The Major Mathematical Contributors to Population Genetics and their Significance

Amelia Lovelace

510 Warren Road
Cockeysville, MD 21030

ahlovelace@gmail.com

Hood College

1 Introduction

Population genetics is an area of study that has evolved greatly over the past couple centuries. More specifically it is the study of how traits are inherited across generations within a population. This subject has been thoroughly researched and debated since its birth in the 19th century. Since then this area of study has expanded as our knowledge of genetics grew. This expansion was made possible due to the collaboration of scientists across many disciplines. Although much has been written about the geneticists and biologists that have contributed to the study of population genetics, little has been said for its mathematical contributors. The purpose of this paper is to evaluate the history of population genetics in order to identify the major mathematical contributors and shed some light on their significance to this area of study.

2 History of Population Genetics

Gregor Mendel (1822 - 1884), a German monk, performed experiments on the hybridization of garden peas. Within his experiments, he had recorded data on the traits that were inherited by each hybridization. Mendel observed that when he crossed a pure white flower with a pure purple flower, the next generation, F1, exhibited purple flowers. He suggested that these paired hereditary units or “factors," which were later termed genes, had a recessive and dominant characteristic *a* and *A* respectively [1, p. 5]. These paired factors separate when the sex cells are formed. When the purple flowers of the F1 generation were self-crossed he found a 3 to 1 ratio of purple to white flowers [1, p. 22]. This seemed to suggest that for each trait of the genes, dominant *A* and recessive *a* could be paired in 3 different ways, *AA*, *Aa*, and *aa* [1, p. 47]. We now know that these paired alternate forms of the “factors" *A* and *a* are actually alleles and a genotype of *AA* or *aa* is referred to as homozygous and the genotype of *Aa* is referred to as heterozygous. Mendel's findings led to the three laws of Mendelian inheritance. The first law is the law of segregation. This law states that every individual has a pair of alleles for a specific gene and one copy of the allele from each parent is passed on to the next generation. The second law, or law of independent assortment, states that each gene with separate alleles is passed down independently of one another and has no effect on one another. The third law or law of dominance states that the dominant trait is always expressed in the offspring. For instance, in a heterozygous genotype, the dominant trait is portrayed. Mendel's findings were published 1866 as Experiments on Plant Hybridization; however, this paper was largely ignored at the time due to the fact that the author was not a reputable scientist.

In 1889, Francis Galton (1822 - 1911) published Natural Inheritance. This work was published for the purpose of explaining the processes of heredity and applying them to the many observations Galton made and the data he collected. His data consisted of “stature" or height measurements across two successive generations of many families. He also generated mechanisms for computing the distribution, frequency, regression, and normal variability of his data. Using the mathematical formulas he generated, Galton could predict the heredity of traits for the proceeding generation [2, p. 134-150]. Galton had developed the statistical concepts of correlation and regression in order to apply it to the study of heredity. He had discovered that the mean value of the inherited trait of stature deviates from the mean of the parent and “regresses" toward the mean of the population [2, p. 211]. For instance, the height of the offspring of a set of relatively tall parents would be slightly less than that of its parents for it will regress to the mean height of the population. Galton was the first to provide experimental evidence to find the surface representing the frequency of error or normal surface of correlation [3, p. 6]. It was Galton's discovery of regression that allowed him to make the connection to the normal curve generated by the population as an attribution of the inheritance of traits from one generation to the next [3, p. 4]. Unlike Mendel's work, the work of Galton was widely accepted in the scientific and mathematical community in England.

3 Biometricians vs. Mendelians

In 1900, Mendel's work was “rediscovered" by English geneticist William Bateson (1861 - 1926). This rediscover sparked a heated controversy between two groups of genetic scientists, the Mendelians and the biometricians. The main protagonists for the Mendelians were Bateson and Reginald Punnett (1875 - 1967). They stood by the three laws that govern Mendelian inheritance. The Mendelians believed that inheritance is governed by discontinuous variations in a population. In his 1897 paper, Bateson states that “I have argued that variations of a discontinuous nature may play a prepondering part in the constitution of a new species" [4, p. 34]. In other words, when a population is allowed to breed freely, there is no complete blending of traits in the next generation; therefore, these traits regress to two distinct forms instead of a mean form. Bateson was fond of Mendel's laws due to the fact that they provided an explanation for the results found in his works. The idea of discontinuity is currently explained by mutations. Mutations allow for discontinuous variations within a population and may even lead to speciation or the formation of two reproductively-isolated populations.

The main protagonists for the biometricians were Udny Yule (1871-1951) and Karl Pearson (1857 - 1936). They stood by the works of Galton. They argued that traits are continuous within a population and they regress to the mean of a population. In other words the evolution of new species was the result of gradual accumulation of the effects of small continuous variations. Pearson was able to support his arguments in this controversy by expanding the statistical methods of correlation and regression developed by Galton [5, p. 221]. Pearson created formulas for analyzing the standard error of coefficients of correlation, correlation between sample means of two correlation traits, and correlation between standard deviations [6, p. 63]. These tools he developed were important in mathematically measuring the influence of heredity. Pearson also developed the Chi squared test for independence [7, p. 457]. This test served to determine if the difference between two sets of data is due to chance. More specifically, this tests for the fit of distribution and independence for a large population with random mating. The Chi squared test is important for determining if a trait is passed on to the next generation by chance.

In 1918 Ronald Aylmer Fisher (1890 - 1962) published a paper “The Correlation Between Relatives on the Supposition of Mendelian Inheritance" which settled the dispute between the Mendelians and biometricians by proving that both ideas were compatible. In this work he proved that a large number of Mendelian factors was a sufficient explanation for traits with continuous variation [8, p. 474]. In other words blending inheritance is due to the effect of many Mendelian traits that are individually insignificant. This concept joins the ideas of both continuous and discontinuous traits. More importantly, Fisher was able to prove this through theory and practice. He mathematically calculated the correlation between relatives and he conducted breeding experiments on mice in order to confirm his findings. Fisher was also able to improve upon Pearson's Chi squared test by applying it to smaller populations greater than two [9, p. 8]. This new test came to be known as Fisher's exact test. Fisher is also known for the development of the “fundamental theorem of natural selection" which states that the increase rate of fitness in any organism is equivalent to the genetic variance in reproductive fitness at that time [9, p. 10]. The importance of dominance was confirmed by this theorem in that Mendelian selection favors the dominance of beneficial genes that increase reproductive fitness.

4 Hardy-Weinberg Equilibrium

It was widely accepted in the 19th century that the phenotypic ratio for offspring of a heterozygous cross, *Aa* x *Aa*, to be 3 dominant to 1 recessive. However, based on Mendel's law of segregation, the gene frequencies remain unchanged for each generation. Based on this result, Punnett was curious as to why individuals expressing the dominant trait do not drive out individuals who express the recessive trait. Pearson had derived Mendel's first law in order to show the connection between parent and offspring. Assuming that the sum of the gene frequencies is equal to one half he was able to demonstrate that the gene frequencies, 1*AA* : 2*Aa* : 1*aa*, will be maintained for each generation using the following formula, where Parent 1 has gene alleles a_1 and a_2 and Parent 2 has gene alleles a_3 and a_4 [10, p. 1146].

$$a_1a_2 \times a_3a_4 \ = \ (a_1 + a_2)(a_3 + a_4) = a_1a_3, a_1a_4, a_2a_3, a_2a_4$$

Pearson's solution is similar to what we would call the Hardy-Weinberg equilibrium which solves this problem for a gene frequency of one. Punnett asked his colleague Godfrey Harold Hardy (1877-1947), a mathematician at Cambridge, for help in regard to this problem. Hardy published his solution in his 1908 paper, “Mendelian proportions in a mixed population." He proved that if mating is random in a large population, and selection is absent, then gene frequencies will not change from generation to generation if and only if the square of the frequency of heterozygotes is equal to the product of the frequency of the homozygotes [12, p. 4952]. In other words suppose the parental gene proportions are *AA* : 2*Aa* : *aa* or *p* : 2*q* : *r* and the sum of these gene frequencies is one such that $p + 2q + r = 1$, then the gene proportion for the offspring will be $(p+q)^2 : 2(p+q)(q+r) : (q+r)^2$ and the sum of their gene frequencies will also be one such that

$$(p+q)^2 + 2(p+q)(q+r) + (q+r)^2 = 1$$

or

$$p_1^2 + 2q_1 + r_1 = 1$$

$$\Leftrightarrow$$

$$q^2 = pr$$

5 Significance and Future Studies

Many of the contributions made towards the study of population genetics are attributed to mathematicians or statisticians such as Pearson, Fisher, and Hardy. The development of population genetics during the 19th and 20th century would be less successful without the help of these mathematicians. In general these mathematicians have helped this area of study by solving difficult problems that geneticists were unable to solve, providing mathematical groundwork for future data analysis, and settling disputes between rival scientific groups.

Pearson's work led to the development of many formulas used for evaluating population data. Pearson was able to derive formulas for standard error in order to better estimate the parameters of the population sampled. Currently, it is standard procedure to include this statistical information for any research involving population data. Although his steps on correlation and regression analyses are not used today, our current methods are built off of his founding ideas. Genetic population studies today rarely use correlation or regression to analyze simple data; however, they are still used for very complicated genetic processes that involve multiple loci. Unlike the regression and correlation analyses, the Chi squared test developed by Pearson is still widely used in genetic population studies today. It is used to test for variance within a large normal population. For example, scientists were able to determine that the ratio of purple to yellow kernels on an ear of corn is not due to chance but is genetically predetermined by using the Chi squared test.

Although Pearson's test can only be used for large populations, Fisher was able to correct for this discrepancy so that it could be used for testing smaller populations as well. Fisher was also able to solve the dispute between the Mendelians and the biometricians. He was able to show that the beliefs held by each side were indeed compatible and could be used together to describe the process of heredity. In solving this dispute, geneticists were able to put their differences aside and get back to work on their research. This dispute hindered the productivity of the scientists at the time for more energy was utilized to prove the other group wrong than to find a common ground to agree upon. Geneticists currently accept both parts of Mendelian and biometrician beliefs. Discontinuous variation within a population is now explained by mutations which were not yet discovered at the time. Also, continuous variation within a population is attributed to a large number of Mendelian traits as predicted by Fisher. Fisher was able to apply his statistical theories to practice by generating genetic experiments and data to provide further evidence towards his mathematical models. This is important due to the fact that it provides further evidence for less mathematically inclined geneticists of the time. Fisher's work on dominance reaffirmed the importance of dominance in evolution and as a result, provided additional evidence towards Darwin's theory of evolution by natural selection which is now widely accepted in the scientific community.

Hardy is considered to be a pure mathematician in that his mathematical findings cannot be applied to a real-world situation. In his 1940 article, A Mathematician's Apology, he states, “I have never done anything useful. No discovery of mine has made or is likely to make, direct or indirectly, for good or ill, the least difference in the amenity of the world" [11, p. 821]. Although Hardy mostly worked in the field of number theory and complex analysis, he did make a useful contribution to the field of population genetics. Hardy's discovery of the Hardy-Weinberg equilibrium is still used today for the purpose of population studies. Geneticists use this equilibrium for the purpose of predicting genotypic frequencies of traits for future generations based on the parental genotype frequencies using population data. While many geneticists and other statisticians were struggling with the problem, Hardy was able to simply prove that the genotypic frequencies between generations remain unchanged. The contributions of the mathematicians, Pearson, Fisher, and Hardy made a “difference in the amenity of the world" by providing the foundation to the field of population genetics within the 19th and 20th century. These men were able to provide mathematical means of solving problems that geneticists were struggling to solve.

References

1. G. Mendel, Versuche über Planzenhybriden, Verhand-lungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, Abhandlungen,(1866) 3-47.
2. F. Galton, Natural Inheritance, (1889) 1-266; available at http://www.mugu.com/galton/books/natural-inheritance/pdf/galton-nat-inh-1up-clean.pdf.
3. D. Denis, The Origins of Correlation and Regression: Francis Galton or Auguste Bravais and the Error Theorists? (2014) 1-10; available at http://www.psychology.sunysb.edu/ewaters/301/11_correlation_coef/origin_corr.pdf.
4. W. Bateson, Progress in the study of variation I, *Science Progress I*, (1897).
5. K. Pearson, Proceedings of the Royal Society of London, Series B, Containing Papers of a Biological Character , **81**:547 (1909) 219-224.
6. K. Pearson, Mathematical contributions to the theory of evolution. XII. On a generalized theory of alternative inheritance, with special reference to Mendel's laws. Philos, *Trans. R. Soc. A* **203** (1904) 53-86.
7. C. Gillispie ed. *Dictionary of Scientific Biography*, “Pearson, Karl," Vol 10. Charles Scribner's Sons: New York 1972, 447-473.
8. J. Crow, Eighty years ago: the beginnings of population genetics, *Genetics* **119** (1988) 473-476 reprinted in Crow and Dove 2000.
9. C. Gillispie ed. *Dictionary of Scientific Biography*, “Fisher, Ronald Aylmer." Vol 5. Charles Scribner's Sons: New York (1972) 7-11.

10. A. Edwards, G. H. Hardy 1908 and HardyWeinberg Equilibrium. *Perspectives*, **179** (2008) 1143-1150.
11. J. Crow, Hardy, Weinberg and language impediments, *Genetics* **152** (1999) 821-825.
12. G. Hardy, Mendelian proportions in a mixed population, *Science* **28** (1908) 4950-4955.