$

Since $m_n = k^{n-1}\mathbb{E}[Y^n]/x$ and $Vx = v/a$, the positivity in Lemma 3.3(ii) gives $D_n \ge 0$. Also $x + Q_{n+1} = \lambda(x+Q_n) + q$ implies $m_{n+1} \le m_n/k$, so $T \le S/k$. This gives the stated bound on $S - \gamma T$.

On a compact subinterval of $(1, m)$ the prefix factors decay geometrically in n, also after differentiation up to a factor of order n. Conditionally on a branching root, $W = \lambda Y$ has integrably bounded derivative by Proposition 2.5. Thus the chain rule applies to the series. □

Lemma 4.5 (Response at the first split). *Let $M = 2\mathbb{E}[(1-W)^2]$, $t = \mathbb{E}W^2/w$ and*

$$c_n = \frac{a^2 n(n^2-1)}{12x\lambda^{3/2}}.$$

Then

$$\partial_x m_1 \ge wA_1\big[(1-t)/x - A_1\big], \qquad \partial_x m_n \ge m_n A_n(c_n - 1 + nA_nM) \quad (n \ge 2).$$

Let Y_i be the transformed escape of child i at the first split and put

$$\delta_i = \left(1 + \sum_{j\ne i}\frac{Y_j}{1-Y_j}\right)^{-1}, \qquad M_* = \mathbb{E}\sum_{i=1}^{J}\delta_i^2.$$

We have $M_* \ge M$. At every bias where $M_* \ge 1$, $\partial_x m_n \ge 0$ for all $n \ge 1$.

Proof. Write $A_n^{-1} = xk^{1-n} + U_n$, where $U_n = \sum_{j=0}^{n-1} k^{2j-n+1}$. Since $x\partial_x\lambda = -a$ and $\log\lambda \ge a/\lambda$, the first term has derivative $k^{1-n}(1 + (n-1)a/(2\lambda)) \le 1$. In U_n pair the exponents r and $-r$. For $r > 0$, the arithmetic–geometric mean inequality gives $k^r - k^{-r} \ge r(k - k^{-1})$. Summing r^2 over the positive exponents gives $n(n^2-1)/6$. Hence

$$-\partial_x U_n = \frac{a}{2x\lambda}\sum_{r>0} r(k^r - k^{-r}) \ge c_n, \qquad \partial_x A_n \ge A_n^2(c_n - 1).$$

Put $B = -s\partial_s W$ and $B_n = \mathbb{E}[W^{n-1}B]$. As $s = 1/x$, $x\partial_x\eta_n = nB_n$. At the split $W = S_0/(s+S_0)$ with S_0 decreasing in s, so $B \ge W(1-W)$. This proves the first bound because $\partial_x A_1 = -A_1^2$.

To use the child responses, differentiate $W/(1-W) = \sum_i Y_i/(1-Y_i)$, with $Y_i = L_iW_i$. This gives

$$B = \sum_i \left(\frac{1-W}{1-Y_i}\right)^2 [L_iB_i + (a/\lambda)G_iY_i].$$

We have $(1-W)/(1-Y_i) \geq \delta_i \geq 1-W$ and $W^{n-1} \geq Y_i^{n-1}$. Conditionally on the siblings and J, the selected child's subtree retains its original law. Since $\mathbb{E}L^n = x\iota_n$ and $(a/\lambda)\mathbb{E}[GL^n] = xA_n^2$, averaging gives

$$B_n \geq M_* x(\iota_n B_n + A_n^2 \eta_n), \qquad M_* \geq 2\mathbb{E}[(1-W)^2] = M.$$

Now $B_n \geq MxA_n^2\eta_n$ gives the second bound. If $M_* \geq 1$, then $(1-x\iota_n)B_n \geq xA_n^2\eta_n$. Since $\partial_x(x\iota_n) = nA_n^2$, this gives $\partial_x[(1-x\iota_n)\eta_n] \geq 0$ at the bias under consideration. Also $\partial_x(1/U_n) \geq 0$ by the paired sum above. The product of these two nonnegative factors is m_n, so $\partial_x m_n \geq 0$. □

Corollary 4.6. *For every transient $\lambda > 1$, if $w \leq 4/9$ then $v'(\lambda) < 0$. In particular the conclusion holds whenever $(m-\lambda)\mathbb{E}J \leq (4/5)m(\lambda-1)$.*

Proof. By Lemma 4.5, $M_* \geq 1$ gives $\partial_x m_n \geq 0$ for every $n \geq 1$. The first term in (4.2) is positive and $\partial_x V = -s^2 v'$, so it suffices to show that $M_* \geq 1$.

For fixed child odds put $z = W$. Jensen applied to $u \mapsto (1 + z/(1-z) - u)^{-2}$ across the J child odds gives $\sum_i \delta_i^2 \geq f_J(z)$, where $f_j(z) = j^3(1-z)^2/(j-z)^2$. The derivative in j has the sign of $j - 3z$, so the minimum over real $j \geq 2$ is at $j = \max(2, 3z)$. Its excess over $2 - 9z/4$ equals

$$\begin{cases} z(3z-2)^2/[4(2-z)^2], & z \leq 2/3, \\ (3z-2)^3/4, & z \geq 2/3. \end{cases}$$

Averaging gives $M_* \geq 2 - 9w/4 \geq 1$.

Finally $W = U/(a+U)$ for $U = \sum_{i=1}^J \beta_i$. Jensen gives $w \leq \mathbb{E}U/(a + \mathbb{E}U)$, where $\mathbb{E}U = \mathbb{E}J\,\mathbb{E}\beta \leq (m-\lambda)\mathbb{E}J/m$ by Remark 2.3. The stated condition gives $\mathbb{E}U \leq 4a/5$, hence $w \leq 4/9$. □

Corollary 4.7 (Offspring laws with an analytic proof)**.** *If $\mathbb{E}[1/\nu] \leq 31/49$, the analytic estimates in this paper give $v'(\lambda) < 0$ throughout the transient interval $0 < \lambda < m$.*

Proof. By Theorem 2.7 and Corollary 4.6 we may assume $\lambda > 1$ and $w > 4/9$. Set $e = \mathbb{E}[1/J]$, $C = 1 - \mathbb{E}[1/\nu] = p(1-e)$ and $b = p_{\text{eff}}$. With $d = 1-w$, Lemma 4.1 gives $(1-e)d \leq se$. Hence $b = psw/(s+d) \geq Cw$, and $b + ae \geq Cw + pse \geq C$. Since $e \leq 1/2$,

$$24b + 10a = 20(b + a/2) + 4b \geq (20 + 4w)C \geq \frac{196}{9}C \geq 8.$$

The last inequality uses $C \geq 18/49$. Theorem 3.7 gives the conclusion. □

Theorem 4.8. *If $1 < \lambda < m$, $\lambda \leq 9/5$ and $s \geq 4$, then $v'(\lambda) < 0$.*

Proof. By Corollary 4.6 we may assume $w > 4/9$. The coefficients D_n in (4.2) are nonnegative, so we may use the lower bounds on $\partial_x m_n$ from Lemma 4.5 term by term. Lemma A.1 then gives $\partial_x V > 0$. Since $\partial_x V = -s^2 v'$, the result follows. □

5 The intermediate range of scaled bias

For intermediate s we need an upper bound on the speed and a lower bound on $-\lambda Z v'$. Both use the first two moments of the escape W at the first split. We write $h_1 = \mathbb{E}L = x/(1+x)$, $k = \sqrt{\lambda}$ and $r = v/a$.

Theorem 5.1. *If $1 < \lambda < m$, $\lambda \le 9/5$ and $1/8 \le s \le 4$, then $v'(\lambda) < 0$.*

5.1 A lower bound in terminal moments

Write $z = kY$, $\mu_j(f) = \mathbb{E}[z^j f(Y)]$ and $P_j = \mu_j(y(1-y))$. We expand the factor $(1-\lambda YY')^{-2}$ to separate the two independent trees. In the factor from the first tree, the first split allows us to average the logarithmic response bound of Lemma 4.2 over the path length G.

Proposition 5.2. *For admissible θ, put*

$$g_\theta(y) = \lambda y(1-y) + 2\theta y(1-\lambda y) - \lambda r(1-y)(2-(1+\lambda)y).$$

Then

$$\mathcal{C} \ge \sum_{j\ge 0} (j+1)\left[P_j \mu_j(g_\theta) + 2F_j \mu_j(f_\theta)\right], \tag{5.1}$$

where, for $n = j+1$,

$$F_j = k^j x \iota_n \,\mathbb{E}\left[W^n \left\{ q\lambda^n \iota_n \iota_{n+1} + \left(\frac{a}{2+a} + q\iota_{n+1}\right)(1-W) + \frac{(1-W)^2}{2}\right\}\right].$$

The coefficients $\mu_j(f_\theta)$ are nonnegative. If $\mu_K(g_\theta) \ge 0$, then $\mu_j(g_\theta) \ge 0$ for every $j \ge K$.

Proof. Use $\underline{u} = \varphi_\lambda(\beta)$ in Theorem 3.5. In the static term $\lambda H + 2\theta R_W$, substitute $\beta = aY/(1-Y)$ and interchange the two trees in the second summand of the numerator of τ/D in Proposition 3.1. With Lemma 3.4, the bound becomes

$$\mathcal{C} \ge \mathbb{E}\left[\frac{Y(1-Y)g_\theta(Y') + 2\varphi_\lambda(\beta)(1-Y)^2 f_\theta(Y')}{(1-\lambda YY')^2}\right].$$

Expanding the inverse square gives the series in (5.1) with F_j replaced by $\mathbb{E}[z^j(1-Y)^2\varphi_\lambda(\beta)]$. The expansion is absolutely integrable since $\lambda YY' \le 1/\lambda$ and $0 \le \varphi_\lambda(\beta) \le u$ with $\mathbb{E}u < \infty$. Single crossing gives $\mu_j(f_\theta) \ge 0$ by weighting $f_\theta(Y)$ by z^j. For the last assertion, the quadratic g_θ has opposite endpoint values $-2\lambda r$ and $(1-1/\lambda)(1-v)$, and the extra weight from K to j is z^{j-K}.

It remains to bound the expectation multiplying $\mu_j(f_\theta)$. Write $d_0 = a/\log\lambda$. By Lemma 4.2 and $Y = LW$,

$$z^j(1-Y)^2\varphi_\lambda(\beta) = k^j L^n W^n [aG - d_0 \log W - \lambda + \lambda LW].$$

Since $\mathbb{E}L^n = x\iota_n$ and $a\mathbb{E}[GL^n]/\mathbb{E}L^n = \lambda^n\iota_n$, averaging the bracket with weight $L^n/\mathbb{E}L^n$ gives

$$q\lambda^n\iota_n\iota_{n+1} - (1-q\iota_{n+1})(1-W) - d_0\log W.$$

Here we used $\lambda\iota_{n+1}/\iota_n = 1 - q\iota_{n+1}$ and $\lambda^n\iota_n - \lambda + 1 - q\iota_{n+1} = q\lambda^n\iota_n\iota_{n+1}$. Now $-\log W \ge (1-W) + (1-W)^2/2$ and $d_0 = \int_0^1 \lambda^t\,\mathrm{d}t \ge \sqrt{\lambda} \ge 2\lambda/(1+\lambda)$. The resulting lower bound is F_j, as required. □

5.2 Bounding the speed at fixed moments

Put $d = \mathbb{E}[1-W] = 1-w$ and $\eta = \mathbb{E}[(1-W)^2]/d$. Keeping just two children at the first split already restricts this pair of moments.

Lemma 5.3 (Comparison with two children)**.** *With $\mu = \mathbb{E}Y$ and $m_2 = \mathbb{E}Y^2$, we have $\eta(1-m_2) \le 1-2\mu+m_2$.*

Proof. Write $A_i = 1-Y_i$ and $X = 1-W$. For $i = 1,2$, let S_i be the sum of the other child odds. Then $X = A_i/(1+A_iS_i)$ and S_i is independent of A_i. Under the law tilted by A_i^2, both $1/A_i$ and $\mathbb{E}[(1+A_iS_i)^{-2} \mid A_i]$ decrease in A_i. Their covariance is nonnegative, which gives

$$\frac{\mathbb{E}[X^2/A_i]}{\mathbb{E}X^2} \ge \frac{\mathbb{E}A_i}{\mathbb{E}A_i^2}.$$

The first two children alone give $1/X \ge 1/A_1 + 1/A_2 - 1$. Multiply by X^2, take expectations, and use the preceding inequality twice. This gives $1/\eta \ge 2\mathbb{E}A_1/\mathbb{E}A_1^2 - 1$. Substituting $\mathbb{E}A_1 = 1-\mu$ and $\mathbb{E}A_1^2 = 1-2\mu+m_2$ proves the claim. □

The first two moments of W also give an upper bound on the speed. For this comparison we only need upper bounds on the moments of W. The same two-point law gives all of them. For the general moment problem see Denuit, De Vylder and Lefèvre [10].

Lemma 5.4. *Suppose $0 \le d_0 \le d$ and $\eta \le \eta_0 \le 1$. Let W_+ take the values $1-\eta_0$ and 1 with respective probabilities d_0/η_0 and $1-d_0/\eta_0$. Form N_+, Z_+ by using independent copies of $Y_+ = LW_+$ in Proposition 3.1, with L independent of W_+. Then $v \le N_+/Z_+$. At $d_0 = \eta_0 = 0$ we set $W_+ = 1$.*

Proof. Under the law tilted by $1-W$, the mean of W is $1-\eta$. Jensen's inequality gives $\mathbb{E}[W^j(1-W)] \ge d(1-\eta)^j$. Consequently

$$\mathbb{E}W^n = 1 - \sum_{j=0}^{n-1} \mathbb{E}[W^j(1-W)] \le 1 - d_0 \sum_{j=0}^{n-1} (1-\eta_0)^j = \mathbb{E}W_+^n.$$

Independence of L gives the same moment comparison for Y and Y_+. For $T = Y$ or Y_+, Lemma 3.3 gives $v \le a\mathbb{E}Y/(1-\mathbb{E}Y) \le a\mathbb{E}T/(1-\mathbb{E}T)$. The nonconstant coefficients in (3.2) with $w = v$ and $\xi = T$ are therefore nonnegative. Replacing the two copies of Y in turn yields $0 = N - vZ \le N_+ - vZ_+$. The displayed moment formula also holds when $d_0 = \eta_0 = 0$. □

Proof of Theorem 5.1. By Corollary 4.6 we may assume $w > 4/9$. The binary comparison restricts d, η, since $\mathbb{E}Y = h_1(1-d)$ and $\mathbb{E}Y^2 = \mathbb{E}L^2(1-2d+d\eta)$. Lemma 5.4 bounds $r = v/a$ in terms of the same parameters. For every $y \in [0, 1/\lambda]$, both $f_\theta(y)$ and $g_\theta(y)$ decrease as r increases. An upper bound on r thus suffices both to choose an admissible θ and to bound the two factors $\mu_j(f_\theta), \mu_j(g_\theta)$ from below.

After averaging in L, the factors in (5.1) are expectations of polynomials in $X = 1-W$. A quadratic lower bound $P(u) \ge c_0 + c_1u + c_2u^2$ on $[0,1]$ gives $\mathbb{E}P(X) \ge c_0 + c_1d + c_2d\eta$. An upper bound works in the same way. For polynomial bounds under moment constraints see Bertsimas and Popescu [7, Section 2].

On each box we bound the products in (5.1) from below. For the nonnegative factors $P_j, F_j, \mu_j(f_\theta)$ we take the larger of zero and the computed lower bound. If the lower bound for $\mu_j(g_\theta)$ is negative, we multiply it by an upper bound for P_j. Once a lower bound for $\mu_K(g_\theta)$ is nonnegative throughout the box, Proposition 5.2 allows us to use zero as a lower bound for $\mu_j(g_\theta)$ at every later rank. A strictly positive lower bound for the partial sum through any rank $j \geq K$ then proves $v'(\lambda) < 0$ for every law and bias in the box. Lemma A.2 combines this criterion with three analytic tests to cover the intermediate range. □

Proof of Theorem 1.1. Theorems 2.7 and 3.7 cover $0 < \lambda \leq 1$ and $9/5 \leq \lambda < m$. In the remaining range put $s = (\lambda - 1)/p$. Theorems 4.3, 5.1 and 4.8 give $v'(\lambda) < 0$ for $s \leq 1/8$, $1/8 \leq s \leq 4$ and $s \geq 4$, respectively. Proposition 2.6 gives the asserted regularity and the right derivative at zero. The mean value theorem now gives strict decrease on $[0, m)$. □

A Reproducing the interval verifications

The programs and data used in this appendix are available in the `ancillary/` directory of the accompanying repository. To reproduce the checks, download the repository at this revision and run `sh REPLAY.sh` from that directory. This requires Python 3 and a C++20 compiler supporting IEEE binary64 arithmetic. The first step checks the printed boundary coefficients in Theorem 3.7 over the rationals. Both interval verifiers build in temporary directories.

Both programs use outward-rounded IEEE binary64 arithmetic in round-to-nearest mode with gradual underflow. After an addition, subtraction, multiplication, division, square root or explicit fused multiply-add, they take an adjacent floating-point value on the required side. The compilation flags suppress other contractions.

The derivative at the first split

Lemma A.1. *At every transient bias with $0 < x \leq 1/4$, $0 < a \leq 4/5$ and $w > 4/9$, the right-hand side of (4.2) is positive.*

Proof. To handle the mean w without subdividing its interval, put $e_n = \eta_n/w$, $u_n = m_n/w = A_n e_n$, $P = S/w^2$, $Q = T/w^2$ and $H = h/w$. Under the law tilted by W, the mean of W is $t = e_2$. Thus $w \leq t$ and, by Jensen's inequality,

$$t^{n-1} \leq e_n \leq t \quad (n \geq 2), \qquad 0 \leq e_n - e_{n+1} \leq 1 - t \quad (n \geq 1).$$

The latter difference is exactly $1 - t$ at $n = 1$.

For $G_n = 2Hu_n - PC_n = 2HD_n/w$, (4.2) gives

$$\frac{2h^2}{w^3}\partial_x V = wP(P - \gamma Q) + \sum_{n \geq 1} G_n \frac{\partial_x m_n}{w}.$$

For the first term, Lemma 4.4 gives $P - \gamma Q \geq aP/(2\lambda)$. At the given law $0 \leq G_n \leq 2u_1 u_n$. By Lemma 4.5 we may therefore bound the sum below by replacing $\partial_x m_n/w$ with

$$\ell_1 = A_1[(1-t)/x - A_1], \qquad \ell_n = u_n A_n[c_n - 1 + 2nA_n(1 - w(2-t))] \quad (n \geq 2).$$

Both G_n and ℓ_n are affine in w once the normalized moments are fixed. Indeed $H = u_1 - xw[(1+\lambda)P/2 - kQ]$ and C_n is affine by its definition.

We retain $K = 16$ terms. For $d_n = u_n - u_{n+1}$,

$$2(P-Q) = u_1^2 + \sum_{n\ge 1} d_n^2, \qquad d_n = (A_n - A_{n+1})e_n + A_{n+1}(e_n - e_{n+1}).$$

Here $A_n - A_{n+1} = A_n(k - 1 + q\iota_{n+1})/k \ge 0$. Telescoping bounds the tail of $\sum d_n^2$ after K by u_{K+1}^2. Likewise $R_K := \sum_{n>K} u_n^2 \le e_{K+1}^2 \iota_K$, since $A_n^2 \le \iota_{n-1} - \iota_n$. As $A_n \le 1/n$ we have $\ell_n \ge -u_n/n$ for $n > K$. The omitted derivative terms thus sum to at least $-2u_1 R_K/(K+1)$.

The program bisects the box $[0, 1/4] \times [0, 13/16] \times [7/16, 1]$ in (x, a, t), visiting both closed children at each split. On a box with upper endpoint t_+ we have $7/16 < w \le t_+$. The retained lower bound is quadratic in w. The three basis functions below are nonnegative and sum to one. Thus, after rescaling $[7/16, t_+]$ to $[0, 1]$, it suffices to bound the three coefficients in the basis $(1-z)^2, 2z(1-z), z^2$. For a product of affine functions with endpoint values f_0, f_1 and g_0, g_1, these coefficients are $f_0 g_0, (f_0 g_1 + f_1 g_0)/2, f_1 g_1$. The prefix factors are enclosed using the monotonicity of U_n in k from Lemma 4.5. In the positive terms proportional to $1/x$ the upper endpoint of x gives a lower bound. If a_- is the lower endpoint of a, we may also replace a^2 by a_-^2 in the numerator of c_n since $G_n u_n A_n \ge 0$. After adding the tail bound, outward rounding gives $2h^2 \partial_x V / w^3 > 2^{-24}$ on all 41 396 leaves. □

The intermediate range

Lemma A.2. *At every transient bias with $1/4 \le x \le 8$, $0 < a \le 4/5$ and $w > 4/9$, the bounds of Section 5 give $v'(\lambda) < 0$.*

Proof. The program starts with $[1/4, 8] \times [0, 13/16]$ in (x, a). It discards rectangles on which $ax > 1$, since $ax = p$. On each remaining rectangle it starts with $0 \le d \le \min(5/9, 1/x)$ and $d \le \eta \le 1$, by Lemma 4.1 and Jensen's inequality. Lemma 5.3 narrows the interval for η.

In each box let d_0 be the lower endpoint for d and η_0 the upper endpoint for η. Lemma 5.4 bounds the speed using only

$$\mathbb{E}W_+^j = 1 - d_0 \sum_{i=0}^{j-1} (1-\eta_0)^i, \qquad \mathbb{E}[W_+^j(1-W_+)] = d_0(1-\eta_0)^j.$$

These formulas extend to $d_0 = \eta_0 = 0$. For the two-point law write $\mu_j^+(f) = \mathbb{E}[(kY_+)^j f(Y_+)]$. The speed computation retains the numerator ranks $1, \dots, K$ and the denominator ranks $0, \dots, K-1$ with $K = 128$ in the expansions

$$N_+/a = \lambda^{-1} \sum_{j\ge 1} \mu_j^+(1)^2, \qquad Z_+ = \sum_{j\ge 0} \mu_j^+(y)\,[\mu_j^+(1-y) + \mu_j^+(1-\lambda y)].$$

For the numerator tail, use $\mathbb{E}W_+^j \le \mathbb{E}W_+^{K+1}$ for $j > K$ and sum $\lambda^{j-1}\iota_j^2 \le \iota_{j-1} - \iota_j$. This bounds the tail in N_+/a by $x^2 \iota_K (\mathbb{E}W_+^{K+1})^2$. For the denominator, multiply the pointwise inequality $\tau/D \ge Y_+$ of Proposition 3.1 by $(\lambda Y_+ Y_+')^K$ and average. The tail in Z_+ is thus at least $\mu_K^+(y)\mu_K^+(1)$. We also use $Z_+ \ge \mathbb{E}Y_+ \ge \mathbb{E}Y$ and $r \le \mathbb{E}Y/(1-\mathbb{E}Y)$. When a box

in (d, η) is split, its upper bound on $R = r/h_1$ remains valid on both children. A child that computes a local bound keeps the smaller of that bound and the inherited one.

Three analytic tests can settle a box. They are $2d\eta = M \geq 1$ by Lemma 4.5 and (4.2), $24b + 10a \geq 8$ with $b = a\mathbb{E}Y/(1 - \mathbb{E}Y)$ by Theorem 3.7, and admissibility of $\theta = 1$ by Corollary 3.6. Elsewhere we use a value of $\theta \in [0, 1]$ certified by the moment bounds, with $\theta = 0$ always admissible. The program divides each of the four factors in (5.1) by $h_1 = x/(1 + x)$, so each partial sum is divided by h_1^2.

For each polynomial P in $X = 1 - W$ write

$$P(u) = \sum_{i=0}^{B} p_i \binom{B}{i} u^i (1 - u)^{B-i}, \qquad B = 28.$$

A quadratic $c_0 + c_1 u + c_2 u^2$ has Bernstein coefficient $c_0 + c_1 i/B + c_2 i(i - 1)/[B(B - 1)]$ at i. If these coefficients are at most p_i for every i, nonnegativity of the basis functions gives $\mathbb{E}P(X) \geq c_0 + c_1 d + c_2 d\eta$. Applying the same comparison to $-P$ gives an upper bound.

For $n + h \leq B$ the Bernstein coefficient of $(1 - u)^n u^h$ at i is $(B - i)_n (i)_h / (B)_{n+h}$, with $(k)_j$ the falling factorial, by multiplying by $[u + (1 - u)]^{B-n-h}$. Writing the quadratic coefficients as $c_0 + li + ci^2$, the program chooses l, c approximately and checks the minorant at all 29 coefficients. It rounds $l + ci$ upward and then $p_i - i(l + ci)$ downward. It tests partial sums through rank 22, stopping as soon as the criterion in the proof of Theorem 5.1 holds.

For $0 \leq k \leq 4$ divide $[2^k/4, 2^{k+1}/4]$ into eight equal x-intervals, and divide $[0, 13/16]$ into thirteen equal a-intervals. On each of the resulting 520 rectangles, the file `joint/subdivisions.txt` records a finite binary tree. Each branch bisects either x or a at its exact dyadic midpoint. Thus the closed leaves cover their root rectangle. The replay verifies every leaf and compares the digest of the outcomes with the recorded value. The partition has 8 405 rectangles, of which 220 have $ax > 1$ throughout. On the other rectangles the verification excludes 3 532 moment boxes and proves strict positivity on 2 102 159, including 12 634 by the three analytic tests. □

References

[1] Elie Aïdékon. Monotonicity for $\lambda \leq 1/2$. Unpublished note, 20 September 2013, 2013.

[2] Elie Aïdékon. Speed of the biased random walk on a Galton–Watson tree. *Probab. Theory Relat. Fields*, 159(3–4):597–617, 2014.

[3] Gérard Ben Arous and Alexander Fribergh. Biased random walks on random graphs. In Vladas Sidoravicius and Stanislav Smirnov, editors, *Probability and Statistical Physics in St. Petersburg*, volume 91 of *Proceedings of Symposia in Pure Mathematics*, pages 99–153. American Mathematical Society, Providence, RI, 2016.

[4] Gérard Ben Arous, Alexander Fribergh, Nina Gantert, and Alan Hammond. Biased random walks on Galton–Watson trees with leaves. *Ann. Probab.*, 40(1):280–338, 2012.

[5] Gérard Ben Arous, Alexander Fribergh, and Vladas Sidoravicius. Lyons–Pemantle–Peres monotonicity problem for high biases. *Commun. Pure Appl. Math.*, 67(4):519–530, 2014.

[6] Gérard Ben Arous, Yueyun Hu, Stefano Olla, and Ofer Zeitouni. Einstein relation for biased random walk on Galton–Watson trees. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 49(3):698–721, 2013.

[7] Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM J. Optim.*, 15(3):780–804, 2005.

[8] Adam Bowditch and Yuki Tokushige. Differentiability of the speed of biased random walks on Galton–Watson trees. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 17(1):609–642, 2020.

[9] Bernard Bru, François Jongmans, and Eugene Seneta. I. J. Bienaymé: Family information and proof of the criticality theorem. *Int. Stat. Rev.*, 60(2):177–183, 1992.

[10] Michel Denuit, Etienne De Vylder, and Claude Lefèvre. Extremal generators and extremal distributions for the continuous s-convex stochastic orderings. *Insur. Math. Econ.*, 24(3):201–217, 1999.

[11] GPT-6 Astra and Heng Ma. Monotonicity of the speed of biased random walks on Galton–Watson trees. Hexagon:2611.00735v1, 2026. Full version and exact computations.

[12] Shen Lin. Harmonic measure for biased random walk in a supercritical Galton–Watson tree. *Bernoulli*, 25(4B):3652–3672, 2019.

[13] Russell Lyons. Random walks and percolation on trees. *Ann. Probab.*, 18(3):931–958, 1990.

[14] Russell Lyons, Robin Pemantle, and Yuval Peres. Ergodic theory on Galton–Watson trees: speed of random walk and dimension of harmonic measure. *Ergodic Theory Dyn. Syst.*, 15(3):593–619, 1995.

[15] Russell Lyons, Robin Pemantle, and Yuval Peres. Biased random walks on Galton–Watson trees. *Probab. Theory Relat. Fields*, 106(2):249–264, 1996.

[16] Russell Lyons, Robin Pemantle, and Yuval Peres. Unsolved problems concerning random walks on trees. In Krishna B. Athreya and Peter Jagers, editors, *Classical and Modern Branching Processes*, volume 84 of *The IMA Volumes in Mathematics and its Applications*, pages 223–237. Springer, New York, 1997.

[17] Madhulatha Mandarapu and Sandeep Kunkunuru. A Computer-Assisted Proof of Speed Monotonicity for the Biased Random Walk on a Galton–Watson Tree Beyond the Known Range. Preprint, arXiv:2609.29894, 2026.

[18] Behzad Mehrdad, Sanchayan Sen, and Lingjiong Zhu. The speed of a biased walk on a Galton–Watson tree without leaves is monotonic with respect to progeny distributions for high values of bias. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 51(1):304–318, 2015.

[19] Pierre Rousselin. Invariant measures, Hausdorff dimension and dimension drop of some harmonic measures on Galton–Watson trees. *Electron. J. Probab.*, 23:1–31, 2018. Paper No. 46.

[20] Felix Schwarzfischer. A proof of the Lyons–Pemantle–Peres monotonicity conjecture for high offspring laws. Preprint, arXiv:2611.25882, 2026.

[21] He Song, Longmin Wang, and Kainan Xiang. The speed of a biased walk on a Galton–Watson tree without leaves is monotonic for low values of bias. *J. Appl. Probab.*, 62(3):1044–1052, 2025.