ON THE GALTON-WATSON PROCESS AND ITS MODELING CAPABILITIES

KEVIN ZHU

Abstract. This paper introduces Galton-Watson processes and the prerequisites necessary to understand it. Topics include probability generating functions, variance, and expected value, properties of generation size, key results regarding the extinction of a branching tree, the process's modeling capabilities for different distributions, and variations of the standard branching process.

1. Introduction

In the late 19th century, there was concern among the British people that certain aristocratic last names would die out. As a result, British Mathematicians Francis Galton and Henry William Watson set out to model the propagation of family names; their model, referencing both of their last names, has subsequently been named the Galton-Watson process. Let us first define this process informally:

- Generation 0 has 1 individual.
- Each individual within the same branching tree has the same distribution for the number of offspring they produce.
- The offspring of Generation n become Generation $n + 1$.

These criteria result in a branching tree that either goes on infinitely or goes extinct at some point in time, one example of which is shown in figure 1. Galton and Watson, interested in figuring out the probability that a certain last name would go extinct, questioned how the probability of a branching tree going extinct would differ given different distributions for the number of offspring an individual produces; the answer to their query is one of the main focuses of this paper and arguably the most important result of the Galton-Watson process.

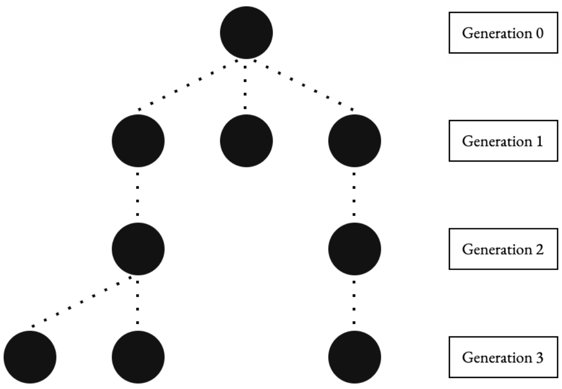

Figure 1. Possible Branching Tree

Before we delve into the meat of the mathematics, though, it is necessary to cover some of the fundamentals of probability theory in order to understand the proofs provided later in the paper.

1.1. **Expected Value.** The expected value of a discrete random variable X is denoted as $\mathbb{E}(X)$ and is defined as

$$\sum_{i=1}^{n} X_i \mathbb{P}(X = X_i)$$

where $\{X_1, X_2, X_3, \ldots, X_n\}$ is the support for X and $\mathbb{P}(X = X_i)$ denotes the probability that the random variable X is equal to X_i. Intuitively, the expected value of X is just the weighted average of all values of X where the weights are the probabilities of attaining those values.

The law of total expectation is helpful to know because it often allows us to simplify calculations for expected value. In the following theorem, Y is an arbitrary random variable.

Theorem 1.1. $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y))$ [Hua20].

It will be applied in a later section to help us find the expected number of individuals in the nth generation.

Expected value is an important concept to know while studying the Galton-Watson process since it plays a significant role in determining the probability of a branching tree going extinct. In addition, finding the expected number of individuals in generation n is often a desired result.

1.2. **Variance.** Intuitively, variance describes how spread out a data set is. If most values are congregated around the mean, the variance will be relatively small. On the other hand, if most values are quite far from the mean, the variance will be relatively large. It is defined quite similarly to standard deviation (σ); in fact

$$\text{Variance} = \sigma^2.$$

The more useful variance expression for analyzing the Galton-Watson process is in terms of expected value though:

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

This equation is especially useful for finding the variance of a known distribution and will help us out in a later section.

The law of total variance is also helpful to know. Again, Y is just an arbitrary random variable.

Theorem 1.2. $\text{Var}(X) = \mathbb{E}(\text{Var}(X \mid Y)) + \text{Var}(\mathbb{E}(X \mid Y))$.

It can be applied to help us find the variance of the number of individuals in the nth generation.

1.3. **Probability Generating Functions.** The probability generating function of a discrete random variable X taking nonnegative integer values is denoted as $g(s)$ and is defined to be

$$\sum_{i=0}^{\infty} \mathbb{P}(X = i) s^i$$

where i takes on the possible values of X.

Understanding probability generating functions is very important while studying the Galton-Watson process, since they are used as a tool to represent the expected number of offspring an individual has and to calculate the probability of a geometric branching tree going extinct.

Before we jump into the next section, to avoid confusion on the reader's end, notation that will be frequently used shall be defined right now:

- X_i: the discrete random variable denoting the number of offspring individual i has
- $g(s)$: the probability generating function of X, $\sum_{i=0}^{\infty} \mathbb{P}(X = i)s^i$
- Z_n: the discrete random variable denoting the number of individuals in the nth generation
- q_n: the probability that the branching tree is extinct by the nth generation, $\mathbb{P}(Z_n = 0)$
- $\lim_{n\to\infty} q_n$: the probability that the branching tree eventually goes extinct
- μ: the expected number of offspring an individual has, $\mathbb{E}(X)$
- σ: the standard deviation of the number of offspring an individual has

2. Properties of Z_n

It is desirable to find the expected value and variance of Z_n if predicting the number of individuals at a later time is of interest. First, let us look at $\mathbb{E}(Z_n)$

Theorem 2.1. $\mathbb{E}(Z_n) = \mu^n$.

Proof. Let us prove this theorem via induction.
Base case ($n = 0$): By the definition of the Galton-Watson process, generation 0 has 1 individual. Thus, $\mathbb{E}(Z_0) = 1 = \mu^0$, showing that the theorem holds when $n = 0$.
Inductive hypothesis: Assume that $\mathbb{E}(Z_n) = \mu^n$ for $n = k$.
Inductive step: Let $n = k + 1$. It follows that

$$Z_{k+1} = \sum_{i=1}^{Z_k} X_i.$$

Taking the expected value of both sides, we get that

$$\mathbb{E}(Z_{k+1}) = \mathbb{E}\left(\sum_{i=1}^{Z_k} X_i\right).$$

Applying theorem 1.1,

$$\begin{aligned}\mathbb{E}(Z_{k+1}) &= \mathbb{E}\left(\sum_{i=1}^{Z_k} \mathbb{E}(X_i \mid Z_k)\right)\\ &= \mathbb{E}(\mu Z_k)\\ &= \mu\mathbb{E}(Z_k)\\ &= \mu(\mu)^k\\ &= \mu^{k+1}.\end{aligned}$$

By the principle of mathematical induction, the theorem holds for all $n \geq 0$. ■

Theorem 2.2. $\mathrm{Var}(Z_n) = \begin{cases} n\sigma^2 & \textit{if } \mu = 1, \\ \dfrac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \textit{if } \mu \neq 1. \end{cases}$

Proof. **Case 1**: $\mathrm{Var}(Z_n) = n\sigma^2$ if $\mu = 1$.
Let us first prove the case where $\mu = 1$ since it is much simpler than the other case. By theorem 1.2 [Tsi18],

$$\begin{aligned} \mathrm{Var}(Z_{n+1}) &= \mathbb{E}\left(\mathrm{Var}(\mathrm{Z_{n+1}} \mid \mathrm{Z_n})\right) + \mathrm{Var}(\mathbb{E}(\mathrm{Z_{n+1}} \mid \mathrm{Z_n})) \\ &= \mathbb{E}(\sigma^2 Z_n) + \mathrm{Var}(\mu Z_n) \\ &= \sigma^2 \mathbb{E}(Z_n) + \mathrm{Var}(\mu Z_n), \end{aligned}$$

and by Theorem 2.1,

$$\mathrm{Var}(Z_{n+1}) = \sigma^2 \mu^n + \mathrm{Var}(\mu Z_n). \tag{2.1}$$

When $\mu = 1$,

$$\mathrm{Var}(Z_{n+1}) = \sigma^2 + \mathrm{Var}(Z_n).$$

Let us finish via induction.
Base case ($n = 0$): By the definition of the Galton-Watson process, generation 0 has a fixed number of individuals. Thus, $\mathrm{Var}(Z_0) = 0 = 0(\sigma^2)$, showing that the theorem holds for this case when $n = 0$.
Inductive hypothesis: Assume that $\mathrm{Var}(Z_n) = n\sigma^2$ for $n = k$ if $\mu = 1$.
Inductive step: Let $n = k + 1$. It follows that

$$\begin{aligned} \mathrm{Var}(Z_{k+1}) &= \sigma^2 + \mathrm{Var}(Z_k) \\ &= \sigma^2 + k\sigma^2 \\ &= (k+1)\sigma^2. \end{aligned}$$

By the principle of mathematical induction, the theorem holds for all $n \geq 0$ if $\mu = 1$.
Case 2: $\mathrm{Var}(Z_n) = \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1}$ if $\mu \neq 1$.
Let us now direct our attention to the case where $\mu \neq 1$. Again, we can use an inductive process to derive the result.
Base case ($n = 1$): By Equation 2.1, $\mathrm{Var}(Z_1) = \sigma^2 \mu^0 + \mathrm{Var}(\mu Z_0) = \sigma^2 = \frac{\sigma^2(\mu^1 - 1)\mu^0}{\mu - 1}$, showing that the theorem holds for this case when $n = 0$.
Inductive hypothesis: Assume that $\mathrm{Var}(Z_n) = \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1}$ for $n = k$ if $\mu \neq 1$.
Inductive step: Let $n = k + 1$. By Equation 2.1,

$$\begin{aligned} \mathrm{Var}(Z_{k+1}) &= \sigma^2 \mu^k + \mathrm{Var}(\mu Z_k) \\ &= \sigma^2 \mu^k + \mu^2 \,\mathrm{Var}(Z_k) \\ &= \sigma^2 \mu^k + \mu^2 \left(\frac{\sigma^2(\mu^k - 1)\mu^{k-1}}{\mu - 1} \right) \\ &= \sigma^2 \mu^k + \left(\frac{\sigma^2(\mu^k - 1)\mu^{k+1}}{\mu - 1} \right), \end{aligned}$$

and by factoring out $\sigma^2\mu^k$,

$$\begin{aligned}\operatorname{Var}(Z_{k+1}) &= \sigma^2\mu^k\left(1+\frac{(\mu^k-1)\mu}{\mu-1}\right)\\ &= \sigma^2\mu^k\left(\frac{\mu-1}{\mu-1}+\frac{(\mu^k-1)\mu}{\mu-1}\right)\\ &= \sigma^2\mu^k\left(\frac{\mu-1}{\mu-1}+\frac{\mu^{k+1}-\mu}{\mu-1}\right)\\ &= \frac{\sigma^2(\mu^{k+1}-1)\mu^k}{\mu-1}.\end{aligned}$$

By the principle of mathematical induction, the theorem holds for all $n \geq 1$ if $\mu \neq 1$. Since we have proved all cases, the theorem holds for all plausible μ. ■

3. Extinction Probability

The probability of a branching tree eventually going extinct was the main result that Galton and Watson wanted to determine in their initial investigation of the process. It turns out, though, that there is only a closed-form expression for this probability for certain distributions of X.

3.1. Generalized Extinction.

Theorem 3.1. Unless $\mathbb{P}(X=1)=1, \lim_{n\to\infty} q_n = 1 \text{ if } \mu \leq 1$.

Remark 3.2. When $\mathbb{P}(X=1)=1$, $Z_n = 1 : n \geq 0$. Thus, $\lim_{n\to\infty} q_n = 0$ even though $\mu \leq 1$.

Proof. Let us begin by looking at the probability generating function of X, which is $g(s)$. As defined previously,

$$g(s) = \sum_{i=0}^{\infty} \mathbb{P}(X=i)s^i.$$

Letting $s = q_{n-1}$,

$$g(q_{n-1}) = \sum_{i=0}^{\infty} \mathbb{P}(X=i)(q_{n-1})^i.$$

Intuitively, this power series represents the probability that the lineages of Generation 0's X offspring all go extinct within another $n-1$ generations or, in other words, the probability that the branching tree goes extinct in n generations. Thus,

$$g(q_{n-1}) = q_n. \tag{3.1}$$

Since we are interested in the probability of eventual extinction, it is logical to look at this expression specifically when n approaches ∞. Because $g(s)$ is continuous,

$$\lim_{n\to\infty} q_n = \lim_{n\to\infty} g(q_{n-1})$$

implies that

$$\lim_{n\to\infty} q_n = g\left(\lim_{n\to\infty} q_n\right).$$

This is a very useful result, as we can see that $\lim_{n\to\infty} q_n$ is simply a solution to the equation

$$s = g(s) : s \in [0,1].$$

A problem arises, though, when it is realized that the equation often has multiple solutions on that interval. Thus, it is important to clarify which of the solutions represents the probability of eventual extinction.

Lemma 3.3. $\lim_{n\to\infty} q_n$ is the smallest solution to the equation $s = g(s) : s \in [0, 1]$.

Proof. Let us define r to be the smallest solution to the equation $s = g(s)$ where $s \in [0, 1]$. Since we already know that $\lim_{n\to\infty} q_n$ is a solution, to prove that $\lim_{n\to\infty} q_n = r$, we just have to show that $q_n \leq r$ for $n \geq 0$.

It is perhaps easiest if we use an inductive proof.

Base Case (n=0): By the definition of the Galton-Watson process, generation 0 has 1 individual. Thus, $q_0 = \mathbb{P}(Z_0 = 0) = 0 \leq r$, showing that the lemma holds when $n = 0$.

Inductive hypothesis: Assume that $q_n \leq r$ for $n = k$.

Inductive step: Let $n = k + 1$. By equation 3.1, proving

$$q_{k+1} \leq r = g(r)$$

is equivalent to proving

$$g(q_k) \leq g(r).$$

Since we have assumed that $q_k \leq r$, we can complete this proof just by showing that $g(s)$ is a nondecreasing function on the specified interval, which is accomplished if it is shown that $g'(s) \geq 0 : s \in [0, 1]$. As defined previously,

$$g(s) = \sum_{i=0}^{\infty} \mathbb{P}(X = i)s^i.$$

Taking the derivative,

$$g'(s) = \sum_{i=1}^{\infty} \mathbb{P}(X = i)is^{i-1}.$$

Since $\mathbb{P}(X = i), i$, and s^{i-1} are all nonnegative, their product will be nonnegative as well, making $g'(s) \geq 0$ and completing the inductive step. By the principle of mathematical induction, $q_n \leq r$ holds for all $n \geq 0$, and the lemma is subsequently proven. ■

To find the smallest solution to the equation $s = g(s) : s \in [0, 1]$, it is helpful to look at some properties of $g(s)$.

Lemma 3.4. $g'(1) = \mu$.

Proof. As derived in the proof of lemma 3.3,

$$g'(s) = \sum_{i=1}^{\infty} \mathbb{P}(X = i)is^{i-1}.$$

Substituting 1 for s,

$$g'(1) = \sum_{i=1}^{\infty} \mathbb{P}(X = i)i,$$

which is precisely the formula for $\mathbb{E}(X)$. Since $\mu = \mathbb{E}(X)$, $g'(1) = \mu$. ■

$g'(1) = \mu$ is a useful equation since it allows us to consider the more nebulous μ as a feature of the graph.

Lemma 3.5. $g(1) = 1$.

Proof. As defined previously,

$$g(s) = \sum_{i=0}^{\infty} \mathbb{P}(X = i)s^i.$$

Substituting 1 for s,

$$g(1) = \sum_{i=0}^{\infty} \mathbb{P}(X = i).$$

Since $\lim_{n\to\infty}\{0, 1, 2, \ldots, n\}$ is the entire sample space of X, $g(1) = 1$. ■

The equation $g(1) = 1$ tells us that $q_n = 1$ if $g(s) - s \neq 0 : s \in [0, 1)$. To prove that $g(s) - s \neq 0 : s \in [0, 1)$ when $\mu = g'(1) \leq 1$, it is helpful to analyze how $g(s) - s$ changes with respect to s on this interval. Let $f(s) = g(s) - s$. Taking the derivative of both sides,

$$\begin{aligned} f'(s) &= g'(s) - 1 \\ &= \sum_{i=1}^{\infty} \mathbb{P}(X = i)is^{i-1} - 1. \end{aligned}$$

Since $s < 1$,

$$\sum_{i=1}^{\infty} \mathbb{P}(X = i)is^{i-1} - 1 < \sum_{i=1}^{\infty} \mathbb{P}(X = i)i - 1 = \mu - 1 : s \in [0, 1),$$

and because $\mu \leq 1$,

$$f'(s) < \mu - 1 \leq 1 - 1 : s \in [0, 1).$$

Thus, $f'(s) < 0 : s \in [0, 1)$, indicating that $f(s)$ is always decreasing on this interval. Therefore,

$$f(s) = g(s) - s > f(1) = g(1) - 1 = 0 : s \in [0, 1),$$

proving that $g(s)-s \neq 0 : s \in [0, 1)$ and the subsequent fact that $\lim_{n\to\infty} q_n = 1$ if $\mu \leq 1$. ■

Theorem 3.6. $\lim_{n\to\infty} q_n < 1$ if $\mu > 1$.

Proof. We can utilize the relationship between $g(s)$ and s to prove this as well. However, the approach is slightly different and another property of $g(s)$ needs to be introduced.

Lemma 3.7. $g(0) = \mathbb{P}(X = 0)$.

Proof. As defined previously,

$$g(s) = \sum_{i=0}^{\infty} \mathbb{P}(X = i)s^i.$$

Substituting 0 for s,

$$\begin{aligned} g(0) &= \sum_{i=0}^{\infty} \mathbb{P}(X = i)0^i \\ &= \mathbb{P}(X = 0)0^0 + \mathbb{P}(X = 1)0^1 + \mathbb{P}(X = 2)0^2 + \cdots \\ &= \mathbb{P}(X = 0). \end{aligned}$$

■

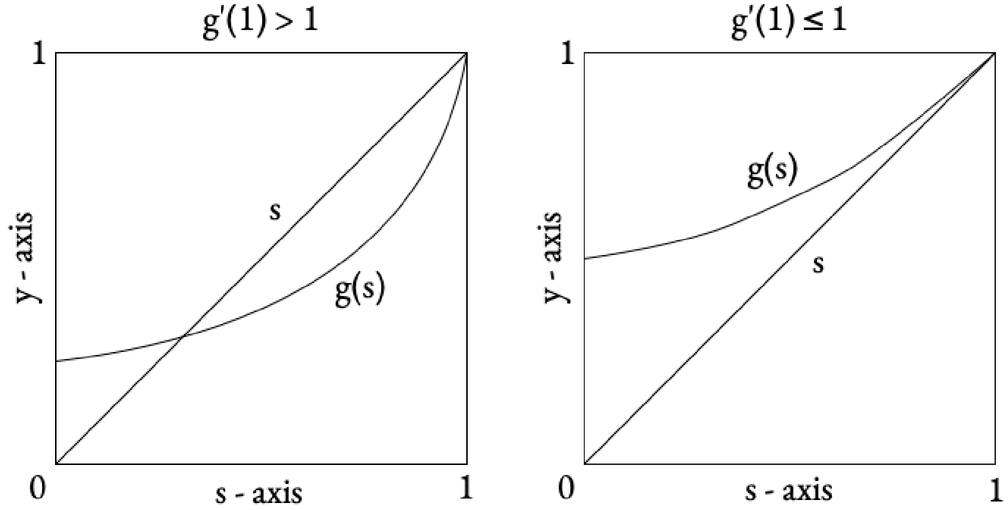

Figure 2. Possible Graphs of $y = s$ and $y = g(s)$

In other words, the y-intercept of $g(s)$ is at $(0, \mathbb{P}(X = 0))$. The usefulness of this lemma only becomes clear when we look at the y-intercept's position with respect to the line $y = s$. Since $\mathbb{P}(X = 0) \geq 0$, we know that it is on the left side of the line $y = s$ or on top of the line. Because $g(s)$ is continuous, it is not hard to see that finding a point $(s, g(s)) : s \in (0, 1)$ on the right side of the line $y = s$ implies that $s = g(s)$ for at least one $s \in (0, 1)$.

Proving that such a point exists is rather easy. To do so, let us revisit the equation $f(s) = g(s) - s$. Because $f(1) = 0$ and $f'(1) > 0$, there exists a value of $s \in (0, 1)$ where $f(s) = g(s) - s < 0$, showing that there exists a point to the right of the line $y = s$ in the specified interval.

Though it is visually obvious, mathematically justifying that the existence of $(0, \mathbb{P}(X = 0))$ and a point right of $y = s : s \in [0, 1)$ implies a solution to $s = g(s) : s \in [0, 1)$ is not too difficult either. The point $(0, \mathbb{P}(X = 0))$ tells us that

$$g(s) - s \geq 0 : s = 0,$$

and the existence of a point right of $y = s$ tells us that

$$\exists s \in (0, 1) \text{ such that } g(s) - s < 0.$$

Since $g(s) - s$ is continuous, as a result of $g(s)$ being continuous, the intermediate value theorem tells us that

$$\exists s \in [0, 1) \text{ such that } g(s) - s = 0,$$

proving that $q_n < 1$ if $\mu > 1$. ■

Visual representations of the two scenarios can be seen in figure 2. With our additional understanding of q_n, let us circle back to our earlier discussion about finding a closed-form expression for it. The reason q_n cannot often be expressed as a closed form expression is that $g(s) - s$, being an arbitrary power series, is unlikely to have roots that can be expressed in a closed form.

Only in special cases where X is distributed a certain way is there a closed form expression for q_n. In the next subsection, we shall take a look at one such case — when X is geometrically distributed.

3.2. **Extinction for Geometric Offspring Distribution.** Normally, geometric distributions define the number of trials required to get the first "success"; all of the trials before the "success" are considered "failures". In the context of the Galton-Watson process, "success" represents the event that an individual stops reproducing and each offspring is considered a "failure".

At any given point in time, let us define the probability of the individual having at least 1 more offspring to be p. Consequently, the probability that the individual stops reproducing is $1-p$. It follows that

$$\mathbb{P}(X=i)=(1-p)p^i.$$

This equation is very helpful for finding $\lim_{n\to\infty} q_n$.

Theorem 3.8. $\lim_{n\to\infty} q_n = 1$ if $p \leq \frac{1}{2}$.

Proof. Substituting $(1-p)p^i$ for $\mathbb{P}(X=i)$,

$$g(s)=\sum_{i=0}^{\infty}(1-p)p^i s^i.$$

Taking the derivative of both sides,

$$g'(s)=\sum_{i=1}^{\infty} i(1-p)p^i s^{i-1}.$$

Substituting 1 for s,

$$g'(1)=\mu=\sum_{i=1}^{\infty} i(1-p)p^i.$$

Thus,

$$\mu=\sum_{i=1}^{\infty} i(1-p)p^i \leq \sum_{i=1}^{\infty} i\left(1-\frac{1}{2}\right)\left(\frac{1}{2}\right)^i : p \leq \frac{1}{2}.$$

Simplifying,

$$\mu=\sum_{i=1}^{\infty} i(1-p)p^i \leq \sum_{i=1}^{\infty} \frac{i}{2^{i+1}} : p \leq \frac{1}{2}.$$

Using the well-known result

$$\sum_{i=1}^{\infty}\frac{i}{2^i}=2 \quad \text{[Wil13]},$$

we can deduce that

$$\begin{aligned}\sum_{i=1}^{\infty}\frac{i}{2^{i+1}} &= \frac{1}{2}\sum_{i=1}^{\infty}\frac{i}{2^i}\\ &= \frac{1}{2}(2)\\ &= 1.\end{aligned}$$

Thus, $\mu \leq 1$ if $p \leq \frac{1}{2}$, and theorem 3.1 tells us that $\lim_{n\to\infty} q_n = 1$ if $\mu \leq 1$. ■

Theorem 3.9. $\lim_{n\to\infty} q_n = \frac{1-p}{p}$ if $p > \frac{1}{2}$

Proof. As derived in the previous proof,

$$g(s) = \sum_{i=0}^{\infty}(1-p)p^i s^i,$$

and since $g(s)$ is an infinite geometric series with first term $1-p$ and ratio ps,

$$g(s) = \frac{1-p}{1-ps}.$$

Because $\lim_{n\to\infty} q_n$ is a solution to $s = g(s)$, we are interested in solving for s when

$$s = \frac{1-p}{1-ps}.$$

Multiplying both sides by $1-ps$,

$$s - ps^2 = 1 - p.$$

Moving the terms with p to the same side of the equation and factoring out p,

$$s - 1 = p(s^2 - 1),$$

and applying the difference of squares identity,

$$s - 1 = p(s+1)(s-1),$$

which becomes

$$1 = ps + p$$

when both sides are divided by $s-1$. Thus,

$$s = \frac{1-p}{p},$$

and subsequently $\lim_{n\to\infty} q_n = \frac{1-p}{p}$ if $p > 1/2$. ■

4. Comparing Different Distributions of X

As shown previously through our analysis of the geometric offspring distribution, properties of Galton-Watson branching trees depend not only on μ, but also on the specific distribution of X. As it turns out, there are many other statistical distributions that produce interesting results when used to model the offspring distribution of a Galton-Watson process. Specifically, in this section, we will introduce the Poisson distribution, binomial distribution, and talk more about the geometric distribution as offspring distributions.

Definition 4.1. Poisson distributions only take on nonnegative integer values. As a result, they are often used to model the number of times an event happens in a given period. Poisson distributions are most notable for always having the same mean and variance — λ. Intuitively, λ represents the expected number of times the event will happen in the period. When λ is small, the Poisson distribution tends to be skewed right. However, as λ increases, its probability mass function increasingly resembles the bell shape of a normal distribution with mean and variance λ. The Poisson distribution has probability mass function

$$\mathbb{P}(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$$

and has infinite support.

Definition 4.2. Binomial distributions also only take on nonnegative integer values. However, they differ from Poisson distributions in that they take 2 parameters, n and p, rather than one. n indicates the number of trials and p indicates the probability of "success" on any given trial; the outcome is the number of times success is attained. The mean is np and the variance is $np(1-p)$. The skewness of the distribution is determined by p. When $p = 0.5$, the distribution is symmetric. When $p < 0.5$, the distribution is skewed right. When $p > 0.5$, the distribution is skewed left. The binomial distribution has probability mass function

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

and has finite support.

Geometric distributions, defined in the previous section, take on only nonnegative integer values as well.

Since individuals can only have nonnegative quantities of offspring, the binomial distribution, Poisson distribution, and geometric distribution are all great choices for being the offspring distribution.

Let's first take a look at the case where $\mu = 1$ for these three distributions. Specifically, let us set $\lambda = 1$ for the Poisson distribution, $n = 10$ and $p = 0.1$ for the binomial distribution, and $p = \frac{1}{2}$ for the geometric distribution. Simulations for 5 runs of each of the distributions can be seen in figure 3.

The variance of the Poisson distribution is

$$\lambda = 1,$$

the variance of the binomial distribution is

$$np(1-p) = 10(0.1)(0.9) = 0.9,$$

and the variance of the geometric distribution is

$$\begin{aligned}\mathbb{E}(X^2) - (\mathbb{E}(X))^2 &= \sum_{i=0}^{\infty} \frac{i^2}{2^{i+1}} - \left(\sum_{i=0}^{\infty} \frac{i}{2^{i+1}}\right)^2 \\ &= 3 - 1 \\ &= 2.\end{aligned}$$

These variances will prove to be useful for comparing graphs of different offspring distributions.

The reason that some graphs appear empty in figure 3 is because they die out almost instantly.

All graphs shown, regardless of offspring distribution, eventually die out. This is expected though, since theorem 3.1 tells us that eventual extinction is imminent if $\mu \leq 1$.

Remark 4.3. To avoid unnecessary wordiness, graphs where X is Poisson distributed, binomially distributed, and geometrically distributed will be referred to as Poisson graphs, binomial graphs, and geometric graphs, respectively, from now on.

Despite sharing the common fate of extinction in the end, the graphs of different offspring distributions display stark differences marked by sporadic behavior.

For one, the binomial graphs all die out almost instantly while there are at least a few Poisson and geometric graphs that exercise brief growth. This difference occurs because the

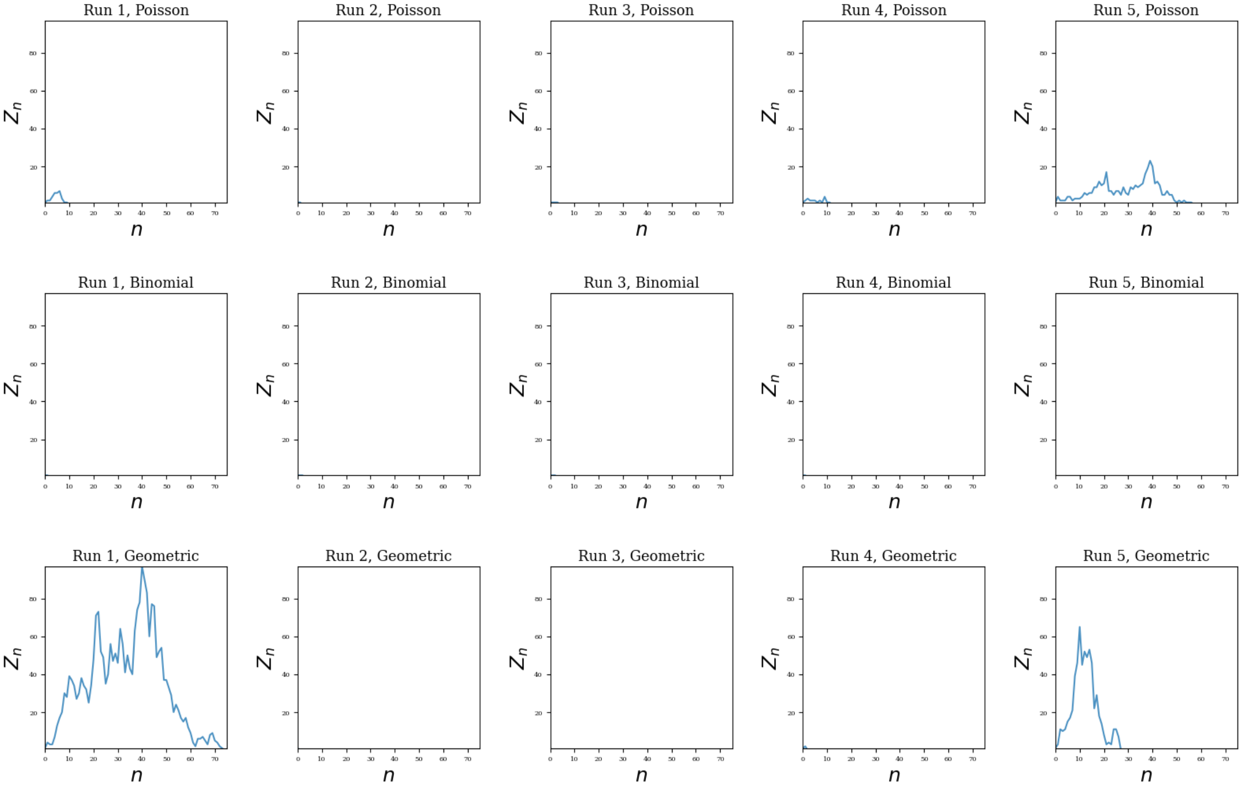

Figure 3. Poisson V. Binomial V. Geometric ($\mu = 1$)

Poisson and geometric distributions have a higher variance than the binomial distribution, resulting in the probability that some individuals have significantly more offspring than the mean being higher. Subsequently, population bursts are also more likely.

It is also worth noticing that the longest surviving geometric graphs seem to grow a lot faster initially than the longest surviving Poisson graphs. This difference is also attributable to variance, as the probability of extreme events, such as rapid growth, increases when variance increases; both the Poisson and geometric distributions have a higher variance than the binomial distribution, but the geometric distribution still has a variance twice that of the Poisson distribution.

Let us also take a look at these graphs when $\mu = 1.5$ to see what happens when extinction is not imminent. Specifically, let us set $\lambda = 1.5$ for the Poisson distribution, $n = 10$ and $p = 0.15$ for the binomial distribution, and $p = \frac{3}{5}$ for the geometric distribution. Simulations for 5 runs of each of the distributions can be seen in figure 4, with the y-axis of each graph now logarithmically scaled to better show behavior in early generations.

The variances, despite taking on different values, still increase from the binomial distribution to the Poisson distribution to the geometric distribution.

Now, a lot more graphs seem to grow exponentially with no signs of stopping regardless of the offspring distribution. This is expected though, since theorem 3.6 tells us that the chance of survival is nonzero if $\mu > 1$. We can assume that graphs reaching $\sim$100 individuals will never die out since the probability of that happening would be equal to the probability of 100 distinct branching trees all dying out — an astronomically small number.

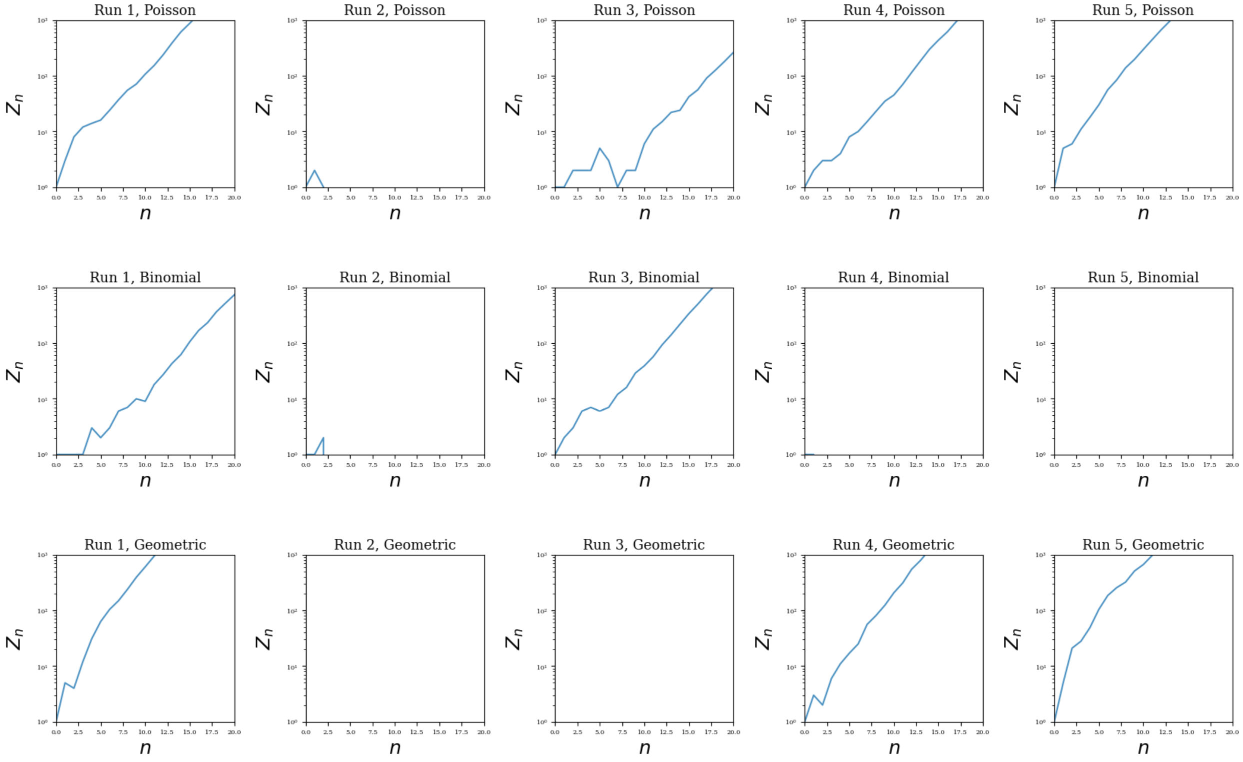

Figure 4. Poisson V. Binomial V. Geometric ($\mu = 1.5$)

The main difference between graphs of different offspring distributions is now growth rate. To compare the growth rates of different graphs, we can look at the generation number where Z_n first crosses 1000.

There are 4 Poisson graphs where eventual extinction does not occur. They reach 1000 individuals in the 13th, 15th, 16th, and ∼22nd generations. On average, the surviving Poisson graphs reached 1000 individuals in the

$$\frac{13+15+16+22}{4} = 16.5\text{th generation.}$$

There are 2 binomial graphs where eventual extinction does not occur. They reach 1000 individuals in the 17th and ∼21st generations. On average, the surviving binomial graphs reached 1000 individuals in the

$$\frac{17+21}{2} = 19\text{th generation.}$$

There are 3 geometric graphs where eventual extinction does not occur. They reach 1000 individuals in the 11th, 11th, and 13th generations. On average, the surviving geometric graphs reached 1000 individuals in the

$$\frac{11+11+13}{3} = 11.\overline{6}\text{th generation.}$$

Again, it is the geometric graphs that have more explosive growth than both the binomial and Poisson graphs due to their higher variance.

4.1. **Choosing a Distribution for Modeling.**

Of course, we do not actually know exact offspring distributions in the real world. The best we can do for making models is to pick a distribution to mimic it. This begs the question: how do we choose between the Poisson, binomial, and geometric distributions when tasked with modeling a branching process?

There is often no trivial answer to this question, as it is not rare to have multiple distributions work equally well. For certain scenarios, though, one distribution shines above the rest. The Poisson distribution, for instance, is very good at modeling disease transmission in the early stage of pandemics since the number of new infections per person is approximately Poisson distributed. Nevertheless, the Poisson distribution has many limitations; its property of having the same mean and variance greatly restricts the number of scenarios where it can be feasibly used. In addition, it has infinite support, making the Poisson distribution a bad choice for modeling human reproduction, where there is a biological upper limit on the number of children a woman can birth. The binomial distribution, on the other hand, has finite support, making it a much better choice for modeling human reproduction and similar events with finite support. What happens when neither finite support nor equal mean and variance is appropriate though? Often, if this is the case, the geometric distribution is a good fit.

5. Variations of the Galton-Watson Process

The standard Galton-Watson process, despite being widely generalizable, is overly simplistic; it does not take into account many factors present in the circumstances that it aims to model. This shortcoming results in the standard process being lackluster even for modeling human population dynamics — its original use.

One of the biggest flaws that comes from using the standard Galton-Watson process to model humans is the negligence of females. Since men were the ones passing on family names to their children in the vast majority of scenarios, the original model considered only men as individuals, disregarding the key role that women play in reproduction. Since it takes one man and one woman to generate an offspring, the number of offspring in generation $n+1$ is not only dependent on the number of males in generation n and the distribution of X; it is also dependent on the number of females in generation n. Thus, researchers in the mid 20th century, realizing this, introduced the bisexual Galton-Watson process.

5.1. **Bisexual Galton-Watson Process.**

As mentioned previously, the key difference between the bisexual Galton-Watson process and the standard Galton-Watson process is that both males and females are considered instead of just males; as a consequence, generation 0 is often defined to have 1 male and 1 female. As a result of this difference, key terms are defined slightly differently.

Definition 5.1. Z_n, instead of representing the number of individuals in generation n, indicates the number of "mating pairs" in generation n. The number of "mating pairs" in generation n is defined to be the maximum number of monogamous heterosexual relationships that can be formed between 2 individuals in generation n. Mathematically,

$$Z_n = \min(M_n, F_n),$$

where M_n represents the number of males in generation n and F_n represents the number of females in generation n.

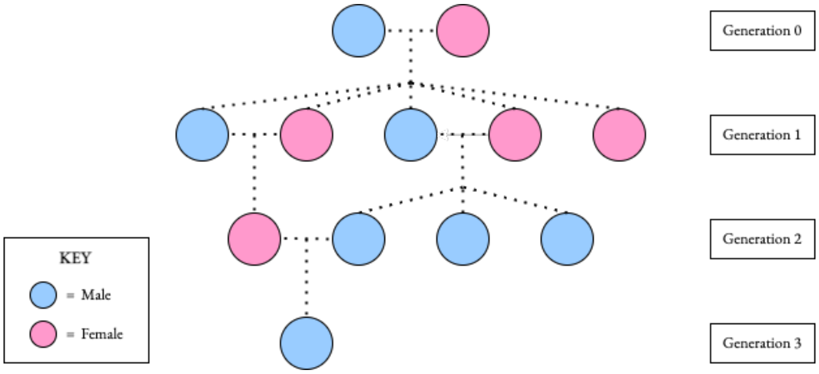

Figure 5. Possible Bisexual Branching Tree

Definition 5.2. μ, instead of representing the expected number of offspring an individual has, indicates the expected number of mating pairs added to Z_{n+1} by a mating pair from generation n. Mathematically,

$$\mu = \mathbb{E}(Z_1).$$

The probability of eventual extinction is the main result for the bisexual Galton-Watson process as well. In a similar fashion to the standard Galton-Watson process, q_n is still dependent on μ for the bisexual Galton-Watson process. In fact, it is still the case that

$$\lim_{n\to\infty} q_n \begin{cases} = 1, & \text{if } \mu \leq 1, \\ < 1, & \text{if } \mu > 1. \end{cases}$$

This is because the bisexual Galton-Watson process can just be considered as a standard Galton-Watson process if mating pairs are thought of as individuals; therefore, the proofs and results covered in the previous section can just be reused.

Although the bisexual Galton-Watson process was originally intended to model human population dynamics, it can also be used to model population dynamics for other monogamous species and even disease models where transmission requires pairwise contact.

Despite eliminating certain shortcomings of the standard Galton-Watson process, the bisexual Galton-Watson process is still not perfect in many ways. For one, the model only considers heterosexual relationships even though homosexuality is at record levels in our modern world. In addition, only individuals from the same generation are allowed to reproduce with one another; this is highly unrealistic in real-world scenarios, as individuals mate with one another under the constraint of the time period they live in rather than their generation number. Thus, some researchers believe that it is more useful if the number of individuals existing at a certain time is considered instead of the number of individuals in a certain generation; the age-dependent Galton-Watson process is the implementation of this.

5.2. **Age-dependent Galton-Watson Process.** The introduction of time results in the concept of generations becoming quite useless; as a result, the age-dependent Galton-Watson process is defined very differently from the other two variations:

- One individual is born at time 0.
- Each individual within the same branching tree has the same distribution for the amount of time they live.

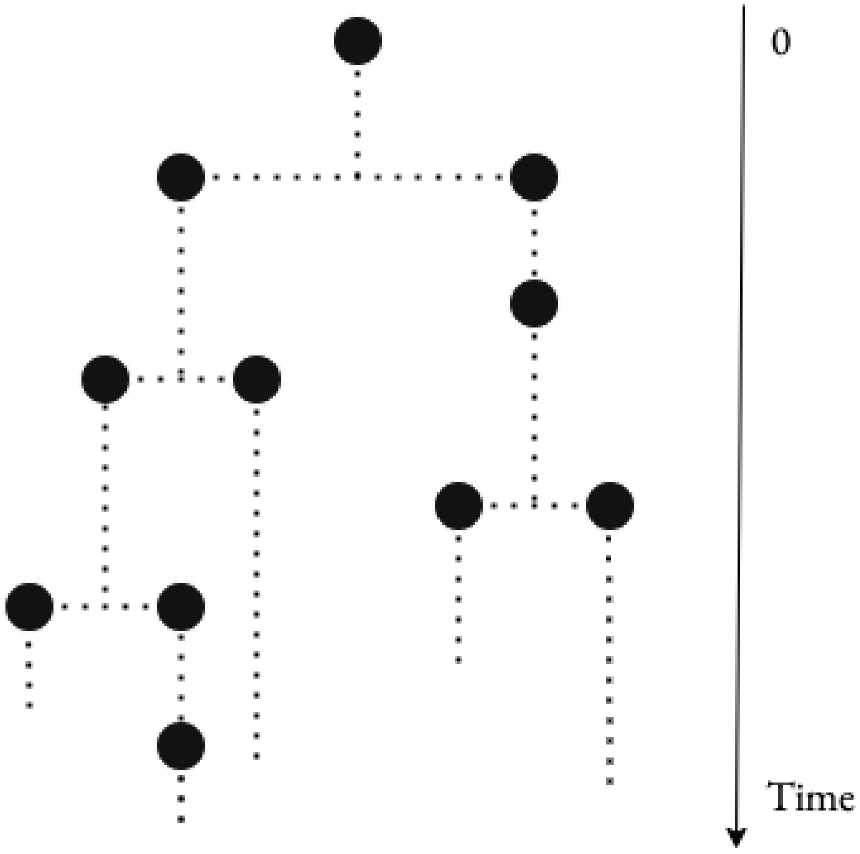

Figure 6. Possible Age-dependent Branching Tree

- Each individual within the same branching tree has the same distribution for the amount of offspring they have.
- The offspring of an individual start their lives at the time of the individual's death.

Since Z_n is now useless, let us replace it with the similar $Z(t)$.

Definition 5.3. $Z(t)$ is the discrete random variable indicating the number of alive individuals at time t.

Definition 5.4. T is the continuous random variable indicating the lifespan of an individual.

Generating functions are also very helpful for analyzing the age-dependent Galton-Watson process. In particular, we like to focus on the generating function of $Z(t)$.

Definition 5.5. $G_t(s)$ is the probability generating function of $Z(t)$. It is given by the following equation:

$$G_t(s) = \int_0^t G(G_{t-u}(s)) f_T(u)\, du + \int_t^\infty s\, f_T(u)\, du \quad \text{[GS01]},$$

where $f_T(u)$ is the probability density function of T.

Even though this equation cannot be solved aside from some special cases, it is useful since it can be used to study the probability of eventual extinction and growth rates over time using techniques similar to the ones used in previous sections.

Although the age-dependent Galton-Watson process is seen as an improvement on the standard Galton-Watson process, some see it as less helpful than the bisexual Galton-Watson process since it still only models 1 gender. In addition, some fault the elementary age-dependent Galton-Watson for doing nothing to account for humans not often giving birth to offspring at their time of death and all at once. Nonetheless, this variation of the Galton-Watson process does have many strengths; for one, it is very good at modeling bacteria populations, especially the ones that split to form new organisms.

6. Acknowledgments

First and foremost, the author would like to thank Simon Rubinstein-Salzedo for creating Euler Circle and the Independent Research and Paper Writing class, giving the author a wonderful opportunity to learn about the Galton-Watson process and other topics in higher mathematics.

The author would also like to thank TA Daniel Naylor for his unwavering support, sharing countless insightful conversations over Zoom and Discord, and his assistance with any mathematical or paper writing queries.

Last but not least, the author would like to thank peers Rowling Yu and Ryan Bansal for having so many interesting conversations with me in breakout room 27.

References

[GS01] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Probability and Random Processes. OUP Oxford, 2001.

[Hua20] Yibi Huang. Stat 244 lecture 10 notes. `https://www.stat.uchicago.edu/~yibi/teaching/stat244/L10.pdf`, 2020. University of Chicago, STAT 244: Statistical Theory and Method I.

[Tsi18] John Tsitsiklis. Lecture 13a: Expectation (part 1). `https://ocw.mit.edu/courses/res-6-012-introduction-to-probability-spring-2018/b343ddecdd393da2822aecd2608eb5f0_MITRES_6_012S18_L13AS.pdf`, 2018. MIT OpenCourseWare, RES.6-012 Introduction to Probability, Spring 2018.

[Wil13] H.S. Wilf. Generating functionology: Second edition. *Generating Functionology: Second Edition*, page 6, 01 2013.