Population Genetics with Statistics

BY YUEJIAN MO

12-21-2017

Abstract

Population genetics is the subfield of genetics, which deals with genetic differences inside and between population, and is primarily found by Sewall Wright, J. B. S Haldane and Ronald Fisher. Fisher also is describled as "a genius who almost single-handedly created the foundations for modern statistical science". So, population genetics has strong relative with statistics. It will be enjoyable to review the work from the devolopment of population genetics and foundation of modern statistics to nowdays. Major fundamental and meaning topics of population genetics and modern statistics are reviewed, escpeically how to modern statistical methods are being developed for answering population genetics proplem. This review attend to review the work of statistists do for biological problems during 20 century.

On the Origin of Species and *The Scope of Biometrika*

In 1859, **Charles Robert Drawin** published ***On the Origin of Species***. Drawin's book introduced idea— **survival of the fittest**. Populations evolve over the course of generations through a process of natural selection: a Ratio of Increase so high as to lead to a Struggle for Life, and a consequence to Natural Selection, entailling Divergence of Character and the Extinction of less-improved forms. The similarity of some specials, showed that modern organism can evolve from organism in the past. For example, different birds in isolative island still are common in anatomy, and different embryon has similary, even human embryo exist tail during the early development period.

However, Drawin can only explain why modern organism exsit. In another words, Drawin can not find a new speicals, which evolutes from old specials in human being history. In UK, there were some mathematicists who appreated Natural Selection. They tried to understand Natural Selection with statistcs, meaning finding the form of a numerical statement. Thus, statistists Francis Galton, Karl Pearson, and Raphael Weldon established journal ***Biometrika*** to promote the study of biometrics in 1901. In the frist issuse, they wrote in ***The Scope of Biometrika***:

> The unit with which such an enquiry must deal, is not an individual out a race, but a stastically representative sample of a race; and the result must take the form of a numerical statement, showing the relative frequency with which the various kinds of individuals composing the race occur. — *The Scope of Biometrika (1901)*

For K. Pearson, only the probility distribution make sence. For example, Drawin's finches is not the object of research, but the distribution of finch population. For a kind of finch, if we can meansure beak's length of total finch, we will obtain four parameters of distribution function of beak's length, which is represente beak's length. The four parameters in K. Pearson system are:

Definition 1. *The mean*

***Definition* 2.** *The standard deviation*

***Definition* 3.** *Symmetry*

***Definition* 4.** *Kurtosis*

K. Pearson said, assuming survival of the fittest, although we couldn't observe new speices's production under environmental pressure in short time, we can find out the change of four parameters of distribution function. The three editors of *Biometrika* published that *Biometrika* would collect data from all over the world, in order to determine there parameters. They expected that parameters of sample vary different environment. In the next 25 years, *Biometrika* published many data from different place. For each group of data, Pearson and his assistants calculated these four paramters and published. For finding out the different, each four parameters will compare with other corresponding data.

At the bgein of 20 century, Weldon designed a experiment to confirm surival of the fittest. During the devolopment of china in Southern England, water outside harbour was less silt than inside. Weldon collected hundred of crabs from Plymouth and Dartmouth, and cultured in bottles. Half of crabs lived in water outside harbour, and other crabs lived in water inside harbour. After some time, Weldon measure the shell of living crabs. Reasonalbily, parameters changed! However, Weldon was dead before finishing this report.

From the articles in *Biometrika*, it is not obvious that they had proved the Natural Selection succefully. After 1925, more and more theory mathematical articals were published. But K.Pearson and other statists provided a lot of anaylsis techniques for biologists. These techniques, which are widely used today for statistical analysis, include the chi-squared test, standard deviation, and correlation and regression coefficents.

Pearson was import in the founding of the school of biometrics, which was a competing theory to describe evolution and population inheritance at the turn of the 20th century.(wikipedia). It is inte rsting that, the biometric school, unlike the Mendelians, focused not on providing a mechanism for inheritance, but rather on providing a mathematical description for inheritance, but rather on providing a mathematical description for inheritance that was not causal in nature. Pearson criticized biologists who did not focus on the statistical validity of their theories, stating that "before we can accept [any cause of a progressive change] as a factor we must have not only shown its plausibility but if possible have demonstrated its quantitative ability"(The grammer of science).

Nowsday, we have many alternative model organism to confirm Drawn's theory. To confirm it, we can do thousand of generations in short time for the short life cycle of model organism. Using this method, Natural Selection can be used to explanate more and more data. So, most scientists think that parameter change in short time is not necessary to prove the evolution process. However, we learn from the revolution of K. Pearson: the object of science research is not itself, but the distribution function in mathematics.[1]

Analysis of Variance and *The Genetical Theory of Natural Selection*

Sir Ronald Aylmer Fisher FRS, borned in 1890, who was famous as a Brithsh statistician and geneticist. In 1919, He began working at the Rothamsted Experimental Station for 14 years, where he analysed its immense data from crop experiments since the 1840s, and developed the anaylsis of variance, data anaylsis method. In genetics, he brought mathematics into population genetics.

In 1918, Fisher wrote the paper *The Correlation between Relatives on the Supposition of Mendelian Inheritance*. This is a milestone work combined Glaton's coefficient of correlation and Mendelian Law, which showed mathematically how continuous variation could result from a number of discrete genetic loci.

Serve excellent pappers *Study in Crop Variation* were published during 1918-1932. In the second paper, *Study in Crop Varation II*, Fisher designed a experiment to determine the fluence of manure. This experiment was different with tranditional methods. Tranditionly, they used one kind of manure in a large farmland. But Fisher divided a large farmland into many small area. Each area used manure or not. Althgoug it solve the bias of different large framland, each small area still are difference in some way. Fisher choosed to determine the action of each area in random. Any possible manure gradient will be zero in total.

For thie example, Fisher develop the method of anaysis of variance, which maybe the most important tool in biology. Fisher also develop testing the goodness of fit, whist p value is used in biology widely.

In book *The Genetical Theroy of Natural Selection*, Fihser showed how Mendelian genetics was consitent with the idea of evolution driven by natural selection.

During 1920s, B.S Haldane also pulished a series of papers, which applied mathematical analysis to real-world examples of natural selection, such as the evolution of industrial melanism in pepered moths.

Works of Fisher, Haldane and others make it firmly based in mathermatical modelling, its predictions confirmed by experiment , Natural selsetion , once considerd hoplessly unverfiable speculation abou history, was becoming predicated, meansurable, and testable. Their work helped to found the discipline of the theortical pupulation genetics.

Modern population genetics and Statistics

Wikipedia do a well summary that:

> "They observed further that there are two groups of chanllegns to the way the modern synthesis viewed inheritance. The first is that other modes such as epigenetic inheritance, phenotypic plasticity, and the materal effect allow new characteristics to arise and be passed on, and for the genes to cath up with the new adptaions later. The second is that all such meacanisms are part, not of an inheritance system but a developmetn system: The fundamental unit is not a discrete selfishly competing gene, bu a collaborating system that works at levels from genes and cells to organisms and cultures to guide evolution."

Bibliography

[1] Karl Pearson. dec 2017. Page Version ID: 814457871.

https://en.wikipedia.org/wiki/Charles_Darwin

(I.) The Scope of Biometrika. (1901). *Biometrika*, *1* (1), 1–2. https://doi.org/10.1093/biomet/1.1.1