The Galton-Watson Branching Process

Mark Takken, Tristan Liu

October 2020

1 Historical Background

In past times, the decay and extinction of aristocratic surnames has been a point of remark and the reasons for this trend a point of interest. Francis Galton, a 19^{th} century statistician, was among those who proposed that this trend was not due to lower fertility of aristocratic families, but rather due to the ordinary laws of probability. Hence, in 1873, he posed the following question in an issue of the "Educational Times":

> *Let $p_0, p_1, p_2, \dots$ be the respective probabilities that a man has $0, 1, 2, \dots$ sons, let each son have the same probability for sons of his own, and so on. What is the probability that the male line is extinct after r generations, and more generally what is the probability for any given number of descendants in the male line in any given generation?*

Reverend H.W. Watson subsequently proposed a model and solution to this problem, leading to a published paper jointly written with Galton. Watson, however, made an algebraic oversight and falsely concluded that all family names eventually die out with probability 1, even if the average number of male offspring is greater than 1.

Nonetheless, the mathematical model discussed in the Galton-Watson paper, today known as the Galton-Watson Branching Process, has been further developed and applied to scientific areas of study such as nuclear chain reactions and epidemiology. In this paper, we discuss basic properties of the branching process, particularly the distribution of the number of males after a certain number of generations, the expectation and variance of this distribution, extinction probability, extinction time and asymptotic growth.

2 The Model

We define Z_n to be the number of people in the nth generation. Unless otherwise specified, we let $Z_0 = 1$. For each individual i in generation n, we let $\xi_i^{(n)}$ be a random variable distributed over the nonnegative integers representing the number of children he has. It follows that

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^n.$$

The $\xi_i^{(n)}$s are independent and identically distributed according to a distribution ξ, which we assume has a finite expected value μ. Define p_k to equal $P(\xi = k)$.

We will discard the case where ξ is a constant distribution (i.e. for some k, $p_k = 1$), because in this case, the process is completely deterministic and hence uninteresting. We will also discard the case where $P(\xi \leq 1) = 1$ because in this case, the $P(Z_n) = 1$ is simply p_1^n.

3 Probability Generating Functions

Definition 3.1. It will be very useful to define a probability generating function for the distribution ξ. We thus define this generating function to be:

$$f(x) = \mathbb{E}(x^\xi) = \sum_{k=0}^{\infty} x^k p_k.$$

Definition 3.2. Let $f_n(x)$ be the probability generating function for Z_n, that is,

$$f_n(x) = \mathbb{E}(x^{Z_n}) = \sum_{k=0}^{\infty} x^k P(Z_n = k),$$

with $f_0(x) = x$. We seek to express f_n in terms of f.

Lemma 3.3. *Suppose we let the root value Z_0 be equal to r instead of 1, for $r \in \mathbb{Z}^+$. Then $\mathbb{E}(x^{Z_n}|Z_0 = r)$, that is, the generating function which expresses the distribution of $Z_n|(Z_0 = r)$, is equal to $f_n(x)^r$.*

Proof. We compute:

$$\begin{aligned}
\mathbb{E}(x^{Z_n}|Z_0 = r) &= \sum_{k\in\mathbb{N}} x^k P(Z_n = k|Z_0 = r) \\
&= \sum_{k\in\mathbb{N}} x^k \sum_{\substack{q_1,q_2,\ldots,q_r\in\mathbb{N} \\ q_1+q_2+\cdots+q_r=k}} \prod_{q=q_1,\ldots,q_r} P(Z_n = q|Z_0 = 1) \\
&= \Big(\sum_{q\in\mathbb{N}} x^q P(Z_n = q|Z_0 = 1)\Big)^r \\
&= f_n(x)^r.
\end{aligned}$$

$\square$

Theorem 3.4. *We have the recurrence $f_{n+1}(x) = f(f_n(x))$.*

Proof. We have:

$$\begin{aligned}
f_{n+1}(x) = \mathbb{E}(x^{Z_{n+1}}) &= \sum_{k=0}^{\infty} \mathbb{E}(x^{Z_{n+1}} | Z_1 = k) P(Z_1 = k) \\
&= \sum_{k=0}^{\infty} \mathbb{E}(x^{Z_n} | Z_0 = k) P(Z_1 = k) \\
&= \sum_{k=0}^{\infty} f_n(x)^k p_k \\
&= f(f_n(x)).
\end{aligned}$$

□

The following corollary immediately follows.

Corollary 3.5. *We have $f_n = f^n$.*

We continue with some basic properties of the generating function f, which we will make us of in the following sections.

Lemma 3.6. *The generating function f satisfies the following properties:*

1. *f is continuous on $[0,1]$ and twice differentiable on $(0,1)$.*
2. *$f'(x) = \mathbb{E}(\xi x^{\xi-1})$ and $f''(x) = \mathbb{E}(\xi(\xi-1)x^{\xi-2})$.*
3. *f is increasing and convex.*
4. *f is differentiable from the left at $x = 1$ with $f'(1) = \mu$.*

Proof. The first statement easily follows from the fact that f is a power series. We proceed to proving the second statement; evaluating $f'(x)$ gives

$$\sum_{k=0}^{\infty} k x^{k-1} p_k = \mathbb{E}(\xi x^{\xi-1}),$$

and evaluating $f''(x)$ gives

$$\sum_{k=0}^{\infty} k(k-1) x^{k-2} p_k = \mathbb{E}(\xi(\xi-1)x^{\xi-2}).$$

The first and second derivatives are clearly positive, so f is convex and increasing. Evaluating $f'(1)$ gives $\sum_{k=0}^{\infty} k p_k = \mathbb{E}(\xi) = \mu$. □

4 Expected Value and Variance

Theorem 4.1. $\mathbb{E}(Z_n) = \mu^n$

Proof. We know from Theorem 3.4 that $f_{n+1} = f(f_n)$. Differentiating (from the left) both sides at $x = 1$ using the chain rule and applying Lemma 3.6 yields:

$$\mathbb{E}(Z_{n+1}) = f'_{n+1}(1) = f'(f_n(1))f'_n(1) = f'(1)f'_n(1) = \mu \cdot \mathbb{E}(Z_n).$$

Since $\mathbb{E}(Z_1) = \mu$ by definition, the theorem follows by induction. □

Definition 4.2. We define $\tau = \min\{n : Z_n = 0\}$ to be the extinction time.

Corollary 4.3. *We have that $P(n < \tau) < \mu^n$.*

Proof. Since the event $\{n < \tau\}$ coincides with the event $\{Z_n \geq 1\}$ and we assume $P(Z_n \leq 1) < 1$, we have

$$P(n < \tau) = P(Z_n \geq 1) < \mathbb{E}(Z_n) = \mu^n.$$

□

We will prove a stronger result in section 6. We now proceed to calculate the variance, but first we need a lemma.

Lemma 4.4. *We have that $\mathbb{E}(Z_n^2) = f''_n(1) + \mu^n$.*

Proof. We calculate:

$$\begin{aligned}
\mathbb{E}(Z_n^2) &= \sum_{k=0}^{\infty} k^2 P(Z_n = k) \\
&= \sum_{k=0}^{\infty} k(k-1)P(Z_n = k) + \sum_{k=0}^{\infty} kP(Z_n = k) \\
&= f''_n(1) + \mathbb{E}(Z_n) = f''_n(1) + \mu^n
\end{aligned}$$

□

Theorem 4.5. *If $\sigma^2 :=$ Variance(ξ) is finite, then the variance of Z_n is given by $n\sigma^2$ if $\mu = 1$ and $\frac{\sigma^2\mu^{n-1}(\mu^n-1)}{\mu-1}$ otherwise.*

Proof. We have the recurrence

$$\begin{aligned}
f''_{n+1}(1) = (f'(f_n) \cdot f'_n)'(1) &= f''_n(1)f'(f_n(1)) + f''(f_n(1))f'_n(1)^2 \\
&= f'(1)f''_n(1) + f''(1)f'_n(1)^2.
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
\text{Variance}(Z_n) &= \mathbb{E}(Z_n^2) - \mathbb{E}(Z_n)^2 \\
&= f''_n(1) + \mu^n - \mu^{2n} \qquad \text{(by the previous lemma)} \\
&= f'(1)f''_{n-1}(1) + f''(1)f'_{n-1}(1)^2 + \mu^n - \mu^{2n} \\
&= f'(1)^2 f''_{n-2}(1) + f''(1)(f'_{n-1}(1)^2 + f'(1)f'_{n-2}(1)^2) + \mu^n - \mu^{2n} = \\
&\dots \\
&= f'(1)^n f''_0(1) + f''(1)\big(f'_{n-1}(1)^2 + f'(1)f'_{n-2}(1)^2 + \cdots + f'(1)^{n-1}f'_0(1)\big) + \mu^n - \mu^{2n} \\
&= f''(1)(\mu^{2n-2} + \mu^{2n-3} + \cdots + \mu^{n-1}) + \mu^n - \mu^{2n} \\
&= (\sigma^2 + \mu^2 - \mu)\mu^{n-1}(1 + \mu + \mu^2 + \cdots + \mu^{n-1}) + \mu^n - \mu^{2n}.
\end{aligned}$$

If $\mu = 1$, then this expression is equal to $n\sigma^2$. Otherwise, it is equal to:

$$\begin{aligned}
&= (\sigma^2 + \mu^2 - \mu)\mu^{n-1}\frac{\mu^n - 1}{\mu - 1} + \mu^n - \mu^{2n} \\
&= \frac{\sigma^2\mu^{n-1}(\mu^n - 1)}{\mu - 1} + \frac{\mu^2 - \mu}{\mu - 1}\mu^{n-1}(\mu^n - 1) + \mu^n - \mu^{2n} \\
&= \frac{\sigma^2\mu^{n-1}(\mu^n - 1)}{\mu - 1}.
\end{aligned}$$

This proves the theorem. In fact, higher moments of Z_n (such as $\mathbb{E}(Z_n^3)$), if they are finite, can be obtained through a similar calculation. □

5 Probability of extinction

Definition 5.1. Let $\eta = P(Z_n = 0 \mid n \in \mathbb{Z}^+)$ be the probability of extinction.

The ultimate goal of this section is to determine how the value of μ affects whether or not $\eta = 1$, that is, whether or not extinction is certain.

Lemma 5.2. *Both η and 1 are solutions of the equation $p = f(p)$ and any other solution $\psi \in [0, 1]$ satisfies $\eta \le \psi \le 1$.*

Proof. Because $f_n(0) = P(Z_n = 0)$, we have $\lim_{n\to\infty} f_n(0) = \eta$. It follows that since $f_{n+1}(0) = f(f_n(0))$ and f is continuous, taking the limit as n goes infinity gives $\eta = f(\eta)$. That $1 = f(1)$ simply follows from the fact that the coefficients of f must add up to 1.

Suppose ψ is also a fixed point of f. Because f is increasing, $f(0) \le f(\psi) = \psi$. Inductively, if $f_n(0) \le \psi$, taking f of both sides gives $f_{n+1}(0) \le f(\psi) = \psi$. Thus, for all n, $f_n(0) \le \psi$, and, taking the limit $n \to \infty$, $\eta \le \psi$. □

Remark 5.3. In fact, by the convexity of f, there are at most two fixed points of f, so η and 1 with $\eta \le 1$ are the only solutions to $p = f(p)$.

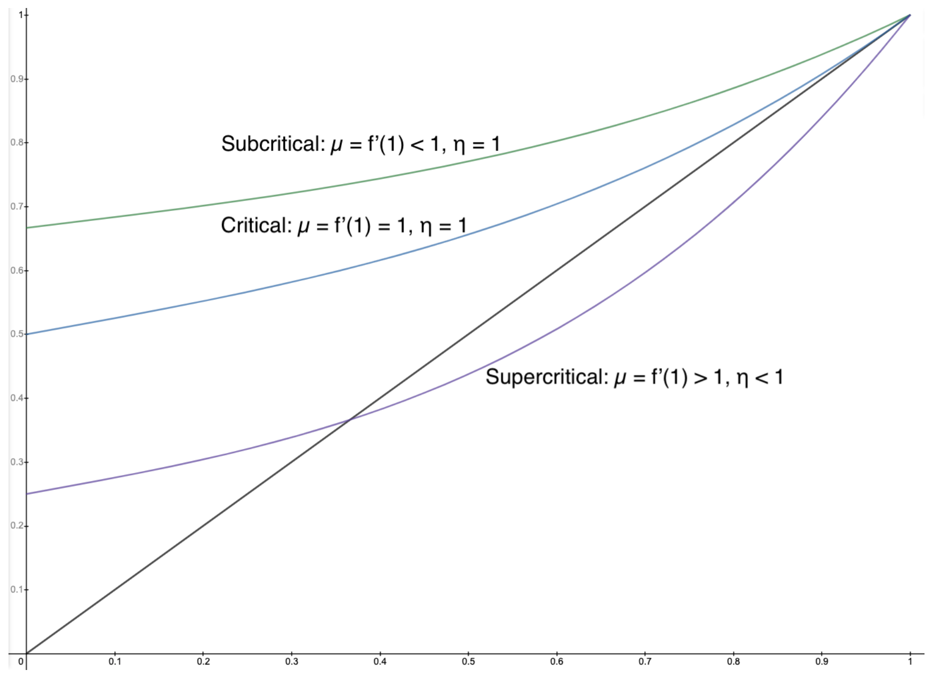
Theorem 5.4. *We have the following:*

- *If $\mu < 1$, $\eta = 1$. (subcritical case)*
- *If $\mu = 1$, $\eta = 1$. (critical case)*
- *If $\mu > 1$, $\eta < 1$. (supercritical case)*

Proof. Consider the function $g(x) = f(x) - x$, and suppose that $\eta < 1$. Then $g(\eta) = 0 = g(1)$, so there exists by the Mean Value Theorem $c \in (\eta, 1)$ such that $g'(c) = 0$ or $f'(c) = 1$. Because f is strictly convex and $c < 1$, $1 = f'(c) < f'(1) = \mu$. Taking the contrapositive, we have that if $\mu \le 1$, then $\eta = 1$.

Now we consider the case where $\mu > 1$. Note that $f(0) = p_0$. If $p_0 = 0$, then clearly extinction is impossible, so $\eta = 0 < 1$. If $p_0 > 0$, then $f(0) - 0 > 0$. Because $f'(1) > 1$ and $f(1) = 1$, there exists some t close to 1 such that $f(t) < t$. Because we have $f(0) - 0 > 0$ and $f(t) - t < 0$, we can find a solution to $f(x) - x$ by the Intermediate Value Theorem, so there must be a fixed point other than 1, which must be η. □

The following graph illustrates the situation pictorially.



6 Extinction Time and Growth Rate

In the critical and subcritical case, the population goes extinct with probability one. The logical question to ask is how long it takes to go extinct. In the supercritical case, when the population doesn't die out, we might ask how it behaves asymptotically.

Theorem 6.1. *Suppose the variance is finite. In the subcritical case, there exists a constant C such that $\mathbb{P}(\tau > n) \sim C\mu^n$. In the critical case, $\mathbb{P}(\tau > n) \sim \frac{2}{\sigma^2 n}$.*

Proof. To begin, recall that $\mathbb{P}(\tau > n) = \mathbb{P}(Z_n > 0) = 1 - f_n(0)$ and that the $f_n(0)$ converge to 1.

We begin with the subcritical case. For large n, $f_n(0)$ is close to 1, so we can approximate $1 - f_{n+1}(0)$ with the first-order Taylor series (i.e. $f(1-x) = 1 - \mu x + O(x^2)$, with $O(x^2)$ positive because f is concave-up) to get

$$\begin{aligned} 1 - f_{n+1}(0) &= 1 - f\big(1 - \big(1 - f_n(0)\big)\big) \\ &= 1 - \Big(1 - \mu\big(1 - f_n(0)\big) + O\big(\big(1 - f_n(0)\big)^2\big)\Big) \\ &= \mu(1 - f_n(0)) - O\big((1 - f_n(0))^2\big). \end{aligned}$$

We see that the first-order Taylor series is an overestimate. Thus for large n, there exists a constant $C > 0$ such that

$$\mu(1 - f_n(0)) - C(1 - f_n(0))^2 \le 1 - f_{n+1}(0) \le \mu(1 - f_n(0)).$$

Note that iterating the upper bound gives us $1 - f_n(0) = P(\tau > n) \le \mu^n$. Alternatively, we can recall a more strict result from Section 4. Dividing by $\mu(1 - f_n(0))$ and redefining C/μ as

C, we get $1 - C(1 - f_n(0)) \leq \frac{1-f_{n+1}(0)}{\mu(1-f_n(0))} \leq 1$. Since $1 - f_n(0) \leq \mu^n$, we equivalently have

$$1 - C\mu^n \leq \frac{\mu^{-(n+1)}(1 - f_{n+1}(0))}{\mu^{-n}(1 - f_n(0))} \leq 1$$

.

This means that the terms $\frac{\mu^{-(n+1)}(1-f_{n+1}(0))}{\mu^{-n}(1-f_n(0))}$ get exponentially close to 1, so $\sum_{n=0}^{\infty} \big(1 - \frac{\mu^{-(n+1)}(1-f_{n+1}(0))}{\mu^{-n}(1-f_n(0))}\big)$ converges. By Weierstrass's Theorem on the convergence of products [1],

$$\prod_{k=0}^{\infty} \frac{\mu^{-(k+1)}(1 - f_{k+1}(0))}{\mu^{-k}(1 - f_k(0))} = \lim_{n\to\infty} \frac{1 - f_n(0)}{\mu^n}$$

converges as well, so there exists some constant C such that $1 - f_n(0) \sim C\mu^n$.

We now consider the critical case, that is, where $\mu = 1$. We approximate using the second-order Taylor series expansion, which gives us

$$\begin{aligned}1 - f_{n+1}(0) &= 1 - f_n(0) - \frac{f''(1)}{2}(1 - f_n(0))^2 + O((1 - f_n(0))^3) \\ &= 1 - f_n(0) - \frac{\sigma^2}{2}(1 - f_n(0))^2 + O((1 - f_n(0))^3). \quad \text{(recalling Lemma 4.4)}\end{aligned}$$

We let $y_n = \frac{1}{1-f_n(0)}$:

$$\begin{aligned}y_{n+1} &= y_n \cdot \frac{1}{1 - \frac{\sigma^2}{2} \cdot \frac{1}{y_n} + O\big(\frac{1}{y_n^2}\big)} \\ &= y_n \cdot \left(1 + \frac{\sigma^2}{2y_n} + O\Big(\frac{1}{y_n^2}\Big)\right) \\ &= y_n + \frac{\sigma^2}{2} + O\Big(\frac{1}{y_n}\Big)\end{aligned}$$

Iterating on n gives us:

$$\begin{aligned}y_n &= y_{n-1} + \frac{\sigma^2}{2} + O\Big(\frac{1}{y_{n-1}}\Big) \\ &= y_{n-2} + 2 \cdot \frac{\sigma^2}{2} + O\Big(\frac{1}{y_{n-1}}\Big) + O\Big(\frac{1}{y_{n-2}}\Big) = \\ &\dots \\ &= 1 + n \cdot \frac{\sigma^2}{2} + \sum_{i=0}^{n-1} O\Big(\frac{1}{y_i}\Big)\end{aligned}$$

[1] If $\sum_{n=1}^{\infty} |a_n| < \infty$, then $\prod_{n=1}^{\infty}(1 + a_n)$ converges.

Now, since from above we know $y_n \geq \frac{n\sigma^2}{2} \implies \frac{1}{y_n} \leq \frac{2}{n\sigma^2}$ we can bound $1 + \sum_{i=0}^{n-1} O\Big(\frac{1}{y_i}\Big)$ in the last line as

$$\begin{aligned} 1 + \sum_{i=0}^{n-1} O\Big(\frac{1}{y_i}\Big) &= 1 + O\Big(\frac{1}{y_0}\Big) + O\Big(\sum_{i=1}^{n-1} \frac{2}{i\sigma^2}\Big) \\ &= O(1) + O\Big(\sum_{i=1}^{n-1} \frac{1}{i}\Big) \\ &= O(\log(n)). \end{aligned}$$

Therefore, we have:

$$y_n = n \cdot \frac{\sigma^2}{2} + O(\log(n)) = n \cdot \frac{\sigma^2}{2} + o(n) \implies y_n \sim n \cdot \frac{\sigma^2}{2}.$$

□

Theorem 6.2. *The sequence $W_n = \frac{Z_n}{\mu^n}$ converges to some random variable W with probability 1. If $\mathbb{E}(Z_1^2) < \infty$, then $\mathbb{E}(W) = 1$ and Variance$(W) = \frac{\sigma^2}{\mu^2 - \mu}$, and $W > 0$ if and only if $Z_n \to \infty$.*

Proof. We have that $\mathbb{E}(Z_{n+1}|Z_n) = \mu Z_n$, and thus that $\mathbb{E}(Z_{n+k}|Z_n) = \mu^k Z_n$. Dividing, we then have $\mathbb{E}(W_{n+k}|W_n) = W_n$. Thus the sequence of W_n is a nonnegative martingale, and by the Martingale Convergence Theorem [2], it converges (almost surely) to a random variable W with finite expectation.

We proceed to compute the expectation and variance of W_n:

$$\begin{aligned} &\mathbb{E}(Z_n) = \mu^n \implies \mathbb{E}(W_n) = 1, \\ &\mathbb{E}(Z_n^2) = \mu^{2n} + \frac{\sigma^2 \mu^{n-1}(\mu^n - 1)}{\mu - 1} \implies \mathbb{E}(W_n^2) = 1 + \frac{\sigma^2}{\mu^2 - \mu}\Big(1 - \frac{1}{\mu^n}\Big) \\ &\implies \text{Variance}(W_n) = \frac{\sigma^2}{\mu^2 - \mu}\Big(1 - \frac{1}{\mu^n}\Big). \end{aligned}$$

Taking the limit as n approaches infinity yields:

$$\begin{aligned} \mathbb{E}(W) &= 1, \\ \text{Variance}(W) &= \frac{\sigma^2}{\mu^2 - \mu}. \end{aligned}$$

We now show that $P(W = 0|Z_n \to \infty) = 0$. Let $\eta^* = P(W = 0)$. We consider the quantity $P(W = 0|Z_1 = k)$ and define W^k to be the limit as $n \to \infty$ of the ratio of the number of offspring of person k in the first generation to μ^n. In order for the event $\{W = 0|Z_1 = k\}$ to occur, each of the W^k must be equal to 0. In addition, since each W^k is equal to $\frac{W}{\mu}$, we have $P(W^k = 0) = P(W = 0)$. It follows that

$$\begin{aligned} &P(W = 0|Z_1 = k) = \prod_k P(W^k = 0) = P(W = 0)^k \\ &\implies \eta^* = P(W = 0) = \sum_{k=0}^{\infty} p_k P(W = 0)^k = f(\eta^*). \end{aligned}$$

[2] A nonnegative supermartingale W_n converges almost surely to some random variable W_∞, which is almost surely finite.

Thus η^* is a fixed point of f, and since it is not equal to 1 (because the expectation of W is positive and its variance is finite), it is equal to η, the extinction probability. Since extinction implies $W = 0$, this means that extinction and $W = 0$ are the same event. Therefore,

$$P(W = 0 | Z_n \to \infty) = P(Z_n \to 0 | Z_n \to \infty) = 0.$$

We have just shown that if $Z_n \to \infty$, then $W > 0$. Since the converse is obvious, we have $W > 0$ if and only if $Z_n \to \infty$, as desired.

$\square$

Sources Used

1. Watson, H. W., and Francis Galton. "On the Probability of the Extinction of Families." The Journal of the Anthropological Institute of Great Britain and Ireland, vol. 4, 1875, pp. 138–144. JSTOR, www.jstor.org/stable/2841222.

2. Lalley, Steve. "Branching Processes." Statistics 312: Stochastic Processes, The University of Chicago. Lecture Notes. http://galton.uchicago.edu/ lalley/Courses/312/Branching.pdf.

3. Harris, Theodore Edward. "The Theory of Branching Process." Springer-Verlag. 1964. https://www.rand.org/pubs/reports/R381.html.