

Ancestry and Relations

Abstract

In this paper, common trends amongst ancestry lines are analyzed and observed. The goal of the paper is to show how common ancestry forms in subsequent generations of offspring. The methods used include unique and non-unique testing. The affect of population growth is also observed with these methods.

Project 3

Mathematics 2130

Submitted by: John Hollett

Submitted to: Ivan Booth

November 13, 2015

1 Introduction

In this paper, how ancestry changes after each subsequent generation will be discussed. Generations refer to the population of each specific generation. The tests range from basic conditions to more selective processes. These processes will include random sampling with and without unique pairs. A pair refers to the duality between ancestors whether uniquely selected or not.

In non-unique tests, the participants of each generation may be assigned the same parent. In unique testing, each participant contains two different parents. The difference between these tests will be explored and compared.

A program was written in C[#] for tracing and tracking ancestry. This utility is required to perform the ancestry tests. The following equation describes the approximate number of generations required to see a common ancestor. This simple simulation is tasked with the responsibility of networking relationships in a decipherable manner.

$$C_{Generation} = \log_2(N_{Population}) \quad (1)$$

This section will include sampling and analysis of non-unique ancestors of varying sizes for each generation. The figure (??) depicts the results of a test.

In this test, an initial population of ten is selected, and a maximum of ten generations of humans. Each test in the paper is iterated 250 times, including this test.

This test serves as a proof in determining the program works correctly. A smaller size test was created and carried out to satisfy this requirement.

In Figure (??), the axes described is equivalent in for each test. The information shown is

the number of common ancestors for each generation starting from the founding generation.

$$C_{Generation} = \log_2(10) = 3.32$$

With the data depicted by figure (??), it is possible for one common ancestor to emerge after the second generation. Using equation (1), the result supports the graph. This information shows the utility created performs sufficiently to produce proper data.

1.1 Further Testing

In this section, several tests are compared. Each test consists of varying initial population sizes. The number of maximum generations remains set to fifty.

Based on Figure (??), trends can be observed from the changing initial size. Each of tests performed shows a common ancestor emerges only after the first or second generation.

The tests using lower initial population show the number of generations required to produce a non-changing common ancestor set changes with initial population. This is also seen in larger founding population sizes.

In Figure (??), the data contradict perceived consistency in the number of generations required to observe a consistent ancestor set across all subsequent generations. Despite these results, the simulation returns the average number of generations it takes to form a common ancestry set for each founding population.

The average of ancestors carried through subsequent generations depicted in the figures (??) and (??) is roughly 80% of the starting population. Inversely, 20% of the founding population drops from the ancestry tree outright. The data depicted in the next section addresses any influences uniquely selected ancestors exhibit.

Uniquely selected ancestors differs in the methods utilized. Parents chosen for the ele-

ments in a new generation are not equal for unique tests.

2 Unique Testing

The data prepared in the previous section contrast the differences of unique and non-unique tests. Tests were re-ran for populations of 10, 25, 50 and 100, 200, 300. The number of maximum generations the tests were ran with remains fifty for this section.

Figure (??) and (??) shows the average amount of common ancestors in relation to each generation. Using unique selection, the analysis reveals similar results. Checking the results with equation (1) for each initial population supports figures (??) and (??).

$$C_{Generation} = \log_2(50) = 5.64$$

Figure (??) depicts a unique ancestor set being formed in relation to the founding size as seen in the previous section. This proves the difference between the two methods has negligible affect on the data.

Figure (??) shows increasing initial populations. The increase in population in subsequent generation is at a constant rate of N . The result of unique and non-unique tests show there is no significant difference between the two. The unique test results, on average, produce negligibly greater amounts of ancestors. Meanwhile the number of generations required to find an ancestor is greater however it is insignificant.

3 Population Growth

The tests done for this section will simulate population growth. Growth is determined by a percentage that is constant amongst each generation. The population of any subsequent

generation is determined by applying the percentage to the prior generation.

For the tests completed in this section, the populations will grow in size for each generation. The initial population for each test is 25 and varying growth percentages.

In figure (??), despite the population growth, the average number of ancestors remains relatively consistent in all tests. The significant difference is seen when the common ancestry set is found.

With increasing population growth, the number of common ancestors increases. This result reflects the improved probability of selecting and thus creating more common ancestors from founding set.

$$C_{\text{Generation}} = \log_2(25) = 4.64$$

The first common ancestor occurs approximately during the 5th generation. Following the 5th generation, the common ancestor set expands until the complete ancestry set is found. Using equation (1), the result above agrees with figure(??).

4 Technical Details

Explanation of the details of the testing methods are outlined in this section. The data collected were created using a simulation. The C# programming language was used to create a iterative function to simulate data.

Two programmed objects named "Human" and "Generation" were used to correlate coherent data. In a generation, a set of "human" elements is stored. Previous sets are accessible in each "Generation". A newly created "Generation" replaces the last as the current set. Using sequences in mathematics this is explained. A sequence defined using the initial element A_1 and subsequent elements are described in relation to the first in the sequence. The

following example describes this process:

$$A_1 = \text{Generation}_0$$

$$A_N = \text{Generation}_{N-1}, \text{ Where } N \geq 2$$

Where Generation_I is the population set at index I

A "Human" is defined as an object containing two "human" ancestors. For subsequent generations, each "Human" contains the union of its two ancestor's ancestors derived from the original population set. This process is described as:

$$P = \{p_0..p_N\}, \text{ Where } P \text{ denotes original ancestry set}$$

$$A = \{a \in P | a_0, a_1, \dots, a_N\}$$

$$B = \{b \in P | b_0, b_1, \dots, b_N\}$$

$$C = A \cup B, \text{ Where } C \text{ is the ancestry set of the current Human}$$

(2)

Testing ancestry in the current generation, the simulation processes the elements in the set. Using a "Human" in the set, a copy is created of its ancestry set. The copy is then intersected by each remaining "human" in the generation. If the list created is not empty, there is at least one common ancestor in the current generation.

New generations are created using parameters describing; number of new elements, previous generation, uniqueness. Creating a new population requires the prior generation. The two methods for creating each "human" in the population is determined by reading the

uniqueness parameter. A non-unique "human" is created by selecting two random parents. Unique describes the same process as non-unique while requiring the parents to be unequal.

Figure (??) depicts method used to randomly select a parent from a population of 25. While there is no definitive trend in this data, the line of best indicates gradual error. As the number of trials approaches completion, the trend line rests in the middle of the range. This suggests most of the error is eliminated through sufficient testing. The error caused by the method used for selecting a parent is negligible but, other studies require more reliable methods.

In a population set containing 25 elements, the randomly selected values lies between 0 and 24. Maximum value in figure (??) is 24. A set of data containing 25 elements in the language chosen begins at 0 and ends at 24. The actual range is 0, ..., 24 inclusive, resulting in 25 elements.

5 Conclusion

The tests performed this paper, show common ancestry occurs consistently after a number of generations. The figures provided show the initial population affects the appearance of a common ancestor. Following the observation of a common ancestor, the figures depict consistently increasing traces of the founding members in each "Human". This trend is seen for unique and non-unique data, with or without population growth.

In a paper on ancestry [1], the results of tests were far more advanced and complex than those used in this test. It takes into account the various levels of segregation in the population, such as migration and life span.

While this paper does not detail such in depth analysis, real world application of the results does merit a meaningful conclusion. It serves as evidence that common ancestry

occurs and it should not surprise the reader that he/she may be distantly related to the writer.

6 C# Code

References

- [1] "On the Common Ancestors of All Living Humans" <http://tedlab.mit.edu/~dr/Papers/Rohde-MRCA-two.pdf>, Douglas L. T. Rhode, Massachusetts Institute of Technology, 11 November 2003, Web 9 November 2015