Galton Watson Processes

John Blake (jblake33@students.kenensaw.edu)

Instructor Dr. Glenn Young, Kennesaw State University

MATH3262 Mathematical Modeling

May 1st 2023

Table of Contents

[Introduction 2](#_Toc133839713)

[Methods/Analysis 2](#_Toc133839714)

[Results 6](#_Toc133839715)

[Additional Code 6](#_Toc133839716)

[Bibliography 7](#_Toc133839717)

# Introduction

This document will derive, discuss, and analyze a model that serves to answer the question, “how many children did your ancestors have, on average, to ensure your last name could be passed down to you?”, among other questions [1]. Questions regarding family lineage and ancestry have been answered by companies – like Ancestry.com - in the genealogy industry, which is an industry evaluated at an estimated $3.4 billion USD in 2021 [3]. As the demand for learning about lineage (and last names) increases [3], the curiosity for where our last names - called surnames, in the scope of this document - were obtained and how they were passed between generations increases as well.

When family lineages are diagrammed, they form a tree structure. In the context of graph theory, a graph consists of nodes (individuals) that are connected through edges (the parent-child relationship). A tree requires that every node have one “parent” node (unless it is the root, or start of the tree), and any whole number of “child” nodes. The tree can’t grow to any future generations if all of the nodes in any generation produce 0 child nodes. The probability that this occurs will depend on the probability that a node has any number of child nodes, which will be explored in this model: known as the (Classical) Galton Watson process.

The Galton Watson process is a discrete-time Markov chain that models how a surname is spread through future generations of a parent (‘’generations’’ refers to their children, grandchildren, …). The discrete timestep for this model is a generation.

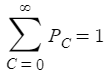
This model makes a handful of assumptions for simplicity.

* Only male offspring are considered; this assumption means that the male’s surname is used when naming their children, and since the female’s surname is always ignored, female offspring are not considered. “Male children” and “children” are used interchangeably in this document.
* A family size can be any whole number (0, 1, 2, 3, …) of children.
* Family sizes are independent of each other.
* Each family has identical probabilities of spreading their surname to their children.

These assumptions aren’t realistic when considering human reproduction, but they do simplify the model to make it clearer and more understandable. The reproduction behavior as outlined in the assumptions is more similar to cell reproduction rather than human reproduction.

# Methods/Analysis

Assume that, at the 0th generation, there is 1 male parent. The number of children this parent has is denoted C, and the probability that this parent has k children is denoted P(C = k), where k = 0, 1, 2, …, ∞. This probability will be denoted PC, where 1 ≥ PC ≥ 0 for all C = 0, 1, 2, … and the sum of all of these probabilities must be 1:



The state Xn is the number of individuals in the nth generation. This makes X0 = 1 an initial condition for the model. Xn = 0 is an absorbing state; if there are no individuals in a generation, the next generation must also have no individuals. This absorbing state will be referred to as the “extinction state”, and the probability that the extinction state will occur at generation n will be denoted en. The goal of this model of the Galton-Watson process will be to find this probability, written as:



This can be read as “the probability that there are 0 individuals in some generation n, given that there was one individual in generation 0”. This can also be expanded out to a sum of probabilities involving the number of individuals in a current generation and their likelihood of extinction. The probability of extinction at generation n is based on the probability of having some number of children in the previous generation and if they all didn’t have any children:

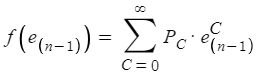


Here, this shows that the probability of extinction is given by the following:





Where e0 = 0 is an initial condition, and is a probability generating function, with the form:



Note that this function is increasing, since:



This derivative must be non-negative, since the ”P’s and e’s” are both probabilities (thus, they necessarily must be between 0 and 1). The only way for it to be 0 is if is an equilibria (discussed below). Otherwise, it is positive (increasing).

More definitions:

Is the sequence of extinction probabilities.

Is the sequence of persistence probabilities (The | notation here reads as ”1 minus the sequence of extinction probabilities”)

This definition of the extinction state probability enhas some noteworthy properties:

For all n, since the derivative of f(en) found above is positive. (the sequence is increasing)

Since en is a probability, it must be between 0 and 1. (the sequence is bounded)

Given that en is both increasing and bounded, must have a limiting value. This value is necessarily in the range of 0 and 1 as well. Let this value be denoted e\*, which is the long-term probability of surname extinction. This also means that 1 – e\* is the long-term probability of surname persistence.

e\* should solve both:

And .

This definition makes e\* an equilibria of f(en). One equilibria value that is always true, regardless of the P’s, is en = 1:



However, we are interested in finding other possible equilibria (anywhere where en= f(en) between 0 and 1), which may or may not exist depending on the values of the P’s.

The probability generating function, f(en-1 ), has already been shown to be increasing. Furthermore, this:



 demonstrates that the function is concave-up (the second derivative is also positive).

A hypothetical function that is increasing and concave up is plotted on the en-f(en) plane below [6, plots]. This function has a f(en) axis intercept of P0, the probability of having 0 children in a generation. As shown above, this function must also pass through the point (1, 1). This function is plotted below with a black color.

Also plotted is the line en= f(en), a line with a slope of 1 that passes through the origin. This line also passes through the point (1, 1). This line is plotted below with a purple color.

|  |  |
| --- | --- |
| **Figure 1.1** | **Figure 1.2** |
| There’s two intersections (equilibria): at point (1, 1), and also at another point (e\*, f(e\*)). This implies there is some probability that the surname will persist. | The only intersection is at the equilibrium point (1, 1). There’s a probability of 1 that the surname will go extinct (it is certain that it will not persist). |

Depending on the values for en and PC for each C = 0, 1, 2, …; there may exist an intersection other than certain extinction (there is a e\* < 1), where the surname is expected to persist rather than go extinct. The shape of the probability generating function’s curve (the black line, as plotted above) determines if/where this intersection occurs. In Figure 1.1, this intersection occurs, which necessarily means that the slope at the point (1, 1) is greater than 1. In Figure 1.2, the intersection does not occur, and the slope at the point (1, 1) is less than 1. The slope of the tangent line of this function at the point (1, 1) can be used to make the following conclusion:





Note that the derivative of the probability generating function when evaluated at 1 (slope of the tangent line at the point of interest (1, 1) ) is the expected value, or number of children, for an individual (with a surname of interest) in this context. Combined with the figures 1.1 and 1.2 above, this helps form a conclusion to the classical Galton-Watson process:

 means the only equilibria is at (1, 1), which also means the surname is expected to go extinct.

means there is an additional equilibria – other than (1, 1) - where the probability of extinction is less than 1, which also means the surname is expected to persist.

Thus, a surname is expected to go extinct if the expected number of children for an individual is less than 1, based on the assumptions made in the introduction. In layman's terms, if an individual with a specific surname can at least replace themselves in a generation, then that surname should persist.

# Results

The below example shows how the model derived from above might be used. Given a set of probabilities P0, P1, P2, …, Pn for the likelihood of any individual with a particular surname having 0, 1, 2, …, n children, it can easily be predicted whether or not the individual’s surname should persist or go extinct.

**Example**: Given each set of probabilities for an individual having some number of children, determine if the surname will likely go extinct (or persist).

Ex. 1: 

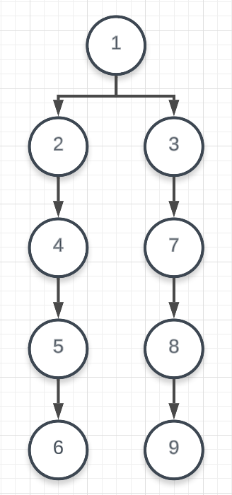
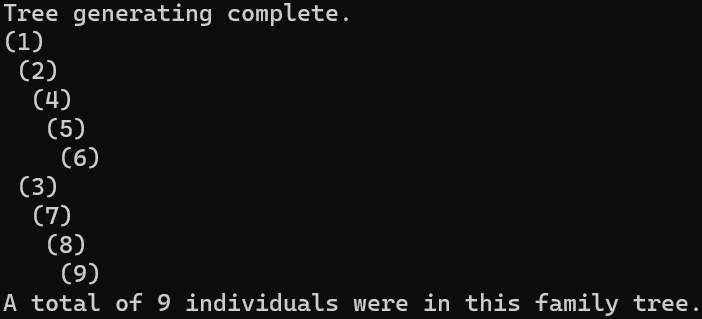
**Solution to Ex. 1**:= <-- Since this is greater than 1, the surname is expected to persist.

Ex. 2: 

**Solution to Ex. 2**:= **0.875** <-- Since this is lesser than 1, the surname is expected to go extinct.

# Additional Code

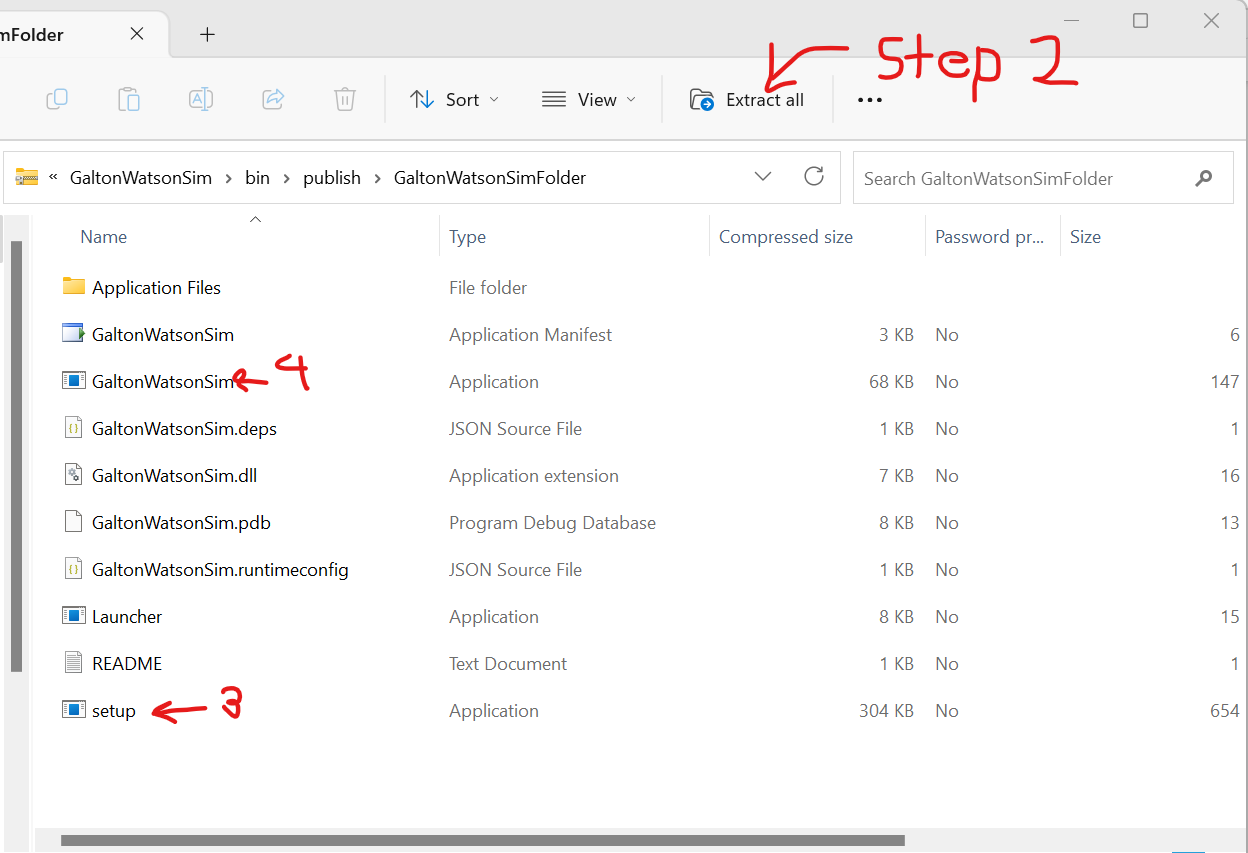
The program, “Galton Watson SIM”, provided in the D2L submission is a small program that performs sample “family tree” generating based on assigned values for the P’s. The program has options for assigning probability values and running a simulation with the values. Help instructions are provided below for how to access and run this program. Below is an example output of running a simulation once:



Source code is available at: https://github.com/jblake33/GaltonWatsonSim

Recommended steps for installing the program are as follows:

1. Download the GaltonWatsonSim.zip file from D2L.
2. Extract the folder. An example of where the button is located in the file explorer is below (on Windows 11):



1. Double-click on “setup” to download any required files.
2. Double-click on “GaltonWatsonSim” (the one with Type “Application” and file size 147KB, compressed size 68KB)

# Bibliography

[1] McKenna, Ian, “The Galton Watson Process” https://corecomputations.wordpress.com/2011/07/26/the-galton-watson-process-part-i/ Accessed April 11, 2023

[2] Lalley, Steve, “Branching Processes” <http://galton.uchicago.edu/~lalley/Courses/312/Branching.pdf> Accessed April 11, 2023

[3] Verified Market Research <https://www.verifiedmarketresearch.com/product/genealogy-products-services-market/> Accessed April 11, 2023

[4] Eager, Eric https://www.youtube.com/watch?v=QPy8fr-KlUA Accessed April 14, 2023

[5] Eager, Eric https://www.youtube.com/watch?v=E6gdPuaoYlc Accessed April 14, 2023

[6] Roch, Sebastien “Branching processes” Figure 6.1 <https://people.math.wisc.edu/~roch/mdp/roch-mdp-chap6.pdf> Accessed April 24, 2023