On The Galton Watson Process

Kavya Venturpalli

June 2024

1. Abstract

In this paper we explain the Galton-Watson process and the mathematical concepts behind it. Why did it originate? How can I apply it? What are the prerequisites I need to know to understand branching processes's and generating functions? This paper answers questions like that and similar ones.

2. Background

A branching process is a sequence of random variables, but what is a random variable? A random variable is the numerical value determined by the outcome of a random event.

2.1 Conditional Expectation

The expected value of a random variable is the average that random variable is expected to take on, knowing what the expected value, certain constraints can be put on them. That is Conditional Expectation.

$$E[X \mid Y]$$

For two random variables X and Y, the conditional expectation is given above. It denotes the expected value of some random variable X when the random variable Y is known.

2.2 Law of Total Expectation

The expected value of a random variable is the average that random variable is expected to take on, knowing what the expected value is we can It states that the expected value of a random variable can be obtained by taking the expected value of its conditional expectation with respect to another random variable. Letting X be a random variable and $\mathcal{G}$ be an outcome of another random variable Y. The Law of Total Expectation can be written as

$$E[X] = E[E[X \mid \mathcal{G}]]$$

The Law states that to find the expected value of a random variable X you can first find the expected value of a random variable X given the variable Y, and use that to find the expected value of X. This is just applying the property of conditional expectation to larger expected value problems.

2.3 Recurrence Relations

A recurrence relation provides a way to compute the terms of a sequence based on the previous terms, for a branching process this is very useful as we are calculating the next generation based on the previous one. The standard form of a recurrence relation can be seen as:

$$a_n = f(a_{n-1}, a_{n-2}, \dots, a_{n-k})$$

where f is some function that defines how the term a_n relates to the previous k terms

2.4 Probability Generating Functions

PGF's are very useful in terms of branching processes as they take the distributions of random variables and turn them into independent functions. The $G_X(s)$ is the probability-generating function of a discrete random variable X which only takes positive values

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} P(X = k)s^k$$

where $P(X = k)$ is the probability that X takes the value k. There are some properties of PGF's that need to be taken into account. What happens when s is equal to 1?

$$G_X(1) = \sum_{k=0}^{\infty} P(X = k) = 1$$

It is established that the sum of all probabilities is 1. What about finding the mean? The expected value $E[X]$ can be found taking the derivative of the generating function.

$$E[X] = G'_X(1)$$

There is also the property of variance which is how spread out the data is from the average. So what if we want to find that? The second and first derivative can be used.

$$\text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$$

What if there are two independent random variables X and Y, the probability generating function of their sum $Z = X + Y$ is the sum of their PGF's.

$$G_Z(s) = G_X(s) \cdot G_Y(s)$$

2.5 Independent and Identically Distributed

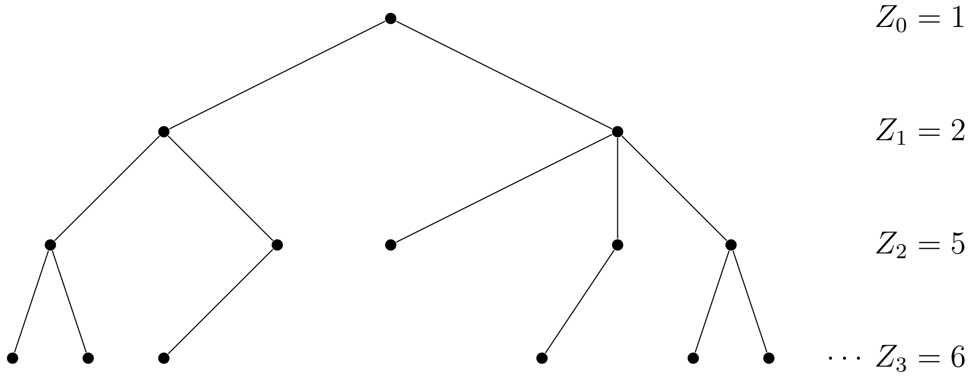
This term appears a lot in stochastic processes. Especially talking about random variables. An independent random variable is one where the events before do not have any effect of the outcome. The identically distributed part means that all these random variables have the same probability distribution. Consider tossing a fair coin n times. Let X_i be the outcome of the i-th toss, where $X_i = 1$ for heads and $X_i = 0$ for tails. The random variables $X_1, X_2, \ldots, X_n$ are i.i.d., each with the same Bernoulli distribution (a distribution with only two outcomes, 0 and 1.):

$$P(X_i = 1) = 0.5, \quad P(X_i = 0) = 0.5$$

3. Introduction

In this paper we do exactly what the title implies, explain the Galton-Watson Process. The process originally appeared in London when many Victorians were concerned about aristocratic last names going extinct. Francis Galton published a question regarding the distribution of surnames. "Their law of population is such that, in each generation, a_0 percent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 has two; and so on up to a5 who have five. Find what proportion of their surnames will have become extinct after r generations, and how many instances there will be of the surname being held by m persons." To this Reverend Henry William Watson replied with a solution. Together Galton and Watson created a simple mathematical model for the propagation of family names. Generalizations of the extinction probability formulas below played a role in the calculation of the critical mass of fissionable material needed for a chain reaction. Galton-Watson processes continue to play a fundamental role in both the theory and applications of stochastic processes.The definition below:

Assume that we have a population of individuals each of which produces a random number of offspring according to a probability distribution $a = (a_0, a_1, a_2, ...)$ An individual gives birth to k children with a probability a_k for $k \geq 0$ independent of other individuals. This makes a a probability distribution.



According to the diagram above Z_n is the number of individuals in the $n-th$ generation. Since we are assuming that $Z_0 = 1$, the sequence $Z_0, Z_1, ...$ is a *branching process*. A branching process is also a Markov chain as the size of a generation depends on the size of a previous generation. [**mishra2009**]

Definition 3.1

A Galton-Watson Process is a stochastic process X_n which evolves according to the formula $X_0 = 1$, X_0 has to have at least 1 offspring in order for reproduction otherwise the whole lineage would die out.

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}$$

We have X which represents the population size of a certain generation. More specifically X_{n+1} denotes the size of the (n+1)-th generation, determined by the offspring distribution of the individuals in the n-th generation. This formula represents the size of the next generation X_{n+1} in a Galton-Watson process. $X_n$is the number of individuals in the n-th generation. $\xi_j^{(n)}$ is the number of offspring produced by the j-th individual in the n-th generation, which are independent and identically distributed random variables. The sum indicates that the total number of individuals in the $(n+1)$-th generation is the sum of the offspring produced by each individual in the n-th generation.

A **generation** refers to a stage in a population model, more precisely X_0 notes some amount of individuals in that generation. Each individual in Generation 0 produces a random number of offspring according to a given probability distribution. The total number of offspring forms Generation 1, denoted X_1. This process continues indefinitely.

4. Mean Generation Size

In a branching process, the size of the $n-th$ generation is the sum of the total number of offspring of the parents in the generation before.

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_i,$$

X_i denotes the number of children born to the $i-th$ person in the (n-1) generation. Similar to our sequence of $Z_0, Z_1, ...$ the sequence $X_1, X_2, ...$ is also a branching process which is independently and identically distributed sequence with the common distribution a. Z_{n-1} is independent of X_i meaning that the number of individuals in the $(n-1)-th$ generation will not have any influence on the number of children an individual i will have in the $n-th$ generation.

Example 4.1

$$\mu = \sum_{k=0}^{\infty} k a_k$$

Letting μ be equal to the sum above $\sum_{k=0}^{\infty} k a_k$, μ is the mean of the offspring distribution. To find the mean of the size of the $n-th$ generation $E(Z_n)$ condition on Z_{n-1}. This statement means that to find the expected value of individuals in the $n-th$ generation $E(Z_n)$ consider the condition that we already know the number of individuals in the $(n-1)-th$ generation Z_{n-1}. By the law of total expectation (2.1)

$$\begin{aligned} E\left(Z_n\right) &= \sum_{k=0}^{\infty} E\left(Z_n \mid Z_{n-1}=k\right) P\left(Z_{n-1}=k\right) \\ &= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{Z_{n-1}} X_i \mid Z_{n-1}=k\right) P\left(Z_{n-1}=k\right) \end{aligned} \tag{1}$$

Finding the overall expected value $E(Z_n)$, considering all arbitrary sizes of k of the earlier generation Z_{n-1} and the sum over the expected values of Z_n given each k. In the second part of the equation where $Z_n = \sum_{i=1}^{k} X_i$ where X_i is the number of children of an $i-th$ individual in the (n-1)-th generation. Substituting Z_n with $\sum_{i=1}^{k} X_i$ This equation shows that the expected number of individuals in the n-th generation, given that there are k individuals in the $(n-1)$ generation, is the sum of the expected number of children X_i for each of the k individuals.

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{k} X_i \mid Z_{n-1} = k\right) P\left(Z_{n-1} = k\right) \\
&= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{k} X_i\right) P\left(Z_{n-1} = k\right) \qquad (2) \\
&= \sum_{k=0}^{\infty} k\mu P\left(Z_{n-1} = k\right) = \mu E\left(Z_{n-1}\right),
\end{aligned}
$$

X_i are i.i.d of random variables which are independent of Z_{n-1}. This makes the expected value of the sum of random variables the sum of each of the expected values added up. $E(X_i) = \mu$ where μ is the average of the offspring distribution, also where k represents the number of individuals in the $(n-1)-$ th generation. Substituting back in the sum μ can be factored out because it is a constant. The expected value of $Z_(n-1)$ is $\sum_{k=0}^{\infty} kP\left(Z_{n-1} = k\right)$ so the expected value of Z_n is the equal to the expected value of Z_{n-1} multiplied by the mean of the offspring distribution.

$$E(Z_n) = \mu E(Z_{n-1})$$

Repeating this relation,

$$E\left(Z_n\right) = \mu E\left(Z_{n-1}\right) = \mu\left(\mu E\left(Z_{n-2}\right)\right)$$

$$= \mu^2 E\left(Z_{n-2}\right) = \cdots = \mu^n E\left(Z_0\right)$$

Assuming that the size of the Z_0 generation is equal to 1,

$$E(Z_n) = \mu^n$$

The expected value of Z_n is the mean of the offspring to the power of the n-th generation.

5. Extinction and other cases

$$\lim_{n\to\infty} E(Z_n) = \lim_{n\to\infty} \mu^n = \begin{cases} 0, & \text{if } \mu < 1, \\ 1, & \text{if } \mu = 1, \\ \infty, & \text{if } \mu > 1. \end{cases}$$

Generations can become very large and n can expand all the way towards infinity. There are long-term cases that need some way of being measurable. A common factor is the mean which can be categorized into intervals, a subcritical case, a critical case, and a supercritical case. The subcritical case is when $\mu < 1$, meaning that the average number of children per individual is less than one. This can eventually lead to the average number of children declining exponentially hitting 0 at some point in time. Making the probability of extinction very high. The critical case is when $\mu = 1$ so the average number of children per individual is exactly 1. The generation neither grows or decays but rather stays constant over time, there may be small possibilities of extinction based on small changes over time. The overall idea remains as the population size stays on average. The supercritical case is exactly what it seems, the population size is $\mu > 1$. The average number of children per individual is greater than one indicating that the population size is growing exponentially and might not go extinct.

Simulating 10 generations of $(Z_0, ..., Z_{10})$ of a branching process using a Poisson offspring distribution where 3 values for the Poisson mean parameter were chosen based on the sub, super, and critical cases. The subcritical case is where $\mu = 0.75$, critical is where $\mu = 1$, and the supercritical is where $\mu = 1.5$ [**Dobrow2016**]

5.1 Extinction

In branching processes, one of the main uses is calculating the probability of eventual extinction.

Example of Extinction probability

Consider a population where each individual has either 0, 1, or 2 offspring with probabilities $p_0 = 0.3$, $p_1 = 0.4$, and $p_2 = 0.3$, respectively. Starting with one individual in generation 0, calculate the expected population size

μ	Z_0	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
0.75	1	2	3	3	2	1	0	0	0	0	0
0.75	1	0	0	0	0	0	0	0	0	0	0
0.75	1	1	0	0	0	0	0	0	0	0	0
0.75	1	3	3	1	0	0	0	0	0	0	0
0.75	1	2	3	2	2	0	0	0	0	0	0
1	1	2	6	10	8	7	8	6	6	0	0
1	1	3	2	2	7	6	7	8	8	0	0
1	1	1	4	4	4	9	8	7	8	0	0
1	1	3	4	1	1	0	0	0	0	0	0
1	1	2	2	22	41	93	173	375	763	1,597	0
1.5	1	1	1	3	7	7	9	11	19	29	0
1.5	1	3	2	5	18	34	68	127	246	521	1,011
1.5	1	2	5	3	2	6	9	17	18	13	19

Table 1: Simulations of a Branching Process for Three Choices of μ

in generation 2 and determine the probability that the population becomes extinct after two generations.

$$\mu = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2$$

The equation above shows how to calculate the average.

$$\mu = 0 \cdot 0.3 + 1 \cdot 0.4 + 2 \cdot 0.3 = 1$$

Applying the concept of generating functions,

$$G(s) = p_0 s^0 + p_1 s^1 + p_2 s^2$$

$$G(s) = 0.3 + 0.4s + 0.3s^2$$

Knowing that we are trying to find the probability of a generation going extinct, using the theorem $q = G(q)$ where q represents the probability that the branching process dies out.

$$q = 0.3 + 0.4q + 0.3q^2$$

Substituting s for q, now all to do is simple algebra, subtracting q from both sides:

$$0.3q^2 - 0.6q + 0.3 = 0$$

$$q^2 - 2q + 1 = 0$$
$$(q-1)^2 = 0$$
$$q = 1$$

The extinction probability is 1 meaning that the population will die out with certainty.

Theorem 5.2 Extinction Probability
Given a branching process, let G be the probability-generating function of the offspring distribution. Then, the probability of eventual extinction is the smallest positive root of the equation $s = G(s)$ If $\mu \leq 1$, that is, in the subcritical and critical cases, the extinction probability is equal to 1.

The theorem provides a method to determine the extinction probability in a branching process using the PGF of the offspring distribution. The extinction probability is the smallest positive root of $s = G(s)$. In subcritical and critical cases ($\mu \leq 1$), the extinction probability is always 1, indicating certain extinction. In the supercritical case ($\mu > 1$), there is a chance of survival, and the extinction probability is found by solving the fixed-point equation.

Example 5.3 Problem of Extinction
Assume $p_0 = 1/2, p_1 = 1/4$, and $p_2 = 1/4$. The tree for these probabilities of the first two generations is shown below.

To solve this example the theory of sums of independent random variables to assign branch probabilities. If there are two offspring in the first generation, the probability that there will be two in the second generation is [**GrinsteadSnell**]

To solve this example, using the theory of independent sums of random variables will be very helpful. If there are two offspring in the first generation, the probability that there will be two in the second is:

$$P(X_1 + X_2 = 2) = p_0p_2 + p_1p_1 + p_2p_0$$
$$= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} = \frac{5}{16}$$

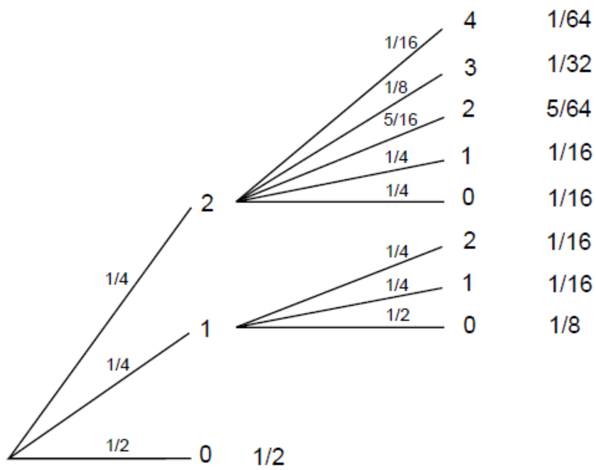


Figure 1: branching process

But what happens if the branching process goes extinct? Let d_n be the probability that the process dies out by the nth generation. Knowing $d_0 = 0$. In our example, $d_1 = \frac{1}{2}$ and $d_2 = \frac{1}{2} + \frac{1}{8} + \frac{1}{16} = \frac{11}{16}$. Note that we must add the probabilities for all paths that lead to 0 by the nth generation. By definition,

$$0 = d_0 \leq d_1 \leq d_2 \leq \cdots \leq 1.$$

Since, d_n converges to a limit d, $0 \leq d \leq 1$. d is the probability that the branching process goes extinct. To find all the probabilities we write d_n as all the possible outcomes in the first generation. If there is j offspring in the first generation, for it to die out by the $n - th$ generation. Each line must die out by the n-1 generation. Since these are independent, the probability is $(d_{n-1})^j$.

$$d_m = p_0 + p_1 d_{m-1} + p_2 (d_{m-1})^2 + p_3 (d_{m-1})^3 + \cdots$$

Let h_z be the generating function for p_i

$$h(z) = p_0 + p_1 z + p_2 z^2 + \cdots$$

Using the generating function we can rewrite $0 = d_0 \leq d_1 \leq d_2 \leq \cdots \leq 1$ as

$$d_m = h(d_{m-1})$$

Since d_n converges to d then by $d_m = p_0+p_1d_{m-1}+p_2(d_{m-1})^2+p_3(d_{m-1})^3+\cdots$ the value d satisfies the equation below:

$$d = h(d)$$

The obvious solution is where $d = 1$ causing,

$$1 = p_0 + p_1 + p_2 + \cdots$$

Now this is where Galton and Watson made a mistake, they directly assumed that the only probability was 1. To understand this better that the solutions of $h(z) = p_0 + p_1z + p_2z^2 + \cdots$ represent the intersection of the graph:

$$y = z$$

and

$$y = h(z) = p_0 + p_1z + p_2z^2 + \cdots$$

Because of this the graph of $y = h(z)$ should be looked at, also remember that $h(0) = p_0$.

$$h'(z) = p_1 + 2p_2z + 3p_3z^2 + \cdots ,$$

and,

$$h''(z) = 2p_2 + 3 \cdot 2p_3z + 4 \cdot 3p_4z^2 + \cdots .$$

From this we see that for $z \geq 0$, $h'(z) \geq 0$ and $h''(z) \geq 0$. For the nonnegative z, h(z) is a concave up and increasing function. The graph $y = h(z)$ can intersect the graph $y = z$ at two points. We know that the graph intersects at (1,1) because of the rule that the generating function is equal to one, so the sum of all probabilities is 1. There are three possible graphs:

The first case, (a), the graph $h(z)$ intersects the line $y = z$ at two points d and 1, where $0 \leq d < 1$. The next case, (b) is where the graph h(z) intersects the line $y = z$ at one point $d = 1$. So the only solution is $z = 1$, so the probability that the process will go extinct is exactly 1. The last case, (c) is where the graph h(z) intersects with the line $y = z$ at two points. 1 and d where $d > 1$. Similar to b the only solution that works, is within $0 \leq z \leq 1$ where $z = 1$ meaning that the population will go extinct with probability 1.

$$h'(1) = p_1 + 2p_2 + 3p_3 + \cdots =,$$

The derivative represents the number of offspring produced by a single parent. Looking back at the previous cases, a, b, and c it is established that

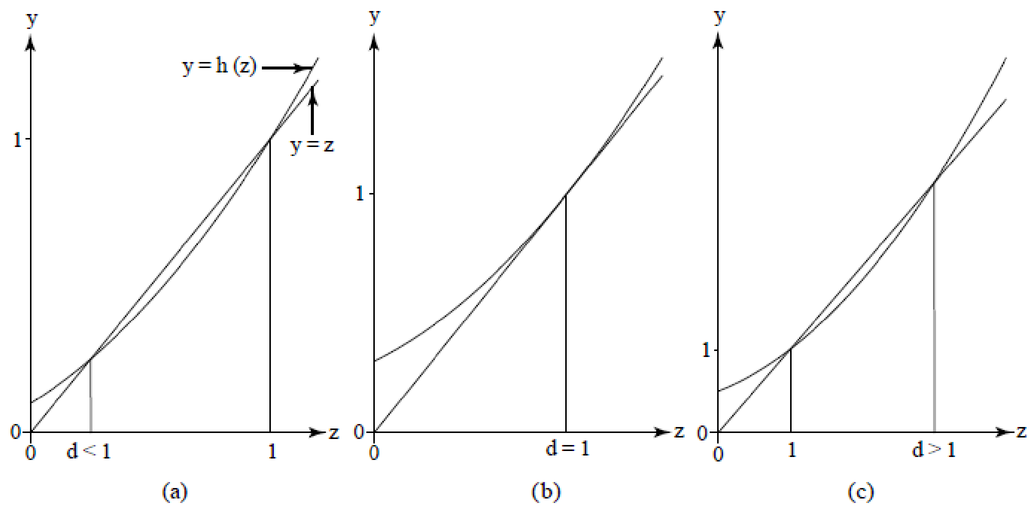


Figure 2: Graphs of $y = z$ and $y = h(z)$

Case(a) represents the supercritical case as $h'(1) > 1$. Case(b) represents the critical case as $h'(1) = 1$, and Case(c) is the subcritical case as $h'(1) < 1$

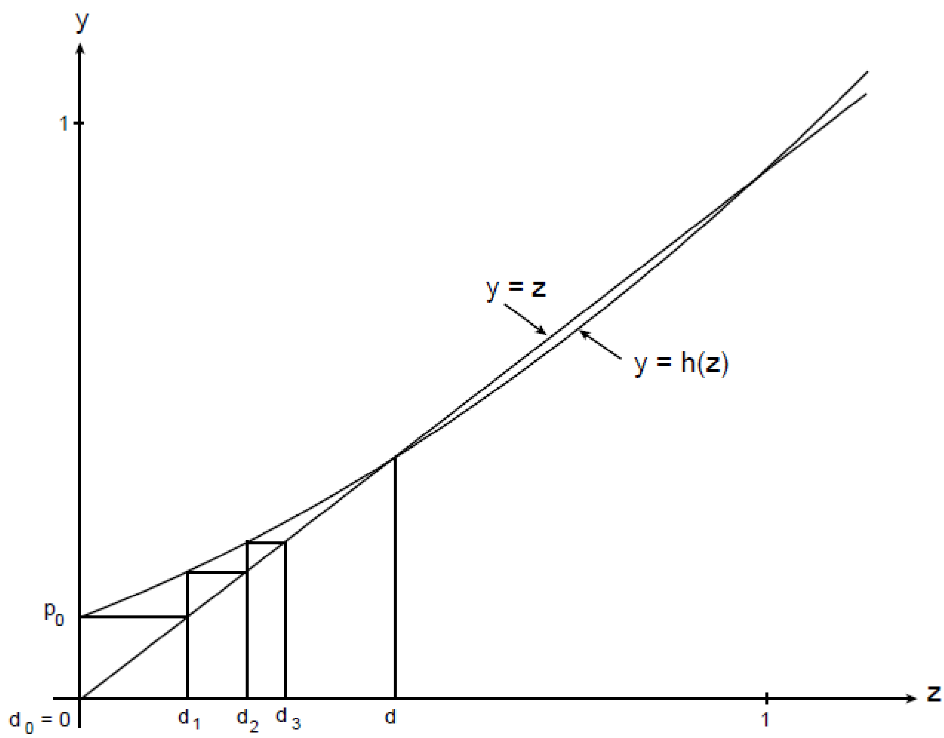


Figure 3: Geometric Construction of d

This depicts the geometric process of finding the fixed point d where the graph $y = h(z)$ and $y = z$ intersect. Recall that $d_0 = 0$, $d_1 = h(d_0) = p_0$, $d_2 = h(d_1)$, $\ldots$, and $d_n = h(d_{n-1})$. So the sequence $d_0, d_1, d_2, \ldots$ is constructed geometrically. This sequence also represents the iterations of h(z) starting from 0. This figure helps to visualize extinction probability and how starting at zero and repeating the function h will lead to the branching process converging to some d, that d represents when the Galton-Watson process will die out.

Similarly Alfred Lotka worked on a problem like this where he tried to find the extinction probability that a male line of descent would go extinct. Looking into the 1920 census, Lotka fitted the distribution of male offspring to a zero-adjusted distribution (starting from a basline or control factor).

This takes the form:

$$a_0 = 0.48235 \quad \text{and} \quad a_k = (0.2126)(0.5893)^{k-1}, \quad \text{for } k \geq 1.$$

The generating function of the offspring distribution is:

$$G(s) = 0.48235 + 0.2126 \sum_{k=1}^{\infty} (0.5893)^{k-1} s^k = 0.48235 + \frac{(0.2126)s}{1-(0.5893)s}.$$

The mean of the male offspring distribution is $\mu = 1.26$. It is interesting that despite a mean number of children (sons and daughters) per individual of about 2.5, the probability of extinction of family surnames is over 80%. [**Lotka1931**]

5.2 Generating Functions in Branching Processes

A generating function is just another way to write a sequence, and referring to the Galton-Watson process which is just a sequence of generations. The two can be related. This proof shows how the generating function $G_n(s)$ evolves based on the generating function of the offspring distribution, G_s

$$G_n(s) = \sum_{k=0}^{\infty} s^k P(Z_n = k)$$

This is the generating function for the distribution Z_n, which is the number of individuals in the n-th generation

$$G(s) = \sum_{k=0}^{\infty} s^k a_k$$

In this equation a_k is the probability that an individual has k offspring

$$G_n(s) = E(s^{Z_n})$$

The generating function $G_n(s)$ can be represented as the expected value of $s^(Z_n)$

$$G_n(s) = E(s^{Z_n}) = E\left(s^{\sum_{i=1}^{Z_{n-1}} X_i}\right)$$

Z_n is expressed as the sum of offspring X_i of each individual in the (n-1)-th generation

$$G_n(s) = E\left(E\left(s^{\sum_{i=1}^{Z_{n-1}} X_i} \middle| Z_{n-1}\right)\right)$$

By the law of total expectation, we take the conditional expectation given Z_{n-1} and then the overall expectation.

$$E\left(s^{\sum_{i=1}^{Z_{n-1}} X_i} \middle| Z_{n-1} = z\right)$$

Given $Z_{n-1} = z$ the sum of offspring $\sum_{i=1}^{z} X_i$ can be taken into understanding.

$$E\left(s^{\sum_{i=1}^{z} X_i}\right) = E\left(\prod_{i=1}^{z} s^{X_i}\right)$$

Because of the independence of X_i this equation can be rewritten as the product of expectations.

$$E\left(\prod_{i=1}^{z} s^{X_i}\right) = \prod_{i=1}^{z} E(s^{X_i})$$

Since X_i is i.i.d, each term is the expected value of s^X, which makes it the generating function G_s.

$$E\left(\prod_{i=1}^{z} s^{X_i}\right) = [G(s)]^z$$

This makes the inner expectation $[(G_s)]^z$.

$$G_n(s) = E\left([G(s)]^{Z_{n-1}}\right) = G_{n-1}(G(s))$$

The generating function $G_n(s)$ can be expressed recursively using the generating function of the (n-1)-th generation and the offspring generating function.

References

[1] P. Mishra, "Introduction to High-Performance Graph Processing", New York University, 2009. [Online]. Available: https://cs.nyu.edu/~mishra/COURSES/09.HPGP/scribe1. [Accessed: 06- Jul- 2024].

[2] R. Dobrow, *Introduction to Stochastic Processes with R*, John Wiley & Sons, 2016, p. 185.

[3] Z. Gordon, "Lecture 7: Branching Processes", University of Texas at Austin, 2024. [Online]. Available: https://web.ma.utexas.edu/users/gordanz/notes/lecture7.pdf. [Accessed: 06- Jul- 2024]. p. 4.

[4] Introductory Probability (Grinstead and Snell), Generating Functions, Branching Processes. https://stats.libretexts.org/Bookshelves/Probability_Theory/Book%3A_Introductory_Probability_%28Grinstead_and_Snell%29/10%3A_Generating_Functions/10.02%3A_Branching_Processes. [Accessed: 11- Jul- 2024].

[5] A. J. Lotka, *Principles of Physical Biology*, Baltimore: Williams and Wilkins, 1931.