hexagon:2610.00681v1

THREE AXES OF COLLATZ DYNAMICS: 2-ADIC MEASURES, THE DIMENSION SPECTRUM OF INVERSE TREES, AND CRITICAL EXPONENTS OF DENSITY PROPAGATION

CHUNPING ZENG

Abstract. This work presents a systematic measurement and critical diagnosis of three research axes on the Collatz $(3n+1)$ problem, with the aim of determining which tools can advance the problem and which are structurally incapable of doing so. The three axes are: the 2-adic statistics of individual orbits, the combinatorial structure of the inverse tree, and the critical exponent of density propagation (the Terras–Krasikov–Lagarias line).

On the first axis we establish a strict identity and a divergence criterion. For the odd-step map $m \mapsto m' = (3m+1)/2^a$ with $a = v_2(3m+1)$, one has $\log m_{K+1} - \log m_1 = K\log 3 - (\sum_i a_i)\log 2 + \sum_i \log(1+\frac{1}{3m_i})$, verified to 10^{-13}. Divergence is therefore equivalent to $\bar{a} < \log_2 3$. Over 298 million odd steps spanning nine orders of magnitude we obtain $\bar{a} = 1.996929$, matching the i.i.d. geometric prediction $E[a] = 2$, with a safety margin of 0.412 (a 20.6% drop would be needed for divergence). The distribution of a is exactly geometric; the autocorrelation is approximately zero; a surrogate control (preserving the marginal, destroying the serial structure) shows no difference between real and shuffled sequences at long windows.

On the second axis we derive a closed form for one-step propagation on the inverse tree. The two extensible residue classes give $\Lambda_1(\beta) = \frac{(3/4)^\beta}{1-4^{-\beta}}$ and $\Lambda_2(\beta) = \frac{(3/2)^\beta}{1-4^{-\beta}}$. Because preimages are uniformly distributed over the three residue classes, the extension matrix has rank one and $\rho(\beta) = \frac{\Lambda_1+\Lambda_2}{3} = \frac{3^{\beta-1}}{2^\beta-1}$. At $\beta = 1$ one has $\Lambda_1 = 1$ (critical) and $\Lambda_2 = 2$ (supercritical), the latter being exactly cancelled by the one-third loss into leaf nodes, so that the total spectral radius is exactly 1. This turns "density propagation stops at $\beta = 1$" from an observation into a corollary of the closed form.

On the third axis the depth-k reachable set is measured as a genuine fractal of dimension γ_k ($\gamma_1 = 0.101$ logarithmic, $\gamma_{12} = 0.603$, $\gamma_{24} = 0.824$, $\gamma_{40} = 0.923$, increasing monotonically to 1). A decisive test rejects the competing polylogarithmic hypothesis (at $k = 24$: measured growth $\times 294$, polylog prediction $\times 1.7\times 10^7$). We further show that the type-transfer spectral radius cannot produce any γ_k: its zero is constantly the asymptotic ceiling 1; obtaining γ_k requires the infinite-branch Ruelle operator on the spatial variable — which we construct and then falsify by a fit-free hard criterion (operator count 5398 > the hard bound $X/2 = 5000$).

Two popular frameworks are falsified: the local-entropy framework (its numerical verification is vacuous and its upper-bound theorem conflicts with the random model) and the conjecture that truncated propagation possesses a finite critical root (the truncated ratio is always below 1 and increases monotonically to 1). All numerics carry finite-size labels; the finite-size deviation is proved to be $O(1/\ln X)$ rather than $O(1/X)$. A methodological ledger records nine recurring errors and their countermeasures.

The three axes are finally unified by the master identity $\rho(\beta) \equiv M(\beta-1)$, where M is the moment generating function of the single-step log-increment: the inverse tree, the stopping-time large deviations, and the orbit drift are three readings of one function. Five corollaries follow, including $\rho'(1) = \delta(3)$ (the spectral slope at criticality equals the drift) and $\beta_c = -\ln \min_\beta \rho(\beta)$ (the far-tail rate is the negative logarithm of the spectral minimum). A three-way shootout against the Bessel(3)/meander hypothesis for the survival-profile shape falsifies that hypothesis ($z = \pm 50$–69), isolating an open phenomenon: Collatz survival trajectories have a definite "initial plateau, mid-concentrated descent" shape that no model with the correct marginal reproduces.

Date: October 2026.

1. Introduction: three axes and the methodological stance

1.1. **The odd-step submap and notation.** The Collatz map is $C(n) = n/2$ (n even) and $C(n) = 3n + 1$ (n odd). Since even steps carry no information, the standard compression keeps only odd values, giving the odd-step map

$$T(m) = \frac{3m+1}{2^{a(m)}}, \qquad a(m) := v_2(3m+1) \geq 1.$$

We write $m_1, m_2, \ldots$ for the odd values along an orbit, $m_{i+1} = T(m_i)$. Here v_2 is the 2-adic valuation, $\bar{a}$ denotes the Cesàro average along an orbit, and log is the natural logarithm except at 2-adic thresholds, where $\log_2$ is used and marked explicitly.

1.2. **The three research axes.** *Axis 1* is the 2-adic statistics of single orbits: from $3m_i + 1 = 2^{a_i} m_{i+1}$ one takes logarithms, and $\sum_i a_i$ controls growth. This is the territory of random models, whose strongest result is Tao's logarithmic-density theorem. *Axis 2* is the combinatorial structure of the inverse tree: running backwards from 1 via $n = (2^a t - 1)/3$ generates all preimages; this deterministic branching structure, independent of any randomness assumption, is where all unconditional theorems (Krasikov–Lagarias's 0.84, Tao's density result) reside. *Axis 3* is density propagation, the quantitative version of axis 2: track the propagation of reachable mass across residue classes mod 2^k and characterize provable density by the critical exponent of a propagation operator.

1.3. **Methodology: shootouts, controls, falsifiable criteria.** This paper follows three disciplines. *First*, every closed form must be verified by an independent numerical path, covering parameter space (several β, several branches, several scales), not a single point. *Second*, every "structural signal" must pass a surrogate control: construct a control sequence preserving the marginal distribution while destroying serial structure; if the real sequence is indistinguishable from the control, the alleged signal is a consequence of distributional shape, not of serial structure. *Third*, every fit must name a falsifiable competing hypothesis and explicitly reject one of the two. These disciplines are not formalism: the nine errors recorded in this project (Section 12.3) are all violations of one of them, and each correction was supplied by another. Every number below carries its finite-size label; any unlabeled quantity must not be used as an asymptotic conclusion.

2. The strict identity and the growth axis

2.1. **Identity and divergence criterion.** From $3m_i+1 = 2^{a_i} m_{i+1}$ one gets $m_{i+1} = \frac{3m_i}{2^{a_i}}\big(1+\frac{1}{3m_i}\big)$. Taking logarithms and summing,

$$\log m_{K+1} - \log m_1 = K \log 3 - \Big(\sum_{i=1}^{K} a_i\Big) \log 2 + \sum_{i=1}^{K} \delta(m_i), \qquad \delta(m) := \log\Big(1 + \frac{1}{3m}\Big). \tag{2.1}$$

The identity is verified numerically to 10^{-13} over 298 million odd steps; it is exact, with no approximation. Defining the per-step net growth $g_i := \log 3 - a_i \log 2 + \delta(m_i)$, one has $\sum_{i=1}^{K} g_i = \log \frac{m_{K+1}}{m_1}$, so divergence is equivalent to $\limsup_K \sum_i g_i = +\infty$. Ignoring δ (whose size is far below the first two terms) gives the classical criterion

$$\bar{a} := \lim_K \tfrac{1}{K} \sum_{i=1}^{K} a_i \;<\; \log_2 3 = 1.584963 \quad \Longleftrightarrow \quad \text{divergence.} \tag{2.2}$$

This is a strict equivalence, not a heuristic. Note the gap of $2 - 1.584963 = 0.415$ between $E[a] = 2$ (uniform model) and the threshold — this gap is the entire difficulty of the conjecture.

2.2. **The correction term and the effective threshold.** The δ term is not negligible decoration. Writing the criterion per step, step i is a growth step iff

$$a_i < A^*(m_i) := \log_2 3 + \frac{\delta(m_i)}{\log 2}.$$

The effective threshold ranges over $[\log_2 3,\, 2]$ as m varies:

m	$\delta(m)$	$A^*(m)$	remark
1	0.2877	2.000	trivial cycle, exactly neutral
3	0.1054	1.737	
5	0.0645	1.678	
31	0.0106	1.600	
103	3.3×10^{-4}	1.5855	correction already negligible
$\to \infty$	$\sim \frac{1}{3m}$	$\to 1.5850$	back to the classical threshold

At $m = 1$ an exact fact is worth recording: $a(1) = v_2(4) = 2$, $A^*(1) = 2$, so $g = 0$ — the trivial cycle is a fixed point of net growth, an exact identity and not a numerical coincidence.

2.3. **Stratified measurement: threshold elevation absorbed by the measured mean.** A natural guess is that small-m regions enjoy elevated thresholds (up to 2), so orbits lingering near small values face looser growth conditions and might form the most dangerous escape window. This guess is falsified by stratified measurement (Table below, 298 million odd steps): the small layer does have higher threshold (1.728 vs 1.585), but the measured $\bar{a}$ of that layer simultaneously soars to 3.927 (vs 2.000 for large layers), and the net effect makes $\bar{g}$ *more* negative in the small layer (-1.894; it descends faster). The growth-step fraction is stable at 36%–37% across all scales. The mechanism is structural: threshold and measured mean are not independent quantities — they are two faces of the same $3m + 1$, and identity (2.1) forces them to move together.

3. The 2-adic distribution and the safety margin

3.1. **Single-step distribution and autocorrelation.** The single-step distribution has an exact shape: for uniformly random odd m,

$$\mathbb{P}\big(a(m) = k\big) = 2^{-k}, \qquad k = 1, 2, 3, \ldots \tag{3.1}$$

(an exact identity: among the 2^k odd residue classes mod 2^{k+1} exactly one has $v_2(3m+1) = k$). Measured ratios $\mathbb{P}(a = k)/2^{-k}$ fall in $[0.97, 1.00]$ for $k \le 5$; the sole exception is $k = 4$ with 1.195. On the 2-adic axis, Collatz orbits are indistinguishable from i.i.d. geometric randomness (to 10^{-3}). One correction is reported: an initial measurement of $\rho(1) = +0.027$ was an uncentred second moment, not an autocorrelation; the correctly centred values are

$$\rho(1) = +0.0007, \qquad \rho(2) = -0.0031, \qquad \rho(3) = -0.0000.$$

The autocorrelation is approximately zero: no exploitable correlation structure between consecutive steps.

3.2. **The surrogate control.** Even with a perfect marginal and zero autocorrelation, the sequence may carry higher-order structure. On 297 million odd steps (5500 orbits, starting points to 10^{130}) we run a surrogate control: keep the marginal distribution of a, shuffle the sequence order globally, destroying all serial structure.

window W	real vs shuffled (standardised diff z)	reading
$W \le 30$	$z = -11$	significant difference
$W = 150$	$z = +0.6$	no difference
$W \ge 200$	$z \to 0$	no difference

The difference is located in non-stationarity, not serial structure: the small-value stage ($\log_2 m < 10$) deviates (measured $\bar{a} \approx 2.36$ vs 1.90–2.00), while the large-value stage converges exactly to 2.00; shuffling destroys this stage. Notably the direction: non-stationarity

makes local growth *easier* ($z < 0$), not harder — there is no hidden "anti-growth" mechanism.

3.3. **Cross-scale stability and the safety margin.** Across nine magnitudes from 10^6 to 10^{15} (1500 orbits per layer), the measured $\bar{a}$ is

$$1.9997,\ 1.9904,\ 1.9947,\ 1.9953,\ 1.9932,\ 1.9992,\ 1.9948,\ 2.0041,\ 1.9977$$

with fluctuation ± 0.007, pure statistical noise: no drift toward the critical value as n grows. The pooled value is $\bar{a} = 1.996929$ against $E[a] = 2$ and the divergence threshold $\log_2 3 = 1.584963$. The safety margin is 0.412: $\bar{a}$ would have to *drop* by 20.6% for divergence. Equivalently: the budgeted quantity $\mathrm{freq}(a \equiv \sum_k c_k \bmod 2^k)$ measures 0.9969 against the required $0.5850 = \log_2 3 - 1$. Finally, the "most dangerous orbit" test: among 6000 orbits the most extreme has $\bar{a} = 1.738$ ($\sigma = 0.153$ marginal), but with $\sigma/\sqrt{313} = 0.082$ the observed deviation $z = 3.2$ matches the extremal expectation $\Phi^{-1}(1 - 1/6000) = 3.4$ — there are no anomalous orbits; observed extremes are fully explained by extreme-value statistics.

4. Extreme-window scaling and the critical window

4.1. **Scaling law of minimal window means.** Even with $\bar{a}$ globally at 2, short windows can show "local growth segments" with $\bar{a} < \log_2 3$; this decides whether numerical search can stumble on a counterexample. For windows of length W, the minimum over all N_w windows satisfies

$$2 - \min_W \bar{a} \;\approx\; 0.72\,\sigma\sqrt{\frac{2\ln N_w}{W}}. \tag{4.1}$$

Measured ratios 0.64–0.73, stable from $W = 30$ to $W = 400$. The coefficient is below the Gaussian 1 because the geometric distribution is right-skewed: the left tail is thinner than a symmetric one, so minima deviate less than the Gaussian prediction — a shape effect, not serial structure (independently confirmed by the surrogate control).

4.2. **The critical window and the required search.** Solving (4.1) for the largest self-sustaining window with $2 - \min_W \bar{a} = 2 - \log_2 3 = 0.415$ gives

$$W^* \approx 12 \ln N.$$

On a sample of 3×10^7 steps the formula gives $W^* = 12\ln(3 \times 10^6) \approx 179$: measured — windows of length ≤ 150 self-sustain $\bar{a} < 1.585$; at $W = 200$ they do not (minimum reaches only 1.615). Theory and measurement agree. This scaling law explains quantitatively why numerical verification never sees a counterexample: sustaining a local-growth window of length W requires search of size $\exp(W/12)$ — $W = 1000$ needs e^{83} steps, and a divergent orbit requires $W = \infty$. These conclusions are descriptive: they prove nothing and exclude nothing; they show only that on the 2-adic axis, no strategy based on "finding local growth segments" can succeed.

5. Collapse and diagnosis of the local-entropy framework

5.1. **The collapse identity.** A family of frameworks defines a "local entropy" for Collatz orbits: let $h_p(n)$ be the fraction of odd steps with $p \mid (3n+1)$, and study the weighted functional $\sum_p h_p(n) \cdot \frac{\log p}{p(p-1)}$. If m is an odd value on the orbit, $3m + 1 = 2^a m'$ with m' the next odd value, so for every odd prime p,

$$p \mid (3m+1) \iff p \mid m'. \tag{5.1}$$

That is, h_p does not measure "prime factors of $3n+1$" at all; it is exactly the frequency with which odd values on the orbit are divisible by p. The whole functional collapses to a Cesàro

average of a nonnegative function:

$$\bar{\mathcal{H}}(n_0, K) - C_0 = \frac{1}{K}\sum_{i=2}^{K+1} g(m_i), \qquad g(m) := \sum_{\substack{p\geq 5\\ p\mid m}} \frac{\log p}{p(p-1)}. \tag{5.2}$$

Since odd values on the orbit are never divisible by 3 ($3m+1 \equiv 1 \pmod 3$) and $g(m) = 0$ iff $m = 1$. Moreover

$$\sum_{p} \frac{\log p}{p(p-1)} = 0.755368,$$

the same constant as $C(1) = \mathbb{E}[\log \frac{n}{\mathrm{rad}(n)}]$ in power-kernel fluctuation theory — two independent research routes collide with the same Euler product.

A conditional-law warning. Equation (5.2) is an unconditional law. In the Collatz situation m is always odd, so $3m+1$ is always even, i.e. $v_2 \geq 1$ holds conditioned. The unconditional law gives $\mathbb{E}[(v_2-1)_+]\log 2 = \frac{\log 2}{2} = 0.3466$, while the conditional law gives $\log 2 = 0.6931$ — a factor of exactly 2. For odd p, by CRT, "being odd" is independent of the class mod p, so the law is unchanged. Hence if the framework's threshold is calibrated by the unconditional law, the threshold is globally misplaced: the true budget is $C(1) + \frac{\log 2}{2} = 1.1019$ vs 0.7554 — and the error sits entirely in the dominant prime $p = 2$ (true budget share 63% vs 46%).

5.2. **Three free consequences.** The collapse identity (5.2) immediately gives three results, one correcting an invalid argument in the original framework. (1) *Nontrivial cycles*: if a nontrivial cycle has odd values $m_1, \ldots, m_k$, each $m_i \neq 1$ is odd and not divisible by 3, so it has a prime factor $p \geq 5$, hence $g(m_i) > 0$. Thus $\bar{\mathcal{H}} = C_0 + \frac{1}{k}\sum_i g(m_i) > C_0$; the excess is termwise positive, so the Diophantine coincidence the original framework worried about is structurally excluded. (2) *Exact characterisation of S_0*: preimages of 1 are $S_0 = \{2^a \cdot \frac{4^j-1}{3} : a \geq 0,\ j \geq 1\}$; independent check: 37 elements with $n \leq 2500$ (12+9+7+5+3+1), matching the original enumeration digit by digit, including the order of the first 25. (3) *Proposition 4 proved rigorously*: the original framework treated the "finitely many prime factors" case by Siegel's theorem, which is inapplicable. The correct proof uses S-unit equations: if only finitely many primes S appear, all m_i are S-units, and each step gives $2^{a_i} m_{i+1} - 3m_i = 1$, i.e. $X - Y = 1$ with X, Y $\overline{S}$-units ($\overline{S} = S \cup \{2, 3\}$); by finiteness of S-unit equation solutions there are only finitely many transitions, only finitely many values m_i, and the orbit is eventually periodic, contradicting divergence. Quantitative version: divergent orbits have pairwise distinct odd values, so $K \leq C^{s_K}$ with s_K the number of distinct primes among the first K steps, i.e. $s_K \gtrsim \log K / \log\log K$. But this yields only $\sim \log K$ "new primes" with frequency $\sim \log K / K \to 0$; it does not exclude "bad-step frequency tending to zero while the orbit still diverges".

5.3. **Three hard injuries and closure of the route.** *Injury 1 (numerical).* The framework's threshold constant was wrong: the correct value is 0.755368, not 0.773156 (a 2.4% bias — small, but it sits in the threshold and throws off every "equality only at the trivial cycle" judgement. *Injury 2 (vacuous verification).* For every orbit converging to 1, "bad steps" ($g(m_i) > 0$) are finite in number, so $\bar{\mathcal{H}} = C_0$ holds automatically as $K \to \infty$, regardless of the entropy threshold's content. For all $n \leq 2500$ (indeed 98.52% converging), $\bar{\mathcal{H}} = C_0$; the original report's "260 samples, excess positive, maximum 0.402" was a transient maximum, not a limsup. The numerical table verified nothing about the entropy threshold — it verified the structure of S_0 (which passes independent re-check). *Injury 3 (route closure).* The framework's endgame needs an upper-bound theorem P4: $\forall n,\ \bar{\mathcal{H}} \leq C_0$. But: if the Collatz conjecture is true, no divergent orbit exists and P4 holds vacuously, contributing nothing; if a divergent orbit exists, P4 requires $\mathrm{avg}\, g(m_i) \to 0$, i.e. the orbit actively avoids all small prime factors — while the random model predicts $E[g] \approx 0.028 > 0$. So P4 directly conflicts with "divergent orbits behave randomly": no equidistribution or soft-analysis tool can prove P4, since equidistribution predicts it false. P4 is, in substance, equivalent to the

Collatz conjecture itself. Hence the framework's "squeeze by upper and lower bounds" strategy is infeasible. This agrees structurally with Section 4: the entropy functional measures the odd part m'; growth is controlled by the 2-adic part a; they are orthogonal faces of the same number. In particular when $a = 1$ growth is forced while m' can be a large prime ($g(m') \approx \log m'/m' \to 0$, invisible to entropy).

6. Change of axis: the inverse tree and density propagation

6.1. **Why leave the p-adic axis.** The conclusion of Sections 3–4 is: no usable structure remains on the 2-adic axis. Changing to another prime does not work, for reasons that can be stated once and for all. $p = 3$ is dead: $3m + 1 \equiv 1 \pmod 3$ always (this is the $h_3 \equiv 0$ in the entropy framework); no structure exists on the 3-adic axis. $p \geq 5$ is equally dead: $m \mapsto (3m + 1)/2^a$ is an affine transformation on $\mathbb{Z}_p$, and its iteration has no dynamical content. The only nontrivial p-adic axis is the 2-adic one, which is exhausted: a follows the i.i.d. geometric law, the correction term is exactly absorbed by the measured mean, and the surrogate control shows no long-range difference. The real change of axis is to leave the p-adics for the *counting axis*: the inverse tree, generating all reachable odd numbers backwards from 1, studied as a deterministic branching structure without any randomness assumption.

6.2. **The inverse formula and the legality condition.** One reverse step: $m = \frac{2^a t-1}{3}$ requires $2^a t \equiv 1 \pmod 3$. Since $2^a \equiv (-1)^a \pmod 3$:

$$t \equiv 1\ (3) \Rightarrow a \in \{2, 4, 6, \dots\}; \qquad t \equiv 2\ (3) \Rightarrow a \in \{1, 3, 5, \dots\}.$$

Another basic fact is uniqueness of preimages: every odd m is determined by its unique successor and $a = v_2(3m+1)$, so the inverse tree has no repeated nodes (crucial for all later counting; the early "multiplicity" report was a statistical-calibre error).

6.3. **The three roles of residue classes.** Modulo 3, odd numbers fall into three classes with completely different roles:

type	legal a	$\Lambda_r(1)$	role
$t \equiv 1$	even $2, 4, 6, \dots$	1	critical (contains $t = 1$)
$t \equiv 2$	odd $1, 3, 5, \dots$	2	drives growth
$t \equiv 0$	none	—	leaf

The class $t \equiv 0$ has no preimages since $3m + 1 \equiv 1 \not\equiv 0 \pmod 3$; they are the leaves of the inverse tree. Direct counting at $X = 10^6$: $M_1 = 166667$ and $M_2 = 333333$ preimage totals, with $M_1 + M_2 = X/2 = 500000$ — a mere arithmetic consequence of the three classes having equal population, not an independent consistency check. Emphasis: M_1, M_2 are preimage *totals* (extensible mass flow), not populations: the classes $t \equiv 1$ and $t \equiv 2$ each occupy one third of all odd numbers (≈ 166667 each); the ratio $M_1/M_2 = 1 : 2$ reflects the branch-mass ratio $\Lambda_1 : \Lambda_2 = 1 : 2$ (last column of the table).

7. Branch-mass closed forms, rank-one structure, and criticality

7.1. **The branch-mass closed form and its verification.** Define the β-mass (partition function) $Z_T(\beta) = \sum_{t \in T} t^{-\beta}$. From $m \approx \frac{2^a}{3} t$, the one-step mass ratio is

$$\Lambda_r(\beta) = 3^\beta \sum_{\text{legal } a} 2^{-a\beta}.$$

Substituting the legality conditions ($r = 1$: $a = 2k$, $k \geq 1$; $r = 2$: $a = 2k + 1$, $k \geq 0$):

$$\Lambda_1(\beta) = \frac{(3/4)^\beta}{1 - 4^{-\beta}}, \qquad \Lambda_2(\beta) = \frac{(3/2)^\beta}{1 - 4^{-\beta}}. \tag{7.1}$$

Verification has two independent levels. First, the closed form (rational evaluation) and direct summation (numerical geometric series) are two different computational paths. Second —

and stronger — the sum rule, an algebraically provable identity linking branch masses to the rank-one spectral radius:

$$\Lambda_1(\beta) + \Lambda_2(\beta) = \frac{(3/4)^\beta + (3/2)^\beta}{1 - 4^{-\beta}} = \frac{3^\beta}{2^\beta - 1} = 3\rho(\beta). \tag{7.2}$$

β	Λ_1 closed	Λ_1 sum	Λ_2 closed	Λ_2 sum	$\Lambda_1 + \Lambda_2$	$3\rho(\beta)$
0.5	1.732051	1.732051	2.449490	2.449490	4.181541	4.181541
0.7	1.316442	1.316442	2.138566	2.138566	3.455008	3.455008
0.84	1.141603	1.141603	2.043527	2.043527	3.185130	3.185130
1.0	1.000000	1.000000	2.000000	2.000000	3.000000	3.000000
1.2	0.873578	0.873578	2.006954	2.006954	2.880532	2.880532
1.5	0.742307	0.742307	2.099563	2.099563	2.841870	2.841870

Two caveats. First, the table compares idealised (infinite-sum, no spatial truncation) quantities; finite measured values are systematically slightly larger, with deviation $O(1/\ln X)$ (Section 8). Second, the first version of this table was wrong: its "theory" and "measured" columns agreed digit by digit on all six rows — proving the two columns came from one array, so the "verification" was a comparison with itself; the value also violated the sum rule (7.2) (at $\beta = 0.5$: $1.366025 + 2.593863 = 3.959888$ vs $3\rho(0.5) = 4.181541$). The table has been recomputed from the closed form with the sum rule as a hard generation-time check (this is ledger entry no. 9, Section 12.3).

7.2. **Complete uniformity of the offspring distribution.** Knowing each branch's total mass is not enough — one needs the distribution of mass across the three residue classes to write the propagation matrix. The answer is complete uniformity. Consider the mod refinement: $2^a \bmod 9$ has period 6, and in each subclass of $t \bmod 9$ the three classes each occupy 1/3; enumerating the ($a \bmod 6$, $t \bmod 9$) combinations gives, for every (r, a),

$$\mathbb{P}(\text{offspring} \equiv s \mid r, a) = \tfrac{1}{3}, \qquad s = 0, 1, 2.$$

Program verification: the offspring distributions of $(r = 1, a = 2)$ and $(r = 2, a = 1)$ are exactly $\{0.3333, 0.3333, 0.3333\}$. Hence the extension matrix is rank one:

$$B(\beta) = \begin{pmatrix} \Lambda_1/3 & \Lambda_1/3 \\ \Lambda_2/3 & \Lambda_2/3 \end{pmatrix}.$$

7.3. **Spectral radius and critical exponent.** The rank-one matrix has spectral radius equal to its trace:

$$\rho(\beta) = \frac{\Lambda_1(\beta) + \Lambda_2(\beta)}{3} = \frac{3^{\beta-1}}{2^\beta - 1}. \tag{7.3}$$

At $\beta = 1$: $B(1) = \left(\begin{smallmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{smallmatrix}\right)$, $\rho(1) = 1.00000$. Bisection gives the lower critical $\beta^* = 1.00000000$. Note $\rho(\beta)$ is not monotone: $\rho = 1$ also at $\beta = 2$; but that second root is a lattice artefact of small t as $\beta \to \infty$ — the truncated version of Section 9 is monotone and has no second root. The physical critical point is the lower one, $\bar{\beta}^* = 1$.

7.4. **How supercriticality is exactly cancelled.** Here is a seemingly paradoxical phenomenon: $\Lambda_2(1) = 2$ means the $t \equiv 2$ branch produces twice its mass per step — strictly supercritical — yet the total spectral radius is exactly 1, neither exploding nor decaying. The mechanism is completely explicit. Of the mass produced by $\Lambda_2 = 2$, one third falls into the $t \equiv 0$ class — leaves, never to extend again. The extensible remainder is $2 \times \frac{2}{3} = \frac{4}{3}$, redistributed by uniformity across the two extensible types. The next generation's extensible total mass is

$$\frac{2}{3}\Lambda_1 + \frac{4}{3}\Lambda_2 \xrightarrow{\beta=1} \tfrac{2}{3}\cdot 1 + \tfrac{4}{3}\cdot 2 = \tfrac{10}{3}, \qquad \text{ratio preserved: } (\Lambda_1 : \Lambda_2) = (1:2) \leftrightarrow M_1 : M_2 = 1:2,$$

fully consistent with the direct counting of Section 6. This gives the mechanistic answer to the opening question of Section 6: density propagation stops at $\beta = 1$ because for $\beta < 1$,

$\rho > 1$ (propagation sustains itself, hence provable), while at $\beta = 1$, $\rho = 1$ exactly at the boundary (the remainder term never clears, hence unprovable). What was previously a numerical observation is now a corollary of the closed form (7.3).

8. Finite size: a quantitative anatomy of the lesson

8.1. **The true order of the deviation.** The closed form of Section 7 is idealised (infinitely many steps, no truncation). Every finite measurement deviates from it, and the *order* of that deviation is itself the object of measurement — misjudging it turns finite-size effects into structure. Re-measured by exhaustive computation (exact, enumerating to $a = 40$, covering all residue classes):

X	$\Lambda_1(1)$ measured	$\rho(1)$ measured	$\rho(1) - 1$	$(\rho(1) - 1) \ln X$
3×10^5	1.0947	1.020877	0.020877	0.263
10^6	1.0885	1.019227	0.019227	0.265
2×10^6	1.0853	1.018388	0.018388	0.267

The last column is constant $(0.263, 0.265, 0.267)$: the deviation is $O(1/\ln X)$, not $O(1/X)$. The reason: the correction $\sum\big(\frac{1}{m} - \frac{3}{2^a t}\big)$ itself converges ($O(1)$ in Z), while $Z \sim \ln X$, so the relative deviation decays as $1/\ln X$. This has a practical consequence: convergence is extremely slow. At $X = 2 \times 10^6$ the deviation is still 1.8%; pushing it below 1% needs about $e^2 \approx$ a factor of 7 in scale. Likewise the numerical stationary composition has not converged to the analytic value $[1/3, 2/3]$ — same $O(1/\ln X)$.

8.2. **The methodological ledger.** Five errors of one type occurred in this study: all take a "stable-looking number" for a structural conclusion when it is an artefact of the measurement. They are listed because they are more transferable than any single result.

#	error	true cause	countermeasure
1	fit artefact: reachable-set exponent 0.833	depth truncation + log-fit	measure density at fixed X
2	statistical calibre: density = 1	generated + covered node counts	dedup + conservation check
3	budget truncation: low density, no saturation	BFS node cap L too small	untruncated reverse closure + bidirectional
4	finite-size misjudgement: $O(1/X)$	actually $O(1/\ln X)$	exhaustive computation + scaling test
5	definition mixing: $\beta_{\text{cut}} \approx 0.975$	per-t ratio $\neq Z_P/Z_T$	define first, then measure

The common form deserves abstraction: on a slowly varying or extremely slowly converging quantity, any finite-interval measurement returns a highly stable-looking number with a beautiful goodness of fit. This is the same trap as the "slowly varying functions manufacture stable false constants" phenomenon in power-kernel fluctuation studies. The only reliable criteria are cross-validation by genuinely independent paths, and an explicitly stated convergence law (such as the $1/\ln X$ of the table above).

9. Truncated propagation: closed form, error series, rootlessness

9.1. **Piecewise-constant structure.** The Λ of Section 7 is the "total mass" version: all preimages of t regardless of size. Density propagation really concerns *truncated* mass — preimages still within X. The truncation condition is $m = \frac{2^a t - 1}{3} \le X$, i.e. $a \le A(t)$ with

$$A(t) = \left\lfloor \log_2 \frac{3X + 1}{t} \right\rfloor.$$

$A(t)$ is constant on $I_A = (\frac{3X+1}{2^{A+1}}, \frac{3X+1}{2^A}]$, so the truncated branch mass is piecewise constant:

$$\rho_{\text{cut}}(\beta) = \frac{1}{Z_T} \sum_{A \ge 1} \Big[M_1(I_A)\, \Lambda_1^{\text{cut}}(A) + M_2(I_A)\, \Lambda_2^{\text{cut}}(A) \Big], \tag{9.1}$$

$$\Lambda_1^{\text{cut}} = 3^\beta \frac{4^{-\beta}\big(1-4^{-\beta\lfloor A/2\rfloor}\big)}{1-4^{-\beta}}, \qquad \Lambda_2^{\text{cut}} = 3^\beta \frac{2^{-\beta}\big(1-4^{-\beta\lfloor (A+1)/2\rfloor}\big)}{1-4^{-\beta}},$$

where $M_r(I_A)$ is the β-mass of class-r targets in I_A. Two structural facts. First, the hard boundary: when $A < 2$ (i.e. $t > \frac{4}{3X}$... precisely t in the top interval), $\Lambda_1^{\text{cut}} = 0$, since that class needs even $a \geq 2$, of which there are none in the interval. Second, the floor identity: writing $A = k+f$ $(0 \leq f < 1)$, $\lfloor A/2\rfloor = \lfloor k/2\rfloor$ holds identically, so continuous approximation and flooring are exactly equivalent — an identity, not a numerical coincidence.

9.2. **Asymptotic series of the error.** The closed form (9.1) derives from $m \approx \frac{2^a t}{3}$ while the exact value is $m = \frac{2^a t - 1}{3}$; the difference expands systematically:

$$m^{-\beta} = \Big(\frac{2^a t}{3}\Big)^{-\beta}\Big(1-\frac{1}{2^a t}\Big)^{-\beta} \approx \Big(\frac{2^a t}{3}\Big)^{-\beta}\Big(1+\frac{\beta}{2^a t}\Big).$$

β	exact	closed form (order 0)	closed form (order 1)	error: order 0 → order 1
0.5	0.908208	0.907964	0.908179	−0.03% → −0.00%
0.84	0.965561	0.954099	0.963466	−1.19% → −0.22%
1.0	0.986561	0.943788	0.977637	−4.34% → −0.90%

The error is an asymptotic series expandable systematically: one order of correction compresses the error from 4.34% to 0.90%. The structural correctness of (9.1) is confirmed; the error source is fully understood.

9.3. **Rootlessness and a correction.** Scanning β over the whole parameter range, ρ_{cut} behaves as

$$\rho_{\text{cut}}: \ 0.908\ (\beta{=}0.5) \to 0.9866\ (1.0) \to 0.99993\ (1.5) \to 1.000\ (2.0^+).$$

It approaches 1 monotonically from below, never exceeds 1, and has no finite root. The physical meaning: truncated one-step propagation within X always contracts. This does not contradict "the reachable set saturates to full density" — saturation is achieved by allowing intermediate nodes to exceed X (untruncated reverse closure), which truncated propagation forbids. *Correction (ledger entry 5).* This study once reported "$\rho_{\text{cut}}(1) - 1 = +0.0053$, $\beta^*_{\text{cut}} \approx 0.975$". That pair is void: it came from mixing the "per-t truncated-mass ratio" with the "Z_P/Z_T" definitions. The true value is $\rho_{\text{cut}}(1) \approx 0.987 < 1$. Consequently the conjecture "truncated propagation has a finite critical root" is falsified. The resulting negative but definite conclusion: closed forms (total or truncated) cannot give the Krasikov–Lagarias exponent. The failure is not inaccuracy — Table above shows they are quite accurate — but object selection: $\rho(\beta)$ measures "one-step contraction within the full set", while 0.84 is the dimension of a finite-depth reachable set from 1 allowing intermediate overflow. They are not the same propagation process.

10. The depth spectrum: from fractal dimension to boundary-layer slope

10.1. **Falsification of the polylog hypothesis.** Section 9's negative conclusion suggests changing the object: instead of "one-step propagation within the full set", study the reachable set from 1 of depth at most k. Write $N_k(X) = \#\{t \leq X,\ t \text{ odd} : D(t) \leq k\}$ with $D(t)$ the number of odd steps from t to 1. Two competing hypotheses must be named and one rejected. The power law: $N_k(X) \sim C_k X^{\gamma_k}$, i.e. S_k is a genuine fractal of dimension $\gamma_k < 1$. The polylog hypothesis: $N_k(X) \sim c_k(\log X)^k$, since reverse steps have $m \approx \frac{2}{3}t$ and $\sum a_i \leq \log_2(3^k X)$, giving $\binom{Y}{k} \approx (\log X)^k/k!$ combinations. The author initially expected the latter; the data falsify it. Decisive test ($k = 24$, X from 10^3 to 10^6, a factor of 1000):

hypothesis	predicted growth	measured growth	verdict
polylog $(\log X)^{24}$	$2^{24} \approx 1.7 \times 10^7$	294	falsified (5 orders)
power $X^{0.824}$	$1000^{0.824} = 294$	294	digit-perfect

A second independent test also supports the power law: the density $N_k(X)/X$ drops from 0.317 to 0.0933 (a factor 3.4); the power law predicts $X^{-0.176}$ (factor 3.24) — agreement; the polylog prediction differs from measurement by five orders of magnitude.

10.2. **Measurement of the dimension spectrum.**

k	$N_k(10^3)$	$N_k(10^4)$	$N_k(10^5)$	$N_k(10^6)$	γ_k	power R^2	polylog b/R^2
1	5	7	9	10	0.101	0.953	1.02/0.985
5	56	141	297	580	0.337	0.994	3.36/0.999
12	195	893	3627	12540	0.603	0.998	6.00/0.997
24	317	2330	15755	93340	0.824	0.999	8.17/0.994
40	391	3319	28908	227325	0.923	1.000	9.14/0.991

The measurements themselves are reliable, but their interpretation needs correction (Section 10.4). Measured conclusion: γ_k increases monotonically to 1; $k = 1$ is the boundary case whose true growth is $\frac{1}{4}\log X$ (the polylog exponent $b = 1.02 \approx 1$ confirms this), looking like a small power law $\gamma = 0.101$ over a finite range — the only point where the polylog fit wins, and exactly the point where theory predicts it. A rescaling law was voided: fitting $1 - \gamma_k \sim c/k^\alpha$ gave $\alpha \approx 1.2$–1.3, but with full coverage the fit fails ($\log(1-\gamma_k)$ vs $\log k$ is not a line and drifts with the window): within finite X, $N_k(X)$ saturates as k grows toward $X/2$ (slope tends to 1) — a finite-size effect. No $1 - \gamma_k \sim c/k^\alpha$ is reported.

10.3. **Relation to the Krasikov–Lagarias exponent.** The Krasikov–Lagarias unconditional bound gives "at least $cX^{0.84}$ reachable odd numbers below X". Interestingly, 0.84 falls on our spectrum at $k \approx 25$: $\gamma_{24} = 0.824$, $\gamma_{28} = 0.874$, interpolating to 0.84 at $k \approx 25$. But this must be stated plainly: for now it is a numerical coincidence, not an established correspondence. The reason is essential — KL's theorem is a lower bound on the total reachable count (all reachable odd numbers, unrestricted depth), while γ_k is the dimension of a fixed-depth set. To establish a correspondence one must prove "the depth to which KL's method actually propagates reliably" corresponds to a specific k; this remains open (and is closed negatively in Section 16.5).

10.4. **Reframing: boundary-layer slope, not fractal dimension.** This section corrects the conclusion of Section 10.2 and withdraws the statement "S_k is a genuine fractal". The correction comes from cross-checking an independent technical record (boundary-layer analysis of the stopping-time distribution); all numbers here were independently re-verified. The stopping-time distribution has the scaling form

$$N_k(X) = \frac{X}{2} \cdot F\Big(\frac{k - \mu(X)}{\sigma(X)}\Big), \qquad \mu(X) = 3.671 \ln X - 8.14, \quad \sigma(X) = 5.06\sqrt{\ln X} + 4.4.$$

Independent re-verification (exact histograms to $X = 10^7$): model vs measured for σ (the $(k_{.84} - k_{.16})/2$ statistic) at $X = 10^3, 10^5, 10^6$: 17.70/18.00, 21.57/21.00, 23.21/23.00 — agreement to 1.5%; the model systematically sits below the measured $\mathbb{E}[D]$ (51.03 vs 53.94), the signature of right skew — μ is a centre, not the mean. In this representation the meaning of γ_k changes completely: it is no longer a dimension but the local logarithmic slope as the boundary-layer window sweeps the fit interval. Three previously isolated observables are unified: (i) $\gamma_k < 1$ increasing to 1: local slope of the crossover window sweeping the fit range (fractal reading withdrawn); (ii) failure of the $1 - \gamma_k \sim c/k^\alpha$ fit: the object is not that kind of power law, so the fit must fail; (iii) γ_k depends on the fit window: a direct consequence of boundary-layer position. Row (iii) is the decisive test: with the same exact data, only the regression set changed:

k	four points (10^3–10^6)	five points (10^3–10^7)	record's correction (incl. 10^8)
12	0.6033	0.5728	0.548
24	0.8237	0.7957	0.772
40	0.9233	0.9095	0.897

The four-point values match Table of Section 10.2 digit by digit (0.603/0.824/0.923), so the measurement is sound; but each extension of the interval lowers γ_k systematically, with larger drops for smaller k. This is incompatible with "dimension" and fully compatible with "boundary-layer local slope". The polylog falsification is unaffected and strengthens out of sample: at $k = 40$, X from 10^6 to 10^7: polylog predicts growth $(\ln 10^7/\ln 10^6)^{40} = 7.64$, measured $616674/227325 = 2.71$; the falsification gap widens. Hence: S_k is withdrawn as "a genuine fractal of dimension γ_k"; γ_k reads as a boundary-layer local slope, a window-dependent effective exponent. The values of the table are kept (they are correct measurements within that window), but the fractal interpretation below them is void. The lesson is the same as the ledger's: on quantities varying extremely slowly with scale, finite-interval fits return stable numbers with beautiful R^2, masquerading as structure. The only test that could falsify the fractal reading was changing the observation window — and that test had not been done.

11. Structural limits of the spectral method

This chapter gives the study's most important negative conclusion; its root is not computational precision but the operator's choice of state space. The only closed form is the type-space propagation matrix $B(\beta)$ with spectral radius $\rho(\beta) = \frac{\Lambda_1+\Lambda_2}{3} = \frac{3^{\beta-1}}{2^\beta-1}$, $\rho(1) = 1$. By the standard thermodynamic-formalism correspondence, the pressure function $P(\beta) = \log \rho(\beta)$ has its zero giving the dimension of the reachable set; here it gives

$$\gamma_\infty = \inf\{\beta : \rho(\beta) < 1\} = 1,$$

independent of k. This closed form can only produce the asymptotic ceiling, never any γ_k, still less 0.84. The root is clear: $B(\beta)$ encodes only mass transfer between mod-types; in its construction the spatial variable of t has been completely integrated out — all that remains is "how much mass of which type", and "where within X that mass sits" is discarded. The latter is exactly what decides γ_k: the dimension of S_k is a property of its spatial filling, not of inter-type mass ratios. Hence the structural criterion:

Spectral radius on type space $\implies$ only the asymptotic ceiling, never any finite-depth γ_k.

To obtain γ_k one must pass to the Ruelle operator on the spatial variable $t \in (0, \infty)$:

$$(\mathcal{L}_\beta f)(t) = \sum_a \Big(\frac{2^a}{3}\Big)^{-\beta} f\Big(\frac{2^a t - 1}{3}\Big),$$

an infinite-branch operator (branches enumerated by a, weights $(2/3)^{a-\beta}$); the convergence threshold of the k-step partition function $L_\beta^k \mathbf{1}$ is γ_k. This route is not a continuation of Section 7 but a new investment with a changed state space. Section 13 executes the change: the correct skew-product form is given, and the result of the shootout against exact BFS is negative — the continuous-space Ruelle operator cannot give γ_k, and its failure mode is decidable by a fit-free hard criterion. Moreover, even if it worked, its correspondence with the Krasikov–Lagarias exponent would remain unestablished — unchanged since Section 10.3.

12. Conclusion, falsification list, and methodological ledger

12.1. **Established results.** *Axis 1 (2-adic measurement)*: identity (2.1) and criterion (2.2) hold strictly; $\bar{a} = 1.996929$, safety margin 0.412, no drift across nine magnitudes; a follows the exact geometric law, autocorrelation approximately zero; the surrogate control confirms no long-range difference, with the difference direction "easier"; the extreme-window scaling law (4.1) and the critical window $W^* \approx 12 \ln N$ explain why numerical verification never meets a counterexample. *Axis 2 (inverse tree and closed forms)*: collapse identities (5.1)–(5.2); exact characterisation of S_0 (37 elements below 2500); the rigorous S-unit proof of Proposition 4 with quantitative version $s_K \gtrsim \log K/\log\log K$; branch-mass closed form (7.1) with rank-one structure; spectral radius (7.3) and criticality at $\beta^* = 1$; the exact cancellation of supercriticality by the one-third leaf tax; finite-size deviation $O(1/\ln X)$.

Axis 3 (depth spectrum): polylog hypothesis falsified decisively; truncated-propagation closed form (9.1) with asymptotic error series; truncated version always contracts, no finite critical root; depth spectrum $\gamma_{12} = 0.603$, $\gamma_{24} = 0.824$, $\gamma_{40} = 0.923$ (four points, 10^3–10^6) kept as measurements within that window, but the "fractal dimension" interpretation withdrawn (Section 10.4) and reread as boundary-layer slope. *Axis 3 continued (Section 14)*: back on single-orbit statistics, the few constants of this study derived from structure rather than fit: the exact geometric law under the uniform measure ($E = 2.00000$, $Var = 2.00001$); the far-tail rate closed form $\beta_c = 0.054979$ and its distinguishable difference from the Brownian approximation; the measured transfer matrix of the modular conditional law ($a_i = 5 \Rightarrow a_{i+1} = 4$ at 4.4–4.9× the independent prediction); and the directionally negative conclusion that the Markov rate lies below the iid rate. *Axis 3 continued (Section 13)*: the construction of the propagation operator itself rather than its type-space projection: the correct skew-product Ruelle operator and the structural reason the type coordinate cannot be erased (Table: the two minimal displacements differ by exactly $\log_2 3 = 1.584963$, i.e. 1.000000); exactness of the transfer law established; and the negative conclusion that continuousisation loses integrality, making operator counts exceed the hard bound — the route is unusable.

12.2. **Falsification list.**

Falsified statement	Disproof basis
Local-entropy upper-bound theorem P4 provable	conflicts with random model; in substance equivalent to the conjecture
Local-entropy numerical verification valid	convergent orbits make $\bar{\mathcal{H}} = C_0$ automatic; vacuous
Threshold constant = 0.773156	correct value 0.755368
Truncated propagation has finite critical root $\beta_{\text{cut}} \approx 0.975$	definition mixing; $\rho_{\text{cut}} < 1$ always, monotone to 1
$N_k(X)$ polylog growth	k=24: measured ×294 vs predicted 1.7×10^7
Small-value lingering gives escape chances	threshold elevation exactly absorbed by measured $\bar{a}$
$1 - \gamma_k \sim c/k^\alpha$	cannot fit with full coverage; finite-size effect
Autocorrelation $\rho(1) = +0.027$	uncentred second moment; true value ≈ 0
Spectral closed form gives KL exponent 0.84	type space lacks spatial variable; zero constantly 1
Reachable-set density = 1 / low density, no saturation	statistical-calibre and budget-truncation artefacts
$P(r' \mid r, a) = \frac{1}{3}$ only approximate	exact; and the 3-adic hierarchy rises forever, no fixed point
Continuous-u Ruelle operator gives γ_k	count 5398 > hard bound $X/2 = 5000$
Spectral-equation root gives γ	$s^* = 0.3486$ gives $\gamma = 0.503$, contradicting $\gamma \to 1$
Coupled operator reproduces small k / k=22 = 205557	counting-convention error; not reproducible
S_k genuine fractal of dimension γ_k	γ_k drifts down with window (0.6033 → 0.5728 . . .); boundary-layer slope
γ_k insensitive to fit window	sensitive to X window; only coverage-insensitive; conflated before
β_c falsified (effective rate 0.065–0.075)	withdrawn: that value is $\beta_c + 1.5/n$, the $n^{-3/2}$ prefactor derivative
$X = 10^6$ regression gives $\beta \approx 0.12$	at that scale $\max D = 195$; fit hits the support boundary
Occupation-measure variance 15% below geometric	actually 9%–11% (1.7869/1.8145 vs 2)
Collatz near-critical (phase-transition machinery usable)	$\delta(3) = \ln \frac{3}{4}$; $q_c = 4$ excluded by parity; no critical point in family

Falsified statement	Disproof basis
β_c confirmed by measurement	second withdrawal: fixed n=100 slope drifts 0.1166 → 0.0827 → 0.0663; tangency only
iid (0.0567) vs Markov (0.0514) distinguishable	z flips sign with n; an unconvergence diagnosis, not a verdict
Nonstationarity by size affects far-tail rate	weighted 0.05643 vs stationary 0.05645; 0.04%, negligible
$\gamma_{24} \approx 0.84$ has a traceable KL correspondence	γ_{24} drifts 0.8237 → 0.7957 → 0.7404; no stable value; window coincidence
Median slope has a +5.6% overshoot constant	smoothed statistic bias tends to 0 (−2.70% → −1.33% → −0.25%); asymptotic slope $= 1/\ln(4/3)$
$n^{-3/2}$ prefactor measurable at $X \le 10^7$	fitted α rises 1.27 → 14.96 across windows; $\max D \approx (\ln X)^2$ boundary dominates
Survival profile from conditioned first-passage paths (meander/Bessel(3))	iid model gives near-uniform descent; measured "plateau+concentrated drop"; ends deviate 0.11, $z = -68.8, +51.2$
First-order transition dependence suffices for the a-sequence	Markov profile vs iid differs by < 0.005; survival 1.355×10^{-2} farther from measured 2.30×10^{-3}; wrong direction

12.3. **Methodological ledger and countermeasures.** Entries 1–5 are in Section 8; entry 6 (counting-convention error: the type distribution of one preimage mistaken for three independent child nodes) in Section 13; entry 7 (wrong comparison baseline: comparing measured slopes against a theory curve that itself varies with n, misreading the $1.5/n$ derivative of the $n^{-3/2}$ prefactor as "another exponent") in Section 14; entry 8 (reading "tangency agreement" as "asymptotic agreement": the measured slope crosses the theory curve at $X = 10^7$ and the good local fit was taken as convergence, until the fixed-n, vary-X comparison exposed the monotone drift 0.1166 → 0.0827 → 0.0663). Entry 9 (Section 16.3): the two columns of a "theory vs measured" table were not independent sources — the first Table 7.1 had its two columns identical digit by digit on all six rows, and the value violated the algebraically provable sum rule (7.2) (at $\beta = 0.5$ by 5.3%); dangerously, the error vanished exactly at $\beta = 0.84$ and $\beta = 1$, the two points most likely to be spot-checked. Countermeasure: any "theory vs measured" table must (a) document that the two columns follow genuinely different computational paths, and (b) attach an algebraically provable internal-consistency identity as a hard generation-time check that is printed. The sum rule (7.2) is such an identity.

The forms differ, showing the countermeasures must cover deepening layers: numerical reading → code semantics → the comparison structure itself → convergence of the measured quantity → independence of the verification's two columns. The reliable protections are only two: cross-validation by genuinely independent paths, and writing the complete asymptotic form (with prefactor) of the compared quantities before comparing; for implementations, the shootout against an exact benchmark. It is worth recording that this trap is not specific to this problem: in power-kernel fluctuation studies the same stable false constants ($\kappa' \approx 1.1$, $\omega \approx 0.66$, $\mathfrak{c} \approx 0.88$) appeared, with the same root cause identified. Two independent research routes arriving at the same methodological lesson is itself a transferable output. A subtler trap deserves naming: verification-set self-selection bias — when a proposition's verification set is itself selected by the proposition's conditions, all numerics can pass while the proposition remains unfalsifiable within the set; the counterexample may slip through the gap.

Finally: this paper does not advance the Collatz conjecture itself. The three axes' conclusions should be read as: certain tools are structurally incapable on this problem, and the

locations of incapacity can be pinned down. This is different from "unprovable", which the paper does not claim.

13. The spatial Ruelle operator: construction, shootout, negative conclusion

Section 11 showed that to get finite-depth dimensions, the operator must move from type space to the spatial variable. This section executes the move, gives the operator's correct form, and tests it against exact BFS. The conclusion is negative: the continuous-space Ruelle operator cannot give γ_k, and its failure mode is decidable by a fit-free hard criterion.

13.1. **The correct skew-product form.** The inverse branches are $\Psi_a(t) = \frac{2^a t-1}{3}$ with legality $2^a t \equiv 1 \pmod 3$. Since $2 \equiv -1 \pmod 3$, $(-1)^a r \equiv 1$ with $r = t \bmod 3$: $r = 1 \Rightarrow a$ even $2, 4, 6, \ldots$; $r = 2 \Rightarrow a$ odd $1, 3, 5, \ldots$. With the logarithmic scale $u = \log_2 t$ and $\Psi_a(t) \approx 2^a t/3$, a branch is a translation on the log scale: $u \mapsto u' = u + a - \log_2 3$. The spatial Ruelle operator is the skew product on (u, r):

$$(\mathcal{L}_\beta f)(u', r') = \sum_a \left|\Psi_a'\right|^{-\beta} f\big(u' - (a - \log_2 3), r\big) \cdot P(r' \mid r, a), \qquad \left|\Psi_a'\right| = \frac{2^a}{3}, \tag{13.1}$$

where $P(r' \mid r, a)$ is the type distribution of the new node; $r' = 0$ nodes are leaves (no preimages): counted but not continued.

13.2. **The type coordinate cannot be erased.** A natural simplification erases r and keeps only u. This is structurally impossible, because the legal minimal step depends on r, hence so does the minimal displacement:

$r = t \bmod 3$	legal a	$a_{\min}$	$\Delta u_{\min} = a_{\min} - \log_2 3$
1	even $2, 4, 6, \ldots$	2	$+0.415037$
2	odd $1, 3, 5, \ldots$	1	-0.584963

The two classes differ by exactly $\log_2 3 = 1.584963$ (1.000000). Erasing r amounts to allowing all $a \geq 1$, hence the contracting step $a = 1$ ($\Delta u = -0.585$) — exactly the mechanism by which nodes sink below $u \leq X$ and rise again, producing many offspring. Section 13.3 shows erasure is catastrophic. The transfer distribution is exact: from $p = \frac{2^a(r+3n)-1}{3} = \frac{2^a r-1}{3} + 2^a n$, as n runs over a complete residue system mod 3 so does $2^a n$ (as $3 \nmid 2^a$), hence

$$P(r' = s \mid r, a) = \tfrac{1}{3} \quad (s = 0, 1, 2), \quad \text{for every } (r, a).$$

Numerical re-verification ($N = 3{\times}10^5$, six groups $(r, a) = (1, 2), (1, 4), (1, 6), (2, 1), (2, 3), (2, 5)$): all give 0.333333. The previous round had treated "uniform" as a first-order approximation to be refined; re-verification shows it is the exact transfer law, with no upgradeable margin.

13.3. **The decisive experiment: shootout against exact BFS.** Take $X = 10^4$. Forward simulation fixes the parent cap $B = 9038$; reverse BFS gives the exact benchmark (141 orbits cumulative, $N_k(X)$ to depth k); the operator is realised on a grid of step $du = 0.05$ and compared level by level:

k	exact BFS	coupled (absorb $u < 0$)	ratio	coupled (no absorb)	ratio	erased r	ratio
1	7	8.0	1.14	8.0	1.14	14.0	2.0
3	48	65.1	1.36	67.0	1.40	370.7	7.7
10	648	1178.2	1.82	1386.7	2.14	425205	656.2
22	2136	5398.2	2.53	7323.1	3.43	2.70×10^9	1.26×10^6

The criterion uses no fit. Odd numbers below X number at most $X/2 = 5000$, a hard bound any subset must obey. The coupled operator gives 5398.2 at $k = 22$, exceeding the hard bound by 1.08×; the erased operator gives 2.70×10^9, exceeding by 5.4×10^5. Neither can be the count of any integer set.

$$\text{Operator count} > X/2 \implies \text{the operator's absolute count is meaningless.}$$

The error compounds with depth (ratios $1.14 \to 1.36 \to 1.82 \to 2.53$), not a calibratable constant factor. Absorbing boundaries only lower the $k = 22$ value from 7323 to 5398 (26%), insufficient to repair.

13.4. **Level ε: 3-adic refinement is a dead end.** Might the deviation come from transfer-law coarseness? Ruled out. (i) The reachable set is completely uniform on the 3-adics: among reachable odd numbers at $X = 10^4$ (2136 orbits, 2.72%), the frequencies of t mod 3 lie in $[0.3249, 0.3418]$ (uniform 0.3333), and of t mod 9 in $[0.1053, 0.1175]$ (uniform 0.1111), deviations ≤ 0.0084 and 0.0064. The 3-adic side carries no information. (ii) p mod 9 depends on t mod 27 (e.g. $t \equiv 1 \pmod 9$, $a = 2$: p mod 9 is $1, 4, 7$ for t mod $27 = 1, 10, 19$). Hence "transfer exact to mod 9" needs mod-27 states, the next level mod 81, and so on — the 3-adic hierarchy rises forever and cannot close in a finite state space. Together with Section 13.2: refining the 3-adics has no information to mine and no terminal to reach.

13.5. **The true source of the deviation: continuousisation loses integrality.** With the transfer law exact and the 3-adics information-free, the deviation can only come from continuousisation itself. In the real problem $u = \log_2 t$ takes only discrete values (those of odd integers), while a continuous grid allows any real value. Every real integer node then inflates into a blob of spurious mass, compounding with depth: continuous-u erases the sparsity of the integer lattice. This is not merely a loss of precision — it deletes the measured object: the dimension of S_k is equivalent to its sparsity among integers; once sparsity is flattened, the operator measures something other than γ_k. The hard-bound violation of Section 13.3 ($5398 > 5000$) is the direct self-testimony of this phenomenon. One correction must be recorded: an exploratory report claimed the coupled operator "reproduces small k exactly" with $k = 22$ equal to 205557; re-check shows these numbers are not reproducible — the original implementation mistook the type distribution of a single preimage (mass 1/3 each type) for three independent child nodes. All numbers in this chapter are freshly and independently measured; per the ledger this is entry no. 6, of the form "an implementation-level convention error masquerading as a computational conclusion"; the countermeasure remains the shootout against an exact benchmark.

13.6. **The spectral equation and its uncredibility.** With the growth pattern $f(u, r) \sim A_r e^{su}$ ($s = \gamma \ln 2$), substitution into the skew-product operator gives, in the continuous limit, the eigenvalue relation

$$\lambda(s) = \frac{e^{-0.415037s} + e^{0.584963s}}{6s},$$

with $A_1 = A_2$ symmetric, consistent with the rank-one structure of Section 7 — but here type space and u-space couple through the $\Delta u_{\min}$ difference and are no longer separable. Solving $\lambda(s) = 1$ gives the small root

$$s^* = 0.348581, \qquad \gamma = \frac{s^*}{\ln 2} = 0.5029,$$

in clear conflict with the measured $\gamma_\infty \to 1$. Since Section 13.5 has already judged continuousisation invalid, this equation should not be credited; it is recorded only to document the derivation chain. *Reading note.* The denominator contains s, i.e. $\lambda(s) = (e^{-0.415037s} + e^{0.584963s})/(6s)$, and the solve target is $\lambda(s) = 1$, giving $s^* = 0.348581$. Misreading the denominator $6s$ as plain 6 changes the root of $\lambda(s) = 1$ to $s \approx 2.978$; and since $\lambda_{/6}(s) := (e^{-0.415037s} + e^{0.584963s})/6$ satisfies $\lambda_{/6}(s^*) = s^*$... precisely: $\lambda_{/(6s)}(s^*) = 1$ while $\lambda_{/6}(s^*) = s^*$, so one can be misled into the impression that "the solve target was mistakenly written as $\lambda(s) = s$". This reading divergence actually occurred once in an external check; it is recorded here to prevent recurrence.

13.7. **Conclusion.** Three obstacles close this route, each independently:

Obstacle	Content	Criterion
Continuousisation loses integrality	u continuous: every integer node inflates into a mass blob	operator count 5398 $>$ hard bound $X/2 = 5000$
3-adics information-free	transfer law exactly uniform; reachable set 3-adically uniform	mod-9 deviation ≤ 0.0064; hierarchy unbounded
discrete operator degenerates to enumeration	keeping integrality forces action on exact odd numbers	equivalent to exact BFS; no longer a theory

Hence: the continuous-space Ruelle operator cannot give γ_k; a discrete operator keeping integrality is correct in principle but equivalent to direct enumeration, providing no explanatory power beyond the measurements of Section 10. This line ends here. We reiterate, unchanged since Section 10.3: even if this route succeeded, its correspondence with the Krasikov–Lagarias exponent would remain unestablished; 0.84 and $\gamma_{k\approx 25}$ coincide numerically only.

14. Far-tail large-deviation rate and the modular conditional law

Section 13 closed the spatial Ruelle operator route. The remaining question: on the most primitive axis — single-orbit statistics — is there structure still unextracted? This chapter studies the far-tail decay of the stopping time $D(t)$, giving a closed-form rate constant and a measured transfer matrix of the modular conditional law, together with one self-withdrawal (the claim that β_c had been falsified).

14.1. **The generalised family and the drift criterion: where the arithmetic transition sits.** Embed Collatz in the one-parameter family $T_q(m) = \frac{qm+1}{2^{v_2(qm+1)}}$, q odd. The key fact: the distribution of $v_2(qm+1)$ is independent of q. For among odd residue classes mod 2^k, the class $qm \equiv -1 \pmod{2^k}$ occupies exactly $2^{-(k-1)}$, so for uniformly random odd m and any odd q:

$$\mathbb{P}\big(v_2(qm+1) = k\big) = 2^{-k} \ (k \geq 1), \qquad \mathbb{E}[v_2] = \sum_{k\geq 1} k\, 2^{-k} = 2.$$

Numerical re-verification ($X = 10^6$ and 10^7, first step over all odd $t \in [3, X]$): $P(a = 1..6) = 0.50000, 0.25000, 0.12500, 0.06250, 0.03125, 0.01563$; $E[a] = 2.00000$, $Var[a] = 2.00001$; the geometric law holds to measurement precision ($\lesssim 10^{-5}$ from 2^{-k}), not as an asymptotic. The per-step logarithmic drift is therefore

$$\delta(q) = \ln q - \mathbb{E}[v_2] \ln 2 = \ln \frac{q}{4}.$$

q	$\delta(q)$	phase	known behaviour
1	$\ln \frac{1}{4} = -1.386$	ordered	must descend
3	$\ln \frac{3}{4} = -0.2877$	ordered	Collatz lives here
4	0	critical	q must be odd; excluded by parity
5	$\ln \frac{5}{4} = +0.2231$	disordered	divergent orbits and nontrivial cycles known

Structural consequence: the order parameter (fraction of divergent orbits) jumps directly from 0 at $q = 3$ to 1 at $q = 5$; no tunable critical point exists within the family — $q_c = 4$ is crossed by parity. This is the fingerprint of a first-order-transition type. Collatz sits at distance $\ln \frac{4}{3} = 0.2877$ from criticality, deep in the ordered phase, not near-critical. This is decisive for the critical-exponent/universality/renormalisation machinery: it has no object at $q = 3$. The Gaussian scaling of the D-distribution, $\mu \sim 3.671 \ln X$, $\sigma \sim 5.06\sqrt{\ln X}$, is linear response of the ordered phase (first-passage of a drifted walk), not critical scaling.

14.2. **Closed form of the far-tail rate β_c.** With increments $\xi_i = \ln 3 - a_i \ln 2$ and the exact geometric law, the moment generating function is

$$M(\theta) = \sum_{k\geq 1} 2^{-k} 3^{\theta} 2^{-k\theta} = \frac{3^{\theta}}{2^{1+\theta} - 1}, \qquad \Lambda(\theta) = \theta \ln 3 - \ln\bigl(2^{1+\theta} - 1\bigr).$$

The Cramér root $\Lambda'(\theta^*) = 0$ gives

$$\ln 3 - \frac{2^{1+\theta^*} \ln 2}{2^{1+\theta^*} - 1} = 0 \iff 2^{1+\theta^*} = \frac{\ln 3}{\ln(3/2)} = 2.709511,$$

(14.1)
$$\boxed{\theta^* = 0.438033, \qquad \beta_c = I(0) = -\Lambda(\theta^*) = 0.054979}$$

— the point where the tilted drift vanishes: ensembles with $\beta < \beta_c$ descend overall (ordered); $\beta \geq \beta_c$ is dominated by non-descending orbits (disordered). Compared with the Brownian approximation $\mu_d^2/(2\sigma_d^2) = \frac{0.2877^2}{2\times 0.9609} = 0.043064$: β_c is 27.7% higher — distinguishable, not the same thing. The closed form depends on only two ingredients: the exact geometric law (theorem-level, Section 14.1) and the constant $\ln 3/\ln 2$; it is one of the few numbers in this study derived from structure rather than fit.

14.3. **The prefactor trap and one withdrawal.** An intermediate stage reported: "β_c falsified by measurement, effective rate 0.065–0.075", claiming "suppression beyond the iid model". Withdrawn. The correct asymptotic of first passage is

$$\mathbb{P}(D > n) \sim C\, n^{-3/2} e^{-\beta_c n}, \qquad \text{so} \qquad -\frac{d}{dn} \ln \mathbb{P}(D > n) = \beta_c + \frac{1.5}{n};$$

the earlier procedure compared the measured slope against a theory curve varying with n (whose θ^* depends on L/n), leaving dirty residuals. With the correct baseline: at $n = 60, 100, 140, 170, 200$, $\beta_c + 1.5/n = 0.0817, 0.0717, 0.0675, 0.0656, 0.0642$, and the measured values ($X = 10^7, 10^8$ narrow band) 0.057–0.074 descending with n — the measured values fall pointwise into the corrected interval. The "0.065–0.075" is no new exponent; it is β_c plus the derivative of the universal first-passage prefactor $n^{-3/2}$. β_c is confirmed. (A second misjudgement: regressing directly at $X = 10^6$ gave $\beta \approx 0.12$ because there $\max D = 195$ and the fit hits the support boundary — a bounded-support artefact.)

14.4. **Modular constraints: conditional law and transfer matrix.** The derivation of β_c used only first marginals and independence of a. Whether independence holds must be measured: $a_i \geq k \iff m_i \equiv c_k \pmod{2^k}$ with $c_k = -3^{-1} \bmod 2^k$, so the residue class of m_{i+1} is biased by that of m_i — dependence should exist and be measurable. First distinguish two measures: the exact geometric law of Section 14.1 is under the uniform measure (first step of all odd t); statistics along orbits give the occupation measure:

measure	$E[a]$	$Var[a]$	remark
geometric 2^{-k}	2	2	
uniform ($X = 10^7$)	2.00000	2.00001	exact agreement
occupation ($X_s = 1.5 \times 10^5$)	1.9903	1.7869	variance 10.7% low
occupation ($X_s = 8 \times 10^5$)	1.9916	1.8145	variance 9.3% low

The occupation-measure variance is 9–11% below geometric (not the previously misreported 15%), and still drifts between samples: unconverged. Sequences along orbits are not free

samples. The measured transfer matrix ($X_s = 8 \times 10^5$, 1768 ten-thousand adjacent pairs):

a_i	j=1	j=2	j=3	j=4	j=5	max rel. dev.
1	0.520	0.258	0.104	0.061	0.039	0.30
2	0.474	0.233	0.156	0.093	0.020	0.26
3	0.568	0.174	0.093	0.129	0.020	0.47
4	0.410	0.287	0.226	0.042	0.018	0.66
5	0.339	0.159	0.092	0.379	0.018	3.45
independent	0.502	0.240	0.124	0.085	0.030	

The strongest single deviation is $P(a_{i+1} = 4 \mid a_i = 5) = 0.379$ (another independent sample: 0.429) versus the independent prediction 0.085–0.088: a factor 4.4–4.9. This is the explicit form of the modular constraint: large a's cluster. The lag-1 correlation measures +0.049–+0.059 (decreasing with sample size; unconverged). Conclusion: the independence assumption used in deriving β_c fails; the dependence is real and strong — but acts only in the tail of large a, with limited effect on the bulk.

14.5. **The Markov rate: the effect's direction is negative.** Treating the measured transfer matrix as a Markov chain and recomputing the large-deviation rate at the zero-drift point $s = \ln 3/\ln 2 = 1.584963$:

model	rate I	θ^*
theory β_c (exact geometric)	0.054979	
iid (measured marginal)	0.0567	−0.315
Markov (measured transfer)	0.0514	−0.285
Brownian approximation	0.043064	

The key sign: $I_{\text{Markov}} = 0.0514 < I_{\text{iid}} = 0.0567$. Positive correlation clusters small a's too; long runs of "low-division" steps are easier than iid, so the tail is heavier and the rate smaller. The modular constraint's effect on the rate is −9% to −10% (heavier tail, easier survival) — the direction is exactly opposite to the "faster decay" withdrawn in Section 14.3. The real system's far tail decays no faster than the iid model; the observed "faster" was purely a prefactor artefact. The modular constraint is real structure, but it cannot explain any suppression effect — its sign is negative.

15. Finite-scale drift of the far-tail slope and the survival-trajectory profile

15.1. **The question.** Section 14 gave the closed form of the far-tail rate and claimed in Section 14.3 that it was "confirmed by measurement". This chapter tests exactly that claim. The criterion is simple: if $\beta_c = 0.054979$ is truly the asymptotic exponential rate, the local slope $-\frac{d}{dn}\ln \mathbb{P}(D > n) = \beta_c + \frac{1.5}{n}$ must converge to that curve as X grows. If it does not converge, any "confirmation" is only tangency at some scale.

15.2. **Evolution of the slope at fixed n with X.** Using exact stopping-time data at $X = 10^5, 10^6, 10^7$, central differences with step $h = 6$:

X	$\mathbb{E}[D]$	$\max D$
10^5	37.98	129
10^6	45.99	195
10^7	53.94	256

X	n=60	n=80	n=100	n=120
10^5	0.0745	0.0887	0.1166	
10^6	0.0524	0.0724	0.0827	0.0933
10^7	0.0388	0.0585	0.0663	0.0747
iid + $1.5/n$	0.0818	0.0756	0.0718	0.0693
Markov + $1.5/n$	0.0764	0.0702	0.0664	0.0639

The slope descends monotonically with X, unconverged: at $n = 100$, $0.1166 \to 0.0827 \to 0.0663$, drops of 29% and 20%, with no flattening. The near-agreement of measured 0.0663 with Markov's 0.0664 at $X = 10^7$ is one crossing in the descent, not convergence.

15.3. **Scale-separation condition.** The drift's root is the bounded support: $\max D$ finite; as n approaches it, the distribution is compressed and the slope inflated. The measured growth law:

X	$\ln X$	$\max D$	$\max D/(\ln X)^2$
10^5	11.51	129	0.973
10^6	13.82	195	1.022
10^7	16.12	256	0.985

$\max D(X) \approx (\ln X)^2$, the ratio constant at 0.97–1.02. The scale-separation condition for measuring the asymptotic rate is therefore $n \ll (\ln X)^2$. At $X = 10^7$, $(\ln X)^2 \approx 260$, so $n \leq 100$ is safe; but $n \leq 100$ is precisely the pre-asymptotic region (at $n = 60$ the measured 0.0388 lies far below the theoretical 0.0818). The two requirements conflict: avoiding boundary effects needs small n; entering the asymptotic tail needs large n. Below $X = 10^7$ no window satisfies both.

15.4. **Resolution test.** Setting aside the finite drift: is the far-tail sample even large enough to distinguish iid's 0.0567 from Markov's 0.0514? Central differences on the $X = 10^7$ counts with Poisson errors at $n = 100, 110, 120, 140, 170, 180$ give measured slopes $0.066, 0.070, 0.075, 0.076, 0.074, 0.056$; the z-scores against the two models run from -20.2 to $+19.0$, flipping sign between adjacent n. The results contradict each other: $n = 100$ agrees with Markov ($z = -0.5$) and rejects iid ($z = -20$); $n = 110$ does the reverse; $n \geq 120$ rejects both. A pattern that flips with n is itself a diagnosis of unconvergence — were a model correct, z would stabilise near 0. Hence the table cannot decide which model is right; it can only decide: the data are not yet sufficient to decide.

15.5. **Nonstationarity by size m: almost no effect.** Another candidate explanation is stratification by the size of m: if layers have different a-distributions, trajectories wandering across magnitudes generate nonstationarity. Bucketing by $\log_2 m$ in units of 10 (all orbit steps at $X = 10^7$), the four layers carry step fractions $0.345, 0.459, 0.196, 0.0005$; their $E[a]$ values $1.984, 1.993, 1.998, 1.987$ agree within 1%, and their rates $0.0611, 0.0538, 0.0552, 0.0515$ lie in a narrow band. Weighting by the measured step fractions: stationary iid $I = 0.05645$ vs nonstationary 0.05643 — a difference of 0.0002 (0.04%). Nonstationarity by magnitude has no observable effect on the rate. The magnitude distribution itself has structure: all steps bucket as $(0.345, 0.459, 0.196, \dots)$, while deep steps (step ≥ 100) bucket as $(0.696, 0.290, 0.013, \dots)$ — the deep part of long trajectories concentrates strongly at small m ($m < 1024$). Correspondingly deep steps have pooled marginal $E[a] = 2.185$, $P(a = 1) = 0.463$, significantly deviating from the global 1.991/0.500; the deviation is explained by the correction term $\log_2(1 + \frac{1}{3m})$ (up to 0.415 at small m, Section 2.2), not an independent dynamical structure.

15.6. **The survival-trajectory profile of a: significantly non-flat.** Taking all trajectories with $D \geq 140$ (11489 of them at $X \leq 10^7$), recording a_i for the first 140 steps in segments of 20:

segment	$E[a]$	s.e.	deviation from flat	z
$[0, 20)$	1.5235	0.0028	-0.064	-22.9
$[20, 40)$	1.6178	0.0028	$+0.030$	$+10.8$
$[40, 60)$	1.6196	0.0028	$+0.032$	$+11.4$
$[60, 80)$	1.7105	0.0028	$+0.123$	$+43.9$
$[80, 100)$	1.8047	0.0028	$+0.217$	$+77.6$
$[100, 120)$	1.8063	0.0028	$+0.219$	$+78.1$
$[120, 140)$	1.7502	0.0028	$+0.163$	$+58.1$

The profile is significantly non-flat: early on it sits below the flat prediction of the iid tilted measure ($E[a] = 1.5876$ at θ^*; $z = -22.9$), rises to 1.80 mid-late ($z = +78$), and recedes slightly at the end. This is the chapter's only strong positive result. The mechanism is double. The early negative deviation (smaller a, more positive drift) matches the shape of

first-passage conditioned paths: trajectories conditioned on "not arriving for a long time" must first rise — consistent with Bessel(3)-type conditioned paths, while the flat prediction comes from endpoint constraints, not path constraints — the difference is precisely first-passage framework (with prefactor) versus bare endpoint large deviations. The late positive deviation is mainly caused by the correction term: trajectories enter the small-m region at the end, where $\log(1+\frac{1}{3m})$ raises the required a. Hence the profile supports the first-passage framework (not the endpoint-constraint framework), but it cannot pin down β_c: it is the shape of a conditioned path, while β_c is the rate function of the input model — objects on different levels. The unconvergence of Section 15.2 stands independently.

15.7. **Conclusion and second withdrawal.** Withdrawn: Section 14.3's claim of confirmation of β_c by "measured slope". The slope descends monotonically and unconvergedly with X (table above), so the confirmation fails. The closed form $\beta_c = 0.054979$ itself is intact (it depends only on Geom(1/2) and $\ln 3/\ln 2$), but its status as the Collatz fartail asymptotic rate is unverified and unfalsified on current data; likewise iid (0.0567) and Markov (0.0514) are indistinguishable at current data (z flips with n). Established: (i) scale-separation $\max D \approx (\ln X)^2$, hence below $X \le 10^7$ no window avoids both boundary effects and the pre-asymptotic region; (ii) magnitude nonstationarity affects the rate by 0.04%, negligible; deep-step vs global marginal difference ($E[a]$ 2.185 vs 1.991) explained by the correction term; (iii) the survival profile is significantly non-flat (1.52 → 1.81, deviations to 78σ), qualitatively supporting first-passage conditioned paths over endpoint constraints. Methodological consequence: the same trap again — model comparison on an unconverged quantity. The countermeasure upgrades once more: any comparison involving a limiting quantity must first verify at two scales that the quantity has converged; otherwise label it "unconverged, undecidable" and draw no conclusion from single-point agreement.

16. Unification of the three axes: the master identity

16.1. **The master identity $\rho(\beta) \equiv M(\beta-1)$.** The first fifteen sections split Collatz dynamics into three axes: the 2-adic increment law (axis 1), forward stopping time (axis 2), inverse tree and density propagation (axis 3). This section gives an identity merging all three — not a numerical coincidence but one line of algebra. On axis 1, the single-step log-increment $\xi = \ln 3 - a\ln 2$ with $a = v_2(3m+1)$ follows the exact geometric law $P(a=k) = 2^{-k}$, $k \ge 1$; its moment generating function is

$$M(\theta) = \mathbb{E}\left[e^{\theta\xi}\right] = \sum_{a\ge 1} 2^{-a} 3^{\theta} 2^{-a\theta} = \frac{3^{\theta}}{2^{1+\theta}-1}. \tag{16.1}$$

On axis 3, the rank-one structure gives the spectral radius

$$\rho(\beta) = \frac{\Lambda_1(\beta)+\Lambda_2(\beta)}{3} = \frac{3^{\beta-1}}{2^{\beta}-1}.$$

Comparing the two yields the master identity:

$$\boxed{\rho(\beta) \equiv M(\beta-1)} \tag{16.2}$$

Its meaning: the β-mass growth factor of the inverse tree at exponent β is the moment generating function of the increment law at tilting parameter $\beta-1$. The two axes measure the same function.

16.2. **Corollary chain: five corollaries, all checkable.** (i) *Normalisation*: $\rho(1) = M(0) = 1$ — the load-bearing critical value of Section 7 now has a second interpretation: the normalisation of the mgf at $\theta = 0$. (ii) *Martingale point*: $\rho(2) = M(1) = 1$, i.e. $\mathbb{E}[\frac{3}{2^a}] = 3\sum_{a\ge1} 4^{-a} = 3\cdot\frac{1/4}{3/4} = 1$. So the two roots of $\rho(\beta)=1$, at $\beta=1$ and $\beta=2$, correspond to normalisation ($\theta=0$) and the martingale point ($\theta=1$): Section 7's non-monotonicity with its second root is no lattice artefact but the martingale point — real structure. (iii)

Critical slope equals drift: $\rho'(1) = M'(0) = \mathbb{E}[\xi] = \ln 3 - 2\ln 2 = \ln\frac{3}{4} = \delta(3) = -0.2876821$, exactly the drift criterion $\delta(q) = \ln(q/4)$ of Section 14.1 at $q = 3$: the spectral slope at criticality equals the single-step drift. The three axes bite for the first time: axis-1 drift, axis-3 spectral slope, axis-2 transition criterion are one number. (iv) *The minimum is the Cramér root*: $\min_\beta \rho = \min_\theta M$ by (16.2), and $M'(\theta^*) = 0$ gives $2^{1+\theta^*} = \frac{\ln 3}{\ln(3/2)} = 2.7095113$, $\theta^* = 0.4380327$ — the same equation as the Cramér root of Section 14.2; the minimum point is no coincidence: $\beta_{\min} = 1 + \theta^* = 1.4380327$. (v) *Stopping-time rate and inverse-tree minimum are mutual negative logarithms*: by Legendre duality

$$\boxed{\beta_c = I(0) = -\Lambda(\theta^*) = -\ln M(\theta^*) = -\ln \min_\beta \rho(\beta)} \tag{16.3}$$

Chain verification (all six numbers agree to 10^{-12}):

quantity	expression	value
$2^{1+\theta^*}$	$\ln 3/\ln(3/2)$	2.709511291
θ^*	Cramér root	0.438032659
$\beta_{\min} = 1 + \theta^*$	$\arg\min \rho$	1.438032659
$\rho(\beta_{\min})$	$3^{\theta^*}/(2^{1+\theta^*} - 1)$	0.946504577
$M(\theta^*)$		0.946504577
$e^{-\beta_c}$	$\exp(-0.054979473)$	0.946504577

Hence the question posed at the opening of Section 6 — "are the three axes independent?" — now has a definite answer: they are not. Axis 1 defines $M(\theta) = 3^\theta/(2^{1+\theta} - 1)$; axis 3 is $\rho = M(\beta - 1)$; axis 2's large-deviation rate is $\beta_c = -\min \ln M$. Three readings of one function. It must be stressed that this unification is structural: it explains why the three axes give mutually consistent numbers, but it does not make any axis easier — each axis's wall (equidistribution unprovable, 0.84 not derivable from closed forms, unconverged far tail) remains standing.

16.3. **Repair of Table 7.1 and the ninth ledger entry.** Verifying the master identity uncovered a substantive error in the first version of Table 7.1. Two hard criteria. First: the original six rows had "theory" and "measured" columns identical digit by digit (0.5: 1.366025/1.366025; 1.5: 0.853553/0.853553; all six rows without exception) — identical columns can only come from one array, so that "verification" compared the table with itself. Second: the original values violate the sum rule (7.2): at $\beta = 0.5$, $1.366025 + 2.593863 = 3.959888$ against $3\rho(0.5) = 4.181541$, a 5.3% gap; being an algebraic identity, this independently disproves the original. Notably the error vanishes exactly at $\beta = 0.84$ and $\beta = 1$ — the very points most likely to be spot-checked. Table 7.1 has been recomputed from the closed form with the sum rule as a hard generation-time check. Entry no. 9, of a form different from the previous eight: not numerical reading, not code semantics, not comparison baseline, but the independence of the verification's two columns never having been confirmed. Countermeasure: any "theory vs measured" table must document genuinely different computational paths for the two columns and attach an algebraically provable internal-consistency identity, printed at generation time.

16.4. **Asymptotic slope of the median law: a correction to an external record.** An independent technical record (boundary-layer analysis of stopping times, v1.3) reported $\mu(X) = 3.671\ln X - 8.14$ and noted its +5.6% excess over the renewal prediction $1/\ln(4/3) = 3.4761$, guessing the ratio corresponds to "an explicit overshoot constant". Tested here with exact data. First at decade points: median stopping times at $X = 10^4, 10^5, 10^6, 10^7$ are $27, 35, 43, 51$ — exactly +8 per decade, giving slope $8/\ln 10 = 3.4744$, only 0.05% from 3.4761. But medians are integer-quantised, polluting the slope. Re-measured with smoothed

statistics (narrow-band local means of $\mathbb{E}[D \mid m \in [X/2, X]]$):

fit window	local-mean slope $[X/2, X]$	vs $1/\ln(4/3)$	narrower $[X/1.1, X]$	vs
10^4–10^7	3.4371	−1.12%	3.3821	−2.70%
10^5–10^7	3.4476	−0.82%	3.4299	−1.33%
10^6–10^7	3.4638	−0.35%	3.4673	−0.25%

Both bandwidths, all four windows: the deviation tends to zero monotonically ($-2.70\% \to -1.33\% \to -0.25\%$) — the typical shape of convergence to the renewal value. Conclusion: the asymptotic slope *is* $1/\ln(4/3) = 3.476059$; there is no overshoot constant; 3.671 is a finite-size bias from curvature when the fit window contains small X (10^3-scale). The record's open problem "the ratio 1.056 should equal an explicit overshoot constant" is dissolved: it was never a constant. The σ-law of the same record is confirmed: measured $(k_{.84} - k_{.16})/2$ at $X = 10^3, 10^5, 10^6$ gives $18.00, 21.00, 24.50$; the model $5.06\sqrt{\ln X} + 4.4$ gives $17.57, 23.21, 24.71$ — within integer quantisation (± 0.5). The constant term is less clean: writing median $= (\ln X - c)/|\delta(3)|$, sixty scales give $c = 1.4063 \pm 0.0167$; among candidates, $1 + \theta^* = 1.438$ deviates 1.9σ, $\ln 4 = 1.386$ by 1.2σ, 1.5 by 5.6σ, 2 by 4σ; no candidate is confirmed — in particular the temptation to identify the constant with $1 + \theta^*$ (it sits at the edge, 2σ) is, per the disciplines of this paper, not credited: recorded as undecided. This is the ledger's eighth countermeasure in action: a "almost-pretty" number without convergence evidence does not enter the library.

16.5. **Window drift of γ_k: the KL coincidence falsified.** Section 10.3 noted $\gamma_{24} \approx 0.824$ near the Krasikov–Lagarias exponent 0.84 and listed "prove that the depth to which KL propagates reliably is exactly $k \approx 25$" as the most valuable open problem. This section gives the negative evidence: the proximity is itself a product of the fit window.

k	window 10^3–10^6	window 10^3–10^7	window 10^5–10^7
12	0.6033	0.5728	0.5103
24	0.8237	0.7957	0.7404
40	0.9233	0.9095	0.8805

Three values, three windows: monotone decrease throughout, no stable value. γ_{24} runs from 0.824 down to 0.740; "$\gamma_{24} \approx 0.84$" holds only in a specific window (10^3–10^6), and its proximity to KL's 0.84 is a numerical coincidence, not a traceable correspondence. The drift direction is itself informative and fully consistent with the boundary-layer reframing of Section 10.4: with $N_k(X) = \frac{X}{2} F(z)$, $z = \frac{k-\mu(X)}{\sigma(X)}$, the local log slope is $\gamma_k^{\mathrm{loc}} = 1 + \frac{d\ln F}{dz} \cdot \frac{dz}{d\ln X}$; since $\mu \sim 3.476 \ln X$ grows and $\sigma \sim \sqrt{\ln X}$ grows slowly, at fixed k increasing X pushes the probe point z into the thin left tail of F, where $\frac{d\ln F}{dz}$ is larger in magnitude and $\frac{dz}{d\ln X} < 0$ — so γ_k decreases monotonically with X, as observed. The "most valuable problem" of Section 10.3 should therefore be closed: γ_k is not a quantity that can correspond to the KL exponent; it is a boundary-layer local slope drifting with the window; KL's 0.84 is the product of a whole-space lower-bound method. No bridge between them currently exists.

16.6. **Unmeasurability of the prefactor $n^{-3/2}$.** The $n^{-3/2}$ prefactor of the first-passage defective distribution is classical (ladder-height theory; the root is the square-root singularity of the Wiener–Hopf factorisation at its convergence radius [17]). Can it be measured on Collatz data? Regress $\ln \mathbb{P}(D > n) + \beta_c n$ on $\ln n$: slope c, expecting $\alpha \to 1.5$. Measured answer is negative: at $X = 10^6$, windows $n \in [60, 100], [80, 120], [100, 140]$ give $\alpha = 1.27, 2.62, 4.72$; at $X = 10^5$: $3.16, 6.05, 14.96$ — α explodes monotonically with the window, unconverged. The cause is Section 15.3: $\max D \approx (\ln X)^2$ (ratios $0.973, 1.022, 0.985$), while "entering the asymptotic tail" needs large n and "avoiding boundary effects" needs $n \ll (\ln X)^2$; below $X \le 10^7$ these conditions never intersect — near $\max D$ truncation dominates and α diverges. So $n^{-3/2}$ is unmeasurable on this data and can only be cited as a literature conclusion, consistent with Section 15 — the same wall appearing once more.

16.7. **Conclusion.** The positive outcome of this chapter is the master identity (16.2) with its five corollaries: three axes merged into a single function, and the strongest cross-axis relation to date, $\beta_c = -\ln \min \rho$. Of the three follow-up tests, one reaches a definite conclusion (the median slope converges to $1/\ln(4/3)$; the external record's +5.6% is a finite-size bias) and two are negative (no bridge to the KL exponent exists; the prefactor is unmeasurable at current scales). A substantive error in Table 7.1 was repaired and entered the ledger as no. 9 — its countermeasure (independent sources + internal-consistency hard check) is transferable. It must be repeated: this unification is descriptive; it explains why the three axes give mutually consistent numbers but does not carry any axis past its own wall.

17. Three-way shootout for the survival profile: the Bessel(3) hypothesis falsified

17.1. **The question.** Section 15.6 reported an unexplained phenomenon: conditioned on $D \geq 140$, the increment profile along survival trajectories is significantly non-flat — early $E[a] = 1.5235$ (net drift $\nu = \ln 3 - E[a] \ln 2 > 0$, i.e. rising), later rising to 1.8063 ($\nu < 0$, rapid descent). That section also noted the tilted measure's "flat" prediction comes from endpoint constraints rather than path constraints, so the two do not directly contradict. This chapter tests the stronger hypothesis — also the candidate in Section 15 deemed "most likely to grow a positive theorem":

Hypothesis B: the profile of conditioned survival trajectories is a universal feature of first passage, determined by the path measure of a drifted random walk conditioned on $\tau > n$; its continuous limit is the Brownian meander (time-changed) or Bessel(3) process; Collatz's arithmetic enters only through effective parameters, not the shape.

Hypothesis B is falsifiable: if it holds, an iid random walk built from the measured marginal law should reproduce the measured profile under the same starting distribution and the same conditioning. This chapter shootouts three objects directly: measurement (exact), iid model (increments exactly $\mathbb{P}(a = k) = 2^{-k}$), and first-order Markov model (increments from the measured transfer matrix).

17.2. **Reconstruction of the measured profile.** From exact stopping-time data at $X = 10^7$, the 11489 odds with $D \geq 140$ ($t \in [156159, 9999913]$); re-iterate each forward, recording the first 140 steps' a-sequence, averaged in segments (35 steps each). Net drift $\nu = \ln 3 - E[a] \ln 2$ per segment:

segment	$[0, 35)$	$[35, 70)$	$[70, 105)$	$[105, 140)$
$E[a]$	1.5630	1.6417	1.7962	1.7606
s.e.	0.0014	0.0016	0.0019	0.0018
ν	+0.0152	−0.0393	−0.1464	−0.1218

The segment reconstruction gives $1.5235, 1.6178, 1.6196, 1.7105, 1.8047, 1.8063, 1.7502$ at segments of 20 — digit-consistent with Section 15.6's table, so the data are same-source and reproducible. The shape: the first segment barely descends ($\nu \approx 0$, slight rise); descent concentrates in the third segment ($\nu = -0.146$); the last segment recedes slightly.

17.3. **Conditional-survival simulation of the iid random walk.** Model: start $S_0 = 0$, $\ln t$ sampled uniformly over odd $t \in [1, 10^7]$; increments $\xi = \ln 3 - a \ln 2$ with $a \sim \text{Geom}(1/2)$ (this law proved exact for random odd numbers in Section 3, re-verified here independently: $E[a] = 2.0437, 2.0059, 2.0043, 1.9997, 1.9983$ across magnitudes, matching theory); stop at $S \leq 0$. Six million simulations, of which 43591 have $\tau > 140$. Note the starting points are sampled unconditionally; the conditioned start distribution emerges naturally from the simulation (measured mean $\ln t = 15.508$, model 15.352 — close). This point is crucial: an initial version re-sampled from the "already-survived pool", doubly conditioning the start

and overestimating survival by 3.5× — voided and redone.

segment	measured	iid	diff	z
$[0, 35)$	1.5630	1.6780	-0.1149	-68.8
$[35, 70)$	1.6417	1.6806	-0.0388	-21.6
$[70, 105)$	1.7962	1.6912	$+0.1050$	$+51.2$
$[105, 140)$	1.7606	1.7402	$+0.0205$	$+10.3$

The model's net drift is $\nu = -0.0645, -0.0663, -0.0736, -0.1076$ — nearly uniform descent, slightly accelerating at the end; the measured profile is a strongly non-uniform "plateau — concentrated descent". The two deviate by ≈ 0.11 at both ends with opposite signs, z reaching -68.8 and $+51.2$.

17.4. **Stratified shootouts: excluding stop-time and start-point mixing.** One confound must be excluded: the measured sample mixes all stop times $D = 140, \ldots, 256$, and the model sample may mix differently; if the D-distributions differ, the profile difference could be pure mixing weight. Stratifying by stop time D (Table: layer $D \in [140, 160)$: measured 8991 vs Markov 24220: first-segment $E[a]$ 1.5779 vs 1.6851, diff -0.1072; third segment 1.8170 vs 1.7013, diff $+0.1157$; layer $D \in [160, 180)$: measured 1946, first segment 1.5223 vs 1.6683) and stratifying by starting point $\ln t$ (three layers $[15.0, 15.6)$, $[15.6, 16.0)$, $[16.0, 16.12)$: measured vs iid first-segment differences $-0.1156, -0.1121, -0.1138$ — stable across layers). Both stratifications show the difference is not a mixing effect; it is a systematic deviation stable within every layer.

17.5. **The first-order Markov model.** The iid model ignores first-order dependence; Section 14.4 confirmed transition structure in the a-sequence, re-measured here with the larger sample ($X = 10^6$... precisely a larger run, 1882 ten-thousand survival-step pairs). The occupation-measure marginal gives $E[a] = 1.9328$, significantly below the exact value for random odd numbers — orbit values are not uniformly random odd numbers (they are images under T, with biased distribution); this is consistent with, not contradictory to, Section 3's exact geometric law for uniformly random m. Building a first-order Markov model from this transfer matrix and marginal (3×10^6 simulations):

quantity	measured	iid	Markov
$E[a]$ of $[0, 35)$ segment	1.5630	1.6780	1.6753
$E[a]$ of $[70, 105)$ segment	1.7962	1.6912	1.6890
peak position (mean)	0.2736	0.1742	0.1823
$\mathbb{P}(\tau > 140)$	2.30×10^{-3}	7.27×10^{-3}	1.355×10^{-2}

The conclusion is clear: first-order Markov dependence is not the source of the difference. The Markov profile differs from iid by at most 0.005, while its difference from measurement remains of order 0.11; worse, the Markov survival probability is farther from measurement than iid's (1.355×10^{-2} vs 7.27×10^{-3} against measured 2.30×10^{-3}) — the wrong direction, not a partial correction.

17.6. **Quantitative characterisation of the difference.**

quantity	measured	iid	Markov	reading
overall $E[a]$ (140 steps)	1.6904	1.6976	1.6953	levels agree, 0.4%
first-segment net drift	$+0.0152$	-0.0645	-0.0626	measured does not descend; models do
third-segment net drift	-0.1464	-0.0736	-0.0721	descent concentrated in measured
peak position (median)	0.2429	0.1071	0.1143	measured 2.3× later
$\mathbb{P}(\tau > 140)$	2.30×10^{-3}	7.27×10^{-3}	1.355×10^{-3}	measured tail 3.2× thinner

Three facts form a self-consistent diagnosis: the levels agree (overall descent equal to 0.4%); what differs is the distribution in time. Measured trajectories descend on aggregate like the models but postpone and concentrate the descent into the middle-to-late segments, holding an initial "high plateau"; the peak sits at $\approx 27\%$ vs the models' $\approx 17\%$ (where it is mostly random fluctuation — the measured first segment has $\nu > 0$, meaning real trajectories

genuinely rise for a period the models do not). And the measured tail is thinner than both models'.

17.7. **Conclusion: hypothesis B falsified, and an open phenomenon.** Hypothesis B is falsified. The profile of conditioned survival trajectories is not a universal feature of drifted-walk first passage: the iid model (whose continuous limit is Brownian-meander/Bessel(3)-type paths) gives a nearly uniformly descending profile, deviating from measurement by ≈ 0.11 at both ends with $z = \pm 50$–69; the first-order Markov correction is ineffective. Any scheme explaining Table of Section 15.6 by "universal shape of conditioned first-passage paths" must first explain deviations of this size.

Positively, this section pins the phenomenon down in statable form. *Open phenomenon P:* Collatz survival trajectories have a definite "initial high plateau + mid-concentrated descent" shape — first-segment net drift $\nu \approx 0$ (slight rise), the third segment fastest at $\nu = -0.146$, peak at $\approx 27\%$; iid/Markov models with the same marginal and the same survival conditioning give nearly uniform descent with peak at $\approx 17\%$; and the measured tail is $3.2\times$ thinner than the models'. On mechanism, the excluded and the non-excluded must be stated separately. Excluded: (1) start distribution (stable across three start layers); (2) stop-time mixing (stable across D layers); (3) marginal-law bias (Markov with the measured marginal leaves the profile unchanged; the conditioning swamps it); (4) first-order transition dependence (Markov profile differs from iid by < 0.005). Not excluded: higher-order modular coupling between a and m. $a_i = v_2(3m_i + 1)$ is decided by the mod-2^{a+1} class of m_i, whose mod-structure is in turn decided by $m_i = (3m_{i-1} + 1)/2^{a_{i-1}}$; this chain is lost both in continuousisation and in first-order Markovisation. The data are compatible with this mechanism: the measured anomaly at small m (layer $\ln m < 2$: $E[a] = 2.96$, $n = 1371$) is the signature of special residue-class structure (e.g. elements of the A-set $m = (4^j - 1)/3$ give $a = 2j$). But this chapter gives no quantitative model of that mechanism; P remains a phenomenon, not a theorem. Finally one item to be checked: Section 14.4's table reported $P(a_{i+1} = 4 \mid a_i = 5) = 0.379$–$0.429$ (enrichment 4.4–4.9×); this chapter's 1882-ten-thousand-pair re-measurement on survival steps gives 0.204 (enrichment 2.3–3.3×). The two disagree; the source is to be found (plausibly: all orbits vs conditioned survival trajectories). Until the Markov rate of Section 14.5 is independently re-verified, its numbers are to be regarded as provisional; this chapter's falsification does not depend on it — Markov and iid profiles nearly coincide, so either transfer matrix leaves the conclusion unchanged.

References

[1] J. C. Lagarias, *The $3x+1$ problem and its generalizations*, Amer. Math. Monthly **92** (1985), 3–23.
[2] J. C. Lagarias (ed.), *The Ultimate Challenge: The $3x+1$ Problem*, Amer. Math. Soc., Providence, 2010.
[3] R. Terras, *A stopping time problem on the positive integers*, Acta Arith. **30** (1976), 241–252.
[4] C. J. Everett, *Iteration of the number-theoretic function $f(2n) = n$, $f(2n+1) = 3n+1$*, Adv. Math. **25** (1977), 42–45.
[5] I. Krasikov, J. C. Lagarias, *Bounds for the $3x+1$ problem using difference inequalities*, Acta Arith. **109** (2003), 237–258.
[6] D. Applegate, J. C. Lagarias, *Density bounds for the $3x+1$ problem I: tree-search method*, Math. Comp. **64** (1995), 411–426.
[7] T. Tao, *Almost all orbits of the Collatz map attain almost bounded values*, Forum Math. Pi **10** (2022), e12.
[8] J.-H. Evertse, *On sums of S-units and linear recurrences*, Compositio Math. **53** (1984), 225–244.
[9] J.-H. Evertse, H. P. Schlickewei, W. M. Schmidt, *Linear equations in variables which lie in a multiplicative group*, Ann. of Math. **155** (2002), 807–836.
[10] G. J. Wirsching, *The Dynamical System Generated by the $3n+1$ Function*, Springer, Berlin, 1998.
[11] Ya. G. Sinai, *Statistical $(3x+1)$ problem*, Comm. Pure Appl. Math. **56** (2003), 1016–1028.
[12] N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular Variation*, Cambridge Univ. Press, 1989 (repr.).
[13] R. R. Bahadur, R. R. Rao, *On deviations of the sample mean*, Ann. Math. Statist. **31** (1960), 1015–1027.
[14] N. G. de Bruijn, *The asymptotic behaviour of a function occurring in the theory of primes*, J. Indian Math. Soc. **15** (1951), 25–32.
[15] D. Ruelle, *Thermodynamic Formalism*, 2nd ed., Cambridge Univ. Press, 2004.

[16] K. Falconer, *Fractal Geometry*, 3rd ed., Wiley, Chichester, 2014.
[17] J. H. B. Kemperman, *The oscillating random walk*, Stochastic Process. Appl. **2** (1974), 1–29; cf. A. A. Borovkov, *On the asymptotic behaviour of first-passage probabilities*, and the Wiener–Hopf factorisation (square-root singularity) literature.