The Galton–Watson Branching Process

A note on the extinction fixed point

Giulio & Klaus

Math Corner — May 2026

A short note

Dear reader,

This little note accompanies the interactive demo at `galton_watson.html` and walks through the elegant piece of mathematics behind it: why the extinction probability of a branching process is the smallest fixed point of the offspring generating function in $[0, 1]$. We tried to keep the prose conversational; the mathematics is what it is. There are several closely related objects in play — an offspring random variable, a population sequence, two families of generating functions — and it is easy to get lost in indices, so §2 pins the notation down once and for all.

— Giulio & Klaus

1 Setup

Let X be a random variable taking values in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, with distribution

$$p_k \;:=\; \mathbb{P}(X = k), \qquad k \in \mathbb{N}_0, \qquad \sum_{k \geq 0} p_k = 1.$$

We call X the *offspring random variable* and $(p_k)_{k \geq 0}$ the *offspring distribution*. To avoid trivialities we assume $p_1 \neq 1$ throughout (no purely deterministic process).

Now build the process. For every $n \geq 1$ and every $i \geq 1$, let

$$X_{n,i} \quad \text{be an independent copy of } X,$$

all jointly independent. Think of $X_{n,i}$ as the offspring count of the i-th individual in generation $n - 1$ (so $X_{n,i}$ contributes to generation n). The population sequence is defined by $Z_0 = 1$ and

$$Z_n \;:=\; \sum_{i=1}^{Z_{n-1}} X_{n,i}, \qquad n \geq 1,$$

where the empty sum is 0 (a dead lineage stays dead). The sequence $(Z_n)_{n \geq 0}$ is the *Galton–Watson process*.

The basic — and also the most romantic — question is:

What is the probability q that the lineage ever dies out?

$$q \;:=\; \mathbb{P}\big(\, \exists\, n \in \mathbb{N} : Z_n = 0 \,\big)\,.$$

2 Notation, pinned down

We are about to juggle two generating functions and a few sequences. Let us name them once and never mix them up again:

Symbol	Definition
X	the offspring random variable, $X \sim (p_k)$
μ	mean number of offspring, $\mu := \mathbb{E}[X]$
Z_n	population at generation n. $Z_0 = 1$
$f(s)$	pgf of a single offspring: $f(s) := \mathbb{E}\big[s^X\big] = \sum_{k\geq 0} p_k s^k$
$f_n(s)$	pgf of generation n: $f_n(s) := \mathbb{E}\big[s^{Z_n}\big]$
q_n	probability of extinction by time n: $q_n := \mathbb{P}(Z_n = 0) = f_n(0)$
q	total extinction probability: $q := \lim_{n\to\infty} q_n$

A few comments to keep things from drifting:

- f is the pgf of *one* individual's offspring count. It does not depend on n.
- f_n is the pgf of *the whole population* at generation n. It does depend on n.
- In particular $f_0(s) = \mathbb{E}[s^{Z_0}] = \mathbb{E}[s^1] = s$ (the identity), and $f_1(s) = \mathbb{E}[s^X] = f(s)$.
- We will prove that $f_n = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ copies}}$ — this is the whole structural content of the process.

Throughout, s is a dummy variable in $[0, 1]$. We never differentiate at $s > 1$; everything we need lives on the unit interval.

3 Three structural facts about f

Before we touch the recursion, note three properties of the offspring pgf f on $[0, 1]$ that we will lean on repeatedly:

(1) $f(0) = p_0$ and $f(1) = \sum_k p_k = 1$.

(2) f is non-decreasing on $[0, 1]$: differentiating term by term, $f'(s) = \sum_{k\geq 1} k p_k s^{k-1} \geq 0$.

(3) f is convex on $[0, 1]$: $f''(s) = \sum_{k\geq 2} k(k-1) p_k s^{k-2} \geq 0$.

The mean offspring count is

$$\mu \;=\; \mathbb{E}[X] \;=\; f'(1) \;=\; \sum_{k\geq 1} k p_k.$$

This single number controls everything that follows.

4 The recursion: $f_n = f^{\circ n}$

Here is the move that turns a hard probability question into a fixed-point problem in one dimension.

Lemma 1 (Iteration of the pgf)**.** *For all $n \geq 0$ and all $s \in [0,1]$:*

$$f_{n+1}(s) \;=\; f_n(f(s)) \;=\; f(f_n(s))\,.$$

Consequently $f_n = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}$, with the convention $f_0 = \mathrm{id}$.

Proof. We prove the first equality, $f_{n+1}(s) = f_n(f(s))$. Recall

$$Z_{n+1} \;=\; \sum_{i=1}^{Z_n} X_{n+1,i},$$

where the $X_{n+1,i}$ are iid copies of X, independent of Z_n. Condition on Z_n:

$$\mathbb{E}\big[\, s^{Z_{n+1}} \,\big|\, Z_n \,\big] \;=\; \mathbb{E}\big[\, s^{X_{n+1,1}+\cdots+X_{n+1,Z_n}} \,\big|\, Z_n \,\big] \;=\; \Big(\mathbb{E}\big[s^X\big]\Big)^{Z_n} \;=\; f(s)^{Z_n}.$$

Now take expectations:

$$f_{n+1}(s) \;=\; \mathbb{E}\big[s^{Z_{n+1}}\big] \;=\; \mathbb{E}\big[f(s)^{Z_n}\big] \;=\; f_n\big(f(s)\big),$$

where the last equality is just the definition $f_n(t) = \mathbb{E}[t^{Z_n}]$ applied at $t = f(s)$.

The second equality, $f_{n+1}(s) = f(f_n(s))$, comes from a different (and equally valid) decomposition of the same tree: split it after the very first generation. The Z_1 children of the root each found their own independent Galton–Watson sub-tree of depth n. Writing $W_n^{(1)}, \ldots, W_n^{(Z_1)}$ for those sub-trees' generation-n populations,

$$Z_{n+1} \;=\; W_n^{(1)} + W_n^{(2)} + \cdots + W_n^{(Z_1)}.$$

The $W_n^{(j)}$ are iid copies of Z_n, independent of Z_1, each with pgf f_n. So conditioning on Z_1 and repeating the calculation above with f_n in place of f and Z_1 in place of Z_n,

$$f_{n+1}(s) \;=\; \mathbb{E}\big[f_n(s)^{Z_1}\big] \;=\; f_1\big(f_n(s)\big) \;=\; f\big(f_n(s)\big).$$

Finally, since $f_1 = f$, induction on n gives $f_n = f^{\circ n}$. □

This is striking: the entire infinite-dimensional probabilistic apparatus of the branching process is recorded by iterating a single one-dimensional function.

5 Extinction as a fixed point

Set $q_n := \mathbb{P}(Z_n = 0) = f_n(0)$. Plug $s = 0$ into Lemma 1:

$$q_{n+1} \;=\; f_{n+1}(0) \;=\; f\big(f_n(0)\big) \;=\; f(q_n).$$

So the sequence of cumulative extinction probabilities is obtained by iterating f starting from 0. Because $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$ (extinction is forever), the sequence (q_n) is non-decreasing in n and bounded above by 1, hence convergent. Its limit equals the probability of *eventual* extinction:

$$q \;:=\; \lim_{n\to\infty} q_n \;=\; \mathbb{P}\left(\bigcup_{n\geq 0} \{Z_n = 0\}\right) \;=\; \mathbb{P}(\text{extinction}).$$

Letting $n \to \infty$ in $q_{n+1} = f(q_n)$ and using continuity of f on $[0,1]$:

$$\boxed{f(q) \;=\; q.}$$

Conclusion: *q is a fixed point of f in $[0,1]$.*

6 Which fixed point?

Note that $s = 1$ is *always* a fixed point of f on $[0, 1]$, since $f(1) = 1$. So just saying $f(q) = q$ does not yet pin q down. The theorem that does:

Theorem 2 (Extinction theorem)**.** *The extinction probability q is the* smallest *fixed point of f in $[0, 1]$. Equivalently:*

- *if $\mu \leq 1$, then $q = 1$ (extinction is almost certain);*
- *if $\mu > 1$, then f has a unique fixed point $q \in [0, 1)$, and the extinction probability equals it.*

Why *smallest*

Let $q^\star \in [0, 1]$ be any fixed point of f. Since f is non-decreasing,

$$q_0 = 0 \;\leq\; q^\star \quad\Longrightarrow\quad q_1 = f(0) \;\leq\; f(q^\star) = q^\star,$$

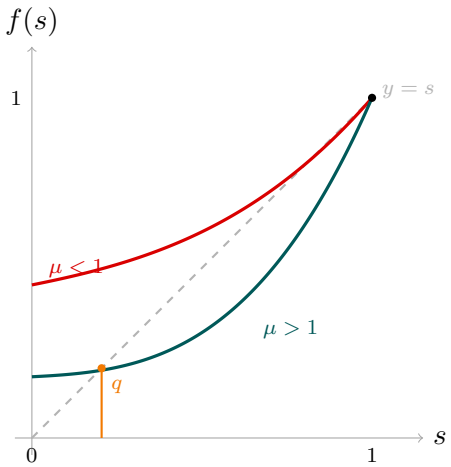
and by induction $q_n \leq q^\star$ for every $n \geq 0$. Letting $n \to \infty$: $q \leq q^\star$. So q sits below every fixed point of f in $[0, 1]$; it is the smallest.

The geometric heart: a picture

Theorem 2 is really a picture. On $[0, 1]$ the graph of f:

- starts at $(0, p_0)$, on or above the horizontal axis;
- ends at $(1, 1)$, on the diagonal $y = s$;
- is convex (curving upward, away from the diagonal);
- arrives at the right-hand endpoint with slope $f'(1) = \mu$.

A fixed point of f in $[0, 1]$ is a place where this graph meets the diagonal. The slope μ controls whether that meeting can happen anywhere other than at $(1, 1)$:



Case $\mu \leq 1$. The curve arrives at $(1, 1)$ with slope at most 1. A convex curve that ends at $(1, 1)$ with slope ≤ 1 at that endpoint cannot lie below the diagonal anywhere in $[0, 1)$ (it would have to

curve back upward too steeply at the right endpoint, in contradiction to convexity). So $f(s) > s$ on $[0, 1)$, and the only fixed point in $[0, 1]$ is $s = 1$. Hence $q = 1$.

Case $\mu > 1$. The curve arrives at $(1, 1)$ with slope strictly greater than 1, so just to the left of $s = 1$ it lies *below* the diagonal. At $s = 0$ the curve is at height $p_0 \geq 0$, on or above the diagonal $y = 0$. (When $p_0 = 0$ the lineage cannot die in one step from a single individual, and indeed $q = 0$ trivially.) By the intermediate value theorem the curve crosses the diagonal somewhere in $(0, 1)$. Convexity forces this crossing to be unique. The crossing is q.

The qualitative story is clean: *the dichotomy at $\mu = 1$ is a statement about the angle at which the curve f enters the corner $(1, 1)$.*

7 Computing q in practice

Because f is non-decreasing and the smallest fixed point is reached from below, the simplest reliable recipe is to iterate from zero:

$$q_0 = 0, \qquad q_{n+1} = f(q_n).$$

This is the same recursion we derived for $\mathbb{P}(Z_n = 0)$, so each iterate has a physical meaning: q_n is the probability the lineage has died out by generation n. The sequence converges to q from below, monotonically, and (in the supercritical case) at geometric rate μ.

In the demo you can watch this convergence empirically: pressing $\times 100$ `batch` runs a hundred independent simulations and the resulting estimate $\hat{q}$ slides toward the theoretical q shown as a gold dot on the generating-function plot.

8 Why this is delightful

The Galton–Watson process is one of the cleanest examples in probability of a recurring pattern in mathematics: *a question about an infinite, branching dynamical system collapses to a one-dimensional fixed-point equation.* The whole long-run behaviour of $(Z_n)_{n \geq 0}$ is read off a single picture — a curve, a diagonal, where they meet.

This pattern reappears across the subject: in percolation, in the giant component of a random graph, in queueing theory, in epidemic models, in evolutionary biology. It is worth holding on to.

Where to read more

- K. Athreya & P. Ney, *Branching Processes*, Springer 1972.
- T. Harris, *The Theory of Branching Processes*, 1963.
- R. Lyons & Y. Peres, *Probability on Trees and Networks*, freely available online — chapter 5.
- D. Williams, *Probability with Martingales*, §0.5 — the cleanest one-page treatment in print.