THE GALTON-WATSON BRANCHING PROCESS

ADITYA SINGH

Abstract. The Galton-Watson process is a mathematical representation of the development of a population whose members reproduce or die according to a probabilistic law. The main goal of this paper is to answer the two key analytical questions about the long-term behavior of the process – firstly, what happens in the limit, and secondly, what is the rate of convergence to the limit. In later sections, we also consider slight variants of the process and examples of Galton-Watson processes embedded within other stochastic processes. The emphasis is not on proving the finest limit theorems, but a clear exposition of the behavior of the process and how it changes based on the exact assumptions of the model.

Contents

1. Introduction

This paper introduces the Galton-Watson branching process, first proposed by Francis Galton in 1889 to model the proliferation of family names. It is the simplest possible model for a random population evolving in time, based on the assumption that all individuals reproduce independently and all with the same distribution.

Date: August 10, 2024.

2. Basic Properties and Extinction Probability

Definition 2.1. A Galton-Watson process is a markov chain $\{Z_n : n = 0, 1, 2, \ldots\}$ on the non-negative integers with transition function

$$P(i,j) = \Pr\{Z_{n+1} = j \mid Z_n = i\} = \begin{cases} p_j^{*i} & \text{if } i \geq 1, \\ \delta_{0j} & \text{if } i = 0 \end{cases}$$

where $\{p_k\}_{k\in\mathbb{N}_0}$, $p_k > 0$, $\sum_k p_k = 1$ is referred to as the offspring distribution. The notation $*i$ means i-fold convolution, that is, conditioned on $Z_n = i$, Z_{n+1} is the sum of i independent random variables with distribution $\{p_k\}_{k\in\mathbb{N}_0}$.

We will typically make the following assumptions:

(1) $Z_0 = 1$
(2) $p_k < 1, \forall k$
(3) $p_0 + p_1 < 1$
(4) $p_0 > 0$

Hence, we avoid any purely deterministic processes, as well as the uninteresting case of the single starting individual reproducing one offspring until it eventually dies. The Galton-Watson process under our assumptions begins with a single individual in the zeroth generation, then produces Z_1 independent Galton-Watson processes, and so on. Extinction ($Z_n = 0$ for some n) is accessible from any state.

One can visualize the Galton-Watson process via a tree-like structure (hence the phrase "branching process") – we start with a single ancestor node in the zeroth generation ($Z_0 = 1$), and then branch out to Z_1 child nodes. These Z_1 first generation nodes then reproduce a sum total of Z_2 child nodes, although we don't necessarily know how these second generation offspring are allocated amongst the Z_1 first generation nodes:

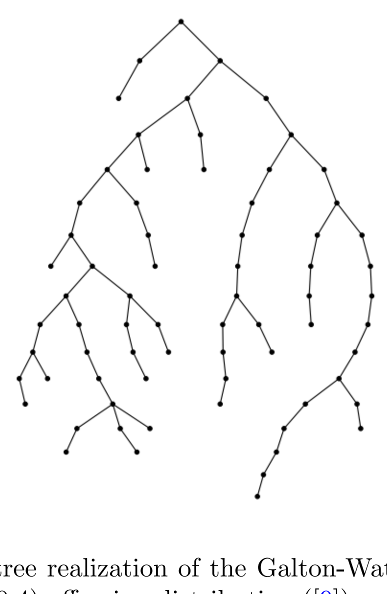

Figure 1. A tree realization of the Galton-Watson process with a $Binomial(3, 0.4)$ offspring distribution ([9]).

To motivate the organization of the paper, we mention a very important historical example of branching processes: the neutron chain reaction.

The discovery of the neutron itself is due to James Chadwick on February 27, 1932. The following year, on September 12, 1933, the Hungarian physicist Leo Szilard was one of the very first to propose that neutrons could be responsible for nuclear chain reactions (nearly a decade later on December 2nd, 1942, Szilard's ideas were confirmed under an enormous tent on the University of Chicago's Stagg Field, where a team headed by Italian scientist Enrico Fermi engineered the first controlled nuclear fission chain reaction).

Here is an excerpt from an American Nuclear Society article, showing hints of a branching process in Szilard's epiphany: "As Szilard later recounted the story, when he reached the intersection of Southampton Row and Russell Square a red light caused him to pause, giving time for his fertile imagination to engage. Then the idea struck him: If a neutron entered an atomic nucleus, and the subsequent reaction released two neutrons, it would be possible to produce a chain reaction. Since neutrons have no charge, each of those newly released neutrons would be able to travel freely through matter until they struck another nucleus.

If there was a sufficiently large mass, with a sufficient purity of the material whose nuclei released two neutrons every time it was hit with one neutron, Szilard realized that there was a distinct potential for industrial-scale power sources. He recognized immediately that there was also a possibility that the reactions could be produced in a manner that was rapid enough to cause an explosion of great force before the material was scattered and the reaction stopped" ([8]).

Note that the branching process Szilard envisions here has an offspring distribution of $p_0 = \alpha$ and $p_2 = 1 - \alpha$ for some $\alpha \geq 0$, as the line "If a neutron entered an atomic nuclei, and the subsequent reaction released two neutrons..." indicates.

Several important questions are made apparent in Szilard's thought process. Firstly is the question of if, once started, the reaction ever stops. After setting up some basic facts about the process, we address the main questions surrounding extinction of the process and rates of extinction in Section 2.3, Section 3.1, and Section 3.2. But Szilard also considers the "possibility that the reactions could be produced in a manner that was rapid enough to cause an explosion of great force before the material was scattered and the reaction stopped." This concerns the total progeny $(Z_1 + Z_2 + \dots)$ born over the course of the process, which we address in Section 4. Indeed, we will see that if we set $\alpha = 1/2$ in Szilard's model, then we have a phenomenon where the process will go extinct with probability one, and yet the expected number of total progeny is infinite (see Section 4.1)!

We begin our analysis of the Galton-Watson process with what is arguably the most basic question:

Question 2.1. What happens to the Galton-Watson process in the limit as n tends to infinity? Does it die out, converge to an equilibrium position, diverge to infinity, or something else?

2.1. **Recursive Structure and Generating Functions.** A central tool for our analysis will be the generating function

$$f(s) := \mathbb{E}[s^{Z_1}] = \sum_{j=0}^{\infty} p_j s^j$$

defined for $s \in [0, 1]$, and its iterates

$$f^{(n)}(s) = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}(s).$$

Since the generating function of the sum of i i.i.d. random variables is the generating function of one raised to the ith power, for any $i \in \mathbb{N}$ we have

$$\sum_{j=0}^{\infty} P(i,j)s^j = \sum_{j=0}^{\infty} p_j^{*i} s^j = [f(s)]^i. \tag{2.2}$$

Let $P_n(i,j) = \Pr\{Z_{n+m} = j \mid Z_m = i\}$ denote the n-step transition probability of the GW process.

Proposition 2.3. *Let $f_n(s) := \mathbb{E}[s^{Z_n}] = \sum_j P_n(1,j)s^j$ be the generating function for Z_n. Then f_n is the n-fold composition $f^{(n)}$ of the generating function f for Z_1.*

Proof. This recursion is of the utmost importance, so we give two proofs – one via the Chapman-Kolmogorov equations, and one via the tower law.

C-K proof: We show $f_{n+1}(s) = f_n[f(s)]$ which may then be iterated to obtain the claim. We have

$$\text{(C-K)} \qquad f_{n+1}(s) = \sum_j P_{n+1}(1,j)s^j = \sum_j \left(\sum_k P_n(1,k)P(k,j) \right) s^j$$

$$\text{(using (2.2))} \qquad = \sum_k P_n(1,k) \sum_j P(k,j)s^j = \sum_k P_n(1,k)[f(s)]^k$$

$$= f_n[f(s)].$$

Tower law proof: We proceed based on the recursive observation that given Z_n, Z_{n+1} is the sum of Z_n independent copies of Z_1. The notation $X_j^{(i)}$ is used when we have a collection of i.i.d. random variables with distribution X_j. We have

$$\begin{aligned} f_{n+1}(s) = \mathbb{E}[s^{Z_{n+1}}] &= \mathbb{E}\mathbb{E}[s^{Z_{n+1}}|Z_n] = \mathbb{E}\mathbb{E}[s^{\sum_{i=1}^{Z_n} Z_1^{(i)}}|Z_n] \\ &= \mathbb{E}[f(s)^{Z_n}] = f_n[f(s)]. \end{aligned}$$

Dually, we could have observed that given Z_1, Z_{n+1} is the sum of Z_1 independent copies of Z_n, so that

$$\begin{aligned} f_{n+1}(s) = \mathbb{E}[s^{Z_{n+1}}] &= \mathbb{E}\mathbb{E}[s^{Z_{n+1}}|Z_1] = \mathbb{E}\mathbb{E}[s^{\sum_{i=1}^{Z_1} Z_n^{(i)}}|Z_1] \\ &= \mathbb{E}[f_n(s)^{Z_1}] = f[f_n(s)]. \end{aligned}$$

□

2.2. **Moments of the Process.** Since $f^{(n)}$ and f_n are the same, we will only use the notation f_n from here onwards. The moments of the process can be computed via the derivatives of the f_n at 1:

$$m \stackrel{\text{def}}{=} \mathbb{E}[Z_1] = \sum_j P(1,j)j = f'(1),$$

and

$$\begin{aligned} \mathbb{E}[Z_n] = \sum_j P_n(1,j)j &= f_n'(1) = f_{n-1}'(f(1))f'(1) \\ &= f_{n-1}'(1)f'(1) = \ldots = [f'(1)]^n = m^n. \end{aligned}$$

Remark 2.4. Markov's inequality gives $\Pr\{Z_n \geq 1\} \leq m^n$, so that in the case of $m < 1$, we see $\lim_{n\to\infty} \Pr\{Z_n \geq 1\} = \Pr\{\bigcap_n (Z_n \geq 1)\} = 0$, that is the process dies out at some n almost surely. We give a second proof of this fact (and extend it to the case $m = 1$) later.

Computing the second moments of the process involves evaluating the second derivative of the generating function at 1. We have

$$f_n''(1) = \sum_j j(j-1)P_n(1,j) = \sum_j j^2 P_n(1,j) - \sum_j jP_n(1,j)$$

$$= \mathbb{E}[Z_n^2] - \mathbb{E}[Z_n],$$

and hence

$$\mathbb{E}[Z_n^2] = f_n''(1) + f_n'(1).$$

Higher moments can be similarly derived by linear combinations of higher order derivatives of f_n at 1.

Proposition 2.5. *Denote $\sigma^2 := Var(Z_1) = f''(1) + m - m^2$. Then the variance of Z_n is given by*

$$\mathrm{Var}(Z_n) = \begin{cases} \frac{\sigma^2 m^{n-1}(m^n-1)}{(m-1)} & \text{if } m \neq 1, \\ n\sigma^2 & \text{if } m = 1 \end{cases}.$$

Proof. $\mathrm{Var}(Z_n) = \mathbb{E}[Z_n^2] - \mathbb{E}[Z_n]^2 = f_n''(1) + f_n'(1) - (f_n'(1))^2$. By the chain rule,

$$f_{n+1}''(1) = f''(1)[f_n'(1)]^2 + f'(1)f_n''(1) = f''(1)m^{2n} + mf_n''(1).$$

Iterating this identity,

$$\begin{aligned} f_n''(1) &= f''(1)m^{2n-2} + mf_{n-1}''(1) \\ &= f''(1)m^{2n-2} + m[f''(1)m^{2n-4} + mf_{n-2}''(1)] \\ &= f''(1)m^{2n-2} + f''(1)m^{2n-3} + m^2 f_{n-2}''(1) \\ &\vdots \\ &= f''(1)m^{2n-2} + \ldots + m^{n-1}f''(1) \\ &= f''(1)[m^{2n-2} + m^{2n-3} + \ldots + m^{n-1}]. \end{aligned}$$

Hence,

$$\mathrm{Var}(Z_n) = f''(1)[m^{2n-2} + m^{2n-3} + \ldots + m^{n-1}] + m^n - m^{2n}. \tag{2.6}$$

For $m = 1$, the formula clearly holds. If $m \neq 1$, then we can factor (2.6) to get

$$\mathrm{Var}(Z_n) = \frac{(f''(1) + m - m^2)m^{n-1}(m^n - 1)}{m-1}.$$

□

2.3. **The Fixed Point Extinction Probability Theorem.** As all the information about the Z_n is contained in the generating functions, the general philosophy will be to analyze the generating functions as precisely as possible. The following basic properties are apparent from the definition of f:

Lemma 2.7. *(Properties of the generating function) Let t be real.*

1. *f is strictly convex and increasing on $[0,1]$*
2. *$f(0) = p_0$ and $f(1) = 1$*
3. *If $m \leq 1$, then $f(t) > t$ for $t \in [0,1)$*
4. *If $m > 1$ then $f(t) = t$ has a unique solution in $[0,1)$*

We may summarize as follows:

Proposition 2.8. *There exists a smallest positive solution q for the equation $f(t) = t$, $t \in [0,1]$. Furthermore, when $m \leq 1$, $q = 1$, and when $m > 1$, $q < 1$.*

The situation is illustrated below ([1], pg. 17):

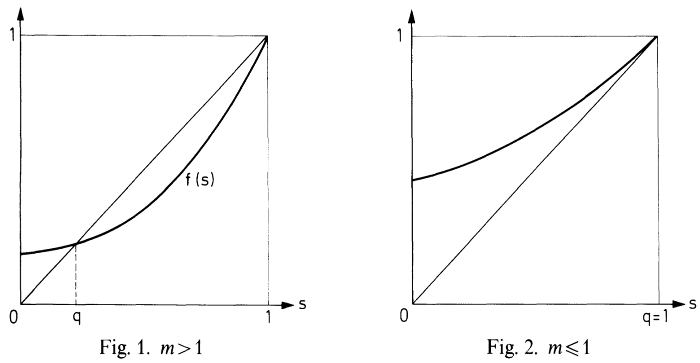

FIGURE 2. Graphs of the generating function f for $m > 1$ and $m \leq 1$.

We also have the Lamerey diagrams ([1], pg. 17):

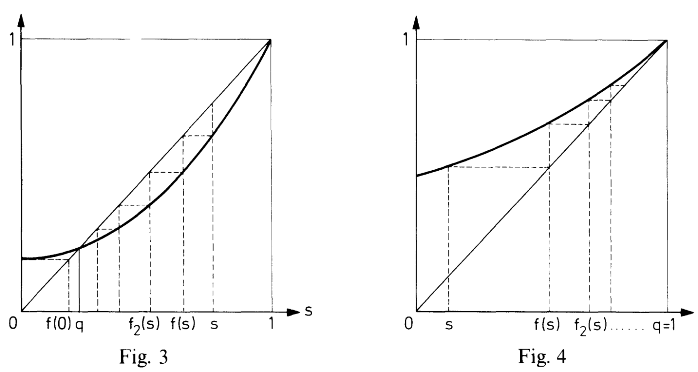

FIGURE 3. Lamerey diagrams of the generating function f.

Figure 3 suggests the following lemma:

Lemma 2.9. *Let q be as in the above proposition.*

1. *If $t \in [0, q)$, then $f_n(t) \uparrow q$ as $n \to \infty$*
2. *If $t \in (q, 1)$, then $f_n(t) \downarrow q$ as $n \to \infty$*
3. *If $t = q$ or $t = 1$, then $f_n(t) = t$, $\forall n$.*

Proof. If $t \in [0, q)$, then

$$\begin{gathered} t < f(t) < f(q) \\ f(t) < f_2(t) < f_2(q) \\ f_2(t) < f_3(t) < f_3(q) \\ \vdots \\ t < f(t) < f_2(t) < \cdots < f_n(t) < f_n(q) = q \end{gathered}$$

for all n. It follows that

$$\lim_{n\to\infty} f_n(t) := L \leq q$$

Since f is continuous, we have

$$L = \lim_{n\to\infty} f_{n+1}(t) = \lim_{n\to\infty} f[f_n(t)] = f[\lim_{n\to\infty} f_n(t)] = f(L)$$

But q is the only solution in $[0, 1)$, so $q = L$.

If $t \in [q, 1)$, then $q < 1$ implies $m > 1$, so that the graph of f lies below $y = x$ on $(q, 1)$. Hence

$$\begin{gathered} f(q) < f(t) < t \\ f_2(q) < f_2(t) < f(t) \\ f_3(q) < f_3(t) < f_2(t) \\ \vdots \\ t > f(t) > f_2(t) > \cdots > f_n(t) > f_n(q) = q \end{gathered}$$

for all n. We can now repeat the same argument as in the case $t \in [0, q)$ to obtain the claim.

For $t = q$ or $t = 1$, the claim is trivial. □

Remark 2.10. Note that on $[0, q)$, the f_n converge uniformly to the constant function q.

We have already done enough work to arrive at an interesting extinction probability theorem. Since $f_n(0) \uparrow q$, we have

$$\lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} \Pr\{Z_n = 0\} = \Pr\{\bigcup_n (Z_n = 0)\} = \Pr\{\textit{extinction}\} = q.$$

Theorem 2.11. *The extinction probability of the Galton-Watson process is the smallest positive solution of the equation $t = f(t)$. It is 1 if $m \leq 1$, and less than 1 if $m > 1$.*

We see that in the subcritical and critical cases ($m < 1$ and $m = 1$), there is only one long-term option – the process dies out almost surely. We now show that in the supercritical case ($m > 1$), there are only two long term options.

Lemma 2.12. *Let $k \in \mathbb{N}$. Then the state k is a transient state of Galton-Watson process Z_n.*

Proof. This is immediate from general markov chain considerations – if a markov chain Z_n has an absorbing state z, and if x is a state such that z is accessible from x, then x is transient. □

Theorem 2.13. *Suppose $m > 1$. Then with probability one, either $Z_n = 0$, for some n, or Z_n escapes to infinity.*

Proof. Observe that Lemma 2.12 implies that for any natural k, Z_n equals k infinitely often with probability zero. Since the naturals are countable, this means the event that the sequence $\{Z_n\}$ takes a natural number k as a limit point occurs with probability zero. Hence, the only two remaining options are that the Z_n die out or diverge to infinity, so we may conclude that $q = \Pr\{\lim_n Z_n = 0\} = 1 - \Pr\{\lim_n Z_n = \infty\}$. □

3. Asymptotic Behavior

Theorem 2.11 and Theorem 2.13 fully answer Question 2.1. But we know next to nothing about the rates of convergence. Call the cases of $m < 1$, $m = 1$, and $m > 1$ the subcritical, critical and supercritical cases respectively. The next analytical questions are:

Question 3.1. In the subcritical and critical cases, how fast does the process tend to extinction?

Question 3.2. In the supercritical case, how fast does the population grow on the event of non-extinction?

The answer to Question 3.2 simply amounts to looking at the right martingale, so we begin there.

3.1. **The Supercritical Case.** Recall that given $Z_n = i$, by definition the process $\{Z_{n+k} : k = 0, 1, 2, \dots\}$ is the sum of i i.i.d. copies of the process $\{Z_0 = 1, Z_1, Z_2, ...\}$. Combining this with the Markov property, we obtain

$$\begin{aligned}\mathbb{E}[Z_{n+k}|Z_n, Z_{n-1}, ..., Z_0] &= \mathbb{E}[Z_{n+k}|Z_n] \\ &= Z_n\mathbb{E}[Z_k|Z_0 = 1] = Z_n m^k.\end{aligned}$$

Hence, scaling the Z_n by m^{-n} gives the following:

Proposition 3.1. *Let $\{Z_n\}_{n\geq 0}$ be a Galton-Watson branching process with mean offspring $m < \infty$. Define the sequence $\{W_n\}_{n\geq 0}$ by $W_n = Z_n m^{-n}$. Let $\mathcal{F}_n$ denote the natural filtration with respect to $\{Z_1, \dots, Z_n\}$. Then, $\{W_n, \mathcal{F}_n\}_{n\geq 0}$ is a martingale.*

Since W_n is a non-negative martingale, and $\mathbb{E}[W_0] = \mathbb{E}[Z_0] = 1$, by the Martingale Convergence Theorem, there exists a random variable W such that

$$\lim_{n\to\infty} W_n = W \text{ a.s.}$$

Also note from Fatou's lemma that

$$\mathbb{E}[W] \leq \liminf_{n\to\infty} \mathbb{E}[W_n] = 1$$

so that W is almost surely finite.

It could happen that m^n shrinks too fast to observe interesting behavior, and so W is degenerate, that is $\Pr\{W = 0\} = 1$ (we know this is always the case for $m \leq 1$). We show this does not happen under an assumption of finite variance:

Theorem 3.2. *If $m > 1$ and $\sigma^2 < \infty$, then*

(1) $\mathbb{E}[W] = 1$
(2) $\Pr\{W = 0\} = q := \Pr\{Z_n = 0,$ *for some* $n\}$

Proof. To prove (1), we show uniform integrability of the W_n so we may interchange limits and the expectation. It suffices to show the second moments of the W_n are uniformly bounded, which is effortless due to the assumption of finite variance and the formula for the variance given in Proposition 2.5. Indeed we have

$$\mathbb{E}[W_n^2] = \frac{\mathbb{E}[Z_n^2]}{m^{2n}} = \frac{\text{Var}(Z_n) - m^{2n}}{m^{2n}} = \frac{\sigma^2 m^{n-1}(m^n - 1)}{(m-1)m^{2n}} - 1$$
$$= \frac{\sigma^2(1 + m^{-n})}{m^2 - m} + 1.$$

Taking suprema over n,

$$\sup_n \mathbb{E}[W_n^2] = \frac{\sigma^2}{m^2 - m} + 1 < \infty.$$

To prove (2), set r $= \Pr\{W = 0\}$. Then $\mathbb{E}[W] = 1$ implies $r < 1$. But

$$r = \sum_k \Pr\{W = 0 | Z_1 = k\} \Pr\{Z_1 = k\} = \sum_k p_k r^k = f(r)$$

and in the supercritical case, we know the equation $f(s) = s$ has a unique solution in $(0, 1)$, so we must have $r = q$. □

Letting the asymptotic symbol $f \sim g$ denote $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$, we now obtain:

Theorem 3.3. *The Z_n grow geometrically like $m^n W$ for large n. More precisely, $Z_n \sim W m^n$.*

Loosely speaking, on the event of non-extinction that $W \neq 0$, we can interpret the constant factor W as the randomness in the first few generations, after which the population grows geometrically. We can see this via a beautiful visualization by Nils Berglund of a G-W process with a *Binomial*$(3, 0.4)$ offspring distribution, in which the center node is the initial ancestor:

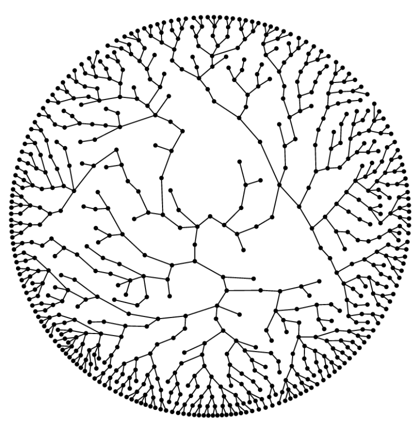

FIGURE 4. A realization of the Galton-Watson process with a *Binomial*$(3, 0.4)$ offspring distribution ([9]).

Remark 3.4. A much deeper version of Theorem 3.2 known as the Kesten-Stigum theorem shows that statements 1 and 2 in Theorem 3.2 are both equivalent to the offspring distribution condition $\mathbb{E}[L\log^+ L] < \infty$, where L denotes a generic random variable with the offspring distribution $\{p_k\}$ (see [2], Chapter 14 for a detailed treatment). Hence, the geometric growth behavior on the event of non-extinction – which Theorem 3.2 establishes only for the case of finite variance – actually occurs precisely when this moment assumption is satisfied.

3.2. **The Subcritical and Critical Cases.** Our analysis so far makes no distinction between the subcritical and critical cases, since both die out with probability one. However, in answering Question 3.1, we will see we can distinguish between the two by the expectation of the extinction time.

Denote the extinction time by

$$\tau := \min\{n \geq 1 : Z_n = 0\}.$$

We will look at the tail $\{\tau > n\}$ of the extinction time for the cases $m < 1$ and $m = 1$.

Theorem 3.5. *Let $\{Z_n\}_{n\geq 0}$ be a Galton Watson process with $m \leq 1$ and $\sigma^2 < \infty$. Then*

(1) If $m < 1$, there exists a $C \in (0, \infty)$ such that $\Pr\{\tau > n\} \sim Cm^n$
(2) (Kolmogorov's Probability Decay Rate) If $m = 1$, then $\Pr\{\tau > n\} \sim \frac{2}{\sigma^2 n}$

Corollary 3.6. *In the subcritical case, the extinction time τ has finite moments of all order. In the critical case, τ has infinite expectation.*

Proof. For any random variable X taking values in $[0, \infty]$, if there exists a constant $c > 0$ such that the tail $\{X > n\}$ is $O(e^{-cn})$, then it has finite moments of all orders, while the tail being $\Omega(1/n)$ implies it has infinite expectation. □

To prove the first part of Theorem 3.5, we will need the following useful method for doing asymptotics: Suppose a_n is a sequence of positive numbers going to infinity, and we want to find a positive function $f(n)$ such that $a_n/f(n)$ converges to a positive constant L. By setting $b_n = a_n/f(n)$ and $\delta_n = (b_{n+1}/b_n) - 1$, a sufficient condition is that $\sum_{n=1}^{\infty} |\delta_n| < \infty$ (for a proof see [4], pg. 13).

Proof of Theorem 3.5, part 1. Observe

$$\Pr\{\tau > n\} = \Pr\{Z_n > 0\} = 1 - \Pr\{Z_n = 0\} = 1 - f_n(0)$$

so the task is to show

$$\lim_{n\to\infty} \frac{1 - f_n(0)}{m^n} := C \in (0, \infty).$$

Set $\delta_n = \frac{1-f_{n+1}(0)}{m(1-f_n(0))} - 1$; we know that it is sufficient to show $\sum_{n=1}^{\infty} |\delta_n| < \infty$. We will show that $\delta_n = O(m^n)$. In the case $m < 1$, we know $f_n(0) \uparrow 1$, so that $O(1 - f_n(0))$ will be small for large n. Thus we will look at the Taylor expansion for f about 1 (note that the radius of convergence of f is not necessarily bigger than 1, but the assumption of finite variance means f is at least twice differentiable at 1, so we may proceed). From Taylor's Theorem with Remainder, we have some $\varepsilon \in (1 - h, 1)$ such that

$$f(1 - h) = f(1) + (-h)f'(1) + \frac{(-h)^2}{2} f''(\varepsilon),$$

or

$$f(1-h) = 1 + (-h)m + O(h^2). \tag{3.7}$$

Substituting $h = 1 - f_n(0)$ into (3.7) gives

$$1 - f_{n+1}(0) = [1 - f_n(0)]m + O([1 - f_n(0)]^2), \tag{3.8}$$

where the precise remainder form is

$$1 - f_{n+1}(0) = [1 - f_n(0)]m - \frac{[1 - f_n(0)]^2}{2} f''(\varepsilon) \tag{3.9}$$

for some $\varepsilon \in (f_n(0), 1)$.

From (3.9) and convexity of f, we see that the first order Taylor expansion is an overestimate, and so (3.8) yields a constant $C \in (0, \infty)$ and N_0 such that for every $n \geq N_0$,

$$[1 - f_n(0)]m - C[1 - f_n(0)]^2 \leq 1 - f_{n+1}(0) \leq [1 - f_n(0)]m \tag{3.10}$$

and we have the upper bound $1 - f_{n+1}(0) \leq [1 - f_n(0)]m$ for all $n \geq 1$.

Dividing (3.10) through by $[1 - f_n(0)]m$, for all $n \geq N_0$ we obtain

$$1 - C[1 - f_n(0)] \leq \delta_n + 1 \leq 1 \tag{3.11}$$

where we have redefined C:=C/m.

Iterating the upper bound $1 - f_{n+1}(0) \leq [1 - f_n(0)]m$ gives $1 - f_n(0) \leq m^n$, and substituting into (3.11), for all $n \geq N_0$ we have

$$0 \leq |\delta_n| \leq Cm^n$$

so that we obtain convergence of $\sum_{n=1}^{\infty} |\delta_n|$ by comparison with the convergent geometric series $\sum_n m^n$. □

We follow the presentation in [2], chapter 9, for the proof of part 2 of Theorem 3.5. The method of proof will be a "proof by comparison" (of generating functions). It proceeds as follows – we will directly verify that the asymptotic holds for a specific example of a Galton Watson process with geometric offspring distribution, in which computing the explicit form of the generating functions turns out to be easy. We then prove an inequality relating the generating functions of an arbitrary critical Galton Watson process (satisfying the assumptions of Theorem 3.5) to the geometric offspring Galton Watson process, which then gives the asymptotic for the arbitrary critical Galton Watson process.

Example 3.12 (Geometric Offspring Galton Watson Process)**.** For $p \in (0, 1)$, define the offspring distribution $\{p_k\}_{k \geq 0}$ of a Galton Watson process by

$$p_0 := p := 1 - q, \quad p_k = q^2 p^{k-1}, \quad k = 1, 2, \ldots$$

This is not exactly the standard geometric distribution, but in the case of $p = 1/2$, we obtain the geometric distribution supported on $\mathbb{N}_0$ for the number of failures before the first success. It is a routine computation to show that $\mathbb{E}[Z_1] = f'(1) = 1$ and $\sigma^2 = f''(1) = \frac{2p}{q}$ (make use of term-by-term differentiation of $f(s)$ on $(0, 1)$). The form of the variance is the key flexibility that will make this Galton-Watson process useful for comparison – based on our choice of p, σ^2 can take any value in $(0, \infty)$. We now compute the first explicit example of generating functions which have been so central to our analysis:

Proposition 3.13. *The generating functions for the geometric offspring Galton Watson process are linear fractional transformations. They take the form*

$$f_n(s) = \frac{np - (np-q)s}{q + np - nps}$$

Proof. Induction on n. For $n = 1$, we have

$$f(s) = \sum_{k=0}^{\infty} p_k s^k = p + \sum_{k=1}^{\infty} q^2 p^{k-1} s^k = p + \frac{(1-p)^2}{p} \sum_{k=1}^{\infty} (ps)^k$$

$$= p + \frac{(1-p)^2}{p}\left[\frac{ps}{1-ps}\right] = \cdots = \frac{p-(p-q)s}{1-ps} \tag{3.14}$$

which aligns with the claim. Now suppose the claim holds for $f_n(s)$. Then

$$f_{n+1}(s) = f(f_n(s)) = f\left(\frac{np-(np-q)s}{q+np-nps}\right).$$

After plugging the input $f_n(s)$ into the base case (3.14) and tediously simplifying, the claim holds. □

Now we verify the asymptotic $\Pr\{\tau > n\} \sim \frac{2}{\sigma^2 n}$ holds for the geometric offspring GW process (for a particular choice of p to be specified below). We have

$$n \Pr\{\tau > n\} = n[1 - f_n(0)] = n\left(1 - \frac{np}{q+np}\right) = \frac{nq}{q+np}$$

Sending n to infinity, we see $\lim_{n\to\infty} n \Pr\{\tau > n\} = \frac{q}{p} = \frac{2}{\sigma^2}$.

The key tool to complete the proof in the general case is the following lemma:

Lemma 3.15 (Spitzer's Comparison lemma)**.** *Suppose that f and g are probability mass functions on $\mathbb{N}_0$, with generating functions $\hat{f}(s) = \sum_{j=0}^{\infty} f(j)s^j$ and $\hat{g}(s) = \sum_{j=0}^{\infty} g(j)s^j$. Assume that*

$$\hat{f}'(1) = \hat{g}'(1) = 1, \quad f''(1) < g''(1) < \infty$$

Then there exists $k, m \in \mathbb{N}$ such that

$$\hat{f}_{n+k}(s) \le \hat{g}_{n+m}(s), \quad \forall\ s \in [0,1], \quad n = 0, 1, 2, \ldots$$

Proof. In this proof, we use the notation f_n (resp. g_n) for the n-fold composition of f (resp. g). Note that all the earlier results about the f_n that we developed in the $f'(1) = 1$ case for the Galton-Watson process apply here - Figure 3 is the picture to look at. We claim $\hat{f} < \hat{g}$ on a one-sided neighborhood about 1. For every $s_0 \in [0,1]$, the first order Taylor expansions about 1 give some $\varepsilon \in (s_0, 1)$ and $\varrho \in (s_0, 1)$ such that

$$\hat{f}(s_0) = \hat{f}(1) + (1-s_0)\hat{f}'(1) + \frac{1}{2}(1-s_0)^2 \hat{f}''(\varepsilon),$$

$$\hat{g}(s_0) = \hat{g}(1) + (1-s_0)\hat{g}'(1) + \frac{1}{2}(1-s_0)^2 \hat{g}''(\varrho).$$

Since $\hat{f}$ and $\hat{g}$ are C^2, we may pick s_0 so that for all $\varepsilon \in [s_0, 1]$ and all $\varrho \in [s_0, 1]$,

$$\hat{f}''(\varepsilon) < \hat{g}''(\varrho)$$

and hence

$$\hat{f}(s) < \hat{g}(s) \tag{3.16}$$

for every $s \in [s_0, 1]$. Since $\hat{f}$ is strictly increasing on $[0,1]$, we may apply $\hat{f}$ to both sides of (3.16) to obtain

$$\hat{f}_2(s) < \hat{f}(\hat{g}(s)) \tag{3.17}$$

for every $s \in [s_0, 1]$.

In the case $\hat{g}'(1) = 1$, we know $\hat{g}(s) \geq s$ for all $s \in [0,1]$, so $s \geq s_0$ implies $\hat{g}(s) \geq s_0$. Thus we can apply the upper bound (3.16) to continue the inequality in (3.17) to obtain

$$\hat{f}_2(s) < \hat{g}_2(s)$$

for all $s \in [s_0, 1]$. As the $\hat{g}_n$ converge upward (in n) on $[0,1]$, $s \in [s_0, 1]$ implies $\hat{g}_n(s) \in [s_0, 1]$ for all n, so we may iterate the above argument to obtain for any $s \in [s_0, 1]$ and for all n,

$$\hat{f}_n(s) \leq \hat{g}_n(s). \tag{3.18}$$

Since $\hat{f}_n(0) \uparrow 1$, there exists $k \in \mathbb{N}$ such that $s_0 < \hat{f}_k(0)$. But $\hat{g}_n(0) \uparrow 1$ as well, so we can find $m > k$ so that $\hat{f}_k(s_0) \leq \hat{g}_m(0)$. All together now – for $s \in [0, s_0]$,

$$s_0 \leq \hat{f}_k(0) \leq \hat{f}_k(s) \leq \hat{f}_k(s_0) \leq \hat{g}_m(0) \leq \hat{g}_m(s).$$

Hence we may apply $\hat{f}_n$ to both sides of the inequality $\hat{f}_k(s) \leq \hat{g}_m(s)$ and then apply (3.18) to obtain the claim for all $s \in [0, s_0]$. Since $\hat{g}$ is strictly increasing and we picked $m > k$, from (3.18) we see the claim also holds on $[s_0, 1]$. □

Remark 3.19. The strongest version of Spitzer's Comparison Lemma replaces our assumption $f''(1) < g''(1) < \infty$ with the slightly weaker $f''(1-) < g''(1-) < \infty$. Notably, our proof's method of second order Taylor approximations is no longer valid since f and g need not be twice differentiable, and the proof is harder.

Proposition 3.20. *Suppose $m = 1$ and $\sigma^2 < \infty$. Then*

$$\frac{1}{n}[\frac{1}{1 - f_n(s)} - \frac{1}{1-s}] \to \frac{\sigma^2}{2}$$

uniformly on $[0,1)$ as $n \to \infty$.

Proof. We have already seens this holds for the geometric offspring distribution generating functions. We bound f between two suitably chosen such functions. Fix $\epsilon > 0$. It suffices to show the limit is bounded – uniformly in s – above by $(1+\epsilon)\frac{\sigma^2}{2}$ and below by $(1-\epsilon)\frac{\sigma^2}{2}$, and then send $\epsilon \to 0$. We seek to compare f to the generating function $g(s) = \frac{p-(2p-1)s}{1-ps}$ of the geometric distribution. Indeed, since we are free to pick the variance, choose p so that $g''(1) = \frac{2p}{1-p} = (1+\epsilon)\sigma^2 > \sigma^2$, and hence the hypotheses to compare f and g in Spitzer are satisfied. We get $k, m \in \mathbb{N}$ such that for any $s \in [0,1)$ and for all n,

$$[1 - f_{n+k}(s)]^{-1} \leq [1 - g_{n+m}(s)]^{-1},$$

so subtracting off $[1-s]^{-1}$ from both sides gives

$$[1 - f_{n+k}(s)]^{-1}] - [1-s]^{-1} \leq [1 - g_{n+m}(s)]^{-1}] - [1-s]^{-1}$$

$$= \frac{(n+m)p}{q} = (n+m)(1+\epsilon)\frac{\sigma^2}{2}.$$

We stress that n is independent of s. Hence,

$$\lim_{n\to\infty} \frac{1}{n}[\frac{1}{1-f_n(s)} - \frac{1}{1-s}] \leq (1+\epsilon)\frac{\sigma^2}{2}. \tag{3.21}$$

Now choose p so that the variance of the geometric offspring Galton-Watson process is $(1-\epsilon)\sigma^2$. From Spitzer we get k' and m' such that for all n we have

$$(k'+n)(1-\epsilon)\frac{\sigma^2}{2} \leq [1-f_{m'+n}(s)]^{-1} - [1-s]^{-1},$$

and hence

$$(1-\epsilon)\frac{\sigma^2}{2} \leq \lim_{n\to\infty} \frac{1}{n}[\frac{1}{1-f_n(s)} - \frac{1}{1-s}]. \tag{3.22}$$

Simultaneously sending $\epsilon \to 0$ in (3.21) and (3.22) gives the claim. □

Proof of Theorem 3.5 Part 2. Plugging in $s = 0$ to the above proposition gives

$$\lim_{n\to\infty} \frac{1}{n[1-f_n(0)]} = \frac{\sigma^2}{2}$$

which is equivalent to $\Pr\{\tau > n\} \sim \frac{2}{\sigma^2 n}$. Hence the proof that the tail of the extinction time decays linearly with n is complete. □

4. Total Progeny

We have seen that both the subcritical and critical Galton-Watson processes die out almost surely. It follows that with probability one, finitely many individuals are born in either case. However, there is a subtle difference in that the expectation of the extinction time is finite in the subcritical case and infinite in the critical case. We now show we can again distinguish the subcritical and critical case by the property that the expected number of individuals in the subcritical case is finite, while the expected number of individuals born in the critical case is infinite. The approach is a fixed point theorem reminiscent of the proof for the extinction probability fixed point theorem (Theorem 2.11), and will allow us to compute explicitly the expected number of individuals born as a function of m.

For convenience, we will count the zero generation ancestor, so that we define the total progeny of a Galton-Watson process $\{Z_n\}$ at the nth stage to be

$$Y_n = 1 + Z_1 + Z_2 + \cdots + Z_n$$

Proposition 4.1. *Let $g_n(s) = \mathbb{E}[s^{Y_n}]$ denote the generating function for Y_n. Then*

$$g_n(s) = sf(g_{n-1}(s)).$$

Proof. Let $n \in \mathbb{N}$. The task is to show

$$\mathbb{E}[s^{Y_n}] = s\mathbb{E}\mathbb{E}[s^{Y_{n-1}}]^{Z_1}.$$

We have

$$\begin{aligned}
\mathbb{E}[s^{Y_n}] &= s\mathbb{E}[s^{Z_1}s^{Z_2}\dots s^{Z_n}] \\
&= s\mathbb{E}\mathbb{E}[s^{Z_1}s^{Z_2}\dots s^{Z_n}|Z_1] \\
&= s\mathbb{E}\mathbb{E}[s^{\sum_{i=1}^{Z_1} 1}s^{\sum_{i=1}^{Z_1} Z_1^{(i)}}\dots s^{\sum_{i=1}^{Z_1} Z_{n-1}^{(i)}}|Z_1] \\
&= s\mathbb{E}\mathbb{E}[s^{\sum_{i=1}^{Z_1} 1+Z_1^{(i)}+\dots+Z_{n-1}^{(i)}}]|Z_1] \\
&= s\mathbb{E}\mathbb{E}[s^{\sum_{i=1}^{Z_1} Y_{n-1}}|Z_1] \\
&= s\mathbb{E}\mathbb{E}[s^{Y_{n-1}}]^{Z_1}
\end{aligned}$$

□

For $s \in (0,1)$, as $g(s) = sf(s)$, and $f(s) < f(1) < 1$, we have $g(s) < s$. Using the fact that f is strictly increasing, we see

$$g_2(s) = sf(g_1(s)) < sf(s) = g_1(s).$$

We can iterate this argument to see for any n and for any $s \in (0,1)$,

$$g_1(s) > g_2(s) > \dots > g_n(s),$$

and hence the g_n converge downward to a limit

$$\rho(s) := \lim_{n\to\infty} g_n(s).$$

Then

$$\rho(s) = \lim_{n\to\infty} sf[g_{n-1}(s)] = sf[\lim_{n\to\infty} g_{n-1}(s)] = sf[\rho(s)] \tag{4.2}$$

so that for fixed $s \in (0,1)$, $\rho(s)$ is a solution to the equation

$$t = sf(t). \tag{4.3}$$

Similar to our analysis with the fixed point equation $t = f(t)$, we can show $\rho(s)$ is the unique solution of (4.3) in $(0,1)$. That there is a unique such solution is immediately apparent if one visualizes a vertically scaled down curve of the graph of $f(t)$ in Figures 1 and 2. A simple geometric argument goes as follows:

Let q be the smallest positive solution of $t = f(t)$ given by Theorem 1. Since the graph of $f(t)$ is a convex curve, the graph of $sf(t)$ is a convex curve, so that it can intersect the line $y = x$ at most twice. But for $t = 0$, $t < sf(t)$, while for $t = q$ and $t = 1$, the inequality is reversed (since $t = f(t)$ implies $t > sf(t)$). Hence, there is only one solution of (4.3) in $(0,1)$, and moreover, this unique solution lies in $(0,q)$. It is clear that this solution is indeed $\rho(s)$ since $\rho(s) < 1$.

Since the PGFs g_n converge pointwise to ρ on $(0,1)$, by the continuity theorem for probability generating functions ([3], section XI.6), we know that $\rho(s)$ is equal to $\sum_{k=0}^{\infty} p(k)s^k$, where the coefficients $p(k)$ are given by $p(k) = \lim_{n\to\infty} \Pr\{Y_n = k\}$. One then sees that $\sum_{k=0}^{\infty} p(k)$ corresponds to the probability Y_n is finite for all n, which is exactly the probability of extinction q. Thus, $\rho(s)$ is a honest probability generating function only in the subcritical and critical cases where $q = 1$. In these two cases, since we know there is an N at which $Z_N = 0$, $\lim_{n\to\infty} \Pr\{Y_n = k\}$ indeed equals $\Pr\{S = k\}$, where $S = 1 + \sum_{n=1}^{\infty} Z_n$ denotes the total progeny. The function $\rho(s)$ now becomes the generating function of the total progeny S.

Theorem 4.4. *The expectation of the total progeny $S = 1 + \sum_{n=1}^{\infty} Z_n$ is given by $\frac{1}{1-m}$ for $m < 1$ and is infinite in the case $m = 1$.*

Proof. Differentiating the generating function identity in (4.2) gives

$$\rho'(1) = f(1) + f'(1)\rho'(1)$$
$$\rho'(1) = 1 + m\rho'(1)$$

and the claim now follows. □

4.1. **Example Computation of the Total Progeny Probabilites.** As an application of the theory of this section, we compute the total progeny probabilites for a particularly simple critical Galton Watson process, characterized by the offspring distribution $p_0 = 1/2$, $p_2 = 1/2$, and $p_k = 0$ for all other $k \in \mathbb{N}$. Even for such a simple process, to compute the total progeny probabilities directly would be prohibitive labor. The generating function for Z_1 is

$$f(s) = \frac{1}{2} + \frac{1}{2}s^2.$$

To solve for the generating function g of the total progeny, we solve equation (4.3), which is the following quadratic in t:

$$t = s[\frac{1}{2} + \frac{1}{2}t^2].$$

Upon solving, the only solution in $(0, 1)$ is

$$g(s) = \frac{1 - \sqrt{1 - s^2}}{s}.$$

We may now use the identity $\sqrt{1 - s^2} = \sum_{n=0}^{\infty} \binom{1/2}{n}(-s^2)^n$ to see that for odd naturals $k = 2n - 1$, $n \in \mathbb{N}$,

$$\Pr\{S = k\} = \binom{1/2}{n}(-1)^{n+1}.$$

For example, the first few probabilities given by the above formula are $\Pr\{S = 1\} = 1/2$, $\Pr\{S = 3\} = 1/8$, and $\Pr\{S = 5\} = 1/16$, which the reader may directly verify.

5. Variants

The single-type offspring Galton-Watson process we have been developing so far is the simplest possible model. In this section, we briefly consider two slightly more complex models. The first is the multi-type Galton-Watson process, in which we allow multiple kinds of offspring. The second is the Galton-Watson process with immigration, in which a random number of individuals who are not offspring of the previous generation enter the population at each generation, and then reproduce according to the offspring distribution.

5.1. **Multi-Type Galton Watson Process.** The Galton-Watson process with k types proceeds as follows: it begins with a single ancestor individual of any one of the k types, but now this individual can reproduce any number of each of the k types, in accordance with its respective offspring distribution. These direct descendants then reproduce independently according to their respective offspring distributions.

Notation: Let $T \subseteq \mathbb{R}^k$ be all vectors whose components are non-negative integers. Let $\{e_i\}_{1\leq i\leq k}$ denote the standard basis of $\mathbb{R}^k$. Let $\mathbf{1}_n$ denote the vector of all ones.

Definition 5.1. The Galton-Watson process with k types is a Markov chain $\{Z_k\}_{k\geq 0}$ on the state space T whose transition law is as follows. Let $f^{(i)}$ denote the generating functions for the direct descendants of an individual of type i, that is

$$f^{(i)}(s_1,\ldots,s_k) = \sum_{(r_1,\ldots,r_k)\in T} p^{(i)}(r_1,\ldots,r_k)s_1^{r_1}\ldots s_k^{r_k}, \quad 1\leq i\leq k$$

where $p^{(i)}(r_1,\ldots,r_k)$ is the probability that an individual of type i has r_1 descendants of type 1, r_2 descendants of type 2, and so on. If $Z_n = (r_1,\ldots r_k)\in T$, then Z_{n+1} is the sum of $r_1+\cdots+r_k$ independent random variables, r_1 of which have generating function $f^{(1)}$, r_2 of which have generating function $f^{(2)}$, and so on.

We always initialize the process with a single individual, which we denote as $Z_0 = e_i$ if that individual is of type i.

Definition 5.2. For $Z_0 = e_i$, the generating function of Z_n is

$$f_n^{(i)}(s_1,\ldots,s_k) = \sum_{(r_1,\ldots,r_k)\in T} p_n^{(i)}(r_1,\ldots,r_k)s_1^{r_1}\ldots s_k^{r_k}$$

where $p_n^{(i)}(r_1,\ldots,r_k)$ is the probability that the nth generation of descendants of a type i individual are comprised of r_1 individuals of type 1, r_2 individuals of type 2, and so on.

Define the vector-valued function $\mathbf{f_n}:\mathbb{R}^k\to\mathbb{R}^k$ by $\mathbf{f_n} := (f_n^{(1)},\ldots,f_n^{(k)})$. The function $\mathbf{f}=\mathbf{f_1}$ contains the entire datum of the population. In the single-type Galton-Watson process, we saw that the generating function f_n of the nth generation descendants was the n-fold composition $f^{\circ n}$ of the direct descendants generating function f. The situation is analogous here:

Proposition 5.3. *The function* $\mathbf{f_n}$ *is the n-fold composition of the function* $\mathbf{f}$.

Proof. It suffices to show $\mathbf{f}_{n+1} = \mathbf{f}[\mathbf{f_n}(\mathbf{s})]$ which may then be iterated to obtain the claim. Componentwise, this means we must check

$$f_{n+1}^{(i)}(\mathbf{s}) = f^{(i)}(f_n^{(1)}(\mathbf{s}),\ldots,f_n^{(k)}(\mathbf{s})).$$

Let $Z_n^{(i)}$ be the number of individuals of type i in the nth generation. By definition, $f_{n+1}^{(i)}(\mathbf{s}) = \mathbb{E}^{\{Z_0=e_i\}}[s_1^{Z_{n+1}^{(1)}}\ldots s_k^{Z_{n+1}^{(k)}}]$, where $Z_{n+1}^{(j)}$ is the number of type j individuals in the $n+1$ generation. To simplify notation, let $\mathbf{s}^{Z_n}$ be shorthand for $s_1^{Z_{n+1}^{(1)}}\ldots s_k^{Z_{n+1}^{(k)}}$.We have

$$f_{n+1}^{(i)}(\mathbf{s}) = \mathbb{E}^{\{Z_0=e_i\}}[\mathbf{s}^{Z_n}] = \mathbb{E}^{\{Z_0=e_i\}}\mathbb{E}[\mathbf{s}^{Z_n}|Z_1]. \tag{5.4}$$

Given $Z_1 = (Z_1^{(1)},\ldots,Z_1^{(k)})$, the distribution of $Z_{n+1}^{(i)}$ is that of the random variable $\sum_{i=1}^{Z_1^{(1)}} Z_{n,1}^{(i)}+\cdots+\sum_{i=1}^{Z_1^{(k)}} Z_{n,k}^{(i)}$, where the notation $Z_{n,j}^{(i)}$ denotes the number of nth generation descendants of type i with initial condition $Z_0=e_j$, $1\leq j\leq k$, and the sums are over *i.i.d.* copies. Substituting into (5.4) and rearranging terms gives

$$f_{n+1}^{(i)}(\mathbf{s}) = \mathbb{E}^{\{Z_0=e_i\}}[(\mathbb{E}^{\{Z_0=e_1\}}\mathbf{s}^{Z_n})^{Z_1^{(1)}}\ldots(\mathbb{E}^{\{Z_0=e_k\}}\mathbf{s}^{Z_n})^{Z_1^{(k)}}]$$
$$= \mathbb{E}^{\{Z_0=e_i\}}[(f_n^{(1)}(\mathbf{s}))^{Z_1^{(1)}}\ldots(f_n^{(k)}(\mathbf{s}))^{Z_1^{(k)}}] = f^{(i)}(f_n^{(1)}(\mathbf{s}),\ldots,f_n^{(k)}(\mathbf{s}))$$

as desired. □

In the single-type case, the recursive structure of the generating functions showed the extinction probability q satisfies the fixed point equation $s = f(s)$. Since the recursive structure of the multi-type process is the same as the single-type case, we have the same result in the multi-type case:

Theorem 5.5. *Let $\{Z_k\}_{k\geq 0}$ be a Galton-Watson process with k types. Then we have k extinction probabilites $q_1, \dots, q_k$ for each of the initial conditions $Z_0 = e_1, \dots, Z_0 = e_k$, which we arrange in the extinction probability vector $\mathbf{q} = (q_1, \dots, q_k)$. Then $\mathbf{q}$ solves the fixed point equation*

$$\mathbf{f}(\mathbf{s}) = \mathbf{s}$$

where $\mathbf{f}$ is the direct sum of the generating functions as defined above.

Proof. As $f_n^{(i)}(0) = \Pr\{Z_n = 0 | Z_0 = e_i\}$, we see $\lim_{n\to\infty} f_n^{(i)}(0) = q_i$, and hence $\lim_{n\to\infty} \mathbf{f}_n(0) = \mathbf{q}$. But then

$$\mathbf{q} = \lim_{n\to\infty} \mathbf{f}_{n+1}(0) = \lim_{n\to\infty} \mathbf{f}[\mathbf{f}_n(0)] = \mathbf{f}[\lim_{n\to\infty} \mathbf{f}_n(0)] = \mathbf{f}(\mathbf{q}).$$

□

The mean number of offspring m in the single-type process was used to distinguish between the cases of extinction with probability one ($m \leq 1$) and a positive probability of infinite growth ($m > 1$). For the multi-type process, there is an analogue of the role the mean m plays for the single-type process under the following condition. If we assume the moment matrix $[M]_{ij} = \mathbb{E}[Z_1^{(j)} | Z_0 = e_i]$ is positive regular, meaning there exists an n such that M^n is strictly positive (each entry $[M^n]_{ij} > 0$), then the Perron-Frobenius theorem guarantees the existence of a maximal eigenvalue ρ. The following theorem, whose proof we omit (see [6], sections II.6 and II.7), shows ρ replaces the role of m for the multi-type process:

Theorem 5.6. *Suppose Z_n is a multi-type Galton-Watson process with positive regular moment matrix M. Also assume the generating function $\mathbf{f}(\mathbf{s})$ does not assume the form $f(s) = As$, for some n by n matrix A of non-negative elements. Let ρ be the maximal eigenvalue guaranteed by the Perron-Frobenius theorem.*

(1) If $\rho \leq 1$, then $\mathbf{q} = \mathbf{1}_n$. If $\rho > 1$, then $\mathbf{q} < \mathbf{1}_n$ (each component is less than 1).
(2) $\lim_{n\to\infty} \mathbf{f}(\mathbf{s}) = \mathbf{q}, \quad \forall \mathbf{s} \in [0, 1)^n$
(3) The solution $\mathbf{q}$ to the fixed point equation $\mathbf{f}(\mathbf{s}) = \mathbf{s}$ is the unique solution in $[0, 1]^n$

5.2. **Galton-Watson Process with Immigration.** We have seen that the standard Galton-Watson process with one-type is unstable, in that it either blows up to infinity or dies out, and all states $Z_n = k$ are transient, for $k \in \mathbb{N}$. However, by introducing an immigration component into the single-type process in which new individuals which are not direct offspring can enter the population, the structure of the process significantly changes as the state $Z_n = 0$ is no longer an absorbing state (as immigrants entering the (n + 1)th generation can "revive" the process). In this section, we show under certain conditions on the offspring distribution and immigration distribution, the Galton-Watson process can have a nontrivial stationary distribution.

The immigration variant works as follows – the process begins as usual with $Z_0 = 1$. But then for every generation $n \geq 1$, a random number I_n of individuals without parents in the (n - 1)th generation are added, which then reproduce as the other individuals. The law for the immigration is the same for each generation and is given by

$$\Pr\{I_n = k\} := q_k, \quad k = 0, 1, 2, \dots$$

We denote the generating function of the number of immigrants added as

$$\beta(s) = \sum_{k=0}^{\infty} q_k s^k.$$

We let f denote the generating function for the number of direct descendants as usual, $f(s) = \sum_{k=0}^{\infty} p_k s^k$. The random variable Z_n counts the total number of individuals in the nth generation, that is, the offspring of individuals in the $(n-1)th$ generation and the immigrants which entered in the nth generation.

Note that the structure of the markov chain Z_n has changed – if $0 < \beta(0) < 1$, then it is possible for Z_n to be irreducible and aperiodic. We will prove existence of a stationary probability distribution for such Z_n.

Let G_n be the generating function for Z_n. We have the following recursive formula for the G_n in terms of β and the f_n:

Proposition 5.7. *Let $n \in \mathbb{N}$. Then*

$$G_n(s) = \beta(s) f_n(s) \prod_{j=1}^{n-1} \beta[f_j(s)].$$

Proof. To count all the different sources for individuals in Z_n, it will be useful to consider the random variable $Z_n^{(m)}$, which denotes the number of individuals in the nth generation who are descendants of immigrants who entered the population in the mth generation. Let $H_n^{(m)}$ be its generating function. Since given I_m, the distribution of $Z_n^{(m)}$ is the sum of I_n independent random variable with the generating function f_{n-m} for $(n-m)th$ generation offspring, we have

$$H_n^{(m)}(s) = \mathbb{E}[s^{Z_n^{(m)}}] = \mathbb{E}\mathbb{E}[s^{Z_n^m} | I_m] = \mathbb{E}[f_{n-m}(s)^{I_m}] = \beta[f_{n-m}(s)] \tag{5.8}$$

Let us denote the number of nth generation offspring of the original zeroth generation ancestor as X_n. Then we can decompose the generating function G_n as

$$G_n(s) = \mathbb{E}[s^{Z_n}] = \mathbb{E}[s^{X_n + Z_n^1 + \cdots + Z_n^{(n-1)} + I_n}]$$

Since everything is independent, the generating functions factor, and combining with (5.8) we get

$$G_n(s) = f_n(s)\beta[f_{n-1}(s)] \dots \beta[f(s)]\beta[s].$$

Recalling that $f_n(s) = f[f_{n-1}(s)]$, we now see

$$G_n(s) = \beta(s) G_{n-1}[f(s)].$$

and iterating this identity gives the desired formula for $G_n(s)$. □

Theorem 5.9. *Suppose that the Markov chain Z_n is irreducible and aperiodic. Also assume that $m = \sum_k kp_k < 1$ and $\lambda = \sum_k kq_k < \infty$. Then there exists a stationary probability distribution $\{\pi_k\}_{k\geq 0}$ for Z_n.*

Proof. The first part of the proof follows [7], page 2. By standard markov chain theory, it suffices to show Z_n is positive recurrent. It is sufficient to show $\lim_{n\to\infty} G_n(0)$ converges to a positive limit, since $G_n(0)$ is the n-step transition probability from state 1 to state 0 of the chain. Plugging in 0 to the formula established in Proposition 5.7,

$$G_n(0) = \beta(0) f_n(0) \prod_{j=1}^{n-1} \beta[f_j(0)].$$

Note that the assumptions imply $\beta(0) > 0$, while we know $f_n(0) \uparrow 1$, so that the $G_n(0)$ converge to a positive limit if and only if

$$\lim_{n\to\infty} \prod_{j=1}^{n} \beta[f_j(0)]$$

converges to a positive limit. Taking logarithms and applying first order taylor expansions, it is sufficient to show

$$\sum_{j=1}^{\infty} 1 - \beta[f_j(0)]$$

converges to a finite limit. By the mean value theorem, $1 - \beta[f_j(0)] < \lambda[1 - f_j(0)]$, so by comparison it suffices to check the convergence of

$$\sum_{j=1}^{\infty} [1 - f_j(0)].$$

But in Theorem 3.5, we established that in the subcritical case $[1 - f_j(0)]$ is $O(m^n)$, and the result follows. □

6. The Galton-Watson Process in Action

We now consider some applications of the Galton-Watson theory developed in this paper to certain processes – since the theory is so simple, this is entirely a matter of properly identifying the Galton-Watson process present.

Example 6.1. The Random Cantor Set - A Galton Watson Process on [0, 1]

We apply supercritical Galton-Watson process theory to verify the existence of a very interesting class of mathematical objects – random fractals. We will look at the random Cantor set. The random Cantor set in $[0, 1]$ takes two parameters, $k \in \mathbb{N}$ and $p \in (0, 1)$, which we denote as

$$C(k, p) = \bigcap_{n=0}^{\infty} C_n$$

defined as follows. Define $C_0 = [0, 1]$, and divide it into k equal intervals

$$[0, 1/k], [1/k, 2/k], \ldots, [(k-1)/k, 1].$$

For each interval, we independently keep it with probability p, and discard it with probability $1 - p$. The union of the kept intervals form C_1. Let Y_1 denote the number of intervals kept in C_1. Split each of these Y_1 pieces into k pieces of length k^{-2}, and again keep each one with probability p, and so on.

Observe that the number of intervals in each generation is a Galton-Watson process with offspring distribution $\mathrm{Binomial}(k, p)$. One might wish to consider standard fractal geometry questions about the set $C(k,p)$ such as computing its Hausdorff dimension. The first order of business is to check that the set is not always the empty set, else the dimension is trivially 0 with probability one. If we choose $kp > 1$, then we know there is a positive probability that we obtain at least one nested sequence of non-empty compacts and so $C(k,p) \neq \emptyset$. In this case, it can be shown that the Hausdorff dimension is $\frac{\log kp}{\log k}$ (for a proof, see [4], Chapter 4). For example, the natural random analog of the classical ternary Cantor set would be given by the process "split in to three subintervals, keep two on average." This would be a $C(3, 2/3)$ random Cantor set, which on the event of being non-empty, has dimension $\frac{\log 2}{\log 3}$ in accordance with the classical ternary Cantor set.

Example 6.2. There's a Galton Watson Process in my Random Walk! ([5], Problem 5)

We give a proof of the familiar fact that simple random walk on the integers is null recurrent via Galton-Watson process theory (assuming Polya's theorem which states that it is recurrent). We also prove a scaling law for the maximum displacement of simple random walk on the integers.

Since null recurrence is a class property, it suffices to show the expected number of steps to go from position 1 to position 0 is infinite.

Let S_n be simple random walk on the integers starting at position 1 ($S_0 = 1$). We denote the time of the first return to the origin (or equivalently, the number of steps taken to go from position 1 to position 0) by

$$T = min\{n \in \mathbb{N} : S_n = 0\}.$$

Random walk is recurrent in one dimension, so we know with probability one T is finite. Let $Z_0 := 1$, and

$$Z_k := \sum_{n=0}^{T-1} \mathbf{1}\{S_n = k, S_{n+1} = k+1\}.$$

That is, Z_k is the number of times the random walker steps from k to $k+1$ before returning to the origin. We claim $\{Z_k\}_{k\geq 0}$ is a Galton-Watson process. That it is a Markov chain is clear. To verify

$$P(i,j) = \Pr\{Z_{n+1} = j \mid Z_n = i\} = \begin{cases} p_j^{*i} & \text{if } i \geq 1, \\ \delta_{0j} & \text{if } i = 0 \end{cases}$$

note that if $Z_n = 0$, the random walker never cross from n to $n+1$ before returning to the origin, which implies it never crossed from $n+1$ to $n+2$. Now suppose $Z_n = i$ for $i \geq 1$. Then the random walker crossed from n to $n+1$ i many times, which gives i copies of the process of the random walker at $n+1$ with the chance to cross from $n+1$ to $n+2$ any number of times before returning to n. Hence,

$$\Pr\{Z_{n+1} = j \mid Z_n = i\} = i \Pr\{Z_1 = j\} = p_j^{*i}.$$

We have $p_0 = 1/2$, since this is the prob the RW steps left to go back to the origin from 1. Similarly, $p_1 = 1/4$, since for the event $\{Z_1 = 1\}$ to occur, we first require the random walker to go right to position 2 from position 1, and then once it returns to position 1 with prob. one, to go left from position 1 to the origin. It is easy to see that $p_n = (\frac{1}{2})^{n+1}$, which we recognize the as the geometric offspring distribution (Example 3.12) with $p = 1/2$. Hence, Z_n is a critical Galton-Watson process.

Observe that $T = 2\sum_{k\geq 0} Z_k$. Recall from the section on total progeny that for a critical Galton-Watson process, the expectation of the total progeny $\sum_{k\geq 0} Z_k$ is infinite, and so we see that $\mathbb{E}[T] = \infty$, that is, the expected number of steps to go from 1 to 0 is infinite.

Now let τ denote the stopping time of Z_n, and let M denote the maximum displacement from the origin obtained by the random walker before its first return to the origin.

From Theorem 3.5 Part 1, one can obtain the following scaling law for the stopping time of critical Galton Watson processes: for all $x > 1$, we have

$$\lim_{n\to\infty} \Pr\{\tau > nx | \tau > n\} = \frac{1}{x}. \tag{6.3}$$

Indeed,

$$\lim_{n\to\infty} \Pr\{\tau > nx | \tau > n\} = \lim_{n\to\infty} \frac{\Pr\{\tau > nx \cap \tau > n\}}{\Pr\{\tau > n\}}$$
$$= \lim_{n\to\infty} \frac{\Pr\{\tau > nx\}}{\Pr\{\tau > n\}} = \frac{1}{x} \lim_{n\to\infty} \frac{nx\Pr\{\tau > nx\}}{n\Pr\{\tau > n\}} = \frac{1}{x}$$

But the stopping time τ of the Galton-Watson process Z_n is the maximum displacement M obtained by the random walker from the origin before its first return! Indeed, the event $\{\tau = n\}$ occurs if and only if $Z_n = 0$ and $Z_{n-1} > 0$, that is, the random walker went from $n-1$ to n at least once, but never from n to $n+1$. Thus we get the scaling law (6.3) for M.

Example 6.4. Waiting Times ([3], pg. 299)

We can apply Galton-Watson process theory to a queuing problem. Suppose there is a server, which can serve one customer at a time. The duration of service is given by the distribution $\{\beta_k\}_{k\geq 0}$, where $\beta_k = \Pr\{\text{service duration is } k\}$. To not overburden the word "time," call points on the time-axis epochs. At epoch 0, the queuing process begins with a single customer at the server, and no customers in line. While the customer is being served, other customers arrive in line. Our model is that customers can arrive only at integral-valued epochs, and can arrive only one at a time. The distribution of customers arriving at an epoch is given by Bernoulli(p), for $p \in (0, 1)$. Suppose the original customer's service duration is m. This means that for epochs $0, 1, \dots, m-1$, they are being served. At the end of epoch $m-1$, their service terminates. At epoch m, if there is a customer in line, then their service immediately begins at this epoch. If there are no customers in line at epoch m, we say the service is interrupted at epoch m. Call the duration of uninterrupted service commencing at epoch 0 the busy period. We use the Galton-Watson process to analyze the duration of the busy period.

We identify a Galton-Watson process as follows: the original customer whose service begins at epoch 0 is the Z_0 ancestor. If they are served up to and including epoch m, then the customers that arrive at epochs $1, \dots, m+1$ are counted as direct descendants of this customer. These individuals comprise the Z_1 generation, and so on. Let f be the generating function of Z_1. If N is the duration of service of the original customer, then $f(s)$ is determined by $X_1 + \cdots + X_N$, where $X_i \sim \text{Bernoulli}(p)$, that is, the probability of k direct descendants given the original customer is served for n epochs is given by the corresponding Binomial(n, p) probability, for $k \leq n$. Let β be the generating function for the service duration,

$\beta(s) = \sum_k \beta_k s^k$. We claim

$$f(s) = \beta(ps + q) \tag{6.5}$$

where $q := 1 - p$. We have

$$f(s) = \sum_k p_k s^k = \sum_k \sum_{n=k}^{\infty} \beta_n \binom{n}{k} p^k q^{n-k} = \sum_n \beta_n \sum_{k=0}^{n} \binom{n}{k} (ps)^k q^{n-k}$$
$$= \sum_n \beta_n (q + ps)^n = \beta(q + ps).$$

Differentiating both sides of (6.5) gives $m = p\sigma$, where $\sigma = \sum_k k\beta_k$ is the expectation of the service duration. Hence, the busy period terminates with probability one if and only if $p\sigma \leq 1$, and for $p\sigma < 1$, the expectation of the total number of customers during the busy period is $\frac{1}{1-p\sigma}$. Clearly the only feasible queues are those with $p\sigma$ well below 1, else congestion will be the order of the day.

Acknowledgments

It is a pleasure to thank my mentor Mark Olson for providing useful references for this paper, supplementing my understanding of basic probability theory, and being highly accessible and supportive. Greg Lawler's lectures on probability and analysis showcased to me a far more intuitive style of doing math than the crisp proofs I had seen in traditional lectures. Thanks to Peter May for organizing the REU.

References

[1] K.B. Athreya and P.E. Ney. Branching Processes. Springer-Verlag Berlin Heidelberg. 1972.
[2] Rabi Bhattacharya, Edward C. Waymire. Random Walk, Brownian Motion, and Martingales. Springer Cham. 2021.
[3] William Feller. An Introduction to Probability Theory and Its Applications. Wiley. 1968.
[4] Gregory F. Lawler. Random Walk and the Heat Equation. American Mathematical Society. 2010.
[5] Steve Lalley. Branching Processes. https://galton.uchicago.edu/~lalley/Courses/312/Branching.pdf
[6] Theodore E. Harris. The Theory of Branching Processes. The RAND Corporation. 1964
[7] E. Seneta. The Stationary Distribution of a Branching Process Allowing Immigration: A Remark on the Critical Case. Journal of the Royal Statistical Society, 1968, Vol. 30, pp. 176 - 179.
[8] Rod Adams. Anniversary - 80 years ago, Leo Szilard envisioned neutron chain reaction. American Nuclear Society. Sept 17, 2013. https://www.ans.org/news/article-1424/anniversary-80-years-ago-leo-szliard-envisioned-neutron-chain-reaction/
[9] Nils Berglund. Martingales and Stochastic Calculus. 2013.